3

# The Modularity of Sentence Processing Reconsidered

FERNANDA FERREIRA AND JAMES NYE

The idea that the sentence processing system is modular has fallen out of fashion. The proposal got off to a promising start with the publication of results in the early to mid-1980s suggesting that word meanings are activated without regard to their global contexts, and that sentence structures are assigned to words at least initially without consideration of whether the structure would map on to a sentence interpretation that made sense given prior knowledge or given the contents of the immediate linguistic or visual context. Eventually, the modular view of sentence processing became strongly associated with what was termed the "two-stage" model of comprehension, a model which assumed that an initial syntactic analysis or parse was created by implementing a couple of simple parsing operations, and that this initial parse was then revised if it either did not lead to a globally connected syntactic structure, or if it led to a meaning that did not fit the comprehender's expectations and goals. By the late 1980s, connectionist approaches to cognition were becoming increasingly popular, and although one of their most salient properties is their flexibility, connectionism became strongly associated with interactive architectures, and those were assumed to be nonmodular. About a quarter century of research has since been directed at trying to show that sentence processing is not modular, and that instead the interpretation assigned to a sentence is influenced by all kinds of knowledge and sources of information ranging from visual context to beliefs about the intentions and even social background of the speaker. The nonmodular view is so widely accepted at this point that it is now almost mandatory to end scholarly papers and presentations with the observation that the findings support a highly interactive system in which knowledge sources freely communicate. It has been a

very long time since anyone in the field came forward with any sort of argument in support of the modularity hypothesis.

In this chapter, we will review the evidence that is meant to support this overall consensus in the field that sentence processing is nonmodular. We will begin by summarizing the original (1983) modularity proposal. We will briefly examine the important features of a module as described in the (1983) book (*The Modularity of Mind*—henceforth, *TMOM*), focusing specifically on how those properties were interpreted by researchers working on sentence processing. Then, we will summarize a large literature that emerged in response to the idea that sentence processing might be modular. The organization will be thematic: We will consider first the debate concerning the use of what might be described as intra-linguistic information, including prosody and lexical information. From there, we will consider the debates focused around the use of context, including both visual and discourse context. We will argue that although some of the simplest and most obvious versions of modularity might be implausible, it is a distortion to assert that the data undermine modularity in sentence processing entirely. Indeed, seen in a fresh light, the results of the bulk of studies conducted over the last 25 years can be taken as evidence for a more refined and detailed view of sentence comprehension, which retains many of the features of a modular system. The point is to use the findings from the studies to inform how we understand what the sources of information are and how they are organized, activated, and combined. We will also suggest that, in many cases, the claims for nonmodularity have simply been exaggerated—particularly those based on experiments using the so-called *visual world paradigm* (VWP).

An interesting new development in the field of sentence processing is the advent of new approaches emphasizing the shallowness of sentence comprehension. These approaches go under a few different names, including *good-enough language processing* (Ferreira et al., 2002), *shallow processing* (Sanford & Sturt, 2002), *late assignment of syntax theory* (Bever, Sanz, & Townsend, 1998), *analysis-by-synthesis* (Fodor, Bever, & Garrett, 1969; Garrett, 2000), and *noisy channel/rational communication* (Levy, 2008; Gibson, Bergen, & Piantadosi, 2013) models of processing. The common assumption is that comprehenders simplify, misinterpret, or alter the input to end up with an interpretation that is more compatible with semantic expectations. These models have been difficult to categorize with respect to the modularity hypothesis. On the one hand, the idea that comprehenders use simple tricks or heuristics to obtain at least an initial interpretation seems compatible with modularity, particularly the features relating to shallowness. In addition, the models are consistent with other approaches to cognition that emphasize the limited use of information for speed and sometimes, even, for more accurate performance (Gigerenzer, 2004; Kahneman, 2011). On the other hand, because these models suggest that the system is biased toward plausibility or, in more current terminology, because they emphasize the role of "priors" in the Bayesian sense, they seem to emphasize nonmodular aspects of

the system; they seem to highlight the idea that the sentence processing system is driven by semantic considerations and above all wants to create interpretations that are semantically or pragmatically compelling. One of our goals will be to try to sort through these possibilities and make the case that these approaches are consistent with a modular approach to sentence processing, if we emphasize shallowness rather than encapsulation.

## THE MODULARITY OF MIND

As is now well known, the modularity thesis assumes that some cognitive systems have the following features. First, there are what we might call the more biological properties: Modular systems are associated with *neural specialization*; for example, specific areas of the brain seem to respond selectively to linguistic input (Fedorenko, Duncan, & Kanwisher, 2012). In addition, modular systems *emerge during development* with little in the way of individual variation. Although recent research on child language has tended to emphasize major differences in vocabulary and some other aspects of language competence in children from different social and economic backgrounds (Hoff, 2006), it remains clear that core language capacities emerge in almost all children at about the same time and in roughly the same sequence (Gleitman & Wanner, 1982). Modules also tend to become *selectively impaired* when an individual suffers from a biologically based disorder such as dyslexia or when a person experiences brain damage (e.g., aphasia; Joanisse, Manis, Keating, & Seidenberg, 2000; Dick, Bates, Wulfeck, Utman, Dronkers, & Gernsbacher, 2001; Sitnikova, Salisbury, Kuperberg, & Holcomb, 2002; Caramazza & Zurif, 1976).

The second set of module properties have to do with what we'll describe as superficiality: Modules deliver *shallow outputs*, which in the case of language can be taken to mean that what the sentence processing system delivers to systems that operate further along the information processing stream is merely the conditions for an interpretation; for example, the system that must determine what action to be performed based on a spoken utterance does not have information about the presence of gaps or traces in the syntactic representation from which the interpretation was derived. Similarly, people have "*limited central access*" to the internal operations of the sentence processing system; they might obtain an interpretation for a sentence, but they can't reason about the sources of that interpretation or the intermediate representations that were constructed to obtain it. This set of properties concerning superficiality have received less attention than the others, but we will argue that they are at least as significant, and that they relate closely to the newer models of sentence processing that were mentioned earlier—models which assume that the sentence processing system often engages in shallow processing.

The final set of properties of a module are the ones that have been the target of the great empirical scrutiny, particularly in the area of sentence processing. These are the features that relate most closely to issues of information flow

in a cognitive system, and map on to the older distinction between so-called "top-down" and "bottom-up" streams of information flow (Zekveld, Heslenfeld, Festen, & Schoonhoven, 2006; Field, 2004). Most important of these is that a modular system must exhibit *information encapsulation*: a module can access its inputs and its own databases and processes, but it cannot access anything outside the module. Its operations are also therefore *domain-specific*: the module consults a narrow range of information and that database is stated in a proprietary vocabulary related to the domain of processing. And because of this domain specificity and information encapsulation, the system can operate *automatically* (mandatory operation) and *quickly*.

Fodor in (2000) reinforces the importance of information encapsulation by describing it as being at the "heart of modularity" (p. 63). For a system to be a module, it must consult only a limited computational database when it analyzes input. It is also perhaps for this reason that most empirical investigations of whether a system is modular, and in particular whether the sentence processing system is modular, have tended to focus on demonstrating that a piece of information assumed to be outside the module does or does not affect processing in that domain. But what the notion of information encapsulation should also highlight is the importance of determining the information sources that are assumed to be used by a particular module. In other words, delineating the representational domain of a putative module is critical to determining whether its operations conform to modularity. In the area of language comprehension, this point was never properly confronted before the claims for anti-modularity started to be made. For example, some of the earliest studies were focused on demonstrating that the sentence processing system takes into account information about prosody when it makes syntactic decisions. The idea was that because prosodic information was stated in a different vocabulary from syntax, it should not be able to affect the computation of a parse tree. The problem with this argument, however, is twofold: First, and more obviously, if a prosodic analysis is *input* to the module that performs syntactic analyses, then prosodic effects on parsing are to be expected and in no way violate the modularity thesis. Second, and perhaps a bit more controversially, if a representational format is proposed which blends syntactic and prosodic information, then again, prosodic influences on syntax are compatible with modularity, as are syntactic influences on prosody. This point will be discussed in more detail.

Finally, it is important to recognize that, in *TMOM*, Fodor also argued that modularity should be construed is a matter of degree: "One would thus expect—what anyhow seems to be desirable—that the notion of modularity ought to admit of degrees. The notion of modularity that I have in mind certainly does" (p. 37). A system is modular "to some interesting extent" if it exhibits some of the properties summarized earlier; not all of them need to be present. At the same time, as we have also seen, the one property that seems necessary for a system to be described as modular is information encapsulation, at least for Fodor.

## THE "TWO-STAGE MODEL" OF SENTENCE PROCESSING

For a variety of historical reasons, almost from the beginning, the idea that the sentence processing system might be modular became almost entirely conflated with testing a particular model of parsing—the so-called "two-stage model" first developed by Lyn Frazier (Frazier & Fodor, 1978) and then elaborated by her colleagues, including the first author (Ferreira & Clifton, 1986; Rayner, Carlson, & Frazier, 1983; Frazier, Pacht, & Rayner, 1999). Thus, in the interests of full disclosure, we acknowledge that the first author is strongly associated with this model, and both authors believe it is a compelling and empirically valid approach to explaining sentence comprehension. Nonetheless, it is important to recognize the historical coincidence that at the same time that *TMOM* was published, the two-stage model was also dominant. That model made several critical architectural assumptions from the perspective of evaluating the modularity hypothesis in this cognitive domain: First, the model assumed that a single parse is constructed for any sentence based on the operation of two simple principles: *Minimal attachment*, which constrains the parser to construct no potentially unnecessary syntactic nodes, and *late closure,* which causes the parser to attach new linguistic input to the current constituent during a parse, rather than going back to a constituent created earlier or postulating the existence of a new constituent. In addition, the two-stage model in its 1980s form assumed that the only information that the parser had access to when building a syntactic structure was its database of phrase structure rules. It therefore could not consult the syntactic information associated with lexical items. For example, in the sequence *Mary knew Bill* the noun phrase (NP) *Bill* would be assigned the role of direct object because that analysis is simpler than the alternative subject-of-complement-clause analysis, and the information that *know* takes sentence complements more frequently than direct objects could not be used to inform the initial parse.

Similarly, decisions concerning the creation of the initial parse could not be influenced by prosodic information either. For example, given something like *Because Mary left Bill*, the NP *Bill* would be syntactically integrated as a direct object, even in the presence of a major intonational phrase boundary after *left*. Of course, during this period when the two-stage model and modularity were both relatively new, the question how prosody might affect parsing had to be put largely on hold because there were few good techniques available for studying the earliest stages of spoken sentence comprehension. And, as was argued in *TMOM*, the modularity of a system cannot be assessed with offline measures or techniques that provide information about the final stages of processing; to assess modularity, it is necessary to tap into early online processing. Yet another historical coincidence is that, in the 1980s, eye movement monitoring systems started to become affordable and easier to use, and so more and more psycholinguistic laboratories acquired some type of eyetracking device. But, at this point, eyetracking was applied almost exclusively to investigations of visual language processing (reading), and reading was assumed not to involve prosody in any

serious way. (This assumption would change, of course, with the "implicit prosody" hypothesis of reading, but that is a topic for a different volume.) Eventually, researchers did venture into the field of spoken language processing and studies examining prosody in parsing were conducted. We will discuss those studies shortly.

In summary, the modularity thesis was tested against a specific model of sentence processing—a model which assumed that the parser proposes analyses serially and consults only phrase structure rules to make syntactic decisions. Eventually, evidence against the two-stage model would be construed as evidence against modularity as well, even though obviously other architectures for sentence processing are conceivable and even plausible. Moreover, findings that challenged assumptions such as the lack of access to subcategory information were not used to inform and update the assumptions about how any hypothetical sentence processing module might be organized or might operate; instead, they were taken as evidence against modularity itself. Having set the stage for the tests of modularity in this way, we now turn to experimental work designed to evaluate the modularity of sentence processing, keeping in mind that they were also, simultaneously, tests of the so-called two-stage model of parsing.

## EVALUATING THE USE OF LANGUAGE-INTERNAL SOURCES OF INFORMATION

We begin with the question whether lexical information, and in particular, information linking elements such as verbs with the kinds of constituents with which they may occur, affects initial parsing. On the surface, it would appear to be rather odd to think this information would not be used, because in many theories of grammar, verb subcategorization information is stated in a syntactic vocabulary (Chomsky, 1965; Gahl & Garnsey, 2006; Hare, McRae, & Elman, 2003). For example, the information that the verb *put* must occur with both a noun phrase and a prepositional phrase can be represented as something like *put[__NP PP].* As *TMOM* emphasizes, to establish whether a system is modular, it is critical to understand what its proprietary databases are. If we assume that a parser builds syntactic structures using syntactic information, then it would not seem unreasonable to assume that verb subcategorization information would be integral to the parser's operations. And, indeed, the earliest studies examining this question suggested that it is. Following from linguistic arguments based mainly on intuition data (Ford, Bresnan, & Kaplan, 1982), Mitchell and Holmes (1985) investigated this question by looking at the processing of sentences such as *The historian suspected the manuscript of his book had been lost*. They found that participants took less time to read the phrase *had been lost* when it co-occurred with *suspected* rather than with a verb such as *read*, which was presumed to occur because *suspected* takes sentential complements more frequently (see also Ford, Bresnan, & Kaplan, 1982). This result could be interpreted as evidence that the parser consults two sources of syntactic information

during construction of its initial parse: phrase structure rules and verb subcategorization frames. It is not obvious that it stands as evidence against the modularity hypothesis.

Soon afterward, however, Ferreira and Henderson (1990) conducted a follow-on study designed to address a limitation of the Mitchell and Holmes (1985) experiments: Because Mitchell and Holmes employed a phrase-by-phrase reading task, it was possible that the reading times conflated initial and reanalysis processes. Self-paced reading requires participants to make a decision on each displayed chunk concerning whether to push a button to receive the next chunk or stay put in order to get more processing time. Ferreira and Henderson therefore designed a similar experiment but used the eye movement monitoring technique, which has exceptional temporal resolution (a sample of the eye position is taken approximately every millisecond) and spatial resolution. They found that verb bias had no effect on early eyetracking measures (e.g., first fixation and gaze durations) but did influence global measures such as total reading time. They concluded that the parser does not consult verb-specific syntactic information, but that such information is used in later stages to revise a misanalysis. They also viewed the results as confirmation of the two-stage model of parsing, which assumed this basic architecture.

Following publication of Ferreira and Henderson (1990), a large number of studies were conducted designed to challenge these conclusions (Wilson & Garnsey, 2009; Trueswell & Kim, 1998). Although some findings consistent with theirs were also reported (Pickering & Traxler, 1998), the field eventually coalesced around the idea that verb information indeed informs initial parsing. Moreover, this idea was also taken as evidence against the original two-stage model, which is appropriate. However, in addition, the finding that verb information influences early parsing processes was also taken as evidence against modularity. But as our arguments thus far should make clear, we believe this conclusion is far too broad. One can easily imagine a modular theory of sentence processing in which the sources of information consulted to derive an initial parse include all the syntactic rules or principles relevant to projecting phrase structure, including verb subcategory information. In short, evidence for lexical guidance of early parsing decisions is not evidence against modularity, because the lexical information is plausibly internal to the syntactic module.

Next, let us consider the question how prosodic information might influence sentence processing. The starting point for most studies published in the topic is that syntactic and prosodic structures are related, and in particular, major syntactic boundaries such as those separating clauses are usually marked by phrase-final lengthening and changes in pitch (Ferreira, 1993). Some clause-internal phrasal boundaries are also marked, although much less reliably (Allbritton, McKoon, & Ratcliff, 1996)—for example, in the sentence *John hit the thief with the baseball bat*, the higher attachment of *with the baseball bat*, which supports the instrument interpretation, is sometimes (but not always) associated with lengthening of *thief*. The logic of the research enterprise was as follows: If certain

prosodic "cues" signal syntactic structure, then the parser might be able to use this information to avoid "going down the garden-path"—that is, it might be able to avoid misanalyzing the sentence structure. Of course, it is not obvious that the use of this information would constitute a violation of modularity, but that was the motivation for some of this research.

One of the earliest studies to consider this question was conducted by Beach (1991), and it claimed to show that prosodic information affects parsing. What the experiments actually demonstrated is that metalinguistic judgments about sentence structure were influenced by the availability of durational and pitch information linked to the final structures of the sentences. The obstacle to drawing any strong inferences concerning modularity at this stage in the history of the field was the unavailability of tasks for measuring online spoken language processing. The phoneme monitoring task had been abandoned in the 1980s (prematurely, as argued by Ferreira & Anes, 1994). The field still awaited the widespread use of electrophysiology to measure online processing of visual and auditory stimuli, and eyetracking had not yet been adapted to the investigation of spoken language. A couple of decades later, these techniques have yielded a wealth of information about the comprehension of utterances, and one of the ideas on which there is now a general consensus in the field is that prosody indeed influences the earliest stages of parsing. To take just one recent example, Nakamura, Arai, and Mazuka (2012) conducted an auditory study using temporarily ambiguous Japanese sentences and the visual world paradigm to investigate how contrastive intonation affected parsing decisions. Their results suggest that prosody can affect early stages of spoken sentence processing, leading comprehenders even to anticipate upcoming structure. Numerous other studies led researchers to similar conclusions (Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991; Kjelgaard & Speer, 1999; Millotte, Wales, & Christophe, 2007).

Now, how shall we evaluate these results and interpretations in light of the modularity hypothesis? If we conflate the two-stage model of parsing and the modularity hypothesis, then we must conclude that sentence processing is nonmodular. But we could instead update a model offered more than 25 years ago in light of this sort of evidence relating to prosody, as indeed the proponents of the two-stage model have (Carlson, Frazier, & Clifton, 2009; Frazier, Carlson, & Clifton, 2006). However, even if evidence is presented to refute specific models of modularity, this should not be taken as evidence against modularity as a whole, but only one potential form of modularity. Our argument is that, when considering modularity, it is important to establish not only what information sources are *internal* to the module, but also what information is *input* to that module. In the case of sentence processing, it seems reasonable to assume that prosodic cues or prosodic representations might be input to the sentence analyzer—that is, in terms of the more traditional bottom-up/top-down processing distinction, it seems plausible that prosodic analysis would take place *before* syntactic parsing. This idea makes some sense, as the flow of information during comprehension seems to be from sensory to conceptual, and prosodic features such as loudness,

duration, and pitch are more sensory/perceptual than information about syntactic categories. Thus, prosody may indeed influence the earliest stages of parsing, but this does not undermine modularity.

## THE USE OF CONTEXT AND PLAUSIBILITY INFORMATION DURING SENTENCE PROCESSING

Although investigations of verb subcategorization information and prosody are important for understanding the nature of sentence processing, it is not clear that they're useful for evaluating the modularity hypothesis, as we have argued. What is clearly relevant and indeed critical is information that certainly appears to be nonsyntactic. One of the earliest analyses came from Crain and Steedman (1985). They observed that many of the sentence forms treated as syntactically dispreferred by the two-stage model are also presuppositionally more complex. For example, consider the sentence *The evidence examined by the lawyers turned out to be unreliable*. According to the two-stage model, minimal attachment leads the parser to initially treat *examined* as a main verb, which causes the parser to be garden-pathed when the *by*-phrase is encountered. The parser must then reanalyze the structure as a reduced relative (see Fodor & Ferreira, 1998, for proposals concerning syntactic reanalysis). Similarly, the prepositional phrase attachment ambiguity in a sentence such as *John hit the thief with the stick* allows for two interpretations: initially, the *with*-phrase is interpreted as an instrument, but the *with*-phrase may instead serve as a modifier. As in the case of the reduced relative ambiguity, in this case too, the more complex syntactic analysis involves modification while the simpler analysis does not.

Crain and Steedman (1985) pointed out that these modification interpretations are not just syntactically more complex; they're presuppositionally more complex as well. Felicitous use of a complex phrase such as *the evidence examined by the lawyer* requires that there be more than one type of evidence in the discourse so that the modifier can be used to pick out the correct referent. This analysis appeals to the Gricean Maxim of Quantity (Grice, 1975), which states that speakers should not include unnecessary information in their utterances (but see Engelhardt, Bailey, & Ferreira, 2006). They argued further that null contexts favor the minimal attachment interpretation because, without a context specifying a set of objects denoted by the head noun, the listener will assume the presuppositionally simpler interpretation. Crain and Steedman presented intuitive evidence that sentences with reduced relative clauses were easy to process in proper contexts, contrary to what the two-stage model would predict.

The problem with the Crain and Steedman (1985) argument, of course, is that offline judgments are not adequate for assessing modularity, because they measure only the output of any putative module. Certainly a sentence such as *The evidence examined turned out to be unreliable* sounds better in context than by itself (as does almost any sentence), but that observation gives us no insight into the processes that support the intuition. For that reason, Ferreira and Clifton (1986) conducted an eyetracking study to assess whether the effect of context

was mainly to influence offline interpretations, or if it indeed intervened in the initial syntactic decisions of the parser. Their data were consistent with the idea that context did not affect initial parsing decisions. Supportive contexts led to shorter global reading times and more accurate question-answering behavior, but early measures of processing revealed that processing times for reduced relative and prepositional modification structures were longer than for their structurally simpler counterparts.

To the best of our knowledge, the findings from this 1986 study still hold. The only serious challenge came from Altmann and Steedman (1988), who elaborated on the Crain and Steedman (1985) proposal and also reported a set of self-paced reading experiments that purported to provide contrary results. This in turn led to a debate between Altmann and Steedman, on the one hand, and Clifton and Ferreira, on the other (1988). However, as Clifton and Ferreira argued, it is unclear that self-paced reading data can trump eyetracking results because the self-paced reading measure has far poorer temporal and spatial resolution, and therefore is biased against detecting early effects of syntactic manipulations.

More interesting than this debate about techniques, however, are the actual details of the Altmann and Steedman (1988) theoretical proposal. We believe the importance of the position they took in that paper has not been adequately appreciated in the 25 years since the paper's publication. Altmann and Steedman argued for a sentence comprehension system with two important properties. The first is that their parser consulted a syntactic database very different from the one assumed in the two-stage model. The important difference is that the representational format for structural information was Steedman's Combinatory Categorial Grammar, which combines syntactic and semantic information (and even some aspects of prosody and intonation; see Steedman, 2000; Steedman & Baldridge, 2011). Thus, if the parser consults a database of structural information contained in that sort of vocabulary, then effects of certain semantic manipulations on initial parsing are not inconsistent with modularity. This argument is the same as the one we made earlier regarding the use of verb subcategorization information: If the information is part of the module's proprietary database, then use of that information cannot constitute a violation of modularity.

But the second property is even more important: Altmann and Steedman (1988) argued for what they termed a weakly interactive architecture. What this architecture amounts to is a system in which "syntax proposes" and "semantics disposes." Crucially, on this model, alternative structural analyses are activated in parallel, and context retains the interpretation that is most contextually appropriate. This sort of mechanism is the same as the one that had been suggested in earlier work to explain the processing of lexical ambiguity (e.g., *bank*), and was specifically discussed in *TMOM* as an example of how a modular system might work. The idea is that, bottom-up, all alternatives are retrieved and made available to subsequent modules that then choose the one that is most suitable. In the case of lexical ambiguity, both meanings of *bank* are activated (and not necessarily equally strongly; modulation of activation according to frequency is also perfectly compatible with bottom-up processing), and the meaning that

fits the context is retained while the other meaning either decays or is inhibited by executive cognitive systems. Similarly, all syntactic structures might be computed or retrieved, and the one that post-sentence processing systems like are retained while the others either decay or are inhibited. The important point, then, is that this type of interaction with context does not violate modularity, as Altmann and Steedman themselves emphasized with their description of their model as merely "weakly interactive."

A related debate has centered around another potential influence on initial parsing decisions—semantic plausibility. Ferreira and Clifton (1986) not only looked at the effects of discourse context on parsing; they also focused on plausibility information linked to animacy. The critical contrasting cases are *the evidence examined* versus *the defendant examined*. With the animate noun *defendant*, the verb *examined* is naturally interpreted as the thing doing the examining; but with the inanimate noun *evidence*, the same syntactic analysis leads to an anomalous interpretation. Ferreira and Clifton reported that the animacy information did not block the garden-path, which led them to argue for a strongly modular architecture. This conclusion has been the target of numerous challenges (Altmann & Steedman, 1988; McClelland, 1987; MacDonald, Pearlmutter, & Seidenberg, 1994; MacDonald, 1993), and at this point, the consensus seems to be that animacy does indeed influence initial parsing (but see Clifton, Traxler, Mohamed, Williams, Morris, & Rayner, 2003). And, in turn, this view is taken to be evidence against modularity. Again, however, animacy is a very basic type of semantic information which some languages treat as a grammatical feature (Dahl & Fraurud, 1996). If the lexical entries for nouns include a simple +/– animacy feature, then it is not implausible to think that a modular parser might be able to access that information in a lexical entry and match it to a lexico-syntactic rule stating that the subject of an agentive verb such as *examine* must be animate. In addition, our arguments concerning the propose/dispose architecture also hold: If syntactic alternatives are constructed in parallel and then selected on the basis of plausibility, then what we have is what Altmann and Steedman (1988) called weak interaction, which is compatible with the modularity thesis. Once again we see that a result incompatible with the two-stage model of parsing (which assumes serial analysis plus reanalysis rather than a propose/dispose architecture) was taken as evidence against modularity itself.

## MODULARITY AND THE VISUAL WORLD PARADIGM

The early 1990s saw the creation of a new paradigm for studying sentence processing—the VWP. The idea behind the paradigm is simple: From reading studies, it was known that what the eyes fixate on and how much time is spent during a fixation are closely tied to attention and processing (Rayner, 1977). The VWP extends this logic to spoken language processing by pairing spoken utterances with simple visual displays containing mentioned and unmentioned objects. The "linking hypothesis" (Tanenhaus, Magnuson, Dahan, & Chambers, 2000) is that as a word is heard, its representation in memory becomes activated,

and this in turn automatically triggers eye movements toward the named object as well as objects semantically and even phonologically associated with it (Huettig & McQueen, 2007). The acceptance and widespread adoption of the task occurred because it lined up with several trends in cognitive science: First, there was an emerging emphasis on cognition and action—that is, on trying to capture how cognitive processes might be used to guide intelligent action and behavior. Second, the idea of multimodal processing was also catching on, with many cognitive scientists wanting to understand the way different cognitive systems might work together—in this case, the auditory language processing system and the visuo-attention system associated with object recognition (Henderson & Ferreira, 2004; Jackendoff, 1996). Third, there was growing interest in auditory language processing generally, and in the investigation of how prosodic information might be used during comprehension (Bear & Price, 1990). And, most relevant to one of the themes of this volume, there was dissatisfaction with the lack of experimental paradigms for empirically evaluating the modularity hypothesis. Reading techniques were of course useful and often quite powerful, but not all questions regarding language processing can be studied with reading (e.g., the use of overt prosody), and some researchers were bothered by the idea that reading is not as fundamental or primary a mode of language as is spoken language. Thus, the VWP was enthusiastically adopted. By now, hundreds of studies have been reported making use of it in one way or another (for summaries, see Huettig, Rommers, & Meyer, 2011; Huettig, Olivers, & Hartsuiker, 2011; Ferreira, Foucart, & Engelhardt, 2013).

The report that triggered the widespread use of the VWP and that is also viewed as having fatally undermined the idea of a modular sentence processing system is Tanenhaus et al. (1995), reported in more detail in Spivey, Tanenhaus, Eberhard, & Sedivy (2002). This study adapted the Altmann and Steedman (1988) ideas concerning presuppositional support to the domain of visual contexts and spoken sentences that could be evaluated against them. To illustrate the study, consider the imperative sentence *Put the apple on the towel in the box*. At the point at which the listener hears *on the towel*, two interpretations are possible: Either *on the towel* is the location to which the apple should be moved, or it is a modifier of *apple*. The phrase *into the box* forces the latter interpretation because it is unambiguously a location. Referential Theory specifies that speakers should provide modifiers only when modification is necessary to establish reference (e.g., we do not generally refer to a big car if only one car is discourse-relevant). From referential theory, it follows that if two apples are present in the visual world and one of them is supposed to be moved, then right from the earliest stages of processing, the phrase *on the towel* will be taken to be a modifier, because the modifier allows a unique apple to be picked out. The listener faced with this visual world containing two referents should therefore immediately interpret the phrase as a modifier and avoid being garden-pathed, and this is indeed what the data seem to show (Farmer, Cargill, Hindy, Dal, & Spivey, 2007; Novick, Thompson-Schill, & Trueswell, 2008; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002; Tanenhaus et al., 1995; Trueswell, Sekerina, Hill, & Logrip, 1999).

However, in recent work we have argued that the VWP is in many ways highly unsuited to the task of assessing modularity (Ferreira, Foucart, & Engelhardt, 2013). Of course, there are numerous other significant questions concerning sentence processing for researchers to ask, and for those questions, the VWP is quite useful (Huettig, Rommers, & Meyer, 2011). But recall once again the argument in *TMOM* that evaluating modularity requires an experimental approach that allows the measurement of online processing, and it should not encourage subjects to adopt atypical strategies for dealing with the experimental situation that might have little to do with normal sentence processing. Now consider how the original Tanenhaus et al. (1995) study was set up. Subjects were allowed to watch as an experimenter laid out a 2 × 2 arrangement of real objects to be manipulated in response to auditory instructions. Two quadrants contained the target and the distractor object and the other two quadrants contained two potential goal locations. Listeners then heard either a syntactically ambiguous or unambiguous instruction containing a prepositional phrase modifier. With this setup, the amount of time available to preview the visual context could be several seconds, and this time interval was not controlled. It seems likely that, during the preview period, listeners might start to generate fairly specific expectations about the form and content of the upcoming utterance, especially since all the utterances consisted of a transitive verb followed by a noun phrase and at least one prepositional phrase. After experience with some trials, the participant may form a template or underspecified form of the upcoming utterance. Thus, both the visual display and the sentences conform to predictable patterns, which participants can learn after a small number of trials (Fine & Jaeger, 2013).

To address these concerns about the suitability of the VWP for evaluating modularity in language processing, we conducted three experiments examining the effects of depriving subjects of a preview of the visual world, and we conducted a production experiment to determine how accurately naïve participants could guess the sentence likely to occur with a particular visual display (Ferreira et al., 2013). We found that participants were not garden-pathed in any condition when they were denied preview of the visual world prior to hearing the sentences, and we also reported that participants were surprisingly good at anticipating which object they would be asked to move and which objects would serve as potential locations. From these results we concluded that listeners engage in a fairly atypical mode of processing in VWP experiments with visual world previews and utterances that are highly similar to each other over all experimental trials: rather than processing utterances incrementally, they instead form an underspecified representation of what they are likely to hear next based on the content of the visual world. They then evaluate that prediction against the utterance itself. Now, it is certainly possible that humans sometimes process language in this way, but most people would agree that typical processing situations are quite a bit more open-ended.

For these reasons, then, we are not convinced that the VWP can provide strong evidence against modularity. Again, the technique is superb for getting at many important questions about how language is processed, but it is not clear

that it is suited for determining to what extent sentence processing is characterized by information encapsulation or domain-specificity.

## MODULARITY AND SHALLOW PROCESSING

In the last fifteen years or so, a new framework for thinking about sentence comprehension has emerged. There are many variants with important distinctions among them, but what they share is the idea that comprehenders sometimes end up with an interpretation that differs from the actual input received—the interpretation is either simpler (construal), somewhat distorted (late assignment of syntax theory; good-enough processing), or outright inconsistent (noisy channel approaches) with the sentence's true content. These models have been difficult to pigeon-hole with respect to the modularity thesis. To try to sort out this issue, we feel it is important to shift the emphasis away from the features of modularity having to do with information encapsulation and toward the features that emphasize shallow outputs and limited central access to the internal operations of a module. Typically, psycholinguists have assumed that the output of any parsing or sentence processing module is a syntactic representation, which is turned over to "central" systems that relate to knowledge and belief. But we could assume instead that the output of the module is an interpretation, with structure-building operations being used to create it. If we adopt these assumptions, then we might not be surprised to discover that people can end up with interpretations that are simpler than the input would seem to mandate, and that might even be nonveridical.

To see how this argument works, let's begin with the mildest form of these models—the ones that assume representations that reduce the input in some way. One implementation is to allow representations to be underspecified (Sanford & Sturt, 2002). Consider construal (Frazier & Clifton Jr, 1997): A major assumption of the construal model is that syntactic structures are not always fully connected—adjunct phrases in particular (e.g., relative clauses) may instead simply get associated with a certain processing domain, "floating" until disambiguating information arrives. The parser thus remains uncommitted (Pickering, McElree, Frisson, Chen, & Traxler, 2006; Traxler, Pickering, & Clifton, 1998) concerning the attachment of the relative clause and the interpretation of the noun phrase and sentence that would follow from any particular attachment (see Frisson & Pickering, 2001; Sanford & Graesser, 2006; Sturt, Sanford, Stewart, & Dawydiak, 2004; Frisson S., 2009 for evidence favoring underspecified representations). A more radical possibility is that the attachment decision is strategically postponed, which is what the good enough language processing (henceforth, GE) theory predicts. Swets, Desmet, Clifton, Ferreira (2008) tested this idea by presenting participants with either fully ambiguous sentences (*the maid of the princess who scratched herself was embarrassed*) or disambiguated controls (*the son of the princess who scratched himself/herself was embarrassed*). The twist they introduced was to manipulate whether participants were required to answer easy or difficult comprehension questions following each sentence.

The rationale was that, with easy questions, readers would not be motivated to resolve the ambiguity; with no interpretive consequences, they would be happy to leave the relative clause unattached. In contrast, with challenging questions, subjects would know they were being "called out" on their understanding of the sentences, and therefore attachment decisions were incentivized. The findings supported these predictions: they found a reading time advantage for sentences with ambiguous relative clauses relative to disambiguated controls when they were followed by easy questions, suggesting that they were easier to process due to the omission of the attachment operation. In contrast, when readers expected to receive questions probing their interpretation of the relative clause, critical regions of the sentences were read more carefully, and the ambiguity advantage was reduced. Other studies support the idea of underspecified representations for global syntactic structures (Tyler & Warren, 1987), semantic information (Frazier & Rayner, 1990), and coercion structures (Pickering, McElree, Frisson, Chen, & Traxler, 2006).

Another line of work explores psycholinguistic analogues of the so-called Moses illusion. The now-famous Moses illusion involves asking people a question such as *How many animals of each sort did Moses take on the ark*. Amusingly, most people answer "two" instead of pointing out that the presupposition behind the question is incorrect (Erickson & Mattson, 1981). The illusion is presumed to occur because Moses and Noah share a large number of semantic features, and semantic processing is often too shallow to allow the distinguishing features to be activated and integrated (see also Barton & Sanford, 1993). Sanford and Sturt (2002) suggest that shallow processing is linked to the focus-presupposition structure of a sentence: elements that are in semantic focus are processed deeply, but those that are assumed or backgrounded are processed more shallowly, leading to these kinds of semantic illusions. This proposal is reminiscent of one offered by Cutler and Fodor (1979), who found in phoneme monitoring studies that phonemes in words which are part of the focus of a sentence are detected more quickly than those that are in words located in the presupposed portion.

More radical variants of shallow processing models are those that allow the comprehension system to generate an interpretation that is even more discrepant from the input. Researchers in the field of text processing and cross-sentence integration have shown that readers are sometimes remarkably insensitive to contradictions in text (Otero & Kintsch, 1992), and also often fail to update their interpretations when later information undermines a fact stated earlier—for example, a character described initially as guilty of a crime but described later as exonerated remains tainted by the original charge in people's memory representations for the story (Albrecht & O'Brien, 1993). These ideas from text processing were exported to the sentence processing literature in a series of experiments showing that people did not seem to fully recover from garden-paths (Christianson, Hollingworth, Halliwell, & Ferreira, 2001). Participants were asked to read sentences such as *While the woman bathed the baby played in the crib* and then they answered a question such as *Did the woman bathe the baby?*. The surprising finding was that most people answered "yes," even though

the meaning of the reflexive verb *bathe* requires that the object be interpreted as coreferential with the subject in an intransitive structure (see also Slattery et al.; Ferreira, 2013). It appears that comprehenders are not entirely up to the task of syntactic reanalysis, and sometimes fail to revise either all pieces of the syntactic structure or all elements of the semantic consequences of the initial, incorrect parse. In addition, the more semantically compelling the original, garden-path interpretation, the more likely people are to want to retain it rather than revise it to the one consistent with the global grammatical form.

Townsend and Bever (2001) offered up a model of sentence comprehension very different from either the traditional two-stage model or the connectionist models of sentence processing that had become popular in the 1990s. The Townsend and Bever model implements an architecture similar to what has been suggested for decision-making (Gigerenzer, 2004; Kahneman, 2003), which distinguishes between a so-called System 1 and System 2 (or Type 1 and Type 2) for reasoning. System 1 is fast, automatic, and operates via the application of simple heuristics—"quick and dirty" rules that usually deliver a reasonably good result. System 2, on the other hand, is slow, attention-demanding, and that is able to consult a wide range of beliefs—essentially anything the organism knows and has stored in memory. Notice how closely this architecture echoes the one suggested in *TMOM*, where System 1 would map on to modular systems and System 2 would map on to the central reasoning system. Of course, one important difference is that Fodorian modules are assumed to be computational—for example, the modular parser consults a detailed, complex syntactic database when building an interpretation, rather than relying on a small set of simple heuristics. Nonetheless, the points of overlap are intriguing.

In Townsend and Bever's (2001) model, which they refer to as LAST (late assignment of syntax theory), sentences are essentially processed twice: first, heuristics are accessed which yield a quick and dirty meaning, and then syntactic computations are performed on the same word string to yield a fully connected, syntactic analysis. The second process ensures that the meaning that is obtained for a sentence is consistent with its actual form. Townsend and Bever also assume that the first stage is nonmodular and the second modular; this is to account for the use of semantics in the first stage, and the use of essentially only syntactic constraints in the second. However, this type of two-stage model can be construed in such a way that the first stage is modular, as long as the heuristics are essentially "reflexes"—as long as they are simple syntactic tricks that are blindly applied to the input without the benefit of consultation with other sources of knowledge. Two models similar in spirit to LAST but which assume a modular architecture for the first stage are the one offered by Ferreira (2003) and Garrett (2000). The Ferreira model assumes that the first stage consults just a couple of heuristics—a version of the "NVN" strategy, in which people assume an agent-patient mapping of semantic roles to syntactic positions, and an animacy heuristic, in which animate entities are biased toward subjecthood. The 2003 Ferreira model captures the results of a series of experiments in which participants appeared to frequently misinterpret passive sentences, particularly when they expressed an implausible

event with reversible semantic roles (e.g., the dog was bitten by the man = the dog bit the man). The application of heuristics in the first stage yields the dog-bit-man interpretation; a proper syntactic parse will deliver the opposite, correct interpretation, but the 2003 model assumes that it is fragile and susceptible to interference from the more frequent interpretation. Garrett (2000) offers a more explicitly analysis-by-synthesis model which incorporates the production system to yield what are widely believed to be top-down effects. A first pass, bottom-up process uses basic syntactic information to yield a simple parse which in turn allows for a rudimentary interpretation; then the language production system takes over and uses that representation to generate the detailed syntactic structure that would support the initial parse and interpretation.

Finally, a family of models has been proposed that assume people engage in rational behavior over what they understand to be a noisy communication channel. The channel is noisy both because listeners sometimes mishear or misread due to processing error or environmental contamination, and because speakers sometimes make mistakes when they talk. Thus, a rational comprehender whose goal is to recover the intention behind the utterance will normalize the input according to Bayesian priors. A body of evidence from research using event-related potentials (ERPs) helped to motivate these ideas (Van Herten, Kolk, & Chwilla, 2005; Kim & Osterhout, 2005). In these experiments, it is reported that subjects who encounter a sentence such as *The fox that hunted the poachers stalked through the woods* experience a P600 rather than an N400 upon encountering the semantically anomalous word, even though an N400 would be expected given that it is presumed to reflect problems with semantic integration. There is still not a great deal of consensus on what triggers P600s, but an idea that has been gaining traction is that it reflects a need to engage in some type of structural reanalysis or revision. The idea, then, is that when a person encounters a sentence that seems to say that the fox hunted the poachers, they "fix" it so it makes sense, resulting in a P600. Other models have taken this idea and developed it further (Gibson, Bergen, & Piantadosi, 2013; Levy, 2011; Levy, Bicknell, Slattery, & Rayner, 2009). These models seem less compatible with modularity than the other "shallow processing" approaches discussed earlier, because the information that is accessed to establish the priors can potentially be anything, ranging from biases related to structural forms all the way to beliefs concerning speaker characteristics (e.g., that a person with an upper-class speech style is unlikely to refer to his tattoo; Van Berkum, van den Brink, Tesink, Kos, & Hagoort, 2008). However, these noisy channel models have not yet been rigorously tested using a methodology that allows early processes to be distinguished from later ones. For example, it remains possible that comprehenders create a simple quick-and-dirty parse in a manner compatible with modularity and then consult information outside the module to revise that interpretation, right down to actually normalizing the input. Indeed, models designed to explain the comprehension of sentences containing self-repairs (*turn left uh right at the light*) assume mechanisms that allow input to be deleted so that the speaker's intended meaning can be recovered in the face of disfluency (Ferreira, Lau, & Bailey, 2004).

## CONCLUSION

We began this chapter on the modularity of sentence processing with a summary of the main features of modules, because it is essential to appreciate that modularity is about more than information encapsulation—other key features include speed, automaticity, shallow outputs, and limited central access. If information encapsulation is treated not as simply one of a cluster of features but rather as "the heart of modularity," then the challenges to the notion that sentence processing is modular will continue to resonate in the cognitive science community, despite the arguments we've made here that many studies purporting to show interactivity can be reconciled with modularity. The key, we argued, is to appreciate two points. First, the so-called "two-stage" model associated with Frazier and colleagues (including the first author) is only one kind of modular model for sentence processing, so evidence against the two-stage model is not evidence against every instantiation of a modular model. And second, whether an influence of some piece of information constitutes a violation of information encapsulation depends critically on what information is contained in the "capsule." If we assume the sentence processing module can consult phrase structure rules only, then effects of even information such as verb subcategorization frames will be construed as disconfirming encapsulation. But if we accept that one of the aims of theory construction in the field of sentence processing is to develop an explanatory model of how the system works, then one key goal will be to determine what sources of information are in fact part of the sentence processing module. The goal would then be to determine what the proprietary databases are that the sentence processing module must consult. Certainly almost everyone would agree that information about what speakers from different social classes are likely to say probably does not belong in a parsing module, but information about verb subcategorization and even animacy are a different matter entirely. Moreover, the assumption of seriality relating to ambiguity resolution should be open to empirical scrutiny and revision as well; as we argued, a system with parallel consideration of alternative parses is compatible with modularity, and indeed mimics the architecture proposed as a bottom-up account of how lexical ambiguity is processed.

We would like to offer a further suggestion, and that is to emphasize the modularity features that cluster around shallowness rather than those that focus on encapsulation. We could assume that the output of the sentence processing module is not a parse in the sense of a detailed syntactic structure, but is rather the conditions for interpretation—a representation that includes information about thematic roles, focus-presupposition structure, and so on, but does not retain highly articulated syntactic forms or traces of movement operations. Complex, detailed syntax might get accessed and used by the module that creates an interpretation, but those detailed syntactic representations also are likely discarded once they serve their role of allowing a propositional interpretation to be built (Sachs, 1967). In addition, the module would be able to consult simple frequency based heuristics such as the NVN strategy (Townsend & Bever, 2001). And if the heuristics deliver a compelling interpretation faster than the

syntactic algorithms do (as in some cases of garden-path reanalysis, which can be time-consuming and often require accessing infrequent forms; MacDonald, Pearlmutter, & Seidenberg, 1994), then the systems subsequent to the sentence processing module may decide to proceed with what they have rather than waiting for more detailed analyses to be performed. These tendencies would result in phenomena such as the Moses illusion, garden-path misinterpretations, and misinterpretations of implausible passives. Moreover, if that interpretation still seems unsatisfactory in a Bayesian sense, then post-sentence processing modules may engage in the sort of normalization and correction that would be expected on a rational view of communication.

We end by returning to our opening observation: Modularity might be out of fashion, but this is not because the evidence against it is particularly compelling. Instead, we suspect that many researchers simply grew weary of the limited set of questions that were being asked in the context of testing modularity against one specific model of sentence processing, and so they decided to shift their energies to broader questions such as dialogue, embodiment, and language-vision interactions. This shift in focus has been positive for the field because so much more is known now than even ten years ago. However, much of what we've learned is not relevant to evaluating modularity, and these new approaches and findings are quite possibly compatible with it.

## REFERENCES

Albrecht, J. E., & O'Brien, E. J. (1993). Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(5), 1061–1070.

Allbritton, D. W., McKoon, G., & Ratcliff, R. (1996). Reliability of prosodic cues for resolving syntactic ambiguity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(3), 714–735.

Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, *30*(3), 191–238.

Barton, S. B., & Sanford, A. J. (1993). A case study of anomaly detection: Shallow semantic processing and cohesion establishment. *Memory & Cognition*, *21*(4), 477–487.

Beach, C. M. (1991). The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. *Journal of Memory and Language*, *30*(6), 644–663.

Bear, J., & Price, P. (1990). Prosody, syntax and parsing. *Proceedings of the 28th annual meeting on Association for Computational Linguistics* (pp. 17–22). Pittsburgh, PA: Association for Computational Linguistics.

Bever, T. G., Sanz, M., & Townsend, D. J. (1998). The emperor's psycholinguistics. *Journal of Psycholinguistic Research*, *27*(2), 261–284.

Caramazza, A., & Zurif, E. B. (1976). Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *Brain and Language*, *3*(4), 572–582.

Carlson, K., Frazier, L., & Clifton, C. (2009). How prosody constrains comprehension: A limited effect of prosodic packaging. *Lingua*, *119*(7), 1066–1082.

Chomsky, N. (1965). *Aspects of the theory of syntax* (Vol. 11). Cambridge, MA: The MIT Press.

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, *42*(4), 368–407.

Clifton, C., Traxler, M. J., Mohamed, M. T., Williams, R. S., Morris, R. K., & Rayner, K. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language*, *49*(3), 317–334.

Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological parser. In D. Dowty, L. Karttunen, & A. Zwicky, *Natural language parsing: Psychological, computational, and theoretical perspectives* (pp. 320–358). Cambridge, UK: Cambridge University Press.

Cutler, A., & Fodor, J. A. (1979). Semantic focus and sentence comprehension. *Cognition*, *7*(1), 49–59.

Dahl, Ö., & Fraurud, K. (1996). Animacy in grammar and discourse. In T. Fretheim & J. Gundel, *Reference and referent accessibility* (pp. 47–64). Amsterdam: John Benjamins.

Dick, F., Bates, E., Wulfeck, B., Utman, J. A., Dronkers, N., & Gernsbacher, M. A. (2001). Language deficits, localization, and grammar: Evidence for a distributive model of language breakdown in aphasic patients and neurologically intact individuals. *Psychological Review*, *108*(4), 759–788.

Engelhardt, P. E., Bailey, K. G., & Ferreira, F. (2006). Do speakers and listeners observe the Gricean Maxim of Quantity? *Journal of Memory and Language*, *54*(4), 554–573.

Erickson, T. D., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior*, *20*(5), 540–551.

Farmer, T. A., Cargill, S. A., Hindy, N. C., Dale, R., & Spivey, M. J. (2007). Tracking the continuity of language comprehension: Computer mouse trajectories suggest parallel syntactic processing. *Cognitive Science*, *31*(5), 889–909.

Fedorenko, E., Duncan, J., & Kanwisher, N. (2012). Language-selective and domain-general regions lie side by side within Broca's area. *Current Biology*, *22*(21), 2059–2062.

Ferreira, F. (1993). Creation of prosody during sentence production. *Psychological Review*, *100*(2), 233–253.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive psychology*, *47*(2), 14–203.

Ferreira, F., & Anes, M. (1994). Why study spoken language? In M. A. Gernsbacher, *Handbook of Psycholinguistics* (pp. 33–56). San Diego, CA: Academic Press.

Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*(1), 11–15.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, *25*(3), 348–368.

Ferreira, F., Foucart, A., & Engelhardt, P. E. (2013). Language processing in the visual world: Effects of preview, visual complexity, and prediction. *Journal of Memory and Language*, *69*(3), 165–182.

Ferreira, F., & Henderson, J. M. (1990). Use of verb information in syntactic parsing: Evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(4), 555–568.

Ferreira, F., Lau, E. F., & Bailey, K. G. (2004). Disfluencies, language comprehension, and tree adjoining grammars. *Cognitive Science*, *28*(5), 721–749.

Field, J. (2004). An insight into listeners' problems: Too much bottom-up or too much top-down? *System*, *32*(3), 363–377.

Fine, A. B., & Jaeger, T. F. (2013). Evidence for implicit learning in syntactic comprehension. *Cognitive Science*, *37*, 578–591.

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

Fodor, J. A. (2000). *The mind doesn't work that way: The scope and limits of computational psychology.* Cambridge, MA: MIT press.

Fodor, J. A., Bever, T. G., & Garrett, M. (1969). *The development of psychological models for speech recognition*. Report ESD-TR-67-633 of the Electronic Systems Division, US Air Force. Bedford, MA: Hanscom Field.

Fodor, J. D., & Ferreira, F. (1998). *Reanalysis in sentence processing.* Dordrecht, The Netherlands: Kluwer.

Ford, M., Bresnan, J., & Kaplan, R. (1982). A competence-based theory of syntactic closure. In J. Bresnan, *The mental representation of grammatical relations* (pp. 727–796). Cambridge, MA: MIT Press.

Frazier, L., & Clifton Jr., C. (1997). Construal: Overview, motivation, and some new evidence. *Journal of Psycholinguistic Research, 26*(3), 277–295.

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

Frazier, L., Carlson, K., & Clifton, C. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences, 10*(6), 244–249.

Frazier, L., Pacht, J. M., & Rayner, K. (1999). Taking on semantic commitments, II: Collective versus distributive readings. *Cognition, 70*(1), 87–104.

Frazier, L., & Rayner, K. (1990). Taking on semantic commitments: Processing multiple meanings vs. multiple senses. *Journal of Memory and Language, 29*(2), 181–200.

Frisson, S. (2009). Semantic underspecification in language processing. *Language and Linguistics Compass, 3*(1), 111–127.

Frisson, S., & Pickering, M. J. (2001). Obtaining a figurative interpretation of a word: Support for underspecification. *Metaphor and Symbol, 16*(3-4), 149–171.

Gahl, S., & Garnsey, S. M. (2006). Knowledge of grammar includes knowledge of syntactic probabilities. *Language, 82*(2), 405–410.

Garrett, M. (2000). Remarks on the architecture of language processing systems. In Y. Grodzinsky, & L. Shapiro, *Language and the Brain: Representation and Processing* (pp. 31–69). San Diego, CA: Academic Press.

Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, *110*(20), 8051–8056.

Gigerenzer, G. (2004). Fast and frugal heuristics: The tools of bounded rationality. *Blackwell Handbook of Judgment and Decision Making* (62–88). Oxford, England: Blackwell.

Gleitman, L. R., & Wanner, E. (1982). *Language acquisition: The state of the art.* Cambridge, England: Cambridge University Press.

Grice, P. (1975). Logic and conversation. In P. Cole & J. Morgan, *Syntax and semantic: Speech acts* (Vol. 3, pp. 41–58). New York: Seminar Press.

Hare, M., McRae, K., & Elman, J. L. (2003). Sense and structure: Meaning as a determinant of verb subcategorization preferences. *Journal of Memory and Language, 48*(2), 281–303.

Henderson, J. M., & Ferreira, F. (2004). Scene perception for psycholinguists. In J. M. Henderson & F. Ferreira, *The interface of language, vision, and action: Eye movements and the visual world* (pp. 1–58). New York, NY: Psychology Press.

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review, 26*(1), 55–88.

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language, 57*(4), 460–482.

Huettig, F., Olivers, C. N., & Hartsuiker, R. J. (2011). Looking, language, and memory: Bridging research from the visual world and visual search paradigms. *Acta Psychologica, 137*(2), 138–150.

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica, 137*(2), 151–171.

Jackendoff, R. (1996). The architecture of the linguistic-spatial interface. In P. Bloom, M. A. Peterson, L. Nadel, & M. F. Garrett (Eds.), *Language and space. Language, speech, and communication* (pp. 1–30). Cambridge, MA: MIT Press.

Joanisse, M. F., Manis, F. R., Keating, P., & Seidenberg, M. S. (2000). Language deficits in dyslexic children: Speech perception, phonology, and morphology. *Journal of Experimental Child Psychology, 77*(1), 30–60.

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *The American Economic Review, 93*(5), 1449–1475.

Kahneman, D. (2011). *Thinking, fast and slow.* New York: Farrar, Straus & Giroux.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language, 52*(2), 205–225.

Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language, 40*(2), 153–194.

Levy, R. (2008). A noisy-channel model of rational human sentence comprehension under uncertain input. In *Proceedings of the conference on empirical methods in natural language processing* (pp. 234–243). Pittsburgh, PA: Association for Computational Linguistics.

Levy, R. (2011). Probabilistic Linguistic Expectations, Uncertain Input, and Implications. *Studies of Psychology and Behavior, 9*(1), 52–63.

Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences*, *106* (50), 21086–21090.

MacDonald, M. C. (1993). The interaction of lexical and syntactic ambiguity. *Journal of Memory and Language, 32*(5), 692–715.

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

McClelland, J. L. (1987). The case for interactionism in language processing. In M. Coltheart, *Attention and performance XII: The psychology of reading* (pp. 1–36). London, England: Erlbaum.

Millotte, S., Wales, R., & Christophe, A. (2007). Phrasal prosody disambiguates syntax. *Language and Cognitive Processes, 22*(6), 898–909.

Mitchell, D. C., & Holmes, V. M. (1985). The role of specific information about the verb in parsing sentences with local structural ambiguity. *Journal of Memory and Language, 24*(5), 542–559.

Nakamura, C., Arai, M., & Mazuka, R. (2012). Immediate use of prosody and context in predicting a syntactic structure. *Cognition, 125*(2), 317–325.

Novick, J. M., Thompson-Schill, S. L., & Trueswell, J. C. (2008). Putting lexical constraints in context into the visual-world paradigm. *Cognition, 107*(3), 850–903.

Otero, J., & Kintsch, W. (1992). Failures to detect contradictions in a text: What readers believe versus what they read. *Psychological Science, 3*(4), 229–235.

Pickering, M. J., McElree, B., Frisson, S., Chen, L., & Traxler, M. J. (2006). Underspecification and aspectual coercion. *Discourse Processes, 42*(2), 131–155.

Pickering, M. J., & Traxler, M. J. (1998). Plausibility and recovery from garden paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(4), 940–961.

Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *The Journal of the Acoustical Society of America, 90*, 2956–2970.

Rayner, K. (1977). Visual attention in reading: Eye movements reflect cognitive processes. *Memory & Cognition, 5*(4), 443–448.

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior, 22*(3), 358–374.

Sachs, J. S. (1967). Recopition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics, 2*(9), 437–442.

Sanford, A. J., & Graesser, A. C. (2006). Shallow processing and underspecification. *Discourse Processes, 42*(2), 99–108.

Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognitive Sciences, 6*(9), 382–386.

Sitnikova, T., Salisbury, D. F., Kuperberg, G., & Holcomb, P. J. (2002). Electrophysiological insights into language processing in schizophrenia. *Psychophysiology, 39*(6), 851–860.

Slattery, T. J., Sturt, P., Christianson, K., Yoshida, M., & Ferreira, F. (2013). Lingering misinterpretations of garden path sentences arise from competing syntactic representations. *Journal of Memory and Language, 69*(2), 104–120.

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology, 45*(4), 447–481.

Steedman, M. (2000). Information structure and the syntax-phonology interface. *Linguistic Inquiry, 31*(4), 649–689.

Steedman, M., & Baldridge, J. (2011). Combinatory categorial grammar. In R. Borsely & K. Borjars, *Non-Transformational Syntax* (pp. 181–224). Oxford, England: Blackwell.

Sturt, P., Sanford, A. J., Stewart, A., & Dawydiak, E. (2004). Linguistic focus and good-enough representations: An application of the change-detection paradigm. *Psychonomic Bulletin & Review, 11*(5), 882–888.

Swets, B., Desmet, T., Clifton, C., & Ferreira, F. (2008). Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory & Cognition, 36*(1), 201–216.

Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. (2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research, 29*(6), 557–580.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*(5217), 1632–1634.

Townsend, D. J., & Bever, T. G. (2001). *Sentence comprehension: The integration of habits and rules* (Vol. 1950). Cambridge, MA: MIT Press.

Traxler, M. J., Pickering, M. J., & Clifton, C. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language, 39*(4), 558–592.

Trueswell, J. C., & Kim, A. E. (1998). How to prune a garden path by nipping it in the bud: Fast priming of verb argument structure. *Journal of Memory and Language, 39*(1), 102–123.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition, 73*(2), 89–134.

Tyler, L. K., & Warren, P. (1987). Local and global structure in spoken language comprehension. *Journal of Memory and Language, 26*(6), 638–657.

Van Berkum, J., van den Brink, D., Tesink, C., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience, 20*(4), 580–591.

Van Herten, M., Kolk, H. H., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research, 22*(2), 241–255.

Wilson, M. P., & Garnsey, S. M. (2009). Making simple sentences hard: Verb bias effects in simple direct object sentences. *Journal of Memory and Language, 60*(3), 368–392.

Zekveld, A. A., Heslenfeld, D. J., Festen, J. M., & Schoonhoven, R. (2006). Top–down and bottom–up processes in speech comprehension. *Neuroimage, 32*(4), 1826–1836.

4

# The Unity of Consciousness and the Consciousness of Unity

THOMAS G. BEVER

> *Truth is stranger than fiction. . . . because fiction is obliged to stick to the possibilities.*
>
> *Truth isn't*
>
> —Mark Twain, *"Following the Equator"*

## A SENTENCE IS LIKE A (MINIATURE) OPERA

Music is often analyzed in relation to language to give perspective on the structural and formal aspects of language. But even the simplest sentence surpasses what music can tell us about it. A sentence in everyday use combines a stream of sound, with rhythm and pitch variations, with memorized units of meaning, an organizing structure that recombines those meaning units into a transcendental unified meaning that includes informational representations, general connotations, and specific pragmatic implications unique to the conversational context.

In other words, each sentence is a miniature opera of nature.

Children grow up surrounded by one opera after another, and miraculously learn to create their own. This is achieved in the context of experiencing only a small number of fully grammatical sentences, many ungrammatical ones, and very little specific feedback on their mistakes. This situation is generally referred to as "the poverty of the stimulus," which is the basis for the argument that much of linguistic structure must be innately prefigured (Chomsky, 1959,1965,1975,1980). Fodor (1981) broadened the implications of this argument beyond language into cognition in general: "The [argument from the poverty of the stimulus] is *the* existence proof for the possibility of a cognitive science"

(p. 258). Nonetheless, the flagship case of the argument continues to be the speed with which children learn language in erratic environments with variable feedback.

In this chapter, I begin with one of the major components of what the child has to discover in learning to understand and then produce language—the perception and comprehension of natural units of composition in the serial string. Interestingly, this problem exists on virtually any grammatical theory, from taxonomic phrase structure all the way up to today's Minimalism. Every view of what language is, going back centuries has some notion of serial hierarchical phrasing as a fundamental component. In phrase structure grammars, describing the phrase is the direct goal and what the child must discover; in generative theories that utilize "structure dependence" the child must discover the phrase in order to have access to the structure. In the next sections, I trace some research on how major language units are perceived, over the past decades, and then turn to the implications of recent studies of the acoustics of normal conversation, which show how deep and puzzling the problem of the poverty of the stimulus really is. The processing of normal conversation reveals a disconnect between the listener's representation of the sound and meaning of utterances. In critical cases it is possible to show that compressed or absent words are unintelligible until the listener hears *later* acoustic information Yet the listener perceives the acoustic presentation of the words as simultaneous with the comprehension of it. This is an instance of *creating* a conscious representation retrospectively.

I draw a number of morals from such facts in language processing: notably, the "poverty of the stimulus problem" is far graver than usually supposed—although the words in some child-directed speech are carefully pronounced, many are not. And children are also surrounded by the same kind of garbled and cue-poor instances from adult speech; this means that structure dependence must guide ongoing comprehension processes of externalized serial input, not used only to decide about the abstract structure of one's language during learning; every level of language experience involves some encoding: this supports the notion that ongoing processing occurs in a set of simultaneous parallel processes in a "computational fractal," that is, each level involves the interaction of associative-serial and structure dependent processes; thus, our conscious experience of language is in part reconstructive in temporarily time-free "psychological moments"—so language comprehension processes move forward *and backward*, even though the phenomenal experience is that it moves only forward.; this reconstructive analysis of our conscious experience of language may be typical of other modalities of our experience.

This leads us to distinguish the computational problem of language acquisition from the acoustic input problem. The computational problem concerns how children generalize in the right way from scant examples of complete, well-formed sentences with clearly presented words, how they alight on the right kind of structure dependent hypotheses. The acoustic input problem is that children (and adults) are often not presented with clear word-by-word inputs to learn and understand from. Rather children must have already solved a large part of the

computational problem in order to resolve the acoustic input problem. This magnifies what we must assume is available to the child at a very young age, and geometrically complicates any attempts to model acquisition with statistical models unadorned by massive and keen prior structures and expectations.

## WHERE IS THE UNIT OF LANGUAGE PROCESSING?

Psychology as a field often depends on resurgent methodology and continually mysterious phenomena: One of the most enduring methods and mysteries is the systematic mislocation of "clicks" presented during auditory presentation of sentences toward "phrasal" boundaries of some kind. The use of click-mislocation was pioneered by Ladefoged and Broadbent (1960), as a way of showing on-line segmentation of syllables. Its utility for exploring on-line complexity and the effect of "phrase" boundaries was initially explored by Garrett (1964). Fodor and Bever (1965) demonstrated the general role of relative depth of surface "phrase" breaks in determining the likelihood of click mislocation to or toward them; Garrett, Bever, and Fodor (1966) showed that the mislocation was not due to local intonational cues, but to the "phrasal" structure that listeners impose on what they are hearing (also demonstrated by Abrams and Bever, 1969, with a different technique). (For a contemporary demonstration of brain spectral activity corresponding to phrase construction without benefit of intonational or statistical cues, see Ding et al., 2016 and further discussion in this chapter).

A revealing aspect relevant for today's discussions is the fact that the citation of the two original click location essays has experienced a "U shaped function" with almost as many citations in the last five years as in the first five, and less than a third of that rate in the intervening years. This reflects the rediscovery of questions about what the "real" unit of ongoing language processing is.

Later studies attempted further to define what perceptual and comprehension units are revealed by click mislocations, "deep" structure units (Bever, Lackner, & Kirk, 1969) or "major" surface phrases (Chapin, Smith, & Abrahamson, 1972). Many click location studies required subjects to write out the sentence and indicate the click location—this invited the interpretation that the click mislocation effect was not perceptual at all, but some form of response strategy related to recapitulating the sentence. Bever (1973) explored this by having listeners mark the click location within a window in the text written out and presented right after hearing the stimulus. In critical cases there was no auditory click at all: to make it plausible that there was a click, the loudness of the actual clicks was varied. When a click was present, the usual effect of a major phrase boundary occurred: when there was no click, subjects' guesses were *not* systematically placed in the major phrase boundary. Using a different method, Dalrymple-Alford (1976) confirmed that click mislocation is not due to a response bias.

Two approaches to the question of the processing unit have continually surfaced and resurfaced over many years: each rests on one of the two ideas dominant in centuries of cognitive science: (a) the currency of mental life is statistically

determined associations; (b) mental life is organized into categorical representations. The argument started with a closer examination of the 'psychological reality of linguistic segments namely the "phrase." During the 1960s much attention was being given to the idea that "phrases" could be defined in terms of serial predictability (Johnson,1970; Osgood,1968). On this view, "phrases" are behaviorally bounded by relatively low points of serial predictability: indeed it is generally the case that phrase-final (content) words are more predictable locally than phrase initial words. So behaviors that seem to reflect categorical phrasing might actually be reflecting variation in serial predictability. However, when syntactic structures are held constant while local predictability is varied, the high serial predictable points actually attract clicks perceptually (Bever et al., 1969). So probability governed segmentation does not account for the online perceptual formation of phrases.

Yet the conflict between some version of association and categorical structural assignment always finds new life. The connectionist postulation of hidden units, back propagation and other statistical devices, along with the rehabilitation of Bayesian statistics, resuscitated notions of mediated associations with complex descriptive power, enabling simulation of categorical structures as their limit (e.g., Rumelhart & McClelland, 1986). In this vein, great attention is given to "feed forward" models of perception in general and sentence processing in particular: the perceptual system is constantly making predictions of what is about to occur, so that much of language interpretation is actually rolling confirmation of specific kinds of rolling perceptual expectations. In the case of language, this can occur simultaneously at various levels from the acoustic to the semantic. The expectations are arguably a blend of probabilistic and categorical features in many domains; phonological, semantic, and syntactic. Canonical demonstrations of this are effects of left → right constraints during processing: something that occurs at point *a* affects the perception of something later at point *b*.

What I will explore in the next few pages is more recent evidence that parsing is not only "forward" it is also "downward," the construction of meaning units within short epochs. The crucial demonstration of this is evidence for backward constraints: something at point *b* in a sentence determines the percept at an earlier point *a*. Most critical to this argument is that the conscious awareness is of a constant forward moving perception, not a period of blank content suddenly filled in by something that comes later. That is, we perceive sentences in "psychological moments" in which the underlying computational processing can move back and forth, or more to the point, forth and back, before "reporting out" to conscious awareness.

## THE UNITY OF PROCESSING UNITS AND THE CONSCIOUS EXPERIENCE OF LANGUAGE

Linguistic and psycholinguistic research on sentence structure and processing has implicitly assumed that the constituent words are given: that is, the

syntactician's (and child's) problem is to determine the regularities that govern how the words and other syntactic units are arranged (and inflected when relevant); the psycholinguist's problem is to determine the processes that underlie how the words and units are composed together in production of sentences, mapped onto representations in comprehension of sentences, and learned in relation to their role in possible syntactic constructions. But outside of syntax classes and psycholinguistic experiments, the words in natural language are rarely clearly or fully presented—the acoustics of one word blends into another, and in many cases, large portions of a word or word sequence are actually not present at all: to borrow a term from phonology, the words are encoded together.

Some well-known facts about serial encoding at the phonological level may help us understand the situation at the syntactic level. First, it is well documented that unvoiced stop consonants in English, may actually not be given any acoustic power of their own. Thus, the final consonant in the words /top/, /tot/, /toc/ may be silent or all converge on glottal stop—yet we hear them quite clearly as distinct—it is the way that the preceding vowel changes as it quickly approaches the articulated position of the consonant. If we could hear the preceding vowels drawn out in time, they would be more like /TOuP/, /TOiT/, /TOaC/: the last bit of the vowel gives the clue as to where the tongue is heading before the vowel goes silent. Yet our conscious percept is that the consonant was actually uttered. This is an example of a "feed forward" activity, in which the material preceding the final silence or glottal stop makes a strong enough prediction about what will be "heard" so that it is actually perceived even when not in the signal itself.

But the influence of one part of a phonological sequence on another is not always "forward," it can be "backward" as well. It is well known that it is the timing of the onset of a post-consonantal vowel that communicates whether the preceding consonant is to be heard as voiced or unvoiced. Even more striking is that in some variants, the initial voiced consonant can also not be explicitly uttered: the difference between /bill, dill, gill/ can be only in the vowel transition following the initial occlusion of the vocal tract, just long enough to indicate voicing—it is the vowel transition away from the silent initial consonant (except for the voicing itself) that indicates what the preceding consonant *was*.

The moral is that at the phonological level, even when a word is uttered in isolated "citation" form, we automatically use early phonetic information to guide the conscious representation of what follows, *and conversely*.

It can be argued that at the level of individual words, this only shows that the unit of word recognition is larger than individual phonemes, for example, that listeners have prepackaged representations of entire syllables, or that different kinds of acoustic features work together in a "cohort" (see e.g., Marlsen-Wilson & Zwitserlood, 1989). This kind of argument may be possible in principle for words and phonology, since there is a finite number of syllables used in any particular language. But as is classically argued, such proposals of memorized units become much harder to rely on at phrasal and sentential levels, since the number of different phrases and sentences is enormous, arguably infinite in the latter case. So we might not expect both forward and backward processing interactions at these

higher levels of language. But in fact, recent evidence suggests that this is the case in normal uses of language outside of the syntax classroom and laboratory.

## UNCONSCIOUS COMPREHENSION PROCESSES WITH BACKWARD INFERENCES

The rapid and unconscious resolution of local ambiguity suggests that corresponding prospective and retrospective processes occur at the syntactic level. For this discussion, the most significant effect is the immediate role of retrospective processing that we are unaware of. If you hear a sentence like the following, in (1a, b), there can be evidence that the ambiguity of the lexically ambiguous phonetic sequence "pair/pear" creates momentary computational complexity reflected for example in decreased accuracy of a click immediately after the word (Garrett, 1964). But you are not aware of it, and have the strong impression that you assigned it the correct interpretation as you heard it. Swinney (1979) showed that both meanings of an ambiguous word facilitate an immediately following lexical decision task, even when there is a preceding disambiguating context, for example, as in (1c, d); but a few words later, only the contextually supported meaning facilitates the task.

(1) a. The pair of doves landed on our porch.
b. The pear and apple landed on our porch.
c. The doves in a pair landed on our porch.
d. The apple and pear landed on our porch.

A series of investigations by Fernanda Ferreira and colleagues (e.g., Christianson, Hollingworth, Halliwell, & Ferreira, 2001; Christianson, Williams, Zacks, & Ferreira 2006).) complements Garrett's (1964) finding at the phrasal level. Even after a garden path in segmentation of a written sentence is corrected by later material in the sentence, listeners retain a semantic representation of the initial (incorrect) segmentation. So for example, in critical trials, they follow the sentence (2a) below with a question, to which the subjects have to report "yes" or "no" to the question in (2b)

(2) a. While Bill hunted the deer ran into the woods.
b. Did Bill hunt the deer?
c. Did the deer run into the woods?

Surprisingly, Christianson et al. (2001) found that about a quarter of the responses were "yes" to (2b) following (2a). At the same time, they found that the subjects almost always answered the question in (2c) correctly: so they argued that "the reanalysis processes got as far as identifying a subject for the main clause verb, but didn't finish up by revising the interpretation on which that same NP was once the object of the verb in the subordinate clause." What is important for

my current focus is that when subjects answered (2b) correctly or not, they were quite confident in their answers. "subjects were quite poor at arriving at an interpretation licensed by the input string, yet surprisingly confident that they had correctly understood the sentences." (p. 380). Christianson et al. take this to be evidence that comprehenders construct representations that are "good enough" to contribute to ongoing comprehension, especially in normal discourse contexts (Ferreira & Henderson, 1991). Since most sentences do not have strong garden paths (especially in auditory mode), "good enough" representations are usually good enough. That is, people arrive at conceptually appropriate interpretations based on incomplete or incorrect analyses of which they are totally unaware. More recent studies support the view that subjects do in fact analyze the correct segmentation in the garden path structures on-line, even though their answers to probe questions indicate that they consciously retain the influence of the incorrect parse (Ferreira et al., 2002; Ferreira & Patson, 2007; Ferreira et al., 2009; Slattery et al., 2013).

A classic line of research on backward influences on processing started with the studies by Connine and colleagues (Connine et al., 1991). They showed that a word with an initial phonetically ambiguous consonant midway between being heard as a voiced or voiceless consonant would be perceptually disambiguated by later context. For example, a sequence phonetically midway between "tent" and "dent," is reported as "tent" when followed by ". . . . in the forest," and as "dent" when followed by ". . . . in the fender." Bicknell et al. (2016) report that the backward influence can extend over more than just the immediately following phrase (e.g., when the following context is either ". . . . was noticed in the forest" vs. ". . . .was noticed in the fender"). It is not clear from the methodologies used whether subjects believe they heard the critical word as disambiguated, or reasoned after the fact as to what the word must have been (for a discussion of these phenomena and related issues, see Bicknell et al., 2016.).

The preceding cases involve the role of apparent "backward" processing in which information that comes later in a sentence is used to specify or revise a prior analysis. A current line of experimental research by Brown, Dilley, and Tanenhaus (2012) complements the study of conversational ellipses and the role of both forward and backward processing. In their study subjects think they "heard" a word that was acoustically ambiguous, or even marginally present at all, based on later acoustic input. Farmer, Brown, and Tanenhaus (2013) apply Clark's (2013) model of hierarchically structured predictions to comprehension: the predictions guide the formation of representations of the world as new information becomes available.

> ". . . . Clark's framework predicts that expectations at higher levels of representation (e.g., syntactic expectations) should constrain interpretation at lower levels of representation (e.g., speech perception). According to this view, listeners develop fine-grained probabilistic expectations about how lexical alternatives are likely to be realized in context. . . . that propagate from top to bottom through the levels of a hierarchically organized system

> representing progressively more fine-grained perceptual information. . . . As the signal unfolds, then, the activation of a particular lexical candidate. . . . [is the one] most congruent with the acoustic signal. . . .” (Farmer, Brown, & Tanenhaus, 2013, p. 211)

This view of language comprehension emphasizes ongoing confirmation of hierarchically organized predictions, with error corrections when a given prediction is disconfirmed, shifting the interpretation of the prior material to an alternate hierarchical analysis. That is, material later in a sequence can revise the organization and interpretation of what came earlier, as a more subtle instance of the garden path phenomena explored by Ferreira et al. (2009). Brown et al. (2013) presented sentences with sequences like (3), and varied the length of the indefinite article, /a/ and the initial /s/ of the last word in the sequence. Using the “visual world” paradigm, they report that when the article /a/ is shortened and the /s/ is lengthened, subjects look at plural target pictures (“raccoons”) even after the /s/, indicating that the interpretation of the ambiguous noun in the sequence /a raccoon s . . . . is/ determined on line by what follows it. That is, when the /s/ is lengthened, subjects first look at the picture with one raccoon; then as the lengthened /s/ is heard, they shift and look at the picture with several raccoons.

Ostensibly this reflects a reanalysis, in which the shortened /a/ is not treated as a separate word; it is attached as part of the final vowel of /saw/, or perhaps reanalyzed as a brief pause. This interpretation is strengthened by the complementary finding that when the /s/ is not lengthened, the shortened definite article is then perceived and interpreted as a word.

The focus of Brown et al. is on how their research shows that listeners are sensitive to variations in local speech rate, but for my purposes the phenomenon is an online demonstration of the use of late information in determining morphological analysis of earlier speech. (See also Farmer, Yan, Bicknell, & Tanenhaus 2015 for general discussion; and Brown et al., 2012 for an example that arguably involves truly “hallucinating a definite article that was not present at all, based on extending the /s/.) Importantly, Tanenhaus et al.’s view of how the comprehension of sentences proceeds is an example of a “top down” application of an interpretation, and perception in which an entire representation can be triggered by information at the end of the signal. This gives great weight to immediate access of contextual cues of a range of kinds, including actual syntactic hierarchical structure. (For more perspective on Tanenhaus’s view on how representational levels interact during sentence comprehension, see Degen & Tanenhaus, 2015.)

(3) . . . . saw uh raccoon Swimming

But in normal conversation, many words aren’t there at all. . . .

The preceding examples assume that all the words in the sentences are present to some degree. But in everyday speech, many acoustic details are slurred or even omitted. This can be demonstrated by showing that fragments several “words”

long are impossible to recognize in isolation, but pop into complete clarity (for native speakers) when heard as part of an entire sentence (Pollack & Pickett, 1964; Greenberg et al., 1996; Greenberg, 1999; Arai, 1999; Arai & Warner, 1999; Johnson, 2004; Warner et al., 2009; Tucker & Warner, 2010).[1] Consider first an approximate transcription of an example from adults talking to each other in a normal conversation (this is an actual example provided by N. Warner, the reader can hear examples like it on her website: http://www.u.arizona.edu/~nwarner/reduction_examples.html).[2]

(4) [tjutʌ̃m]

(Hint: this corresponds to four words). It is completely incomprehensible by itself, but when a latter portion of the longer sequence is included it is comprehensible:

(5) [tju tʌ̃m ɾɨ tʰak̃ tɾ̃ mi]

Everyone immediately hears this as:

(6) Do you have time to talk to me?

The striking significance of this is that phenomologically listeners think they simultaneously hear the fragment and assign it its three word analysis. But we know this cannot be true since the fragment in isolation is incomprehensible. This suggests that backward processing at a local acoustic level is a normal part of comprehension and building representations of conscious experience of language.

But this example was the beginning of a sentence, so perhaps it is a special case, where there is no preceding context. However in an experimental paradigm Van de Ven (2011) found that the following context can contribute importantly to recognition of material in the middle of a sentence. In fact, the following example from a natural conversation supports the view that in some cases, the following context alone is sufficient to clarify a reduced word, while the preceding context alone is not sufficient.

(7) [tʃũn:]

Try pronouncing this to yourself (hint: the production intent is 2 syllables). Now look at a longer sequence in which the example was embedded:

(8) [ɔ̝: ʌ: thɨzdɛ nʌik̃ (pause) ʌ̰ mn wɪɹ tʃũnɾ̃n:(ɨ) spa]

When listeners hear the surrounding material, the excerpt immediately pops into consciousness and what one "hears" is:

(9) . . . err Tuesday night, when we were chillin' in the spa.

Recently we tested this further: it turns out that even with all the material preceding [tʃũn:] (as in "and err Tuesday night when we were. . . .") almost no one perceives it correctly. But if only the following material ("in the spa") is heard along with the sequence, then [tʃũn:] is heard clearly as "chillin." First, such facts support the view that in everyday comprehension the minimal phonetic unit of comprehension is not the word, and that comprehension must be operating with parallel hypotheses at several interactive levels—syntactic and phonetic computations proceed in parallel with frequent cross checks at specific points. One can expect that where those cross checks occur will be the focus of ongoing research, now that we have tools that can chop running speech into a full range of possible units. An initial hypothesis is the *phase*, the unit of syntactic structure that has just enough content for semantic analysis (Chomsky, 1995, 2000, 2008). Phase theory is an active research area in linguistics, so the reader should be skeptical about details by the time this chapter is published, never mind a few years later. (See Boeckx, 2006 for a lucid explication of the technical issues and Citko, 2014 for a recent introduction.) So we can start with a particular hypothesis, as the latest idea on how different levels of a sentence are integrated in working units:

(10) *The unit over which local acoustic/phrasal/meaning integration occurs is the phase.*

However, we must note that "chillin" is involved in two prima facie phases: (a) the preceding material which includes a WH, subject, and auxiliary, which embeds the verb in a complex structure with at least several levels of hierarchical organization; (b) the following material, which embeds the verb in a more compact verbphrase only. The unique effectiveness of the following material leads to a hypothesis for further investigation, based on a single case, but one with some intuitive appeal:

(11) *The effectiveness of a phase in integrating distinct language levels is proportional to its structural simplicity.*

Further research will (I hope) verify or falsify these working hypotheses. A particular question is whether the role of the less complex phases is unique in the comprehension processes, or whether it reflects different degrees of reduction in the production processes. For example in (9) the failure of the preceding material to clarify the excerpt may be because as a NP-Verb phase it is actually less reduced in speech. So it is now an interesting research question whether phases are the "true" units of comprehension that the many "click" experiments attempted to define (Fodor & Bever, 1965; Garret et al., 1966; Bever et al., 1969), whether those effects depend on production processes, or whether the phase in fact is not the relevant factor that elicits segmentation effects. For example, there is new interest in how speakers maintain the predictability, (aka "information density") of their sentence output (e.g., Jaeger, 2006; Levy & Jaeger, 2007; Jaeger, 2010; Frank & Jaeger, 2008). This principle extends both to choice of phrases

and words, and to use of contractions. For example, Frank and Jaeger show that local predictability can determine whether "you are" is contracted to "you're" in sentence production. Dell and Chang (2014) recently proposed a model that combines this approach with Macdonald's ideas that production patterns condition comprehension processes (Macdonald, 1999, 2013). Within a connectionist model of syntax production, they unify the processes of acquisition, production and comprehension based on serial predictability of words. The examples I have mentioned in this chapter suggest that for such a model to be adequate, the unit of predictability is not only serial word-by-word, but ranges within a larger unit. It stands to reason that more complex phases (e.g., NP-Verb) have more information and hence less internal predictability than simpler phases (e.g., VprepP). Thus, increased phonetic reduction in smaller phases (if true in general) could be due to structural or statistical factors in production. These alternatives open up the usual kind of research program in which a structural hypothesis (e.g., that the phase regulates speech production and phonetic reduction) competes with a statistical hypothesis (e.g., that units of mutual predictability regulate speech production and phonetic reduction). Specific experimental predictions are going to interact with each candidate theory of what phases are, so it is too rich an area to explore further here. But it does promise the possibility of an informative interaction between comprehension research, production research and using behavioral data to constrain theories of phases.

Implications for stages of comprehension and assigning syntax during processing: There is an intriguing interaction between the idea of analyzing serial sequences in whole chunks and Townsend's and my proposal about logical stages of alternating between associative and derivational processes during comprehension (Bever & Townsend, 2001; Townsend & Bever, 2001, chapters 5 and 8). We argued and reviewed evidence that comprehension processes necessarily integrate statistically valid patterns with computationally applied derivations, within an "analysis by synthesis" framework. On this model, pattern recognition templates can apply quickly to assign a likely meaning, to be complemented by derivational processes. This raised a question about when the derivational reconstruction of that input occurs: we answered this in the acronym for the model, LAST—late assignment of structure theory—making the point in contradistinction to other models, which either assume that structure must be assigned before meaning, or that derivational structures are actually not assigned at all. In that work, most attention was given to the analysis of sentence level comprehension and syntactic structure assignment. The discussion in this chapter gives some further organizational shape to the units within which pattern recognition and derivational processes can apply to assign meaning—our initial hypothesis for this is the phase, as described in (10). The demonstration of backward processes within such a unit supports the idea that comprehension proceeds in bursts that integrate learned patterns and composed structures.

The disconnect between unconscious processing and our conscious experience of normal conversational language calls into question the *immediacy assumption*—the theoretical preconception that a complete hierarchical layering

of grammatical analyses is applied to language input *as we hear it* (Just and Carpenter, 1980; Marslen-Wilson, 1973, 1975). This assumption has been the bedrock of many distinct kinds of comprehension models (see Christiansen & Chater, 2016 for a review). The importance of backward processing of information I have reviewed shows that the assumption is false. I have focused on the importance of such processing for discriminating the speech signal. However, recent discussions have given a computational motivation for allowing indeterminate sequences to be held in immediate memory to be disambiguated or clarified by following input. On this model, there can be uncertainty immediately after each subsequence as to what it was: the subsequence is held in memory until the following material completes a larger pattern of analysis that embraces the critical subsequence (Levy et al., 2009; Kleinschmidt & Jaeger, 2015; for general discussions see Kuperberg & Jaeger, 2015; K. Bicknell et al., 2016). The critical point of consistency with the model in LAST is the notion that larger units immediately organize the local structure and ultimately the meaning of a lexical sequence. In the critical cases, an early indeterminacy is informed by its role in a larger unit of structure and meaning.

But this cannot be the whole story in the LAST framework. In our proposals, we noted that there must be a hierarchy of parallel levels during ongoing processing, each of which can involve integration of associative cues and structural assignments: this includes individual sounds, words, short phrases, phases, sentences and, arguably so-called "discourses" (see Townsend & Bever, 2001, chapters 5 and 8; Bever & Poeppel, 2010; Poeppel et al., 2007). Integrating Clark's notion of parallel hierarchical processes with analysis-by-synthesis, we can think of these parallel computations as organized into a "computational fractal" in which the same alternation and integration of the two major kinds of information occur within each local linguistic unit (e.g., syllable, word, phrase, phase. . . .): separate study of the processes at each level is a matter of "grain"—the size of each domain over which analysis by synthesis processing can occur.

This reinterpretation of our Analysis by Synthesis model moves toward a reconciliation between our view and the view that syntactic derivational structures are assigned serially from "left" to "right," as sentences are experienced. In this vein, Colin Philips has adduced arguments that such immediate structural assignment occurs, and also counter arguments to examples used by us to demonstrate the original analysis by synthesis proposals (for a review of his model and critique of ours, see e.g., Phillips & Lewis, 2013; Lewis & Tanenhaus, 2015). In discussing our proposal, Philips also notes that an important issue is one of "grain." Our proposal here is that such processes occur in units of layered levels starting with individual sounds, overlapping with those of increasing size—that is, the processing is simultaneously multigrained. As it stands, this proposal offers a resolution of the theoretical conflicts, in principle, though much remains to be spelled out. And of course, it is important to review how Philips' positive research findings that support his model might also fit within the modified, "computational fractal" framework I am presenting here: but that will have to await a new thorough analysis.

## IMPLICATIONS FOR NOTIONS OF CONSCIOUS EXPERIENCE

A related phenomenon is our conscious, but apparently false perception in many cases, that we understand the speech as we hear it serially. This has been long noted in phonology, but most of the effects are extremely local, and hence subject to solution by simply enlarging the scope of the initial input to a bigger chunk, e.g., the syllable, or word, as I mentioned. However, even in this case there is a puzzle: listeners "think" consciously that they heard the individual sounds in words uttered in a citation form, in the order that they occurred. So even at the most basic level of speech perception, our conscious experience of a series of stimuli, actually involves some "backward" processing.

The significance of this sort of phenomenon is magnified in the case of phrasal and sentence level processing. For example, in the cases of "tyuv," and'chilln', where the critical (and incomprehensible) isolated sequence is *followed* by the crucial contextual material, we are not aware that we could not have analyzed the initial sequence until the later material was heard: rather we are convinced that we understood it as it was phonetically presented. This simple fact demonstrates that language comprehension may proceed in sequences of "psychological moments" in which actual processing moves both forward and backward, with some definition of phases specifying the domain of the interaction. This phenomenon has barely been touched in the language sciences, but is clearly fascinating and will have profound implications for consciousness theories, once it is better understood. Prima facie, it is an ultimate demonstration that even in language behavior (i.e., "externalization" of timeless linguistic structures) serial order may be less important than structure dependent organization.

There is a methodological dividend of the notion that there is a decoupling of the perceptual and comprehension processes and our consciousness of when they occurred. Throughout the literature on the post sentence location of clicks, when the reported location is not a phrase boundary, it systematically precedes the actual location. (This started with Fodor & Bever, 1965, and it has popped up several times; see also Townsend & Bever, 1991.) At first blush, this might be interpreted as a simple demonstration of the notion of "prior entry" (Titchener, 1908; Spence & Parise, 2009): an attended to stimulus is perceived earlier than others. It is possibly also related to demonstrations of "chronostasis" in which a more complex stimulus is slowed down relative to a simpler one. For example, Wundt reported a study in which a bell is perceived earlier than its actual location relative to a moving arrow across a series of numbers on a clock-face display. Wundt referred to the relative delay of the numbers as "positive time displacement" (Wundt, 1897, 1918). Correspondingly, in our studies, the subject's task in locating the clicks is to locate the piece of the sentence and the click together, while attending to the entire sentence. To explain the preposition effect, we may refer to a Helmholtzian unconscious inference. Our conscious reconstruction of perceiving and understanding the speech stream as it was presented, leaves the click unanalyzed within the reconstruction of the speech. If it is the case that the click is perceived without the reconstruction processes, the unconscious

inference is that it occurred earlier than it actually did. If one insists that this is merely an explanation of a well-known "positive time displacement" or prior entry effect, at least it is an *explanation*.

The notion that conscious awareness of serial order can involve reconstruction is not novel. There is a distinguished line of research, stimulated by Husserl's (1917/1990) considerations of the conscious perception of time, and most famously re-introduced by Fraisse (1967, 1974). However, most of the research in this vein involves relatively short intervals or rapid sequences of short and simple stimuli. For example, in demonstrations of metacontrast, a later stimulus will "absorb" an earlier one into an "exploding" or moving single object—indeed, this is a large part of how continuous motion is perceived in cinematic projections of at least 1 every tenth of a second. However, the language sequence cases described involve much longer and more complex prospective and retrospective reconstructions. Thus, we have a potential demonstration that the "psychological moment" is itself determined by the perceptual units required: as they become more complex and hierarchical, the physical size of the "moment" can expand dramatically.

Up to now, I have emphasized evidence for retrospective processing of language, because it is the most dramatic demonstration of the reconstructive nature of our conscious experiences. But as I have mentioned, various researchers have suggested that most processing is prospective, that is predictive templates are generated early during each utterance, and the remaining act of perception is actually based on confirmation of an already formed structure. Certainly, we can experience this with close friends and spouses—we often have a strong expectation of what they are about to say and are just waiting for confirmation of it.

While I think it dubious that comprehension of novel discourses always proceeds in this way, let us suppose for a moment that it does. It would not change the implications of for our proposal that during comprehension, conscious awareness is sometimes retrospective. In that case, later input triggers confirmation of a waiting hypothesis, rather than triggering fresh computational processes. Either way, the conscious awareness of the prior input depends on later input.

This concept returns us to the flagship issue of modularity in perceptual processing and representation, which Fodor famously explored. The corresponding puzzle for present and future research is how the distinct levels/modules of representation are actually integrated into the conscious experience of continuous integrated processing. That is, when I understand the sentence "a sentence is like a (miniature) opera" spoken conversationally, my conscious experience is that I hear and interpret the input as a coherent continuous object that unifies the acoustic input and the representational analysis; this occurs even though detailed examination of the sort I have reviewed here shows that the computational details belie this belief. In Fodor's formulation, the "central processor" is the mental cloaca where inputs and outputs to the different modules can meet: but, by definition, the central processor is relatively slow and woolgathering. So it remains to be spelled out how it could create the introspective belief that

we understand sentences synchronously with their presentation. In Fodorian terminology, maybe it will turn out that consciousness itself is made up of the simultaneous output of a number of modules that interconnect with some degree of automaticity. As Fodor might say, stranger things have turned out to be true.

Thus, in this exploration, the study of language may become a theory-rich touchstone for yet another aspect of cognitive science—the nature of conscious experience.

## THE *REAL* POVERTY OF THE STIMULUS

I began this discussion noting the significance of "the poverty of the stimulus" for all of cognitive science, as discussed by Fodor (1981).

Now consider the implications for the language-learning child of how sentences are acoustically mangled in normal conversation. There is evidence that child-directed "motherese" is often clearer than normal conversations in many cases (Bernstein-Ratner, 1996; Bernstein-Ratner & Rooney, 2001), but not all (see Van de Weijer, 1998); it may use devices to clarify word boundaries (e.g., Aslin et al., 1996) and it may be that infants prefer motherese when they have a choice (e.g., Fernald, 1985; Cooper et al., 1997). In any case, it is likely that the vast majority of speech that children hear is between adults, or older children, and there are considerable cultural differences in whether motherese is used at all (Lieven, 1994). Furthermore, various studies have shown that the syntactic or phonetic quality of the child's input may bear little relation to the child's emerging language (C. Chomsky, 1986; McColgan, 2011). In any event, well-articulated motherese is not always dominant even in child-directed speech. Consider a transcribed example from a real motherese sentence. First, attempt to understand the following fragment (five words!), taken from an actual utterance by a mother to her child:

(12) [ɾ̃nʷːɹɨpə̃m]

Now see the whole utterance that follows; (if you are a phonetician) try sounding out the phonetic version alone to see if you can (suddenly) understand the whole utterance. In the acoustic version, adults cannot understand this sentence excerpt; but it immediately pops into perfect comprehension, with the conscious intuition that the entire utterance was reasonably clearly pronounced, which is immediately heard as in (14).

(13) [o gɹe(t) mamɪ mu ðoz mæɣəzɾ̃ns si jy kʰɾ̃n: gɪɾ̃mɾ̃nʷːɹɨpə̃m]

(14) Oh great, mummy put those magazines away so you can't get them and rip them

It is amazing enough that adults can understand conversational speech like this. For a child the problem is doubly compounded, since its grammatical knowledge is incomplete, and it has not yet had time to build up complex language patterns. This simple fact vastly increases the poverty of the stimulus problem, since in

many cases the child may not be able to even encode the utterance in enough detail to serve as a learning model.

There is an important implication of these analyses for how sophisticated the child's comprehension system must be. Over many years, it has been argued that linguistic processes are *structure dependent* (Chomsky, 1980): rules are characteristically sensitive to hierarchical structure. This part of Universal Grammar has been shown to account for pathways to language in first language acquisition (e.g., Crain & Nakayama, 1987 and many later discussions). Recent attempts have been made to show that serial learning models can converge on such sensitivity but such models fail to generalize realistically, omit structure dependence in fact (Perfors et al., 2006), or focus on simulating structure dependence (Reali & Christansen, 2005; see Berwick et al., 2011, for general discussion). It has been shown that adults can learn serial rules but in so doing they utilize different brain areas than those characteristic of language (Musso et al., 2003; Moro, 2011). In the current "minimalist" treatments of language, hierarchical trees are constructed as sets, that is, without serial order constraints (Chomsky, 2013, 2015). On this view, the surface order in language is imposed by how it interfaces with our systems of input and output: but many actual computation of linguistic rules operate strictly on the hierarchical structures without reference to the serial structure of overt language sequences: thus, the comprehension system is building chunks of hierarchically organized structures which themselves may be internally order-free, corresponding to order free processing of the input.

Consider now, the implications of our idea that during language processing, there are "time free" processing zones that mediate between the serial input, structural analysis and simultaneous consciousness of the serial input and its meaning. Earlier, I suggest that the simplest available phase is the unit in which processing can occur both forward and backward. But this is to say in its strong form, that in certain defined domains, serial order is unconsciously suspended during sentence comprehension—allowing for structural dependencies to take precedence. In brief, within certain domains, even the externalization of language as serial may be ignored during behavior in favor of pure structure dependence.

A moment's thought suggests that this must be so, as part of the solution to how the child manages to comprehend normal conversations and build up linguistic knowledge from them: s/he must be listening for phrasal categories that integrate and organize local word sequences. How else could s/he latch onto meanings and structural regularities so automatically and quickly? So the argument that structure dependence appears spontaneously in children's learning language structure applies perforce to early stages of language processing itself (Christophe et al., 2008; for related discussion of models of how the language learning child might benefit from unsegmented input, see Pearl & Phillips, 2016).

These considerations are consistent with an analysis by synthesis model of language acquisition, proposed in general terms in Bever (1970), developed more specifically in Townsend and Bever (2001), and elaborated in later writings

(e.g., Bever, 2008, 2013). On this model, children alternate (logically) between accessing available structures/representational constraints and building generalizations over the language it experiences as represented by those categorical structures. The role of the generalizations is to provide form-meaning pairs for sentences that have not yet been assigned a full grammatical representation. These pairs can then be the input data for further elaboration of grammatical analysis, accessing the categorical structures. The categorical structures are in part innate—unique to language, in part innate as a part of general thought and perceptual processes. The categorical framework itself becomes more elaborate and uniquely adapted to language structure in particular. (See Bever, 2008, for further discussion and examples of this iterative process; see Lidz and Gagliardi, 2015 for a discussion of the interaction of generalizations and structure building during learning; see Bever, 2013 for a general discussion of this model of language acquisition as an instance of intrinsically motivated human problem solving.)

The significant feature of this model is the dynamic integration of probabilistic and categorical information to yield both a repertoire of statistically valid generalizations and a constructed grammatical representation for all the sentences in the language and many of the semi-sentences. While the model has some general support from acquisition data, it is not sufficiently precise to be adequately testable in detail: in part this is because it is a framework for how associative and structural processes can interact, but allows for considerable differences between individuals and the data they experience.

Of course, this is not the first attempt to create a model of language behavior and acquisition that combines both associative and symbolic information. Indeed, the initial flowering of "psycholinguistics" under the leadership of Charles Osgood (Osgood & Sebeok, 1954; Osgood, 1968) was an explicit attempt to show that the then current model of mediated stimulus-response learning could account for the then current phrase structure model of language structure. (Alas, both models were inadequate for their respective goals, but were consonant with each other because the inadequacies corresponded well; see Bever, 1968, 1988 for discussions). In recent years, a class of explicit computational models has appeared that instantiates a dynamic integration of available categorical structures/processes and Bayesian inference algorithms. These models ostensibly connect to Fodor's notion of the language of thought (LoT), the set of symbols and processes that manipulate symbols. The recent models add a Bayesian statistical component to LoT, and recast it as the probabilistic language of thought (pLoT). Researchers in this vein show that many graded kinds of category knowledge can be accounted for as well as apparent category and concept formation. (See Perfors et al., 2006, Goodman & Lassiter 2014; Piantadosi & Jacobs, 2016, for representative discussions among the many articles now appearing on pLoT.) It remains to be seen if such models can actually learn or even render grammatical representations, including processes that involve structure dependent constraints. At the moment these models do not generally address such problems.

This is not to say that no attention is given to how statistically non-categorical input can result in arriving at grammars appropriate to the child's native language. A number of models have also used Bayesian and other statistical techniques of how variable input data may discriminate between candidate grammars. This includes many different target architectures, but all in the general method of using statistically variable input to reinforce or distill out candidate rules or grammars. (For example, see Yang, 2002, 2004; Yang & Roeper 2011; Pearl & Goldwater 2016; Lidz & Gagliardi, 2015.) The critical feature that seems to discriminate these approaches from the emerging pLoT variants of Fodor's LoT is that these approaches presuppose the availability of candidate grammars or rules, both in the child and as the ultimate goal of language learning.

## IMPLICATIONS FOR OLD AND NEW RESEARCH

A cautionary note on the issue of how children and adults deal with normal conversational speech: *sometimes* our spoken utterance may be clear enough, with adequate serial cues for a diligent listener to develop immediate representations of what s/he is hearing. This *may* be especially true of instances of so called child-directed "motherese." But what is important in our examples is that this is not always the case, and may not even be true of the majority of cases. If indeed, most of our comprehension has to deal with compressed and cue-poor input, this also calls into question the generalizability of the many studies of carefully pronounced "laboratory speech" that comprise the overwhelming majority of experimental studies, never mind the use of complete word-by-word visual presentation.

The reader will note that I have extrapolated very far ahead of a very small number of facts, but I hope in ways that are amenable to empirical investigation. For example, one can use the backward reconstruction phenomenon as a tool to study what units are the relevant bridges between serial input and structural output. Here is a(n in principle) simple way to do this. Take conversational corpora and analyze the transcripts (which presumably already have interpreted the conversations into complete words, phrases and sentences); pick out candidate phases according to a theory of what phases are relevant [e.g., as postulated in (10)]; test gated increments of each candidate from its beginning for recognition of the input by subjects (that is, start with an initial fragment, then successively longer ones to see when the initial fragment becomes (retrospectively) clearly interpretable; do the corresponding testing starting from the final part of such fragments. The same kind of procedure can be applied to child-directed speech to examine empirically the claim that a great deal of it is also heavily encoded and dependent on both forward and backward processing. No doubt, these are big projects, but the payoff could be even bigger in leading to a theoretical understanding of how serially presented units build up hierarchical structures and meaning in comprehension and language learning, and to information about normal speaking with many practical applications.

Such research programs can be viewed as the latest step in making good on the implications of the original discoveries by Fodor and his colleagues that in ongoing speech comprehension, sentences are automatically segmented into natural units.

## AUTHOR'S NOTE

It should be obvious to the reader how much this paper owes to Jerry Fodor. Along with Merrill Garrett, we pioneered click mislocation as a method to demonstrate the active online use of syntax during comprehension: this is the foundation of many subsequent explorations of the initial compositional strategies of comprehension. More personally, my many conversations with Jerry, co-teaching a course in the early 1960s and co-authoring our 1974 book (*The Psychology of Language*), gave me wide-ranging instructions in how to think about the general problems of cognitive science. We did discuss the poverty of the stimulus, both in relation to adult comprehension and language acquisition. But we did not discuss consciousness at all, to my recollection: it was viewed at the time as a slightly embarrassing romantic problem not a scientific one. But as Jerry noted in his 2007 review of Strawson's edited book on consciousness, "[it] is all the rage just now. . . . What everybody worries about most [is] what philosophers have come to call "the hard problem." The hard problem is this: it is widely supposed that the world is made entirely of mere matter, but how could mere matter be conscious? How, in particular, could a couple of pounds of grey tissue have experiences?" In considering this question, I (TGB) follow the general approach in "biolinguistics" to an understanding of the biology and genetics of language: to discover what makes consciousness possible, we first have to determine what consciousness *is*, how it is acquired as a habit from the alleged "blooming buzzing confusion" of infancy, how it is represented, how it works. This chapter is not a solution to all that, but a pointer to a problem that I hope will attract the interest of today's graduate students. Without them, our science will be lost to the world.

I am indebted to Roberto de Almeida and Lila R. Gleitman, editors of the volume where this appeared, for many helpful early criticisms and comments; also to Mahmoud Azaz for conceptual advice, to David Poeppel for reminding me about the broad evidence for what I coin the "computational fractal" in language processing, and especially to Michael Tanenhaus for deeply considered advice. Caitlyn Antal and Valerie Kula were invaluable bibliographic assistants. Other helpful comments are due to Stuart Hameroff, Al Bergesen, Felice Bedford, Virginia Valian, Rudi Troike, Louann Gerken, Billl Idsardi, Gary Dell, Florian Jaeger, Mark Pitts, Maryellen MacDonald, Lisa Pearl, Massimo Piattelli-Palmarini, and Noam Chomsky. Most of all, I am indebted to my colleague Natasha Warner, who not only developed a methodology of collecting natural conversations, but also made available her materials, her bibliographic advice, her perspective, and all the phonetic transcriptions in this chapter.

## NOTES

1. For other discussions of reduction in casual speech, see Ernestus (2000), Tucker & Warner (2011), Ernestus & Warner (2011), Dilley & Pitt (2010), Gahl et al. (2012), and chapters in the special 2011 issue of the *Journal of Phonetics*, edited by Ernestus and Warner.
2. Readers interested in the examples discussed in this chapter can email me for a PowerPoint file with sound. tgb@:mail.arizona.edu

## REFERENCES

Abrams, K., & Bever, T. G. (1969). Syntactic structure modifies attention during speech perception and recognition. *Quarterly Journal of Experimental Psychology*, *21*, 280–290.

Arai, T. (1999). A case study of spontaneous speech in Japanese. In *Proceedings of the International Congress of Phonetic Sciences, San Francisco, 1*, 615–618.

Arai, T., & Warner, N. (1999). Word level timing in spontaneous Japanese speech. Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco, August 1999. 1055–1058.

Aslin, R. N., Woodward, J. Z., LaMendola, N. P., & Bever, T. G. (1996). Models of word segmentation in fluent maternal speech to infants. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 117–134). Hillsdale, NJ: Erlbaum.

Bernstein-Ratner, N. (1996). From signal to syntax: But what is the nature of the signal. In J. L. Morgan, & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 135–150). Hillsdale, NJ: Erlbaum.

Bernstein-Ratner, N., & Rooney, B. (2001). How accessible is the lexicon in motherese? In J. Weissenborn., & B. Höhle (Eds.), *Approaches to bootstrapping: Phonological, lexical, syntactic and neurophysiological aspects of early language acquisition Vol. I.* (pp.71–78). Philadelphia: Benjamins.

Berwick, R. C., Pietroski, P., Yankama, B., & Chomsky, N. (2011). Poverty of the stimulus revisited. *Cognitive Science*, *35*(7), 1207–1242.

Bever, T. G. (1968). Associations to stimulus-response theories of language. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory* (478–494). Upper Saddle River, NJ: Prentice-Hall, Inc.

Bever, T. G., (1970/2013). The cognitive basis for linguistic structures. Reprinted in Laka, I., Sanz, M. and Tanenhaus, M. (Eds.), *Language down the garden path: the cognitive and biological bases for linguistic structures* (pp.1–70). New York, NY: Oxford University Press.

Bever, T. G. (1973). Serial position and response biases do not account for the effect of syntactic structure on the location of brief noises during sentences. *Journal of Psycholinguistic Research. 2*, 287–288.

Bever, T. G. (1988). The psychological reality of grammar: A student's eye view of cognitive science. In W. Hirst (Ed.), *The making of cognitive science.* Cambridge, England: Cambridge University Press.

Bever, T. G. (2008). The canonical form constraint: Language acquisition via a general theory of learning. In Guo et al. (Eds.), *Cross-linguistic approaches to the psychology of language* (pp.475–492). New York, NY: Oxford University Press.

Bever, T. G. (2013/2015). The biolinguistics of language universals: The next years. In M. Sanz, I. Laka, & M. K. Tanenhaus (Eds.), *Language down the garden path: The cognitive and biological basis of linguistic structures* (pp. 235–405). New York NY: Oxford University Press.

Bever, T. G., Lackner, J. R., & Kirk, R. (1969). The underlying structures of sentences are the primary units of immediate speech processing. *Perception & Psychophysics*, *5*(4), 225–234.

Bever, T. G., Lackner, J. R., & Stolz, W. (1969). Transitional probability is not a general mechanism for the segmentation of speech. *Journal of Experimental Psychology*, *79*(3), 387–394.

Bever, T. G., & Townsend, D. J. (2001). Some sentences on our consciousness of sentences. In R. Dupoux (Ed.), *Language, Brain and Cognitive Development* (pp. 145–155). Cambridge, MA, MIT Press.

Bever, T. G., & Poeppel, D. (2010). Analysis by synthesis: A (re-)emerging program of research for language and vision. *Biolinguistics*, *43*(2), 174–200.

Bicknell, K., Florian Jaeger, T., & Tanenhaus, M. K. (2016/1). Now or . . . later: Perceptual data are not immediately forgotten during language processing. *Behavioral and Brain Sciences*, *39*, e67. Publisher Cambridge University Press.

Bicknell, K., Tanenhaus, M., and Jaeger, T., Rationally updating uncertainly about previous words. Unpublished manuscript.

Boeckx, C. (2006). *Linguistic minimalism: origins, concepts, methods, and aims*. New York, NY: Oxford University Press.

Brown, M., Dilley L. C., & Tanenhaus, M. K. (2012). Real-time expectations based on context speech rate can cause words to appear or disappear. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 1374–1379). Austin, TX: Cognitive Science Society.

Citko, B. (2014). *Phase theory: An introduction*. Cambridge, UK: Cambridge University Press.

Chapin, P. G., Smith, T. S., & Abrahamson, A. A. (1972). Two factors in perceptual segmentation of speech. *Journal of Verbal Learning and Verbal Behavior*, *11*(2), 164–173.

Chomsky, C. (1986). Analytic study of the tadoma method: Language abilities of three deaf/blind children. *Journal of Speech, Language and Hearing Research*, *29*(3), 332–347.

Chomsky, N. (1959). A review of BF Skinner's verbal behavior. *Language*, *35*(1), 26–59.

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

Chomsky, N. (1975). *Reflections on language*. NewYork, NY: Pantheon Books.

Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: Massachusetts Institute of Technology.

Chomsky, N. (1980). *Rules and representations*. New York, NY: Columbia University Press.

Chomsky, N. (2000). *New horizons in the study of language and mind*. Cambridge, England: Cambridge University Press.

Chomsky, N. (2008). On phases. In R. Freidin, C. Peregrín Otero, & M. L. Zubizarreta (Eds.), *Foundational issues in linguistic theory. Essays in honor of Jean-Roger Vergnaud* (pp. 133–166). Cambridge, MA: MIT Press.

Chomsky, N. (2013). Problems of Projection. *Lingua*, *130*, 33–49.

Chomsky, N. (2015). Problems of projection: Extensions. In E. Di Domenico, C. Hamann, & S. Matteini (Eds.), *Structures, Strategies and Beyond. Studies in Honour of Adriana Belletti* (pp. 1–16). Amsterdam: John Benjamins.

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, *42*, 368–407.

Christianson, K., Williams, C. C., Zacks, R. T., & Ferreira, F. (2006). Younger and older adults' "good-enough" interpretations of garden-path sentences. *Discourse Processes*, *42*(2), 205–238.

Christophe, A., Millotte, S., Bernal, S., & Lidz, J. (2008). Bootstrapping lexical and syntactic acquisition. *Language and speech*, *51*(1-2), 61–75.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(03), 181–204.

Cooper, R. P., Abraham, J., Berman, S., & Staska, M. (1997). The development of infants' preference for motherese. *Infant Behavior and Development*, *20*(4), 477–488.

Connine, C. M., Blasko, D. G., & Hall, M. (1991). Effects of subsequent sentence context in auditory word recognition: Temporal and linguistic constraints. *Journal of Memory and Language, 30*, 234–250.

Crain, S., & Nakayama, M. (1987). Structure dependence in grammar formation. *Language*, *63*(3), 522–543.

Dalrymple-Alford, E. C. (1976). Response bias and judgments of the location of clicks in sentences. *Perception & Psychophysics*, *19*(4), 303–308.

Degen, J., & Tanenhaus, M. K. (2015). Processing Scalar Implicature: A Constraint-Based Approach. *Cognitive Science*, *39*, 667–710. doi: 10.1111/cogs.12171.

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 Ser. B*, *369*, 20120394. http://dx.doi.org/10.1098/rstb.2012.0394

Dilley, L., & Pitt, M. A. (2010). Altering context speech rate can cause words to appear and disappear. *Psychological Science*, *21*(11), 1664–1670.

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience, 19*(1), 158–164.

Ernestus, M. (2000). *Voice assimilation and segment reduction in casual Dutch: A corpus-based study of the phonology-phonetics interface* (Dissertation). Utrecht: LOT.

Ernestus, M., & Warner, N., Eds. (2011). Speech reduction. *Journal of Phonetics* [Special issue], *39*(3), 1–14.

Farmer, T. A., Brown, M., & Tanenhaus, M. K. (2013). Prediction, explanation, and the role of generative models. *Behavioral and Brain Sciences*, *36*, 211–212.

Farmer, T. A., Yan, S., Bicknell, K., & Tanenhaus, M. K. (2015). Form-to-expectation matching effects on first-pass eye movement measures during reading. *Journal of Experimental Psychology: Human Perception and Performance, 41*, 958–976.

Ferreira, F., Bailey, K. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*(1), 11–15. doi:10.1111/1467-8721.00158.

Ferreira, F., Engelhardt, P. E., & Jones, M. W. (2009). Good enough language processing: A satisficing approach. In N. Taatgen, H. Rijn, J. Nerbonne, & L. Schomaker

(Eds.), *Proceedings of the 31st Annual conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Ferreira, F., & Henderson, M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language*, *31*, 725–745. doi:10.1016/0749-596X(91)90034-H.

Ferreira, F., & Patson, N. D. (2007). The "good enough" approach to language comprehension. *Language and Linguistics Compass*, *1*(1-2), 71–83.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, *8*(2), 181–195.

Fodor, J. A. (1981). *Representations: Philosophical essays on the foundations of cognitive science*. Cambridge, MA: MIT Press.

Fodor, J. A., & Bever, T. G. (1965). The psychological reality of linguistic segments. *Journal of Verbal Learning & Verbal Behavior*, *4*, 414–420.

Fodor, J. A., Bever., T. G., & Garrett, M. F. (1974). The psychology of language: An introduction to psycholinguistics and generative grammar. New York: McGraw-Hill.

Fraisse, P. (1967). *Psychologie du temps*. Paris: Presses universitaires de France.

Fraisse, P. (1974). *Psychologie du rhythme*, Paris: Presses universitaires de France.

Frank, A., & Jaeger, T. F. (2008). Speaking rationally: Uniform information density as an optimal strategy for language production. *In The 30th annual meeting of the Cognitive Science Society (CogSci08)* (pp. 939–944), Washington, DC.

Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*, *66*(4), 789–806.

Garrett, M. F. (1964). *Structure and sequence in judgments of auditory events* (Unpublished doctoral dissertation). University of Illinois, Chicago.

Garrett, M. F., Bever, T. G., & Fodor, J. A. (1966). The active use of grammar in speech perception. *Perception & Psychophysics*, *1*, 30–32.

Goodman, N. D., & Lassiter, D. (2014). Probabilistic semantics and pragmatics: Uncertainty in language and thought. In S. Lappin & C. Fox (Eds.), *Handbook of contemporary semantics*. Hoboken, NJ: Wiley-Blackwell.

Greenberg, S. (1999). Speaking in shorthand—a syllable-centric perspective for understanding pronunciation variation. *Speech Communication*, *29*, 159–176.

Greenberg, S., Hollenback, J., & Ellis, D. (1996). Insights into spoken language gleaned from phonetic transcription of the Switchboard corpus. In *Proceedings of ICSLP* (pp. 24–27). Philadelphia: ICSLP.

Husserl, E. (1917). *On the phenomenology of the consciousness of internal time* (1893–1917), Vol. 4, J. B. Brough, Trans. Dordrecht: Kluwer (Original work re-published in 1991).

Jaeger, T. F. (2006). *Redundancy and syntactic reduction in spontaneous speech*. (Unpublished doctoral dissertation), Stanford University, Palo Alto, CA.

Jaeger T. F. (2010. Redundancy and reduction: Speakers manage syntactic information density. Cognitive Psychology, *61*, 23–62.

Johnson, K. (2004). Massive reduction in conversational American English. In K. Yoneyama & K. Maekawa (Eds.), *Spontaneous Speech: Data and Analysis. Proceedings of the 1st Session of the 10th International Symposium* (pp. 29–54). Tokyo, Japan: The National International Institute for Japanese Language.

Johnson, N. F. (1970). The role of chunking and organization in the process of recall. *Psychology of Learning and Motivation*, *4*, 171–247.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*(4), 329–354.

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.

Ladefoged, P., & Broadbent, D. (1960). Perception of sequence in auditory events. *Quarterly Journal of Experimental Psychology, 12*, 162–170.

Levy, R., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In B. Schlo ̈kopf, J. Platt, & T. Hoffman (Eds.), *Advances in neural information processing systems* (NIPS) 19, 849–856. Cambridge, MA: MIT Press.

Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences, 106*(50), 21086–21090.

Levy, R., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In B. Scholkopf, J. Platt, & T. Hoffman (Eds.), *Advances in Neural Information Processing Systems 19: Proceedings of the 2006 Conference* (pp. 849–856). Cambridge, MA: MIT Press.

Lewis, S., & Tanenhaus, M. K. (2015). Aligning Grammatical Theories and Language Processing Models. *Journal of Psycholinguistic Research*, *44*(1), 27–46.

Lidz, J., & Gagliardi, A. (2015). How nature meets nurture: Universal grammar and statistical learning. *Annual Review of Linguistics, 1*, 333–353.

Lieven, E. V. M. (1994). Crosslinguistic and crosscultural aspects of language addressed to children. In C. Gallaway & B. J. Richards (Eds.), *Input and Interaction in Language Acquisition* (pp. 56–73). Cambridge, England: Cambridge University Press.

MacDonald, M. C. (1999). Distributional information in language comprehension, production, and acquisition: Three puzzles and a moral. In B MacWhinney (Ed.), *The emergence of language* (pp. 177–196). Hillsdale, NJ: Erlbaum.

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, *4*, 226.

Marslen-Wilson, W. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature, 244*, 522–523.

Marslen-Wilson, W. (1975). Sentence perception as an interactive parallel process. *Science*, *189*, 226–228.

Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(3), 576–585.

McColgan, K. (2011). *The Relationship between Maternal Language Outcomes* (Unpublished master's thesis). University of Maryland, College Park, MD.

Miller, G. A. (1962). Some psychological studies of grammar. *American Psychologist*, *17*(11), 748–762.

Moro, A. (2011). A closer look at the turtle's eyes. *Proceedings of the National Academy of Sciences*, *108*(6), 2177–2178.

Musso, M., Moro, A., Glauche, V., Rijntjes, M., Reichenbach, J., Büchel, C., & Weiller, C. (2003). Broca's area and the language instinct. *Nature Neuroscience*, *6*(7), 774–781.

Osgood, C. E. (1968). Toward a wedding of insufficiencies. In T. R. Dixon, & D. L. Horton (Eds.), *Verbal learning and general behavior theory* (pp. 495–519). Englewood Cliffs: NJ: Prentice-Hall.

Osgood, C., & Sebeok, T. (1954). Psycholinguistics: A survey of theory and research problems. *Journal of Abnormal and Social Psychology, 49*(4, part 2): 1–203.

Pearl, L. & Goldwater, S. (2016). Statistical learning, inductive bias, and Bayesian inference in language acquisition. In J. Lidz, W. Snyder, & C. Pater (Eds.), *The Oxford Handbook of Developmental Linguistics* (pp. 664–695). New York, NY: Oxford University Press.

Pearl, L. & Phillips, L. (in press). Evaluating language acquisition models: A utility-based look at Bayesian segmentation. In A. Villavicencio & T. Poibeau (Eds.), *Language, cognition and computational models.* Cambridge, England: Cambridge University Press.

Perfors, A., Tenenbaum, J., & Regier, T. (2006). Poverty of the stimulus? A rational approach. In R Sun and N Miyake (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 663–668). Mahwah, NJ: Erlbaum.

Phillips, C., & Lewis, S. (2013). Derivational order in syntax: Evidence and architectural consequences. *Studies in Linguistics*, *6*, 11–47.

Piantadosi, S., & Jacobs, R. (2016). Four problems solved by the probabilistic language of thought. *Current Directions in Psychological Science, 25*(1), 54–59.

Poeppel, D., Idsardi, W., & von Wassenhjove, V. (2007). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society Series B*. doi:10.1098/rstb.

Pollack, I., & Pickett, J. M. (1964). Intelligibility of excerpts from fluent speech: Auditory vs. structural context. *Journal of Verbal Learning and Verbal Behavior*, *3*, 79–84.

Reali, F., & Christiansen, M. H. (2005). Uncovering the richness of the stimulus: Structural dependence and indirect statistical evidence. *Cognitive Science*, *29*(6), 1007–1028.

Rumelhart, D. E., McClelland, J. L., & PDP Research Group (1986). *Parallel distributed processing: explorations in the microstructure of cognition*, Volumes I, II. Cambridge, MA: MIT Press.

Slattery, T. J., Sturt, P., Christianson, K., Yoshida, M., & Ferreira, F. (2013). Lingering misinterpretations of garden path sentences arise from competing syntactic representations. *Journal of Memory and Language*, *69*(2), 104–120.

Spence, C., & Parise, C. (2009). Prior-entry: A review. *Consciousness and Cognition*, *19*(1), 364–379.

Swinney, D. (1979). Lexical access during sentence comprehension: (Re) consideration of context effects. *Journal of Verbal Learning and Verbal Behavior, 18*, 645–659.

Titchener, E. B. (1908). *Lectures on the elementary psychology of feeling and attention*. New York, NY: Macmillan.

Townsend, D., & Bever, T. G. (1991). The use of higher-level constraints in monitoring for a change in speaker demonstrates functionally distinct levels of representation in discourse comprehension. *Language and Cognitive Processes*, *6*(1), 49–77.

Townsend, D. J., & Bever, T. G. (2001). *Sentence Comprehension: The integration of habits and rules*. Cambridge, MA: MIT Press.

Tucker, B. V., & Warner, N. (2010). What it means to be phonetic or phonological: The case of Romanian devoiced nasal. *Phonology*, *27*(2), 1–36.

Van de Ven, M. A. M. (2011). *The role of acoustic detail and context in the comprehension of reduced pronunciation variants (*PhD thesis*)*, Radboud Universiteit Nijmegen, Nijmegen, Netherlands.

Van de Weijer, J. (1998). *Language input for word discovery* (Ph.D. thesis). Nijmegen, Netherlands: Max Planck Series in Psycholinguistics, 9.

Warner, N., Fountain, A., & Tucker, B. V. (2009). Cues to perception of reduced flaps. *Journal of the Acoustical Society of America*, *125*(5), 3317–3327.

Wundt, W. (1918). *Grundriss der Psychologie*. Leipzig, Germany: Alfred Kroener Verlag.

Yang, C. (2002). *Knowledge and learning in natural language*. New York, NY: Oxford University Press.

Yang, C. (2004). Universal grammar, statistics or both? *Trends in Cognitive Sciences 8*(10), 451–456.

Yang, C., & Roeper, T. (2011). Minimalism and language acquisition. In C. Boeckx, (Ed.) *The Oxford handbook of linguistic minimalism*. New York, NY: Oxford University Press.