doi:10.1017/S0140525X1500031X, e62

The Now-or-Never bottleneck: A fundamental constraint on language

Morten H. Christiansen
Department of Psychology, Cornell University, Ithaca, NY 14853
The Interacting Minds Centre, Aarhus University, 8000 Aarhus C, Denmark
Haskins Laboratories, New Haven, CT 06511
christiansen@cornell.edu

Nick Chater
Behavioural Science Group, Warwick Business School, University of Warwick, Coventry, CV4 7AL, United Kingdom
nick.chater@wbs.ac.uk

Abstract: Memory is fleeting. New material rapidly obliterates previous material. How, then, can the brain deal successfully with the continual deluge of linguistic input? We argue that, to deal with this "Now-or-Never" bottleneck, the brain must compress and recode linguistic input as rapidly as possible. This observation has strong implications for the nature of language processing: (1) the language system must "eagerly" recode and compress linguistic input; (2) as the bottleneck recurs at each new representational level, the language system must build a multilevel linguistic representation; and (3) the language system must deploy all available information predictively to ensure that local linguistic ambiguities are dealt with "Right-First-Time"; once the original input is lost, there is no way for the language system to recover. This is "Chunk-and-Pass" processing. Similarly, language learning must also occur in the here and now, which implies that language acquisition is learning to process, rather than inducing, a grammar. Moreover, this perspective provides a cognitive foundation for grammaticalization and other aspects of language change. Chunk-and-Pass processing also helps explain a variety of core properties of language, including its multilevel representational structure and duality of patterning. This approach promises to create a direct relationship between psycholinguistics and linguistic theory. More generally, we outline a framework within which to integrate often disconnected inquiries into language processing, language acquisition, and language change and evolution.

Keywords: chunking; grammaticalization; incremental interpretation; language acquisition; language evolution; language processing; online learning; prediction; processing bottleneck; psycholinguistics

1. Introduction

Language is fleeting. As we hear a sentence unfold, we rapidly lose our memory for preceding material. Speakers, too, soon lose track of the details of what they have just said. Language processing is therefore "Now-or-Never": If linguistic information is not processed rapidly, that information is lost for good. Importantly, though, while fundamentally shaping language, the Now-or-Never bottleneck[1] is not specific to language but instead arises from general principles of perceptuo-motor processing and memory.

The existence of a Now-or-Never bottleneck is relatively uncontroversial, although its precise character may be debated. However, in this article we argue that the *consequences* of this constraint for language are remarkably far-reaching, touching on the following issues:

1. The multilevel organization of language into sound-based units, lexical and phrasal units, and beyond;
2. The prevalence of *local* linguistic relations (e.g., in phonology and syntax);
3. The incrementality of language processing;
4. The use of prediction in language interpretation and production;
5. The nature of what is learned during language acquisition;
6. The degree to which language acquisition involves item-based generalization;
7. The degree to which language *change* proceeds item-by-item;
8. The connection between grammar and lexical knowledge;
9. The relationships between syntax, semantics, and pragmatics.

Thus, we argue that the Now-or-Never bottleneck has fundamental implications for key questions in the language sciences. The consequences of this constraint are, moreover, incompatible with many theoretical positions in linguistic, psycholinguistic, and language acquisition research.

Note, however, that arguing that a phenomenon arises from the Now-or-Never bottleneck does not necessarily undermine alternative explanations of that phenomenon (although it may). Many phenomena in language may simply be overdetermined. For example, we argue that incrementality (point 3, above) follows from the Now-or-Never bottleneck. But it is also possible that, irrespective of memory constraints, language understanding would still

 0140-525X/16

be incremental on functional grounds, to extract the linguistic message as rapidly as possible. Such counterfactuals are, of course, difficult to evaluate. By contrast, the properties of the Now-or-Never bottleneck arise from basic information processing limitations that are directly testable by experiment. Moreover, the Now-or-Never bottleneck should, we suggest, have methodological priority to the extent that it provides an *integrated* framework for explaining many aspects of language structure, acquisition, processing, and evolution that have previously been treated separately.

In Figure 1, we illustrate the overall structure of the argument in this article. We begin, in the next section, by briefly making the case for the Now-or-Never bottleneck as a general constraint on perception and action. We then discuss the implications of this constraint for language processing, arguing that both comprehension and production involve what we call "Chunk-and-Pass" processing: incrementally building chunks at all levels of linguistic structure as rapidly as possible, using all available information predictively to process current input before new information arrives (sect. 3). From this perspective, language acquisition involves learning to process: that is, learning rapidly to create and use chunks appropriately for the language being learned (sect. 4). Consequently, short-term language change and longer-term processes of language evolution arise through variation in the system of chunks and their composition, suggesting an item-based theory of language change (sect. 5). This approach points to a processing-based interpretation of construction grammar, in which constructions correspond to chunks, and where grammatical structure is fundamentally the history of language processing operations within the individual speaker/hearer (sect. 6). We conclude by briefly summarizing the main points of our argument.

Morten H. Christiansen is Professor of Psychology and Co-Director of the Cognitive Science Program at Cornell University as well as Senior Scientist at the Haskins Labs and Professor of Child Language at the Interacting Minds Centre at Aarhus University. He is the author of more than 170 scientific papers and has written or edited five books. His research focuses on the interaction of biological and environmental constraints in the processing, acquisition, and evolution of language, using a combination of computational, behavioral, and cognitive neuroscience methods. He is a Fellow of the Association for Psychological Science, and he delivered the 2009 Nijmegen Lectures.

Nick Chater is Professor of Behavioural Science at Warwick Business School, United Kingdom. He is the author of more than 250 scientific publications in psychology, philosophy, linguistics, and cognitive science, and he has written or edited ten books. He has served as Associate Editor for *Cognitive Science*, *Psychological Review*, *Psychological Science*, and *Management Science*. His research explores the cognitive and social foundations of human rationality, focusing on formal models of inference, choice, and language. He is a Fellow of the Cognitive Science Society, the Association for Psychological Science, and the British Academy.

2. The Now-or-Never bottleneck

Language input is highly transient. Speech sounds, like other auditory signals, are short-lived. Classic speech perception studies have shown that very little of the auditory trace remains after 100 ms (Elliott 1962), with more recent studies indicating that much acoustic information already is lost after just 50 ms (Remez et al. 2010). Similarly, and of relevance for the perception of sign language, studies of visual change detection suggest that the ability to maintain visual information beyond 60–70 ms is very limited (Pashler 1988). Thus, sensory memory for language input is quickly overwritten, or interfered with, by new incoming information, *unless* the perceiver in some way processes what is heard or seen.

The problem of the rapid loss of the speech or sign signal is further exacerbated by the sheer speed of the incoming linguistic input. At a normal speech rate, speakers produce about 10–15 phonemes per second, corresponding to roughly 5–6 syllables every second or 150 words per minute (Studdert-Kennedy 1986). However, the resolution of the human auditory system for discrete auditory events is only about 10 sounds per second, beyond which the sounds fuse into a continuous buzz (Miller & Taylor 1948). Consequently, even at normal rates of speech, the language system needs to work beyond the limits of auditory temporal resolution for nonspeech stimuli. Remarkably, listeners can learn to process speech in their native language at up to twice the normal rate without much decrement in comprehension (Orr et al. 1965). Although the production of signs appears to be slower than the production of speech (at least when comparing the production of ASL signs and spoken English; Bellugi & Fischer 1972), signed words are still very brief visual events, with the duration of an ASL syllable being about a quarter of a second (Wilbur & Nolkn 1986).[2]

Making matters even worse, our memory for sequences of auditory input is also very limited. For example, it has been known for more than four decades that naïve listeners are unable to correctly recall the temporal order of just four distinct sounds – for example, hisses, buzzes, and tones – even when they are perfectly able to recognize and label each individual sound in isolation (Warren et al. 1969). Our ability to recall well-known auditory stimuli is not substantially better, ranging from 7 ± 2 (Miller 1956) to 4 ± 1 (Cowan 2000). A similar limitation applies to visual memory for sign language (Wilson & Emmorey 2006). The poor memory for auditory and visual information, combined with the fast and fleeting nature of linguistic input, imposes a fundamental constraint on the language system: the *Now-or-Never bottleneck*. If the input is not processed immediately, new information will quickly overwrite it.

Importantly, the Now-or-Never bottleneck is not unique to language but applies to other aspects of perception and action as well. Sensory memory is rich in detail but decays rapidly unless it is further processed (e.g., Cherry 1953; Coltheart 1980; Sperling 1960). Likewise, short-term memory for auditory, visual, and haptic information is also limited and subject to interference from new input (e.g., Gallace et al. 2006; Haber 1983; Pavani & Turatto 2008). Moreover, our cognitive ability to respond to sensory input is further constrained in a serial (Sigman & Dehaene 2005) or near-serial (Navon & Miller 2002) manner, severely restricting our capacity for processing

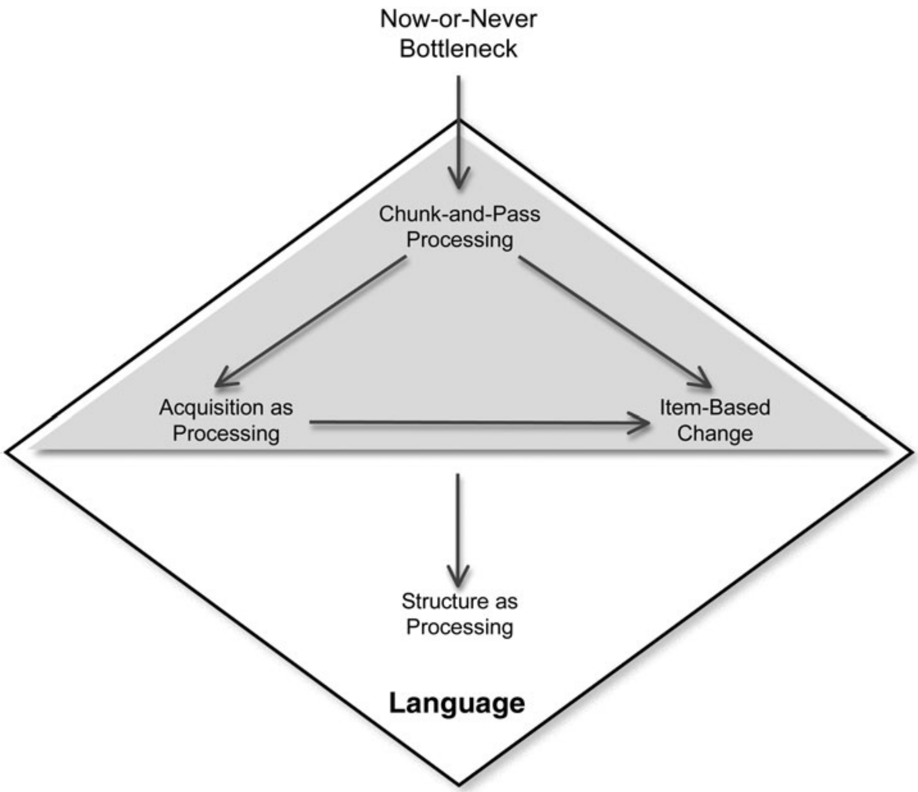

Figure 1. The structure of our argument, in which implicational relations between claims are denoted by arrows. The Now-or-Never bottleneck provides a fundamental constraint on perception and action that is independent of its application to the language system (and hence outside the diamond in the figure). Specific implications for language (indicated inside the diamond) stem from the Now-or-Never bottleneck's necessitating of Chunk-and-Pass language processing, with key consequences for language acquisition. The impact of the Now-or-Never bottleneck on both processing and acquisition together further shapes language change. All three of these interlinked claims concerning Chunk-and-Pass processing, acquisition as processing, and item-based language change (grouped together in the shaded upper triangle) combine to shape the structure of language itself.

multiple inputs arriving in quick succession. Similar limitations apply to the production of behavior: The cognitive system cannot plan detailed sequences of movements – a long sequence of commands planned far in advance would lead to severe interference and be forgotten before it could be carried out (Cooper & Shallice 2006; Miller et al. 1960). However, the cognitive system adopts several processing strategies to ameliorate the effects of the Now-or-Never bottleneck on perception and action.

First, the cognitive system engages in *eager processing:* It must recode the rich perceptual input as it arrives to capture the key elements of the sensory information as *economically*, and as *distinctively*, as possible (e.g., Brown et al. 2007; Crowder & Neath 1991); and it must do so rapidly, before new input overwrites or interferes with the sensory information. This notion is a traditional one, dating back to early work on attention and sensory memory (e.g., Broadbent 1958; Coltheart 1980; Haber 1983; Sperling 1960; Treisman 1964). The resulting compressed representations are lossy: They provide only an abstract summary of the input, from which the rich sensory input cannot be recovered (e.g., Pani 2000). Evidence from the phenomena of change and inattentional blindness suggests that these compressed representations can be very selective (see Jensen et al. 2011 for a review), as exemplified by a study in which half of the participants failed to notice that someone to whom they were giving directions, face-to-face, was surreptitiously exchanged for a completely different person (Simons & Levin 1998). Information not encoded in the short amount of time during which the sensory information is available will be lost.

Second, because memory limitations also apply to recoded representations, the cognitive system further chunks the compressed encodings into *multiple levels of representation* of increasing abstraction in perception, and decreasing levels of abstraction in action. Consider, for example, memory for serially ordered symbolic information, such as sequences of digits. Typically, people are quickly overloaded and can recall accurately only the last three or four items in a sequence (e.g., Murdock 1968). But it is possible to learn to rapidly encode, and recall, long random sequences of digits, by successively chunking such sequences into larger units, chunking those chunks into still larger units, and so on. Indeed, an extended study of a single individual, SF (Ericsson et al. 1980), showed that repeated chunking in this manner makes it possible to recall with high accuracy sequences containing as many as 79 digits. But, crucially, this strategy requires learning to encode the input into multiple, successive, and distinct levels of representations – each sequence of chunks at one level must be shifted as a single chunk to a higher level before more chunks interfere with or overwrite the initial chunks. Indeed, SF chunked sequences of three or four digits, the natural chunk size in human memory

(Cowan 2000), into a single unit (corresponding to running times, dates, or human ages), and then grouped sequences of three to four of those chunks into larger chunks. Interestingly, SF also verbally produced items in overtly discernible chunks, interleaved with pauses, indicating how action also follows the reverse process (e.g., Lashley 1951; Miller 1956). The case of SF further demonstrates that low-level information is far better recalled when organized into higher-level structures than merely coded as an unorganized stream. Note, though, that lower-level information is typically forgotten; it seems unlikely that even SF could recall the specific visual details of the digits with which he was presented. More generally, the notion that perception and action involve representational recoding at a succession of distinct representational levels also fits with a long tradition of theoretical and computational models in cognitive science and computer vision (e.g., Bregman 1990; Marr 1982; Miller et al. 1960; Zhu et al. 2010; see Gobet et al. 2001 for a review). Our perspective on repeated multilevel compression is also consistent with data from functional magnetic resonance imaging (fMRI) and intracranial recordings, suggesting cortical hierarchies across vision and audition – from low-level sensory to high-level perceptual and cognitive areas – integrating information at progressively longer temporal windows (Hasson et al. 2008; Honey et al. 2012; Lerner et al. 2011).

Third, to facilitate speedy chunking and hierarchical compression, the cognitive system employs *anticipation*, using prior information to constrain the recoding of current perceptual input (for reviews see Bar 2007; Clark 2013). For example, people see the exact same collection of pixels either as a hair dryer (when viewed as part of a bathroom scene) or as a drill (when embedded in a picture of a workbench) (Bar 2004). Therefore, using prior information to *predict* future input is likely to be essential *to successfully encoding that future input* (as well as helping us to react faster to such input). Anticipation allows faster, and hence more effective, recoding when oncoming information creates considerable time urgency. Such predictive processing will be most effective to the extent that the greatest possible amount of available information (across different types and levels of abstraction) is integrated as fast as possible. Similarly, anticipation is important for action as well. For example, manipulating an object requires anticipating the grip force required to deal with the loads generated by the accelerations of the object. Grip force is adjusted too rapidly during the manipulation of an object to rely on sensory feedback (Flanagan & Wing 1997). Indeed, the rapid prediction of the sensory consequences of actions (e.g., Poulet & Hedwig 2006) suggests the existence of so-called forward models, which allow the brain to predict the consequence of its actions in real time. Many have argued (e.g., Wolpert et al. 2011; see also Clark 2013; Pickering & Garrod 2013a) that forward models are a ubiquitous feature of the computational machinery of motor control and more broadly of cognition.

The three processing strategies we mention here – eager processing, computing multiple representational levels, and anticipation – provide the cognitive system with important means to cope with the Now-or-Never bottleneck. Next, we argue that the language system implements similar strategies for dealing with the here-and-now nature of linguistic input and output, with wide-reaching and fundamental implications for language processing, acquisition and change as well as for the structure of language itself. Specifically, we propose that our ability to deal with sequences of linguistic information is the result of what we call "Chunk-and-Pass" processing, by which the language system can ameliorate the effects of the Now-or-Never bottleneck. More generally, our perspective offers a framework within which to approach language comprehension and production. Table 1 summarizes the impact of the Now-or-Never bottleneck on perception/action and language.

The style of explanation outlined here, focusing on processing limitations, contrasts with a widespread interest in rational, rather processing-based, explanations in cognitive science (e.g., Anderson 1990; Chater et al. 2006 Griffiths & Tenenbaum 2009; Oaksford & Chater 1998; 2007; Tenenbaum et al. 2011), including language processing (Gibson et al. 2013; Hale 2001; 2006; Piantadosi et al. 2011). Given the fundamental nature of the Now-or-Never bottleneck, we suggest that such explanations will be relevant only for explaining language use insofar as they incorporate processing constraints. For example, in the spirit of rational analysis (Anderson 1990) and bounded rationality (Simon 1982), it is natural to view aspects of language processing and structure, as described below, as "optimal" responses

Table 1. *Summary of the Now-or-Never bottleneck's implications for perception/action and language*

Strategies	Mechanisms	Perception and action	Language
Eager processing	Lossy chunking	Chunking in memory and action (Lashley 1951; Miller 1956); lossy descriptions (Pani 2000)	Incremental interpretation (Bever 1970) and production (Meyer 1996); multiple constraints satisfaction (MacDonald et al. 1994)
Multiple levels of representation	Hierarchical compression	Hierarchical memory (Ericsson et al. 1980), action (Miller et al. 1960), problem solving (Gobet et al. 2001)	Multiple levels of linguistic structure (e.g., sound-based, lexical, phrasal, discourse); local dependencies (Hawkins 2004)
Anticipation	Predictive processing	Fast, top-down visual processing (Bar 2004); forward models in motor control (Wolpert et al. 2011); predictive coding (Clark 2013)	Syntactic prediction (Jurafsky 1996); multiple-cue integration (Farmer et al. 2006); visual world (Altmann & Kamide 1999)

to specific processing limitations, such as the Now-or-Never bottleneck (for this style of approach, see, e.g., Chater et al. 1998; Levy 2008). Here, though, our focus is primarily on mechanism rather than rationality.

3. Chunk-and-Pass language processing

The fleeting nature of linguistic input, in combination with the impressive speed with which words and signs are produced, imposes a severe constraint on the language system: the Now-or-Never bottleneck. Each new incoming word or sign will quickly interfere with previous heard and seen input, providing a naturalistic version of the masking used in psychophysical experiments. How, then, is language comprehension possible? Why doesn't interference between successive sounds (or signs) obliterate linguistic input before it can be understood? The answer, we suggest, is that our language system rapidly recodes this input into chunks, which are immediately passed to a higher level of linguistic representation. The chunks at this higher level are then themselves subject to the same Chunk-and-Pass procedure, resulting in progressively larger chunks of increasing linguistic abstraction. Crucially, given that the chunks recode increasingly larger stretches of input from lower levels of representation, the chunking process enables input to be maintained over ever-larger temporal windows. It is this repeated chunking of lower-level information that makes it possible for the language system to deal with the continuous deluge of input that, if not recoded, is rapidly lost. This chunking process is also what allows us to perceive speech at a much faster rate than nonspeech sounds (Warren et al. 1969): We have learned to chunk the speech stream. Indeed, we can easily understand (and sometimes even repeat back) sentences consisting of many tens of phonemes, despite our severe memory limitations for sequences of nonspeech sounds.

What we are proposing is that during comprehension, the language system – similar to SF – must keep on chunking the incoming information into increasingly abstract levels of representation to avoid being overwhelmed by the input. That is, the language system engages in *eager* processing when creating chunks. Chunks must be built right away, or memory for the input will be obliterated by interference from subsequent material. If a phoneme or syllable is recognized, then it is recoded as a chunk and passed to a higher level of linguistic abstraction. And once recoded, the information is no longer subject to interference from further auditory input. A general principle of perception and memory is that interference arises primarily between overlapping representations (Crowder & Neath 1991; Treisman & Schmidt 1982); crucially, recoding avoids such overlap. For example, phonemes interfere with each other, but phonemes interfere very little with words. At each level of chunking, information from the previous level(s) is compressed and passed up as chunks to the next level of linguistic representation, from sound-based chunks up to complex discourse elements.[3] As a consequence, the rich detail of the original input can no longer be recovered from the chunks, although some key information remains (e.g., certain speaker characteristics; Nygaard et al. 1994; Remez et al. 1997).

In production, the process is reversed: Discourse-level chunks are recursively broken down into subchunks of decreasing linguistic abstraction until the system arrives at chunks with sufficient information to drive the articulators (either the vocal apparatus or the hands). As in comprehension, memory is limited within a given level of representation, resulting in potential interference between the items to be produced (e.g., Dell et al. 1997). Thus, higher-level chunks tend to be passed down immediately to the level below as soon as they are "ready," leading to a bias toward producing easy-to-retrieve utterance components before harder-to-retrieve ones (e.g., Bock 1982; MacDonald 2013). For example, if there is a competition between two possible words to describe an object, the word that is retrieved more fluently will immediately be passed on to lower-level articulatory processes. To further facilitate production, speakers often reuse chunks from the ongoing conversation, and those will be particularly rapidly available from memory. This phenomenon is reflected by the evidence for lexical (e.g., Meyer & Schvaneveldt 1971) and structural priming (e.g., Bock 1986; Bock & Loebell 1990; Pickering & Branigan 1998; Potter & Lombardi 1998) within individuals as well as alignment across conversational partners (Branigan et al. 2000; Pickering & Garrod 2004); priming is also extensively observed in text corpora (Hoey 2005). As noted by MacDonald (2013), these memory-related factors provide key constraints on the production of language and contribute to cross-linguistic patterns of language use.[4]

A useful analogy for language production is the notion of "just-in-time"[5] stock control, in which stock inventories are kept to a bare minimum during the manufacturing process (Ohno & Mito 1988). Similarly, the Now-or-Never bottleneck requires that, for example, low-level phonetic or articulatory decisions not be made and stored far in advance and then reeled off during speech production, because any buffer in which such decisions can safely be stored would quickly be subject to interference from subsequent material. So the Now-or-Never bottleneck requires that once detailed production information has been assembled, it be executed straightaway, before it can be obliterated by the oncoming stream of later low-level decisions, similar to what has been suggested for motor planning (Norman & Shallice 1986; see also MacDonald 2013). We call this proposal Just-in-Time language production.

3.1. *Implications of Strategy 1: Incremental processing*

Chunk-and-Pass processing has important implications for comprehension and production: It requires that both take place *incrementally*. In incremental processing, representations are built up as rapidly as possible as the input is encountered. By contrast, one might, for example, imagine a parser that waits until the end of a sentence before beginning syntactic analysis, or that meaning is computed only once syntax has been established. However, such processing would require storing a stream of information at a single level of representation, and processing it later; but given the Now-or-Never bottleneck, this is not possible because of severe interference between such representations. Therefore, incremental interpretation and production follow directly from the Now-or-Never constraint on language.

To get a sense of the implications of Chunk-and-Pass processing, it is interesting to relate this perspective to specific computational principles and models. How, for

https://doi.org/10.1017/S0140525X1500031X

example, do classic models of parsing fit within this framework? A wide range of psychologically inspired models involves some degree of incrementality of syntactic analysis, which can potentially support incremental interpretation (e.g., Phillips 1996; 2003; Winograd 1972). For example, the sausage machine parsing model (Frazier & Fodor 1978) proposes that a preliminary syntactic analysis is carried out phrase-by-phrase, but in complete isolation from semantic or pragmatic factors. But for a right-branching language such as English, chunks cannot be built left-to-right, because the leftmost chunks are incomplete until later material has been encountered. Frameworks from Kimball (1973) onward imply "stacking up" incomplete constituents that may then all be resolved at the end of the clause. This approach runs counter to the memory constraints imposed by the Now-or-Never bottleneck. Reconciling right-branching with incremental chunking and processing is one motivation for the flexible constituency of combinatory categorial grammar (e.g., Steedman 1987; 2000; see also Johnson-Laird 1983).

With respect to comprehension, considerable evidence supports incremental interpretation, going back more than four decades (e.g., Bever 1970; Marslen-Wilson 1975). The language system uses all available information to rapidly integrate incoming information as quickly as possible to update the current interpretation of what has been said so far. This process includes not only sentence-internal information about lexical and structural biases (e.g., Farmer et al. 2006; MacDonald 1994; Trueswell et al. 1993), but also extra-sentential cues from the referential and pragmatic context (e.g., Altmann & Steedman 1988; Thornton et al. 1999) as well as the visual environment and world knowledge (e.g., Altmann & Kamide 1999; Tanenhaus et al. 1995). As the incoming acoustic information is chunked, it is rapidly integrated with contextual information to recognize words, consistent with a variety of data on spoken word recognition (e.g., Marslen-Wilson 1975; van den Brink et al. 2001). These words are then, in turn, chunked into larger multiword units, as evidenced by recent studies showing sensitivity to multiword sequences in online processing (e.g., Arnon & Snider 2010; Reali & Christiansen 2007b; Siyanova-Chanturia et al. 2011; Tremblay & Baayen 2010; Tremblay et al. 2011), and subsequently further integrated with pragmatic context into discourse-level structures.

Turning to production, we start by noting the powerful intuition that we speak "into the void" – that is, that we plan only a short distance ahead. Indeed, experimental studies suggest that, for example, when producing an utterance involving several noun phrases, people plan just one (Smith & Wheeldon 1999), or perhaps two, noun phrases ahead (Konopka 2012), and they can modify a message during production in the light of new perceptual input (Brown-Schmidt & Konopka 2015). Moreover, speech-error data (e.g., Cutler 1982) reveal that, across representational levels, errors tend to be highly local: Phonological, morphemic, and syntactic errors apply to neighboring chunks within each level (where material may be moved, swapped, or deleted). Consequently, speech planning appears to involve just a small number of chunks – the number of which may be similar across linguistic levels – but which covers different amounts of time depending on the linguistic level in question. For example, planning involving chunks at the level of intonational bursts stretches over considerably longer periods of time than planning at the syllabic level. Similarly, processes of reduction to facilitate production (e.g., modifying the speech signal to make it easier to produce, such as reducing a vowel to a schwa, or shortening or eliminating phonemes) can be observed across different levels of linguistic representation, from individual words (e.g., Gahl & Garnsey 2004; Jurafsky et al. 2001) to frequent multiword sequences (e.g., Arnon & Cohen Priva 2013; Bybee & Schiebman 1999).

Some may object that the Chunk-and-Pass perspective's strict notion of incremental interpretation and production leaves the language system vulnerable to the rather substantial ambiguity that exists across many levels of linguistic representation (e.g., lexical, syntactic, pragmatic). So-called garden path sentences such as the famous "The horse raced past the barn fell" (Bever 1970) show that people are vulnerable to at least some local ambiguities: They invite comprehenders to take the wrong interpretive path by treating *raced* as the main verb, which leads them to a dead end. Only when the final word, *fell,* is encountered does it become clear that something is wrong: *raced* should be interpreted as a past participle that begins a reduced relative clause (i.e., the horse [that was] raced past the barn fell). The difficulty of recovery in such garden path sentences indicates how strongly the language system is geared toward incremental interpretation.

Viewed as a processing problem, garden paths occur when the language system resolves an ambiguity incorrectly. But in many cases, it is possible for an underspecified representation to be constructed online, and for the ambiguity to be resolved later when further linguistic input arrives. This type of case is consistent with Marr's (1976) proposal of the "principle of least commitment," that the perceptual system resolves ambiguous perceptual input only when it has sufficient data to make it unlikely that such decisions will subsequently have to be reversed. Given the ubiquity of local ambiguity in language, such underspecification may be used very widely in language processing. Note, however, that because of the severe constraints the Now-or-Never bottleneck imposes, the language system cannot adopt broad parallelism to further minimize the effect of ambiguity (as in many current probabilistic theories of parsing, e.g., Hale 2006; Jurafsky 1996; Levy 2008). Rather, within the Chunk-and-Pass account, the sole role for parallelism in the processing system is in deciding how the input should be chunked; only when conflicts concerning chunking are resolved can the input be passed on to a higher-level representation. In particular, we suggest that competing higher-level codes cannot be activated in parallel. This picture is analogous to Marr's principle of least commitment of vision: Although there might be temporary parallelism to resolve conflicts about, say, correspondence between dots in a random-dot stereogram, it is not possible to create two conflicting three-dimensional surfaces in parallel, and whereas there may be parallelism over the interpretation of lines and dots in an image, it is not possible to see something as both a duck and a rabbit simultaneously. More broadly, higher-level representations are constructed only when sufficient evidence has accrued that they are unlikely later to need to be replaced (for stimuli outside the psychological laboratory, at least).

Maintaining, and later resolving, an underspecified representation will create local memory and processing demands that may slow down processing, as is observed, for example, by increased reading times (e.g., Trueswell et al. 1994) and distinctive patterns of brain activity (as measured by ERPs; Swaab et al. 2003). Accordingly, when the input is ambiguous, the language system may require later input to recognize previous elements of the speech stream successfully. The Now-or-Never bottleneck requires that such online "right-context effects" be highly local because raw perceptual input will be lost if it is not rapidly identified (e.g., Dahan 2010). Right-context effects may arise where the language system can delay resolution of ambiguity or use underspecified representations that do not require resolving the ambiguity right away. Similarly, cataphora, in which, for example, a referential pronoun occurs before its referent (e.g., "He is a nice guy, that John") require the creation of an underspecified entity (male, animate) when *he* is encountered, which is resolved to be coreferential with John only later in the sentence (e.g., van Gompel & Liversedge 2003). Overall, the Now-or-Never bottleneck implies that the processing system will build the most abstract and complete representation that is justified, given the linguistic input.[6]

Of course, outside of experimental studies, background knowledge, visual context, and prior discourse will provide powerful cues to help resolve ambiguities in the signal, allowing the system rapidly to resolve many apparent ambiguities without incurring a substantial danger of "garden-pathing." Indeed, although syntactic and lexical ambiguities have been much studied in psycholinguistics, increasing evidence indicates that garden paths are not a major source of processing difficulty in practice (e.g., Ferreira 2008; Jaeger 2010; Wasow & Arnold 2003).[7] For example, Roland et al. (2006) reported corpus analyses showing that, in naturally occurring language, there is generally sufficient information in the sentential context before the occurrence of an ambiguous verb to specify the correct interpretation of that verb. Moreover, eye-tracking studies have demonstrated that dialogue partners exploit both conversational context and task demands to constrain interpretations to the appropriate referents, thereby side-stepping effects of phonological and referential competitors (Brown-Schmidt & Konopka 2011) that have otherwise been shown to impede language processing (e.g., Allopenna et al. 1998). These dialogue-based constraints also mitigate syntactic ambiguities that might otherwise disrupt processing (Brown-Schmidt & Tanenhaus 2008). This information may be further combined with other probabilistic sources of information such as prosody (e.g., Kraljic & Brennan 2005; Snedeker & Trueswell 2003) to resolve potential ambiguities within a minimal temporal window. Finally, it is not clear that undetected garden path errors are costly in normal language use, because if communication appears to break down, the listener can repair the communication by requesting clarification from the dialogue partner.

3.2. Implications of Strategy 2: Multiple levels of linguistic structure

The Now-or-Never bottleneck forces the language system to compress input into increasingly abstract chunks that cover progressively longer temporal intervals. As an example, consider the chunking of the input illustrated in Figure 2. The acoustic signal is first chunked into higher-level sound units at the phonological level. To avoid interference between local sound-based units, such as phonemes or syllables, these units are further recoded as rapidly as possible into higher-level units such as morphemes or words. The same phenomenon occurs at the

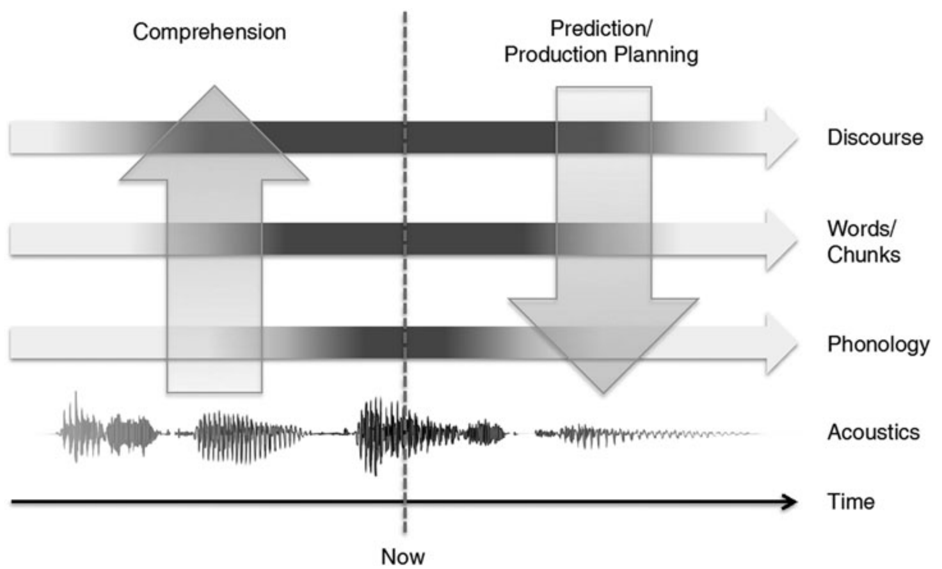

Figure 2. Chunk-and-Pass processing across a variety of linguistic levels in spoken language. As input is chunked and passed up to increasingly abstract levels of linguistic representations in comprehension, from acoustics to discourse, the temporal window over which information can be maintained increases, as indicated by the shaded portion of the bars associated with each linguistic level. This process is reversed in production planning, in which chunks are broken down into sequences of increasingly short and concrete units, from a discourse-level message to the motor commands for producing a specific articulatory output. More-abstract representations correspond to longer chunks of linguistic material, with greater look-ahead in production at higher levels of abstraction. Production processes may further serve as the basis for predictions to facilitate comprehension and thus provide top-down information in comprehension. (Note that the names and number of levels are for illustrative purposes only.)

next level up: Local groups of words must be chunked into larger units, possibly phrases or other forms of multiword sequences. Subsequent chunking then recodes these representations into higher-level discourse structures (that may themselves be chunked further into even more abstract representational structures beyond that). Similarly, production requires running the process in reverse, starting with the intended message and gradually decoding it into increasingly more specific chunks, eventually resulting in the motor programs necessary for producing the relevant speech or sign output. As we discuss in section 3.3, the production process may further serve as the basis for prediction during comprehension (allowing higher-level information to influence the processing of current input). More generally, our account is agnostic with respect to the specific characterization of the various levels of linguistic representation[8] (e.g., whether sound-based chunks take the form of phonemes, syllables, etc.). What is central for the Chunk-and-Pass account: some form of sound-based level of chunking (or visual-based in the case of sign language), and a sequence of increasingly abstract levels of chunked representations into which the input is continually recoded.

A key theoretical implication of Chunk-and-Pass processing is that the multiple levels of linguistic representation, typically assumed in the language sciences, are a necessary by-product of the Now-or-Never bottleneck. Only by compressing the input into chunks and passing them to increasingly abstract levels of linguistic representation can the language system deal with the rapid onslaught of incoming information. Crucially, though, our perspective also suggests that the different levels of linguistic representations do not have a true part–whole relationship with one another. Unlike in the case of SF, who learned strategies to perfectly unpack chunks from within chunks to reproduce the original string of digits, language comprehension typically employs lossy compression to chunk the input. That is, higher-level chunks will not in general contain complete copies of lower-level chunks. Indeed, as speech input is encoded into ever more abstract chunks, increasing amounts of low-level information will typically be lost. Instead, as in perception (e.g., Haber 1983), there is greater representational underspecification with higher levels of representation because of the repeated process of lossy compression.[9] Thus, we would expect a growing involvement of extralinguistic information, such as perceptual input and world knowledge, in processing higher levels of linguistic representation (see, e.g., Altmann & Kamide 2009).

Whereas our account proposes a lossy hierarchy across levels of linguistic representation, only a very small number of chunks are represented *within* a level: otherwise, information is rapidly lost due to interference. This has the crucial implication that chunks within a given level can interact only *locally*. For example, acoustic information must rapidly be coded in a non-acoustic form, say, in terms of phonemes; but this is only possible if phonemes correspond to local chunks of acoustic input. The processing bottleneck therefore enforces a strong pressure toward local dependencies within a given linguistic level. Importantly, though, this does not imply that linguistic relations are restricted only to adjacent elements but, instead, that they may be formed between any of the small number of elements maintained at a given level of representation. Such representational locality is exemplified across different linguistic levels by the local nature of phonological processes from reduction, assimilation, and fronting, including more elaborate phenomena such as vowel harmony (e.g., Nevins 2010), speech errors (e.g., Cutler 1982), the immediate proximity of inflectional morphemes and the verbs to which they apply, and the vast literature on the processing difficulties associated with non-local dependencies in sentence comprehension (e.g., Gibson 1998; Hawkins 2004). As noted earlier, the higher the level of linguistic representation, the longer the limited time window within which information can be chunked. Whereas dealing with just two center-embeddings at the sentential level is prohibitively difficult (e.g., de Vries et al. 2011; Karlsson 2007), we are able to deal with up to four to six embeddings at the multi-utterance discourse level (Levinson 2013). This is because chunking takes place at a much longer time course at the discourse level compared with the sentence level, providing more time to resolve the relevant dependency relations before they are subject to interference.

Finally, as indicated by Figure 2, processing within each level of linguistic representation takes place in parallel – but with a clear temporal component – as chunks are passed between levels. Note that, in the Chunk-and-Pass framework, it is entirely possible that linguistic input can simultaneously, and perhaps redundantly, be chunked in more than one way. For example, syntactic chunks and intonational contours may be somewhat independent (Jackendoff 2007). Moreover, we should expect further chunking across different "channels" of communication, including visual input such as gesture and facial expressions.

The Chunk-and-Pass perspective is compatible with a number of recent theoretical models of sentence comprehension, including constraint-based approaches (e.g., MacDonald et al. 1994; Trueswell & Tanenhaus 1994) and certain generative accounts (e.g., Jackendoff's [2007] parallel architecture). Intriguingly, fMRI data from adults (Dehaene-Lambertz et al. 2006a) and infants (Dehaene-Lambertz et al. 2006b) indicate that activation responses to a single sentence systematically slows down when moving away from the primary auditory cortex, either back toward Wernicke's area or forward toward Broca's area, consistent with increasing temporal windows for chunking when moving from phonemes to words to phrases. Indeed, the cortical circuits processing auditory input, from lower (sensory) to higher (cognitive) areas, follow different temporal windows, sensitive to more and more abstract levels of linguistic information, from phonemes and words to sentences and discourse (Lerner et al. 2011; Stephens et al. 2013). Similarly, the reverse process, going from a discourse-level representation of the intended message to the production of speech (or sign) across parallel linguistic levels, is compatible with several current models of language production (e.g., Chang et al. 2006; Dell et al. 1997; Levelt 2001). Data from intracranial recordings during language production are consistent with different temporal windows for chunk decoding at the word, morphemic, and phonological levels, separated by just over a tenth of a second (Sahin et al. 2009). These results are compatible with our proposal that incremental processing in comprehension and production takes place in parallel across multiple levels of linguistic representation, each with a characteristic temporal window.

3.3. Implications of Strategy 3: Predictive language processing

We have already noted that, to be able to chunk incoming information as fast and as accurately as possible, the language system exploits multiple constraints in parallel across the different levels of linguistic representation. Such cues may be used not only to help disambiguate previous input, but also to generate expectations for what may come next, potentially further speeding up Chunk-and-Pass processing. Computational considerations indicate that simple statistical information gleaned from sentences provides powerful predictive constraints on language comprehension and can explain many human processing results (e.g., Christiansen & Chater 1999; Christiansen & MacDonald 2009; Elman 1990; Hale 2006; Jurafsky 1996; Levy 2008; Padó et al. 2009). Similarly, eye-tracking data suggest that comprehenders routinely use a variety of sources of probabilistic information – from phonological cues to syntactic context and real-world knowledge – to anticipate the processing of upcoming words (e.g., Altmann & Kamide 1999; Farmer et al. 2011; Staub & Clifton 2006). Results from event-related potential experiments indicate that rather specific predictions are made for upcoming input, including its lexical category (Hinojosa et al. 2005), grammatical gender (Van Berkum et al. 2005; Wicha et al. 2004), and even its onset phoneme (DeLong et al. 2005) and visual form (Dikker et al. 2010). Accordingly, there is a growing body of evidence for a substantial role of prediction in language processing (for reviews, see, e.g., Federmeier 2007; Hagoort 2009; Kamide 2008; Kutas et al. 2014; Pickering & Garrod 2007) and evidence that such language prediction occurs in children as young as 2 years of age (Mani & Huettig 2012). Importantly, as well as exploiting statistical relations within a representational level, predictive processing allows top-down information from higher levels of linguistic representation to rapidly constrain the processing of the input at lower levels.[10]

From the viewpoint of the Now-or-Never bottleneck, prediction provides an opportunity to begin Chunk-and-Pass processing as early as possible: to constrain representations of new linguistic material as it is encountered, and even incrementally to begin recoding predictable linguistic input before it arrives. This viewpoint is consistent with recent suggestions that the production system may be pressed into service to anticipate upcoming input (e.g., Pickering & Garrod 2007; 2013a). Chunk-and-Pass processing implies that there is practically no possibility for going back once a chunk is created because such backtracking tends to derail processing (e.g., as in the classic garden path phenomena mentioned above). This imposes a *Right-First-Time* pressure on the language system in the face of linguistic input that is highly locally ambiguous.[11] The contribution of predictive modeling to comprehension is that it facilitates local ambiguity resolution while the stimulus is still available. Only by recruiting multiple cues and integrating these with predictive modeling is it possible to resolve local ambiguities quickly and correctly.

Right-First-Time parsing fits with proposals such as that by Marcus (1980), where local ambiguity resolution is delayed until later disambiguating information arrives, and models in which aspects of syntactic structure may be underspecified, therefore not requiring the ambiguity to be resolved (e.g., Gorrell 1995; Sturt & Crocker 1996). It also parallels Marr's (1976) principle of least commitment, as we mentioned earlier, according to which the perceptual system should, as far as possible, only resolve perceptual ambiguities when sufficiently confident that they will not need to be undone. Moreover, it is compatible with the fine-grained weakly parallel interactive model (Altmann & Steedman 1988) in which possible chunks are proposed, word-by-word, by an autonomous parser and one is rapidly chosen using top-down information.

To facilitate chunking across multiple levels of representation, prediction takes place in parallel across the different levels but at varying timescales. Predictions for higher-level chunks may run ahead of those for lower-level chunks. For example, most people simply answer "two" in response to the question "How many animals of each kind did Moses take on the Ark?" – failing to notice the semantic anomaly (i.e., it was Noah's Ark, not Moses' Ark) even in the absence of time pressure and when made aware that the sentence may be anomalous (Erickson & Matteson 1981). That is, anticipatory pragmatic and communicative considerations relating to the required response appear to trump lexical semantics. More generally, the time course of normal conversation may lead to an emphasis on more temporally extended higher-level predictions over lower-level ones. This may facilitate the rapid turn-taking that has been observed cross-culturally (Stivers et al. 2009) and which seems to require that listeners make quite specific predictions about when the speaker's current turn will finish (Magyari & De Ruiter 2012), as well as being able to quickly adapt their expectations to specific linguistic environments (Fine et al. 2013).

We view the anticipation of turn-taking as one instance of the broader alignment that takes place between dialogue partners across all levels of linguistic representation (for a review, see Pickering & Garrod 2004). This dovetails with fMRI analyses indicating that although there are some comprehension- and production-specific brain areas, spatiotemporal patterns of brain activity are in general closely coupled between speakers and listeners (e.g., Silbert et al. 2014). In particular, Stephens et al. (2010) observed close synchrony between neural activations in speakers and listeners in early auditory areas. Speaker activations preceded those of listeners in posterior brain regions (including parts of Wernicke's area), whereas listener activations preceded those of speakers in the striatum and anterior frontal areas. In the Chunk-and-Pass framework, the listener lag primarily derives from delays caused by the chunking process across the various levels of linguistic representation, whereas the speaker lag predominantly reflects the listener's anticipation of upcoming input, especially at the higher levels of representation (e.g., pragmatics and discourse). Strikingly, the extent of the listener's anticipatory brain responses were strongly correlated with successful comprehension, further underscoring the importance of prediction-based alignment for language processing. Indeed, analyses of real-time interactions show that alignment increases when the communicative task becomes more difficult (Louwerse et al. 2012). By decreasing the impact of potential ambiguities, alignment thus makes processing as well as production easier in the face of the Now-or-Never bottleneck.

We have suggested that only an incremental, predictive language system, continually building and passing on new chunks of linguistic material, encoded at increasingly abstract levels of representation, can deal with the onslaught of linguistic input in the face of the severe memory constraints of the Now-or-Never bottleneck. We suggest that a productive line of future work is to consider the extent to which existing models of language are compatible with these constraints, and to use these properties to guide the creation of new theories of language processing.

4. Acquisition is learning to process

If speaking and understanding language involves Chunk-and-Pass processing, then acquiring a language requires *learning* how to create and integrate the right chunks rapidly, before current information is overwritten by new input. Indeed, the ability to quickly process linguistic input – which has been proposed as an indicator of chunking ability (Jones 2012) – is a strong predictor of language acquisition outcomes from infancy to middle childhood (Marchman & Fernald 2008). The importance of this process is also introspectively evident to anyone acquiring a second language: Initially, even segmenting the speech stream into recognizable sounds can be challenging, let alone parsing it into words or processing morphology and grammatical relations rapidly enough to build a semantic interpretation. The ability to acquire and rapidly deploy a hierarchy of chunks at different linguistic scales is parallel to the ability to chunk sequences of motor movements, numbers, or chess positions: It is a *skill*, built up by continual practice.

Viewing language acquisition as continuous with other types of skill learning is very different from the standard formulation of the problem of language acquisition in linguistics. There, the child is viewed as a *linguistic theorist* who has the goal of inferring an abstract grammar from a corpus of example sentences (e.g., Chomsky 1957; 1965) and only secondarily learning the *skill* of generating and understanding language. But perhaps the child is not a mini-linguist. Instead, we suggest that *language acquisition is nothing more than learning to process*: to turn meanings into streams of sound or sign (when generating language), and to turn streams of sound or sign back into meanings (when understanding language).

If linguistic input is available only fleetingly, then any learning must occur while that information is present; that is, learning must occur in real time, as the Chunk-and-Pass process takes place. Accordingly, any modifications to the learner's cognitive system in light of processing must, according to the Now-or-Never bottleneck, occur *at the time of processing*. The learner must learn to *chunk* the input appropriately – to learn to recode the input at successively more abstract linguistic levels; and to do this requires, of course, learning the structure of the language being spoken. But how is this structure learned?

We suggest that, in language acquisition, as in other areas of perceptual-motor learning, people learn *by processing*, and that past processing leaves traces that can facilitate future processing. What, then, is retained, so that language processing gradually improves? We can consider various possibilities: For example, the weights of a connectionist network can be updated online in the light of current processing (Rumelhart et al. 1986a); in an exemplar-based model, traces of past examples can be reused in the future (e.g., Hintzman 1988; Logan 1988; Nosofsky 1986). Whatever the appropriate computational framework, the Now-or-Never bottleneck requires that language acquisition be viewed as a type of skill learning, such as learning to drive, juggle, play the violin, or play chess. Such skills appear to be learned through *practicing* the skill, using online feedback during the practice itself, although the consolidation of learning occurs subsequently (Schmidt & Wrisberg 2004). The challenge of language acquisition is to learn a dazzling sequence of rapid processing operations, rather than conjecturing a correct "linguistic theory."

4.1. Implications of Strategy 1: Online learning

The Now-or-Never bottleneck implies that learning can depend only on material currently being processed. As we have seen, this implication requires a processing strategy according to which modification to current representations (in this context, learning) occurs right away; in machine-learning terminology, learning is *online.* If learning does not occur at the time of processing, the representation of linguistic material will be obliterated, and the opportunity for learning will be gone forever. To facilitate such online learning, the child must learn to use all available information to help constrain processing. The integration of multiple constraints – or *cues* – is a fundamental component of many current theories of language acquisition (see, e.g., contributions in Golinkoff et al. 2000; Morgan & Demuth 1996; Weissenborn & Höhle 2001; for a review, see Monaghan & Christiansen 2008). For example, second-graders' initial guesses about whether a novel word refers to an object or an action are affected by that word's phonological properties (Fitneva et al. 2009); 7-year-olds use visual context to constrain online sentence interpretation (Trueswell et al. 1999); and preschoolers' language production and comprehension is constrained by pragmatic factors (Nadig & Sedivy 2002). Thus, children learn rapidly to apply the multiple constraints used in incremental adult processing (Borovsky et al. 2012).

Nonetheless, online learning contrasts with traditional approaches in which the structure of the language is learned offline by the cognitive system acquiring a corpus of past linguistic inputs and choosing the grammar or other model of the language that best fits with those inputs. For example, in both mathematical and theoretical analysis (e.g., Gold 1967; Hsu et al. 2011; Pinker 1984) and in grammar-induction algorithms in machine learning and cognitive science, it is typically assumed that a corpus of language can be held in memory, and that the candidate grammar is successively adjusted to fit the corpus as well as possible (e.g., Manning & Schütze 1999; Pereira & Schabes 1992; Redington et al. 1998). However, this approach involves learning linguistic regularities (at, say, the morphological level), by storing and later surveying relevant linguistic input at a lower level of analysis (e.g., involving strings of phonemes); and then attempting to determine which higher-level regularities best fit the database of lower-level examples. There are a number of difficulties with this type of proposal – for example, that only a very rich lower-level representation (perhaps combined with annotations concerning relevant syntactic and semantic context) is likely to be a useful basis for later analysis. But more fundamentally, the Now-or-Never bottleneck requires that information be retained only if it is recoded *at* *processing time*: Phonological information that is not chunked at the morphological level and beyond will be obliterated by oncoming phonological material.[12]

So, if learning is shaped by the Now-or-Never bottleneck, then linguistic input must, when it is encountered, be recoded successively at increasingly abstract linguistic levels if it is to be retained at all – a constraint imposed, we argue, by basic principles of memory. Crucially, such information is not, therefore, in a suitably "neutral" format to allow for the discovery of previously unsuspected linguistic regularities. In a nutshell, the lossy compression of the linguistic input is achieved by applying the learner's *current* model of the language. But information that would point toward a better model of the language (if examined in retrospect) will typically be lost (or, at best, badly obscured) by this compression, precisely because those regularities are *not* captured by the current model of the language. Suppose, for example, that we create a lossy encoding of language using a simple, context-free phrase structure grammar that cannot handle, say, noun-verb agreement. The lossy encoding of the linguistic input produced using this grammar will provide a poor basis for learning a more sophisticated grammar that includes agreement – precisely because agreement information will have been thrown away. So the Now-or-Never bottleneck rules out the possibility that the learner can survey a neutral database of linguistic material, to optimize its model of the language.

The emphasis on online learning does not, of course, rule out the possibility that any linguistic material that *is* remembered may subsequently be used to inform learning. But according to the present viewpoint, any further learning requires *reprocessing* that material. So if a child comes to learn a poem, song, or story verbatim, the child might extract more structure from that material by mental rehearsal (or, indeed, by saying it aloud). The online learning constraint is that material is learned *only* when it is being processed – ruling out any putative learning processes that involve carrying out linguistic analyses or compiling statistics over a stored corpus of linguistic material.

If this general picture of acquisition as learning-to-process is correct, then we should expect the exploitation of memory to require "replaying" learned material, so that it can be re-processed. Thus, the application of memory itself requires passing through the Now-or-Never bottleneck – there is no way of directly interrogating an internal database of past experience; indeed, this viewpoint fits with our subjective sense that we need to "bring to mind" past experiences or rehearse verbal material to process it further. Interestingly, there is now also substantial neuroscientific evidence that replay does occur (e.g., in rat spatial learning, Carr et al. 2011). Moreover, it has long been suggested that dreaming may have a related function (here using "reverse" learning over "fictional" input to eliminate spurious relationships identified by the brain, Crick & Mitchison 1983; see Hinton & Sejnowski 1986, for a closely related computational model). Deficits in the ability to replay material would, in this view, lead to consequent deficits in memory and inference; consistent with this viewpoint, Martin and colleagues have argued that rehearsal deficits for phonological pattern and semantic information may lead to difficulties in the long-term acquisition and retention of word forms and word meanings, respectively, and their use in language processing (e.g., Martin & He 2004; Martin et al. 1994). In summary, then, language acquisition involves learning to process, and generalizations can only be made over past processing episodes.

4.2. Implications of Strategy 2: Local learning

Online learning faces a particularly acute version of a general learning problem: the stability-plasticity dilemma (e.g., Mermillod et al. 2013). How can new information be acquired without interfering with prior information? The problem is especially challenging because reviewing prior information is typically difficult (because recalling earlier information interferes with new input) or impossible (where prior input has been forgotten). Thus, to a good approximation, the learner can only update its model of the language in a way that responds to current linguistic input, without being able to review whether any updates are inconsistent with prior input. Specifically, if the learner has a global model of the entire language (e.g., a traditional grammar), the learner runs the risk of overfitting that model to capture regularities in the momentary linguistic input at the expense of damaging the match with past linguistic input.

Avoiding this problem, we suggest, requires that learning be highly *local*, consisting of learning about specific relationships between particular linguistic representations. New items can be acquired, with implications for later processing of similar items; but learning current items does not thereby create changes to the entire model of the language, thus potentially interfering with what was learned from past input. One way to learn in a local fashion is to store individual examples (this requires, in our framework, that those examples have been abstractly recoded by successive Chunk-and-Pass operations, of course), and then to generalize, piecemeal, from these examples. This standpoint is consistent with the idea that the "priority of the specific," as observed in other areas of cognition (e.g., Jacoby et al. 1989), also applies to language acquisition. For example, children seem to be highly sensitive to multiword chunks (Arnon & Clark 2011; Bannard & Matthews 2008; see Arnon & Christiansen, submitted, for a review[13]). More generally, learning based on past traces of processing will typically be sensitive to details of that processing, as is observed across phonetics, phonology, lexical access, syntax, and semantics (e.g., Bybee 2006; Goldinger 1998; Pierrehumbert 2002; Tomasello 1992).

That learning is local provides a powerful constraint, incompatible with typical computational models of how the child might infer the grammar of the language – because these models typically do not operate incrementally but range across the input corpus, evaluating alternative grammatical hypotheses (so-called *batch* learning). But, given the Now-or-Never bottleneck, the "unprocessed" corpus, so readily available to the linguistic theorist, or to a computer model, is lost to the human learner almost as soon as it is encountered. Where such information has been memorized (as in the case of SF's encoding of streams of digits), recall and processing is slow and effortful. Moreover, because information is encoded in terms of the current encoding, it becomes difficult to neutrally review that input to create a better encoding, and cross-check past data to test wide-ranging grammatical hypotheses.[14] So, as we have already noted, the Now-or-Never bottleneck seems incompatible with the view of a child as a mini-linguist.

By contrast, the principle of local learning is respected by other approaches. For example, item-based (Tomasello 2003), connectionist (e.g., Chang et al. 1999; Elman 1990; MacDonald & Christiansen 2002),[15] exemplar-based (e.g., Bod 2009), and other usage-based (e.g., Arnon & Snider 2010; Bybee 2006) accounts of language acquisition tie learning and processing together – and assume that language is acquired piecemeal, in the absence of an underlying *Bauplan*. Such accounts, based on local learning, provide a possible explanation of the frequency effects that are found at all levels of language processing and acquisition (e.g., Bybee 2007; Bybee & Hopper 2001; Ellis 2002; Tomasello 2003), analogous to exemplar-based theories of how performance speeds up with practice (Logan 1988).

The local nature of learning need not, though, imply that language has no integrated structure. Just as in perception and action, local chunks can be defined at many different levels of abstraction, including highly abstract patterns, for example, governing subject, verb, and object; and generalizations from past processing to present processing will operate across all of these levels. Therefore, in generating or understanding a new sentence, the language user will be influenced by the interaction of multiple constraints from innumerable traces of past processing, across different linguistic levels. This view of language processing involving the parallel interaction of multiple local constraints is embodied in a variety of influential approaches to language (e.g., Jackendoff 2007; Seidenberg 1997).

4.3. Implications of Strategy 3: Learning to predict

If language processing involves prediction – to make the encoding of new linguistic material sufficiently rapid – then a critical aspect of language acquisition is *learning* to make such predictions successfully (Altmann & Mirkovic 2009). Perhaps the most natural approach to predictive learning is to compare predictions with subsequent reality, thus creating an "error signal," and then to modify the predictive model to systematically reduce this error. Throughout many areas of cognition, such error-driven learning has been widely explored in a range of computational frameworks (e.g., from connectionist networks, to reinforcement learning, to support vector machines) and has considerable behavioral (e.g., Kamin 1969) and neurobiological support (e.g., Schultz et al. 1997).

Predictive learning can, in principle, take a number of forms: For example, predictive errors can be accumulated over many samples, and then modifications made to the predictive model to minimize the overall error over those samples (i.e., batch learning). But this is ruled out by the Now-or-Never bottleneck: Linguistic input, and the predictions concerning it, is present only fleetingly. But error-driven learning can also be "online" – each prediction error leads to an immediate, though typically small, modification of the predictive model; and the accumulation of these small modifications gradually reduces prediction errors on future input.

A number of computational models adhere to these principles: Learning involves creating a predictive model of the language, using online error-driven learning. Such models, limited though they are, may provide a starting point for creating an increasingly realistic account of language acquisition and processing. For example, a connectionist model which embodies these principles is the simple recurrent network (Altmann 2002; Christiansen & Chater 1999; Elman 1990), which learns to map from the current input on to the next element in a continuous sequence of linguistic (or other) input; and which learns, online, by adjusting its parameters (the "weights" of the network) to reduce the observed prediction error, using the back-propagation learning algorithm. Using a very different framework, in the spirit of construction grammar (e.g., Croft 2001; Goldberg 2006), McCauley and Christiansen (2011) recently developed a psychologically based, online chunking model of incremental language acquisition and processing , incorporating prediction to generalize to new chunk combinations. Exemplar-based analogical models of language acquisition and processing may also be constructed, which build and predict language structure online, by incrementally creating a database of possible structures, and dynamically using online computation of similarity to recruit these structures to process and predict new linguistic input.

Importantly, prediction allows for top-down information to influence current processing across different levels of linguistic representation, from phonology to discourse, and at different temporal windows (as indicated by Fig. 2). We see the ability to use such top-down information as emerging gradually across development, building on bottom-up information. That is, children gradually learn to apply top-down knowledge to facilitate processing via prediction, as higher-level information becomes more entrenched and allows for anticipatory generalizations to be made.

In this section, we have argued that the child should not be viewed as a mini-linguist, attempting to infer the abstract structure of grammar, but as learning to process: that is, learning to alleviate the severe constraints imposed by the Now-or-Never bottleneck. Next, we discuss how chunk-based language acquisition and processing have shaped linguistic change and, ultimately, the evolution of language.

5. Language change is item-based

Like language, human culture constantly changes. We continually tinker with all aspects of culture, from social conventions and rituals to technology and everyday artifacts (see contributions in Richerson & Christiansen 2013). Perhaps language, too, is a result of cultural evolution – a product of piecemeal tinkering – with the long-term evolution of language resulting from the compounding of myriad local short-term processes of language change. This hypothesis figures prominently in many recent theories of language evolution (e.g., Arbib 2005; Beckner et al. 2009; Christiansen & Chater 2008; Hurford 1999; Smith & Kirby 2008; Tomasello 2003; for a review of these theories, see Dediu et al. 2013). Language is construed as a complex evolving system in its own right; linguistic forms that are easier to use and learn, or are more communicatively efficient, will tend to proliferate, whereas those that are not will be prone to die out. Over time, processes of cultural evolution involving repeated cycles of learning and use are hypothesized to have shaped the languages we observe today.

If aspects of language survive only when they are easy to produce and understand, then moment-by-moment processing will shape not only the structure of language (see also Hawkins 2004; O'Grady 2005), but also the learning problem that the child faces. Thus, from the perspective of language as an evolving system, language processing at the timescale of seconds has implications for the longer timescales of language acquisition and evolution. Figure 3 illustrates how the effects of the Now-or-Never bottleneck flow from the timescale of processing to those of acquisition and evolution.

Chunk-and-Pass processing carves the input (or output) into chunks at different levels of linguistic representation at the timescale of the utterance (seconds). These chunks constitute the comprehension and production events from which children and adults learn and update their ability to process their native language over the timescale of the individual (tens of years). Each learner, in turn, is part of a population of language users that shape the cultural evolution of language across a historical timescale (hundreds or thousands of years): Language will be shaped by the linguistic patterns learners find easiest to acquire and process. And the learners will, of course, be strongly constrained by the basic cognitive limitation that is the Now-or-Never bottleneck – and, hence, through cultural evolution, linguistic patterns, which can be processed through that bottleneck, will be strongly selected. Moreover, if acquiring a language is learning to process and processing involves incremental Chunk-and-Pass operations, then language change will operate through changes driven by Chunk-and-Pass processing, both within and between individuals. But this, in turn, implies that processes of language change should be *item-based*, driven by processing/acquisition mechanisms defined over Chunk-and-Pass representations (rather than, for example, being defined over abstract linguistic parameters, with diverse structural consequences across the entire language).

We noted earlier that a consequence of Chunk-and-Pass processing for production is a tendency toward reduction, especially of more frequently used forms, and this constitutes one of several pressures on language change (see also MacDonald 2013). Because reduction minimizes articulatory processing effort for the speaker but may increase processing effort for the hearer and learner, this pressure can in extreme cases lead to a communicative collapse. This is exemplified by a lab-based analogue of the game of "telephone," in which participants were exposed to a miniature artificial language consisting of simple form-meaning mappings (Kirby et al. 2008). The initial language contained random mappings between syllable strings and pictures of moving geometric figures in different colors. After exposure, participants were asked to produce linguistic forms corresponding to specific pictures. Importantly, the participants saw only a subset of the language but nonetheless had to generalize to the full language. The productions of the initial learner were then used as the input language for the next learner, and so on for a total of 10 "generations." In the absence of other communicative pressures (such as the avoidance of ambiguity; Grice 1967), the language collapsed into just a few different forms that allowed for systematic, albeit semantically underspecified, generalization to unseen items. In natural language, however, the pressure toward reduction is normally kept in balance by the need to maintain effective communication.

Expanding on the notion of reduction and erosion, we suggest that constraints from Chunk-and-Pass processing can provide a cognitive foundation for grammaticalization (Hopper & Traugott 1993). Specifically, chunks at different levels of linguistic structure – discourse, syntax, morphology, and phonology – are potentially subject to reduction. Consequently, we can distinguish between different types of grammaticalization, from discourse syntacticization and

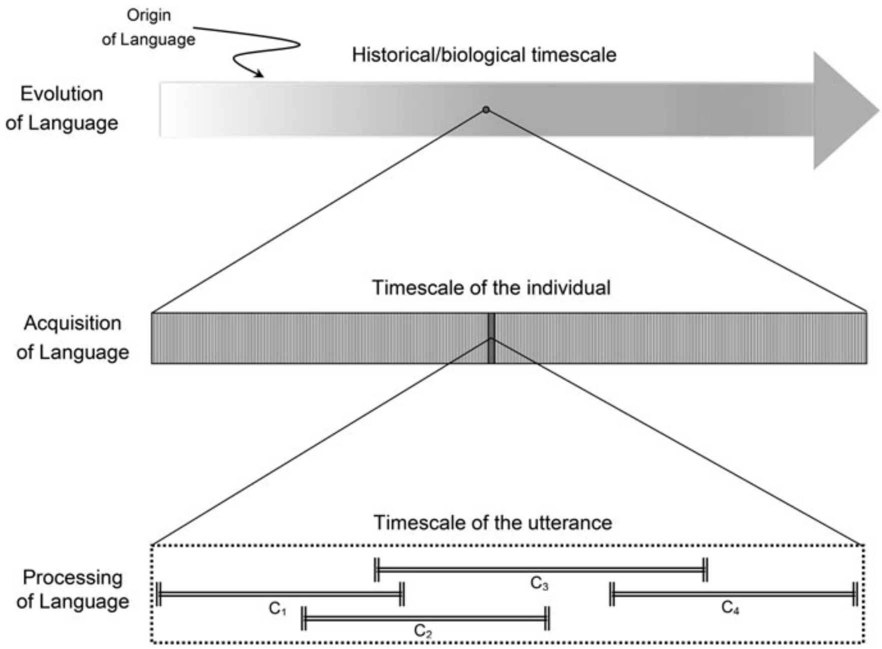

Figure 3. Illustration of how Chunk-and-Pass processing at the utterance level (with the C_i referring to chunks) constrains the acquisition of language by the individual, which, in turn, influences how language evolves through learning and use by groups of individuals on a historical timescale.

semantic bleaching to morphological reduction and phonetic erosion. Repeated chunking of loose discourse structures may result in their reduction into more rigid syntactic constructions, reflecting Givón's (1979) hypothesis that today's syntax is yesterday's discourse.[16] For example, the resultative construction *He pulled the window open* might derive from syntacticization of a loose discourse sequence such as *He pulled the window and it opened* (Tomasello 2003). As a further by-product of chunking, some words that occur frequently in certain kinds of construction may gradually become "bleached" of meaning and ultimately signal only general syntactic properties. Consider, as an example, the construction *be going to*, which was originally used exclusively to indicate movement in space (e.g., *I'm going to Ithaca*) but which is now also used as an intention or future marker when followed by a verb (as in *I'm going to eat at seven*; Bybee et al. 1994). Additionally, a chunked linguistic expression may further be subject to morphological reduction, resulting in further loss of morphological (or syntactic) elements. For instance, the demonstrative *that* in English (e.g., *that window*) lost the grammatical category of number (*that*$_{\text{sing}}$ vs. *those*$_{\text{plur}}$) when it came to be used as a complementizer, as in *the window/windows that is/are dirty* (Hopper & Traugott 1993). Finally, as noted earlier, frequently chunked elements are likely to become phonologically reduced, leading to the emergence of new shortened grammaticalized forms, such as the phonetic erosion of *going to* into *gonna* (Bybee et al. 1994). Thus, the Now-or-Never bottleneck provides a constant pressure toward reduction and erosion across the different levels of linguistic representation, providing a possible explanation for why grammaticalization tends to be a largely unidirectional process (e.g., Bybee et al. 1994; Haspelmath 1999; Heine & Kuteva 2002; Hopper & Traugott 1993).

Beyond grammaticalization, we suggest that language change, more broadly, will be local at the level of individual chunks. At the level of sound change, our perspective is consistent with lexical diffusion theory (e.g., Wang 1969; 1977; Wang & Cheng 1977), suggesting that sound change originates with a small set of words and then gradually spreads to other words with a similar phonological make-up. The extent and speed of such sound change is affected by a number of factors, including frequency, word class, and phonological environment (e.g., Bybee 2002; Phillips 2006). Similarly, morpho-syntactic change is also predicted to be local in nature: what we might call "constructional diffusion." Accordingly, we interpret the cross-linguistic evidence indicating the effects of processing constraints on grammatical structure (e.g., Hawkins 2004; Kempson et al. 2001; O'Grady 2005; see Jaeger & Tily 2011, for a review) as a process of gradual change over individual constructions, instead of wholesale changes to grammatical rules. Note, though, that because chunks are not independent of one another but form a system within a given level of linguistic representation, a change to a highly productive chunk may have cascading effects to other chunks at that level (and similarly for representations at other levels of abstraction). For example, if a frequently used construction changes, then constructional diffusion could in principle lead to rapid, and far-reaching, change throughout the language.

On this account, another ubiquitous process of language change, regularization, whereby representations at a particular linguistic level become more patterned, should also be a piecemeal process. This is exemplified by another of Kirby et al.'s (2008) game-of-telephone experiments, showing that when ambiguity is avoided, a highly structured linguistic system emerges across generations of learners, with morpheme-like substrings indicating different semantic properties (color, shape, and movement). Another similar, lab-based cultural evolution experiment showed that this process of regularization does not result in the elimination of variability but, rather, in increased predictability through lexicalized patterns (Smith & Wonnacott 2010). Whereas the initial language contained unpredictable pairings of nouns with plural markers, each noun became chunked with a specific marker in the final languages.

These examples illustrate how Chunk-and-Pass processing over time may lead to so-called *obligatorification*, whereby a pattern that was initially flexible or optional becomes obligatory (e.g., Heine & Kuteva 2007). This process is one of the ways in which new chunks may be created. So, although chunks at each linguistic level can lose information through grammaticalization, and although they cannot regain it, a countervailing process exists by which complex chunks are constructed by "gluing together" existing chunks.[17] That is, in Bybee's (2002) phrase, "items that are used together fuse together." For example, auxiliary verbs (e.g., to have, to go) can become fused with main verbs to create new morphological patterns, as in many Romance languages, in which the future tense is signaled by an auxiliary tacked on as a suffix to the infinitive. In Spanish, the future tense endings *-é*, *-ás*, *-á*, *-emos*, *-éis*, *-án* derive from the present tense of the auxiliary *haber*, namely, *he*, *has*, *ha*, *hemos*, *habéis*, *han*; and in French, the corresponding endings *-ai*, *-as*, *-a*, *-on*, *-ez*, *-on* derive from the present tense of the auxiliary *avoir*, namely, *ai*, *as*, *a*, *avon*, *avez*, *ont* (Fleischman 1982). Such complex new chunks are then subject to erosion (e.g., as is implicit in the example above, the Spanish for *you*$_{\text{informal, plural}}$ *will eat* is *comeréis*, rather than **comerhabéis*; the first syllable of the auxiliary has been stripped away).

Importantly, the present viewpoint is neutral regarding the extent to which children are the primary source of innovation (e.g., Bickerton 1984) or regularization (e.g., Hudson et al. 2005) of linguistic material, although constraints from child language acquisition likely play some role (e.g., in the emergence of regular subject-object-verb word order in the Al-Sayyid Bedouin Sign Language; Sandler et al. 2005). In general, we would expect that multiple forces influence language change in parallel (for reviews, see Dediu et al. 2013; Hruschka et al. 2009), including sociolinguistic factors (e.g., Trudgill 2011), language contact (e.g., Mufwene 2008), and use of language as an ethnic marker (e.g., Boyd & Richerson 1987).

Because language change, like processing and acquisition, is driven by multiple competing factors, which are amplified by cultural evolution, linguistic diversity will be the norm. Accordingly, we would expect few, if any, "true" language universals to exist in the sense of constraints that can be explained only in purely linguistic terms (Christiansen & Chater 2008). Nonetheless, domain-general processing constraints are likely to significantly constrain the set of possible languages (see, e.g., Cann & Kempson 2008). This picture is consistent with linguistic arguments suggesting that there may be no *strict* language universals (Bybee 2009; Evans & Levinson 2009). For example, computational phylogenetic analyses indicate that word order correlations are lineage-specific (Dunn et al. 2011), shaped by particular histories of cultural evolution rather than following universal patterns as would be expected if they were the result of innate linguistic constraints (e.g., Baker 2001) or language-specific performance limitations (e.g., Hawkins 2009). Thus, the process of piecemeal tinkering that drives item-based language change is subject to constraints deriving not only from Chunk-and-Pass processing and multiple-cue integration but also from the specific trajectory of cultural evolution that a language follows. More generally, in this perspective, there is no sharp distinction between language evolution and language change: Language evolution is just the result of language change over a long timescale (see also Heine & Kuteva 2007), obviating the need for separate theories of language evolution and change (e.g., Berwick et al. 2013; Hauser et al. 2002; Pinker 1994).[18]

6. Structure as processing

The Now-or-Never bottleneck implies, we have argued, that language comprehension involves incrementally chunking linguistic material and immediately passing the result for further processing, and production involves a similar cascade of Just-in-Time processing operations in the opposite direction. And language will be shaped through cultural evolution to be easy to learn and process by generations of speakers/hearers, who are forced to chunk and pass the oncoming stream of linguistic material. What are the resulting implications for the structure of language and its mental representation? In this section, we first show that certain key properties of language follow naturally from this framework; we then reconceptualize certain important notions in the language sciences.

6.1. Explaining key properties of language

6.1.1. The bounded nature of linguistic units. In nonlinguistic sequential tasks, memory constraints are so severe that chunks of more than a few items are rare. People typically encode phone numbers, number plates, postal codes, and Social Security numbers into sequences of between two and four digits or letters; memory recall deteriorates rapidly for unchunked item-sequences longer than about four elements (Cowan 2000), and memory recall typically breaks into short chunk-like phrases. Similar chunking processes are thought to govern nonlinguistic sequences of actions (e.g., Graybiel 1998). As we have argued previously in this article, the same constraints apply throughout language processing, from sound to discourse.

Across different levels of linguistic representation, units also tend to have only a few component elements. Even though the nature of sound-based units in speech is theoretically contentious, all proposals capture the sharply bounded nature of such units. For example, a traditional perspective on English phonology would postulate phonemes, short sequences of which are grouped into syllables, with multisyllabic words being organized by intonational or perhaps morphological groupings. Indeed, the tendency toward few-element units is so strong that long, nonsense words with many syllables such as *supercalifragilisticexpialidocious* is chunked successively, for example, as tentatively indicated:

[[[Super][cali]][[fragi][listic]][[expi][ali]][docious]]

Similarly, agglutinating languages, such as Turkish, chunk complex multimorphemic words using local grouping mechanisms that include formulaic morpheme expressions (Durrant 2013). Likewise, at higher levels of linguistic representation, verbs normally have only two or three arguments at most. Across linguistic theories of different persuasions, syntactic phrases typically consist of only a few constituents. Thus, the Now-or-Never bottleneck provides a strong bias toward bounded linguistic units across various levels of linguistic representations.

6.1.2. The local nature of linguistic dependencies. Just as we have argued that Chunk-and-Pass processing leads to simple linguistic units with only a small number of components, so it produces a powerful tendency toward local dependencies. Dependencies between linguistic elements will primarily be adjacent or separated by only a few other elements. For example, at the phonological level, processes are highly local, as reflected by data on coarticulation, assimilation, and phonotactic constraints (e.g., Clark et al. 2007). Similarly, we expect word formation processes to be highly local in nature, which is in line with a variety of different linguistic perspectives on the prominence of adjacency in morphological composition (e.g., Carstairs-McCarthy 1992; Hay 2000; Siegel 1978). Strikingly, adjacency even appears to be a key characteristic of multimorphemic formulaic units in an agglutinating language such as Turkish (Durrant 2013).

At the syntactic level, there is also a strong bias toward local dependencies. For example, when processing the sentence "*The key to the cabinets was* …" comprehenders experience local interference from the plural *cabinets*, although the verb *was* needs to agree with the singular *key* (Nicol et al.1997; Pearlmutter et al. 1999). Indeed, individuals who are good at picking up adjacent dependencies among sequence elements in a serial-reaction time task also experience greater local interference effects in sentence processing (Misyak & Christiansen 2010). Moreover, similar local interference effects have been observed in production when people are asked to continue the above sentence after *cabinets* (Bock & Miller 1991).

More generally, analyses of Romanian and Czech (Ferrer-i-Cancho 2004) as well as Catalan, Basque, and Spanish (Ferrer-i-Cancho & Liu 2014) point to a pressure toward minimization of the distance between syntactically related words. This tendency toward local dependencies seems to be particularly strongly expressed in strict-word-order languages such as English, but somewhat less so for more flexible languages such as German (Gildea & Temperley 2010). However, the use of case marking in German may provide a cue to overcome this by indicating who does what to whom, as suggested by simulations of the learnability of different word orders with or without case markings (e.g., Lupyan & Christiansen 2002; Van Everbroeck 1999). This highlights the importance not only of distributional information (e.g., regarding word order) but also of other types of cues (e.g., involving phonological, semantic, or pragmatic information), as discussed previously.

We want to stress, however, that we are not denying the existence of long-distance syntactic dependencies; rather, we are suggesting that our ability to process such dependencies will be bounded by the number of chunks that can be kept in memory at a given level of linguistic representation. In many cases, chunking may help to minimize the distance over which a dependency has to remain in memory. For example, the use of personal pronouns can facilitate the processing of otherwise difficult object relative clauses because they are more easily chunked (e.g., *People [you know] are more fun*; Reali & Christiansen 2007a). Similarly, the processing of long-distance dependencies is eased when they are separated by highly frequent word combinations that can be readily chunked (e.g., Reali & Christiansen 2007b). More generally, the Chunk-and-Pass account is in line with other approaches that assign processing limitations and complexity as primary constraints on long-distance dependencies, thus potentially providing explanations for linguistic phenomena, such as subjacency (e.g., Berwick & Weinberg 1984; Kluender & Kutas 1993), island constraints (e.g., Hofmeister & Sag 2010), referential binding (e.g., Culicover 2013), and scope effects (e.g., O'Grady 2013). Crucially, though, as we argued earlier, the impact of these processing constraints may be lessened to some degree by the integration of multiple sources of information (e.g., from pragmatics, discourse context, and world knowledge) to support the ongoing interpretation of the input (e.g., Altmann & Steedman 1988; Heider et al. 2014; Tanenhaus et al. 1995).

6.1.3. Multiple levels of linguistic representation. Speech allows us to transmit a digital, symbolic code over a serial, analog channel using time variation in sound pressure (or using analog movements, in sign language). How might we expect this digital-analog-digital conversion to be tuned, to optimize the amount of information transmitted?

The problem of encoding and decoding digital signals over an analog serial channel is well studied in communication theory (Shannon 1948) – and, interestingly, the solutions typically adopted look very different from those employed by natural language. Crucially, to maximize the rate of transfer of information it is generally best to transform the message to be conveyed across the analog signal in a very nonlocal way. That is, rather than matching up portions of the information to be conveyed (e.g., in an engineering context, these might be the contents of a database) to particular portions of the analog signal, the best strategy is to encrypt the entire digital message using the entire analog signal, so that the message is coded as a block (e.g., MacKay 2003). But why is the engineering solution to information transmission so very different from that used by natural language, in which distinct portions of the analog signal correspond to meaningful units in the digital code (e.g., phonemes, words)? The Now-or-Never bottleneck provides a natural explanation.

A block-based code requires decoding a stored memory trace for the *entire* analog signal (for language, typically, acoustic) – that is, the whole block. This is straightforward for artificial computing systems, where memory interference is no obstacle. But this type of block coding is, of course, precisely what the Now-or-Never bottleneck rules out. The human perceptual system must turn the acoustic input into a (lossy) compressed form right away, or else the acoustic signal is lost forever. Similarly, the speech production system cannot decide to send a single, lengthy analog signal, and then successfully reel off the lengthy corresponding sequence of articulatory instructions, because this will vastly exceed our memory capacity for sequences of actions. Instead, the acoustic signal must be generated and decoded *incrementally* so that the symbolic information to be transmitted maps, fairly locally, to portions of the acoustic signal. Thus, to an approximation, whereas individual phonemes acoustically exhibit enormous contextual variation, diphones (pairs of phonemes) are a fairly stable acoustic signal, as evident by their use in tolerably good speech synthesis and recognition (e.g., Jurafsky et al. 2000). Overall, then, each successive segment of the analog acoustic input must correspond to a part of the symbolic code being transmitted. This is not because of considerations of informational efficiency but because of the brain's processing limitations in encoding and decoding: specifically, by the Now-or-Never bottleneck.

The need rapidly to encode and decode implies that spoken language will consist of a sequence of short sound-based units (the precise nature of these units may be controversial, and may even differ between languages, but units could include diphones, phonemes, mora, syllables, etc.). Similarly, in speech production, the Now-or-Never bottleneck rules out planning and executing a long articulatory sequence (as in a block-code used in communication technology); rather, speech must be planned incrementally, in the Just-in-Time fashion, requiring that the speech signal corresponds to sequences of discrete sound-based units.

6.1.4. Duality of patterning. Our perspective has yet further intriguing implications. Because the Now-or-Never bottleneck requires that symbolic information must rapidly be read off the analog signal, the number of such symbols will be severely limited – and in particular, may be much smaller than the vocabulary of a typical speaker (many thousands or tens of thousands of items). This implies that the short symbolic sequences into which the acoustic signal is initially recoded cannot, in general, be bearers of meaning; instead, the primary bearers of meaning, lexical items, and morphemes, will be composed out of these smaller units.

Thus, the Now-or-Never bottleneck provides a potential explanation for a puzzling but ubiquitous feature of human languages, including signed languages. This is *duality of patterning*: the existence of (one or more) level(s) of symbolically encoded sound structure (whether described in terms of phonemes, mora, or syllables) from which the level of words and morphemes (over which meanings are defined) are composed. Such patterning arises, in the present analysis, as a consequence of rapid online multilevel chunking in both speech production and perception. In the absence of duality of patterning, the acoustic signal corresponding, say, to a single noun, could not be recoded incrementally as it is received (Warren & Marslen-Wilson 1987) – but would have to be processed as a whole, thus dramatically overloading sensory memory.

It is, perhaps, also of interest to note that the other domain in which people process enormously complex acoustic input – music – also typically consists of multiple layers of structure (notes, phrases, and so on, see, e.g., Lerdahl & Jackendoff 1983; Orwin et al. 2013). We may conjecture that Chunk-and-Pass processing operates for music as well as language, thus helping to explain why our ability to process musical input spectacularly exceeds our ability to process arbitrary sequential acoustic material (Clément et al. 1999).

6.1.5. The quasi-regularity of linguistic structure. We have argued that the Now-or-Never bottleneck implies that language processing involves applying highly local Chunk-and-Pass operations across a range of representational levels; and that language acquisition involves learning to perform such operations. But, as in the acquisition of other skills, learning from such specific instances does not operate by rote but leads to generalization and hence modification from one instance to another (Goldberg 2006). Indeed, such processes of local generalization are ubiquitous in language change, as we have noted above. From this standpoint, we should expect the rule-like patterns in language to emerge from generalizations across specific instances (see, e.g., Hahn & Nakisa 2000, for an example of this approach to inflectional morphology in German); once entrenched, such rule-like patterns can, of course, be applied quite broadly to newly encountered cases. Thus, patterns of regularity in language will emerge locally and bottom-up, from generalizations across individual instances, through processes of language use, acquisition, and change.

We should therefore expect language to be quasi-regular across phonology, morphology, syntax, and semantics – to be an amalgam of overlapping and partially incompatible patterns, involving generalizations from the variety of linguistic forms from which successive language learners generalize. For example, English past tense morphology has, famously, the regular *–ed* ending, a range of subregularities (*sing* → *sang*, *ring* → *rang*, *spring* → *sprang*, but *fling* → *flung*, *wring* → *wrung*; and even *bring* → *brought*; with some verbs having the same present and past tense forms, e.g., *cost* → *cost*, *hit* → *hit*, *split* → *split*; whereas others differ wildly, e.g., *go* → *went*; *am* → *was*; see, e.g., Bybee & Slobin 1982; Pinker & Prince 1988; Rumelhart & McClelland 1986). This quasi-regular structure (Seidenberg & McClelland 1989) does indeed seem to be widespread throughout many aspects of language (e.g., Culicover 1999; Goldberg 2006; Pierrehumbert 2002).

From a traditional, generative perspective on language, such quasi-regularities are puzzling: Natural language is assimilated, somewhat by force, to the structure of a formal language with a precisely defined syntax and semantics – the ubiquitous departures from such regularities are mysterious. From the present standpoint, by contrast, the quasi-regular structure of language arises in rather the same way that a partially regular pattern of tracks were laid down across a forest, through the overlaid traces of an endless number of agents finding the path of local least resistance; and where each language processing episode tends to facilitate future, similar, processing episodes, just as an animal's choice of path facilitates the use of that path for animals that follow.

6.2. What is linguistic structure?

Chunk-and-Pass processing can be viewed as having an interesting connection with traditional linguistic notions. In both production and comprehension, the language system creates a sequence of chunking operations, which link different linguistic units together across multiple levels of structure. That is, the syntactic structure of a given utterance is reflected in its processing history. This conception is reminiscent of previous proposals, in which syntax is viewed as a control structure for guiding semantic interpretation (e.g., Ford et al. 1982; Kempson et al. 2001; Morrill 2010). For example, in describing his incremental parser-interpreter, Pulman (1985) noted, "Syntactic information is used to build up the interpretation and to guide the parse, but does not result in the construction of an independent level of representation" (p. 132). Steedman (2000) adopted a closely related perspective when introducing his combinatory categorial grammar, which aims to map surface structure directly onto logic-based semantic interpretations, given rich lexical representations of words that include information about phonological structure, syntactic category, and meaning: "… syntactic structure is merely the characterization of the process of constructing a logical form, rather than a representational level of structure that actually needs to be built …" (p. xi). Thus, in these accounts, the syntactic structure of a sentence is not explicitly represented by the language system but plays the role of a processing "trace" of the operations used to create or interpret the sentence (see also O'Grady 2005).

To take an analogy from constructing objects, rather than sentences, the process by which components of an IKEA-style flat-pack cabinet are combined provides a "history" (combine a board, handle, and screws to construct the doors; combine frame and shelf to construct the body; combine doors, body, and legs to create the finished cabinet). The history by which the cabinet was constructed may thus reveal the intricate structure of the finished item, but this structure need not be explicitly represented during the construction process. Thus, we can "read off" the syntactic structure of a sentence from its processing history, revealing the syntactic relations between various constituents (likely with a "flat" structure; Frank et al. 2012). Syntactic representations are neither computed during comprehension nor in production; instead, there is just a history of processing operations. That is, we view *linguistic structure as processing history.* Importantly, this means that syntax is not privileged but is only one part of the system – and it is not independent of the other parts (see also Fig. 2).

In this view, a rather minimal notion of grammar specifies how the chunks from which a sentence is built can be composed. There may be several, or indeed many, orders in which such combinations can occur, just as operations for furniture assembly may be carried out somewhat flexibly (but not completely without constraints – it might turn out that the body must be screwed together before a shelf can be attached). In the context of producing and understanding language, the process of construction is likely to be much more constrained: Each new "component" is presented in turn, and it must be used immediately or it will be lost. Moreover, viewing Chunk-and-Pass processing as an aspect of skill acquisition, we might expect that the precise nature of chunks may change with expertise: Highly overlearned material might, for example, gradually come to be treated as a single chunk (see Arnon & Christiansen, submitted, for a review).

Crucially, as with other skills, the cognitive system will tend to be a *cognitive miser* (Fiske & Taylor 1984), generally following a principle of least effort (Zipf 1949). As processing proceeds, there is an intricate interplay of top-down and bottom-up processing to alight on the message as rapidly as possible. The language system need only construct enough chunk structure so that, when combined with prior discourse and background knowledge, the intended message can be inferred incrementally. This observation relates to some interesting contemporary linguistic proposals. For example, from a generative perspective, Culicover (2013) highlighted the importance of incremental processing, arguing that the interpretation of a pronoun depends on which discourse elements are available when it is encountered. This implies that the linear order of words in a sentence (rather than hierarchical structure) plays an important role in many apparently grammatical phenomena, including weak cross-over effects in referential binding. From an emergentist perspective, O'Grady (2015) similarly emphasized the importance of real-time processing constraints for explaining differences in the interpretation of reflexive pronouns (*himself, themselves*) and plain pronouns (*him, them*). The former are resolved locally, and thus almost instantly, whereas the antecedents for the latter are searched for within a broader domain (causing problems in acquisition because of a bias toward local information).

More generally, our view of linguistic structure as processing history offers a way to integrate the formal linguistic contributions of construction grammar (e.g., Croft 2001; Goldberg 2006) with the psychological insights from usage-based approaches to language acquisition and processing (e.g., Bybee & McClelland 2005; Tomasello 2003). Specifically, we propose to view constructions as *computational procedures*[19] – specifying how to process and produce a particular chunk – where we take a broad view of constructions as involving chunks at different levels of linguistic representation, from morphemes to multiword sequences. A procedure may integrate several different aspects of language processing or production, including chunking acoustic input into sound-based units (phonemes, syllables), mapping a chunk onto meaning (or vice versa), incorporating pragmatic or discourse information, and associating a chunk with specific arguments (see also O'Grady 2005; 2013). As with other skills (e.g., Heathcote et al. 2000; Newell & Rosenbloom 1981), there will be practice effects, where the repeated use of a given chunk results in faster processing and reduced demands on cognitive resources, and with sufficient use, leading to a high degree of automaticity (e.g., Logan 1988; see Bybee & McClelland 2005, for a linguistic perspective).

In terms of our previous forest track analogy, the more a particular chunk is comprehended or produced, the more entrenched it becomes, resulting in easier access and faster processing; tracks become more established with use. With sufficiently frequent use, adjacent tracks may blend together, creating somewhat wider paths. For example, the frequent processing of simple transitive sentences, processed individually as multiword chunks, such as "I want milk," "I want candy," might first lead to a wider track involving the item-based template "I want X." Repeated use of this template along with others (e.g., "I like X," "I see X") might eventually give rise to a more abstract transitive generalization along the lines of *N V N* (a highway in terms of our track analogy). Similar proposals for the emergence of basic word order patterns have been proposed both within emergentist (e.g., O'Grady 2005; 2013; Tomasello 2003) and generative perspectives (e.g., Townsend & Bever 2001). Importantly, however, just as with generalizations in perception and motor skills, the grammatical abstractions are not explicitly represented but result from the merging of item-based procedures for chunking. Consequently, there is no representation of grammatical structure separate from processing. Learning to process *is* learning the grammar.

7. Conclusion

The perspective developed in this article sees language as composed of a myriad of specific processing episodes, where particular messages are conveyed and understood. Like other action sequences, linguistic acts have their structure in virtue of the cognitive mechanisms that produce and perceive them. We have argued that the structure of language is, in particular, strongly affected by a severe limitation on human memory: the Now-or-Never bottleneck. Sequential information, at many levels of analysis, must rapidly be recoded to avoid being interfered with or overwritten by the deluge of subsequent material. To cope with the Now-or-Never bottleneck, the language system chunks new material as rapidly as possible at a range of increasingly abstract levels of representation. As a consequence, Chunk-and-Pass processing induces a multilevel structure over linguistic input. The history of the process of chunk building can be viewed as analogous to a shallow surface structure in linguistics, and the repertoire of possible chunking mechanisms and the principles by which they can be combined can be viewed as defining a grammar. Indeed, we have suggested that chunking procedures may be one interpretation of the constructions that are at the core of linguistic theories of construction grammar. More broadly, the Now-or-Never bottleneck promises to provide a framework within which to reintegrate the language sciences, from the psychology of language comprehension and production, to language acquisition, language change, and evolution, to the study of language structure.

ACKNOWLEDGMENTS
We would like to thank Inbal Arnon, Amui Chong, Brandon Conley, Thomas Farmer, Adele Goldberg, Ruth Kempson, Stewart McCauley, Michael Schramm, and Julia Ying, as well as seven anonymous reviewers for comments on previous versions of this article. This work was partially supported by BSF grant number 2011107, awarded to MHC (and Inbal Arnon), and ERC grant 295917-RATIONALITY, the ESRC Network for Integrated Behavioural Science, the Leverhulme Trust, Research Councils UK Grant EP/K039830/1 to NC.

NOTES

1. Levinson (2000) used the term *bottleneck* in a different, though interestingly related, way to refer to an asymmetry between the speed of language production and comprehension processes. Slower production processes are in this sense a "bottleneck" to communication.

2. Moreover, the rate of information transfer per unit of time appears to be quite similar across spoken and signed languages (Bellugi & Fischer 1972). Indeed, information transfer also appears to be roughly constant across a variety of spoken languages (Pellegrino et al. 2011).

3. Note that the Chunk-and-Pass framework does not take a stand on whether "coded meaning" is necessarily computed before "enriched meaning" (for discussion, see Noveck & Reboul 2008). Indeed, to the extent that familiar "chunks" can be "gestalts" associated with standardized enriched meanings, then the coded meaning could, in principle, be bypassed. So, for example, *could you pass the salt* might be directly interpreted as a request, bypassing any putative initial representation as a yes/no question. Similarly, an idiom such as *kick the bucket* may directly be associated with the meaning *die*. The same appears to be true for non-idiomatic compositional "chunks" such as *to the edge* (Jolsvai et al. 2013). This viewpoint is compatible with a variety of perspectives in linguistics that treat multiword chunks in the same way as traditional lexical items (e.g., Croft 2001; Goldberg 2006).

4. Our framework is neutral about competing proposals concerning how precisely production and comprehension processes are entwined (e.g., Cann et al. 2012; Dell & Chang 2014; Pickering & Garrod 2013a) – but see Chater et al. (2016).

5. The phrase "just-in-time" has been used in the engineering field of speech synthesis in a similar way (Baumann & Schlangen 2012).

6. It is likely that some more detailed information is also maintained and accessible to the language system. For example, Levy et al. (2009) found more eye-movement regressions when people read *The coach smiled at the player tossed the Frisbee* compared with *The coach smiled toward the player tossed the Frisbee*. They argued that this is because *at* has contextually plausible neighbors (*as*, *and*) whereas *toward* does not. The regression suggests that, on encountering processing difficulty, the language system "checks back" to see whether it recognized an earlier word correctly – but does so only when other plausible alternative interpretations may be possible. This pattern requires that higher-level representations do not throw away lower-level information entirely. Some information about the visual (or perhaps phonological) form of *at* and *toward* must be maintained, to determine whether or not there are contextually plausible neighbors that might be the correct interpretation. This pattern is compatible with the present account: Indeed, the example of SF in section 2 indicates how high-level organization may be critical to retaining lower-level information (e.g., interpreting random digits as running times makes them more memorable).

7. It is conceivable that the presence of ambiguities may be a necessary component of an efficient communication system, in which easy-to-produce chunks are reused – thus becoming ambiguous – in a trade-off between ease of production and difficulty of comprehension (Piantadosi et al. 2012).

8. Although our account is consistent with the standard linguistic levels, from phonology through syntax to pragmatics, we envisage that a complete model may include finer-grained levels, distinguishing, for example, multiple levels of discourse representation. One interesting proposal along these lines, developed from the work of Austin (1962) and Clark (1996), is outlined in Enfield (2013).

9. Note that, in particular, the present viewpoint does not rule out the possibility that some detailed information is retained in processing (e.g., Goldinger 1998; Gurevich et al. 2010; Pierrehumbert 2002). But such detailed information can be retained only because the original input has been chunked successfully, rather than being stored in raw form.

10. It is often difficult empirically to distinguish bottom-up and top-down effects. Bottom-up statistics across large corpora of low-level representations can mimic the operation of high-level representations in many cases; indeed, the power of such statistics is central to the success of much statistical natural language processing, including speech recognition and machine translation (e.g., Manning & Schütze 1999). However, rapid sensitivity to background knowledge and nonlinguistic context suggests that there is also an important top-down flow of information in human language processing (e.g., Altmann 2004; Altmann & Kamide 1999; Marslen-Wilson et al. 1993) as well as in cognition, more generally (e.g., Bar 2004).

11. Strictly speaking, "good-enough first time" may be a more appropriate description. As may be true across cognition (e.g., Simon 1956), the language system may be a satisficer rather than a maximizer (Ferreira et al. 2002).

12. Some classes of learning algorithm can be converted from "batch-learning" to "incremental" or "online" form, including connectionist learning (Saad 1998) and the widely used expectation-maximization (EM) algorithm (Neal & Hinton 1998), typically with diminished learning performance. How far it is possible to create viable online versions of existing language acquisition algorithms is an important question for future research.

13. Nonetheless, as would be expected from a statistical model of learning, some early productivity is observed at the word level, where words are fairly independent and may not form reliable chunks (e.g., children's determiner-noun combinations, Valian et al. 2009; Yang 2013; though see McCauley & Christiansen 2014b; Pine et al. 2013, for evidence that such productivity is not driven by syntactic categories).

14. Interestingly, though, the notion of "triggers" in the principles and parameters model (Chomsky 1981) potentially fits with the online learning framework outlined here (Berwick 1985; Fodor 1998; Lightfoot 1989): Parameters are presumed to be set when crucial "triggering" information is observed in the child's input (for discussion, see Gibson & Wexler 1994; Niyogi & Berwick 1996; Yang 2002). However, this model is very difficult to reconcile with incremental processing and, moreover, it does not provide a good fit with empirical linguistic data (Boeckx & Leivada 2013).

15. Note that the stability–plasticity dilemma arises in connectionist modelling: models that globally modify their weights, in response to new items, often learn only very slowly, to avoid "catastrophic interference" with prior items (e.g., French 1999; McCloskey & Cohen 1989; Ratcliff 1990). Notably, though, catastrophic interference may occur only if the old input rarely reappears later in learning.

16. Although Givón (1979) discussed how syntactic constructions might derive from previous pragmatic discourse structure, he did not coin the phrase "today's syntax is yesterday's discourse." Instead, it has been ascribed to him through paraphrasings of his maxim that "today's morphology is yesterday's syntax" from Givón (1971), an idea he attributed to the Chinese philosopher Lao Tse.

17. Apparent counterexamples to the general unidirectionality of grammaticalization – such as the compound verb *to up the ante* (e.g., Campbell 2000) – are entirely compatible with the present approach: They correspond to the creation of new idiomatic chunks, from other pre-existing chunks, and thus do not violate our principle that chunks generally decay.

18. It remains, of course, of great interest to understand the biological evolutionary history that led to the cognitive pre-requisites for the cultural evolution of language. Candidate mechanisms include joint attention, large long-term memory, sequence processing ability, appropriate articulatory machinery, auditory processing systems, and so on. But this is the study not of language evolution but of the evolution of the biological precursors of language (Christiansen & Chater 2008; for an opposing perspective, see Pinker & Bloom 1990).

19. The term "computational procedure" is also used by Sagarra and Herschensohn (2010), but they viewed such procedures as developing "in tandem with the growing grammatical competence" (p. 2022). Likewise, Townsend and Bever (2001) discussed "frequency-based perceptual templates that assign the initial meaning" (p. 6). However, they argued that this only results in "pseudosyntactic" structure, which is later checked against a complete derivational structure. In contrast, we argue that these computational procedures are all there is to grammar, a proposal that dovetails with O'Grady's (2005; 2013) notion of "computational routines," but with a focus on chunking in our case.

Open Peer Commentary

The ideomotor recycling theory for language

doi:10.1017/S0140525X15000680, e63

Arnaud Badets
CNRS, Institut de Neurosciences Cognitives et Intégratives d'Aquitaine (UMR 5287), Université de Bordeaux, France.
arnaud.badets@u-bordeaux.fr
http://www.incia.u-bordeaux1.fr/spip.php?article255

Abstract: For language acquisition and processing, the ideomotor theory predicts that the comprehension and the production of language are functionally based on their expected perceptual effects (i.e., linguistic events). This anticipative mechanism is central for action–perception behaviors in human and nonhuman animals, but a recent ideomotor recycling theory has emphasized a language account throughout an evolutionary perspective.

The Now-or-Never bottleneck, according to Christiansen & Chater (C&C), is, in a broad-spectrum view, a convincing constraint in language acquisition and processing. From general action–perception principles, this bottleneck deals with a myriad of linguistic inputs to recode them by chunks as rapidly as possible. Accordingly, language processing involves a prediction (or anticipation) mechanism that encodes new linguistic feature very rapidly. I agree with this general position, but the described predictive mechanism in charge of such anticipation does not seem theoretically conclusive in regard to a recent ideomotor recycling theory (Badets et al. 2016).

Sensorimotor and predictive mechanisms have been clearly theorized in the last 40 years (Adams 1971; Shin et al. 2010; Wolpert et al. 2011). For example, as suggested in the Now-or-Never bottleneck framework, the computational modeling approaches of motor control assume that two kinds of internal models are in charge of producing goal-directed behaviors (e.g., Wolpert et al. 2001). The first is the forward model, which predicts the expected sensory consequences as a function of the motor command. The second is the inverse model, a mechanism that transforms the expected sensory consequences into motor commands. Basically, the inverse model is related to a motor plan to reach the expected goal, and the forward model is in charge of monitoring an action by comparing the expected sensory consequences to the actual sensory consequences. Differences can cause an adaptation to the motor mechanism in order to attain the goal. For efficient regulations of goal-directed actions, the forward and inverse models are equally central. However, this theoretical framework assumes an equivalent weight for the representation underlying the expected perceptual effects and the representation of the behavior to achieve these effects. In contrast, the ideomotor theory does not deny the involvement of a movement system but assumes a primary role for expected perceptual events, which could be central in language production and comprehension (Badets et al. 2016 see also Kashima et al. 2013).

Ideomotor theory predicts that behaviors are functionally linked to their sensory consequences (Greenwald 1970; Hommel et al. 2001). The core mechanism is that actions are represented mainly by the expected perceptual consequences (or effects) they aim to produce in the environment. From an evolutionary account, it is obvious that such an action–perception mechanism dedicated to situated interaction is present for millions of years, since ancestral animals (Cisek & Kalaska 2001). Moreover, Badets et al. (2016 have recently suggested "we can easily assume that there is a reuse of cognitive function from mechanisms of simple motor control to more elaborated communication and language processing" (p. 11). In this theory based on the concept of exaptation (Gould & Vrba 1982), the ideomotor mechanism is recycled (i.e., exapted) during evolution or normal development in order to manage new world interaction like the human language (see Anderson 2010 for a neuronal reuse perspective).

According to the ideomotor recycling theory, the expected consequences of abstract meanings are simulated in an anticipative way in order to retrieve the appropriate and concrete words and sentences during the production (≈ action) and the comprehension (≈ perception) of language. Importantly, and as suggested by Greenwald (1972), "it ought to be possible to select a response very directly, perhaps totally bypassing any limited-capacity process, by presenting a stimulus that closely resembles the response's sensory feedback" (p. 52). Consequently, we can easily clarify the close alignment of linguistic meanings during a dialogue between two persons (see also Pickering & Garrod 2013a). In this context, an utterance is represented by the expected consequences of abstract meanings for speaker, which can be processed (≈ stimulus processing) very rapidly, as expected meanings for the subsequent utterance in listener (≈ sensory feedback). For the ideomotor recycling theory, there are common representational formats between shared abstract meanings during a dialogue.

Finally, there is another piece, but indirect, of evolutionary evidence for an ideomotor account in language processing. Indeed, for Gärdenfors (2004) "there has been a co-evolution of cooperation about future goals and symbolic communications" (p. 243). Corballis (2009) suggested the same mutual mechanism between the capacity to envision the far future and language processing. Consequently, if the ideomotor recycling theory can explain some parts of human language (Badets et al. 2016), it could be argued that the same recycled mechanism can also be in charge for the representation of the future (see also Badets & Rensonnet 2015). In this view, Badets and Osiurak (2015) have recently suggested that such an anticipative mechanism could be central for the representation of future scenarios. From different paradigms and domains like tool use, action memory, prospective memory, or motor skill learning, compelling evidence highlights that the ideomotor mechanism can predict far-future-situated events to adapt different and efficient behaviors. For Corballis (2009), language has the capacity to improve the representation of such future scenarios. For instance, in telling a person what will happen next week (associated with predicted storm weather) if he or she practices sailing, it is possible to form a coherent accident representation that can be avoided in the future. This mutual ideomotor mechanism between language and the capacity to envision the future gives an evident evolutionary advantage for humans.

To conclude, it seems that, from an evolutionary perspective, the ideomotor mechanism has been recycled in order to spread its influence on human behavior beyond simple motor acts. The ideomotor recycling theory can apply to language processing and other higher cognitive functions as foresight. For language, common representational formats between shared and expected abstract meanings during a dialogue can explain very rapid and efficient language skills.

Language processing is not a race against time

doi:10.1017/S0140525X15000692, e64

Giosuè Baggio[a] and Carmelo M. Vicario[b]
[a]*Language Acquisition and Language Processing Lab, Department of Language and Literature, Norwegian University of Science and Technology, 7491 Trondheim, Norway;* [b]*School of Psychology, Bangor University, Bangor, Gwynedd LL57, 2AS, United Kingdom.*
giosue.baggio@ntnu.no c.m.vicario@bangor.ac.uk
http://www.ntnu.edu/employees/giosue.baggio
http://www.bangor.ac.uk/psychology/people/profiles/carmelo_vicario.php.en

Abstract: We agree with Christiansen & Chater (C&C) that language processing and acquisition are tightly constrained by the limits of sensory and memory systems. However, the human brain supports a range of cognitive functions that mitigate the effects of information processing bottlenecks. The language system is partly organised around

these moderating factors, not just around restrictions on storage and computation.

Christiansen & Chater's (C&C's) theory builds upon the notion that linguistic structures and processes are specific responses to the limitations of sensory and memory systems. Language relies on three main strategies – incrementality, hierarchical representation, and prediction – for coping with those limitations. We think this list is incomplete, and that it should also include inference, the ability to read and write, pragmatic devices for coordinating speaker and hearer, and the mutual tuning of speech comprehension and production systems in the brain. We aim to show that this is more than merely adding items to a list: Our argument points to a different balance between restrictions on storage and computation, and the full range of cognitive functions that have a mitigating effect on them. Indeed, C&C's concise inventory satisfies all constraints, but no more: Language processing remains a race against time. We argue instead that the moderating factors widely offset the constraints, suggesting a different picture of language than the one envisaged by C&C.

Hearing is the main input channel for language. C&C discuss constraints on auditory analysis, but not the mechanisms by which the brain recovers lost information. Sensory systems rely heavily on perceptual inference. A classic example is phonemic restoration (PhR) (Warren 1970): Deleting auditory segments from speech reduces comprehension, but if the deleted segment is replaced with noise, comprehension is restored. PhR is not the creation of an illusory percept but the reorganisation of input: Because PhR arises in auditory cortices, it requires that energy be present at the relevant frequencies (Petkov et al. 2007). Short segments of unprocessed speech are not necessarily lost but may often be reconstructed. Probabilistic and logical inference to richer structures based on sparse data is available at all levels of representation in language (Graesser et al. 1994; Swinney & Osterhout 1990).

Vision is next in line in importance. In C&C's theory, vision has largely a supporting role: It may provide cues that disambiguate speech, but it is itself subject to constraints like auditory processing. However, the human brain can translate information across modalities, such that constraints that apply to one modality are weaker or absent in another. This applies to some innovations in recent human evolutionary history, including reading and writing. By the nature of texts as static visual objects, the effects of temporal constraints on information intake may be reduced or abolished. This is not to say there are no temporal constraints on reading: Processing the fine temporal structure of speech is crucial for reading acquisition (Dehaene 2009; Goswami 2015). Written information, though, is often freely accessible in ways that auditory information is not. We acquire a portion of our vocabulary and grammar through written language, and we massively use text to communicate. Therefore, it seems that C&C's premise that "language is fleeting" applies to spoken language only, and not to language in general.

But even auditory information can often be freely re-accessed. Misperception and the loss of information in conversation pose coordination problems. These can be solved by deploying a number of pragmatic devices that allow speaker and hearer to realign: echo questions are one example (A: "I just returned from Kyrgyzstan." B: "You just returned from where?"). Information is re-accessed by manipulating the source of the input, with (implicit) requests to slow down production, or to deliver a new token of the same type. Language use relies on a two-track system (Clark 1996): (1) communication about "stuff" and (2) communication about communication. Track 2 allows us to recover from failure to process information on Track 1, and to focus attention on what matters from Track 1. Signals from both tracks are subject to bottlenecks and constraints; nonetheless, Track 2 alleviates the effects of restrictions on Track 1 processing. Interestingly, infants are able to engage in repair of failed messages (Golinkoff 1986). This capacity develops early in childhood, in parallel with language growth (Brinton et al. 1986a; Saxton et al. 2005; Tomasello et al. 1990; Yonata 1999).

Finally, C&C claim that different levels of linguistic representation are mapped onto a hierarchy of cortical circuits. Each circuit chunks and passes elements at increasingly larger timescales. But research indicates the picture is rather more complicated. Most brain regions can work at *multiple* timescales. Frontal and temporal language cortices can represent and manipulate information delivered at different rates, and over intervals of different duration (Fuster 1995; Pallier et al. 2011; Ding et al. 2016). Furthermore, the left parietal lobe is a critical region for both temporal processing (e.g., Vicario et al. 2013; see Wiener et al. 2010 for a review) and amodal (spoken and written) word comprehension (Mesulam 1998). The left inferior parietal cortex is a core area for speech comprehension and production because of its connections with wide portions of Wernicke's (superior temporal cortex [STC]) and Broca's (left inferior frontal gyrus) areas (Catani et al. 2005). The temporal cortex processes speech at different scales: at shorter windows (25–50 ms) in the left STC, and at longer windows (150–250 ms) in the right STC (Boemio et al. 2005; Giraud et al. 2007; Giraud & Poeppel 2012). This asymmetry might result from mutual tuning of primary auditory and motor cortices in the left hemisphere (Morillon et al. 2010). If speech production and perception indeed share some of the constraints described by C&C, then neither system should be expected to lag behind the speed or the resolution of the other.

A more comprehensive theory of language processing would arise from taking into account constraints of the kind discussed by C&C, plus a wider array of cognitive mechanisms mitigating the effect of these constraints, including (but not limited to) Chunk-and-Pass processing and its corollaries. The human brain's ubiquitous capacity to infer, recover, and re-access unprocessed, lost, or degraded information is as much part of the "design" of the language system as incrementality, hierarchical representation, and prediction. The joint effect of these strategies is to make language processing much less prone to information loss and much less subject to time pressures than C&C seem to imply.

Pro and con: Internal speech and the evolution of complex language

doi:10.1017/S0140525X15000709, e65

Christina Behme
Department of Philosophy, Dalhousie University, Halifax, NS B3H 4R2, Canada.
christinabehme@gmail.com

Abstract: The target article by Christiansen & Chater (C&C) offers an integrated framework for the study of language acquisition and, possibly, a novel role for internal speech in language acquisition. However, the "Now-or-Never bottleneck" raises a paradox for language evolution. It seems to imply that language complexity has been either reduced over time or has remained the same. How, then, could languages as complex as ours have evolved in prelinguistic ancestors? Linguistic Platonism could offer a solution to this paradox.

Christiansen & Chater (C&C) promise to provide "an integrated framework for explaining many aspects of language structure, acquisition, processing, and evolution that have previously been treated separately" (sect. 1, para. 5). This integration results in a plausible language acquisition model. Citing a wealth of compelling empirical evidence, C&C propose that language is learned like other skills and that linguistic abilities are not isolated biological traits, as suggested by Hauser et al. (2014), but continuous with other motor and cognitive skills. Rejecting the Chomskyan dogma that language learning is effortless and (virtually) instantaneous (Chomsky 1975; 1980; 1986; 2012), C&C propose that "the Now-or-Never bottleneck requires that language acquisition be viewed as a type of skill learning, such as learning to drive, juggle, play the violin, or play chess. Such skills appear to be learned through *practicing* the skill, using online feedback during the practice itself …" (sect. 4, para. 4). This view integrates language naturally within cognition and does not require the postulation of domain-specific cognitive modules. Additionally, C&C's account casts doubt on Chomsky's claim that the fact that we frequently talk silently to ourselves supports his view that the function of language is not communication (e.g., Chomsky 2000; 2002; 2012). A more parsimonious explanation would assume that frequent internal monologues arose from the habitual "practice" (fine-tuning by [silently] doing) of language learning. C&C argue that "we should expect the exploitation of memory to require 'replaying' learned material, so that it can be reprocessed" (sect. 4.1, para. 5). They cite substantial neuroscientific evidence that such replay occurs and propose that dreaming may have a related function. Given that especially the integration of available information across different types and levels of abstraction and the anticipation of responses might require more practice than the motor-execution of (audible) speech, silent self-conversation might initially provide an additional medium for language learning. Later in life, such internal monologue could be recruited to the function Chomsky envisioned. Future research could uncover at what age children begin using internal monologue, to what degree second-language acquisition is assisted by learners switching their internal monologue from L1 to L2, and whether the lack of internal monologue (e.g., Grandin 2005) has negative effects on fluency in production.

Although C&C's account offers an attractive language acquisition model, it seems to create a paradox for language evolution. C&C argue that there are strong pressures toward simplification and reduction. For example, when a very simple artificial toy language was simulated, it "collapsed into just a few different forms that allowed for systematic, albeit semantically underspecified, generalization … In natural language, however, the pressure toward reduction is normally kept in balance by the need to maintain effective communication" (sect. 5, para. 4). This observation raises the following problem: For an existing, fairly complex system, simplification may indeed lead to the kinds of changes C&C discuss (e.g., that "chunks at each level of linguistic structure – discourse, syntax, morphology, and phonology – are potentially subject to reduction" [sect. 5, para. 5]). But in this view there is a strong pressure toward simplification and virtually no possibility of increasing complexity. Yet it is not clear why the language of our distant ancestors would have been more complex than or at least as complex as modern languages. It has been argued convincingly that the complexity of grammar actually needed to support most daily activities of humans living in complex contemporary societies is substantially less than that exhibited by any contemporary human language (Gil 2009, p. 19), and it seems implausible that existing language complexity is functionally motivated.

If the Now-or-Never bottleneck has the power C&C attribute to it, it must have constrained language learning and use for our distant ancestors in the same way as it does for us. Presumably these ancestors had cognitive capacities that were not superior to ours, and their culture would have imposed even fewer demands for linguistic complexity than contemporary culture. So how could they have evolved a highly complex language system that in turn could be reduced to provide the cognitive foundation for grammaticalization? C&C suggest analogies between language and other cognitive processes (e.g., vision). This is problematic because the visual system evolved to perceive objects that exist independently of this system. On purely naturalist accounts, languages have no existence independent of human brains or human culture. Therefore, both the cognitive abilities underwriting linguistic competence and the language that is learned must have evolved. Decades ago it was suggested that many of the problems that bedevil Chomskyan linguistics can be eliminated if one adopts linguistic Platonism and draws a distinction between the knowledge speakers have of their language and the languages that speakers have knowledge of. Platonism considers as distinct (1) the study of semantic properties and relations like ambiguity, synonymy, meaningfulness, and analyticity, and (2) the study of the neuropsychological brain-states of a person who has acquired knowledge about these semantic properties (e.g., Behme 2014a; Katz 1984; 1996; 1998; Katz & Postal 1991; Neef 2014; Postal 2003; 2009). In such a view, languages and brains that have acquired knowledge of languages are two distinct ontological systems.

In addition to eliminating many problems for contemporary linguistics, such a view also might resolve the language evolution paradox because languages have an independent existence, and only human cognitive capacity evolves. It might be argued that the epistemology of linguistic Platonism is hopeless. Although this is not the place to defend linguistic Platonism, one should remember that in mathematics it is widely accepted that the number systems exist independently of human brains and human culture, and are discovered, just as are other objects of scientific discovery. It has been argued that if one accepts the possibility of mathematical realism, there is no a priori reason to reject the possibility of linguistic realism (e.g., Behme 2014b; Katz 1998). Before rejecting linguistic Platonism out of hand, one ought to remember that

> For psychology, AI, and the related cognitive sciences, the question of what a grammar is a theory of is important because its answer can resolve troublesome issues about where the linguist's work ends and the cognitive scientist's begins. A Platonist answer to this question would clearly divide linguistics and cognitive sciences so that the wasteful and unnecessary quarrels of the past can be put behind us. (Katz 1984, p. 28)

ACKNOWLEDGMENTS
I thank Vyvyan Evans, David Johnson, and Paul Postal for comments on earlier drafts and accept full responsibility for remaining errors.

Socio-demographic influences on language structure and change: Not all learners are the same

doi:10.1017/S0140525X15000710, e66

Till Bergmann,[a] Rick Dale,[a] and Gary Lupyan[b]
[a]*Cognitive and Information Sciences, University of California, Merced, Merced, CA 95343;* [b]*Department of Psychology, University of Wisconsin–Madison, Madison, WI 53706.*
tbergmann@ucmerced.edu www.tillbergmann.com
rdale@ucmerced.edu http://cognaction.org/rick/
glupyan@gmail.com http://sapir.psych.wisc.edu/

Abstract: The Now-or-Never bottleneck has important consequence for understanding why languages have the structures they do. However, not addressed by C&C is that the bottleneck may interact with *who* is doing the learning: While some languages are mostly learned by infants, others have a large share of adult learners. We argue that such socio-demographic differences extend and qualify C&C's thesis.

We wholeheartedly agree with Christiansen & Chater (C&C) that "acquiring a language is learning to process" (sect. 5, para. 3) and that "there is no representation of grammatical structure separate from processing" (sect. 6.2, para. 6). We also agree with C&C's more general thesis that the structure of language cannot be understood without taking into account the constraints and biases of the language learners and users. Although the Now-or-Never cognitive bottleneck is an unavoidable constraint on language comprehension and production, fully understanding its consequences requires taking into account socio-demographic realities, namely *who* is doing the language learning.

C&C write that "Language will be shaped by the linguistic patterns learners find easiest to acquire and process" (sect. 5, para. 3), but what is *easiest* may importantly depend on *who* is doing the learning. Some languages are learned exclusively by infants and used in small, culturally homogeneous communities. For example, half of all languages have fewer than 7,000 speakers. Other languages have substantial populations of non-native speakers and are used in large, culturally and linguistically heterogeneous communities. For example, at present, about 70% of English speakers are non-native speakers (Gordon 2005).

The socio-demographic niche in which a language is learned and used can influence its grammar insofar as different kinds of learners have different learning biases. Languages with many adult learners may adapt to their socio-demographic niche by eschewing features difficult for adults to process. Indeed, as Lupyan and Dale (2010) have shown in an analysis of more than 2,000 languages, languages spoken in larger and more diverse communities (those that tend to have more non-native learners) have simpler morphologies and fewer obligatory markings (see also Bentz & Winter 2013). In contrast, languages used in a socio-demographic niche containing predominantly infant learners tend to have many more obligatory markings – for example, they are more likely to encode tense, aspect, evidentiality, and modality as part of the grammar, and to have more complex forms of agreement (Dale & Lupyan 2012; see also Dahl 2004; McWhorter 2001; Trudgill 2011).

Such influences of the socio-demographic environment on language structure are important caveats to C&C's thesis because the Now-or-Never bottleneck, although present in all learners, depends on the knowledge that a learner brings to the language-learning task.

On C&C's account, successful language processing depends on recoding the input into progressively higher-level (more abstract) chunks. As an analogy, C&C give the example of how remembering strings of numbers is aided by chunking (re-representing) numbers as running times or dates (sect. 2, para. 7). But, of course, this recoding is only possible if the learner knows about reasonable running times and the format of dates. The ability to remember the numbers depends on the ability to chunk them, and the ability to chunk them depends on prior knowledge.

In the case of language learning, recoding of linguistic input is "achieved by applying the learner's *current* model of the language" (sect. 4.1, para. 3) and further constrained by memory and other domain-general processes. But both the learner's language model and domain-general constraints vary depending on *who* the learner is.

Infants come to the language-learning task with a less developed memory and ability to use pragmatic and other extralinguistic cues to figure out the meaning of an utterance. As a consequence, the Now-or-Never bottleneck is strongly in place. The language adapts through increased grammaticalization that binds units of meaning more tightly, thereby increasing redundancy. For example, grammatical gender of the sort found in many Indo-European languages increases redundancy by enforcing agreement of nouns, adjectives, and pronouns, making one more predictable from the other and – arguably – reducing the memory load required for processing the utterances.

Adults come to the language-learning task with more developed memories, and ability for pragmatic inference, but at the same time they are biased by pre-existing chunks that may interfere with chunks that most efficiently convey the meaning in the new language. The greater memory capacities and ability to use contextual and other pragmatic cues to infer meanings, may relax the Now-or-Never bottleneck, nudging grammars toward morphological simplification with its accompanying decrease in obligatory markings (i.e., decrease in redundancy) and increase in compositionality (Lupyan & Dale 2015).

This reasoning helps explain *how* the Now-or-Never bottleneck can create "obligatorification" (sect. 5, para. 8) and also why some languages have more obligatory markings than other languages.

In summary, although we agree with C&C that "multiple forces influence language change in parallel" (sect. 5, para. 9), we emphasize the force constituted by the learning community. Languages adapt to the specific (cognitive) learning constraints and communicative needs of the learners and speakers.

Now or … later: Perceptual data are *not* immediately forgotten during language processing

doi:10.1017/S0140525X15000734, e67

Klinton Bicknell,[a] T. Florian Jaeger,[b,c,d] and Michael K. Tanenhaus[b,d]

[a]*Department of Linguistics, Northwestern University, Evanston, IL 60208-0854;* [b]*Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY 14627-0268;* [c]*Department of Computer Science, University of Rochester, Rochester, NY 14627-0226;* [d]*Department of Linguistics, University of Rochester, Rochester, NY 14627-0096.*

kbicknell@northwestern.edu fjaeger@mail.bcs.rochester.edu
mtan@mail.bcs.rochester.edu
http://faculty.wcas.northwestern.edu/kbicknell/
http://www.bcs.rochester.edu/people/fjaeger/
https://www.bcs.rochester.edu/people/mtan/mtan.html

Abstract: Christiansen & Chater (C&C) propose that language comprehenders must immediately compress perceptual data by "chunking" them into higher-level categories. Effective language understanding, however, requires maintaining perceptual information long enough to integrate it with downstream cues. Indeed, recent results suggest comprehenders do this. Although cognitive systems are undoubtedly limited, frameworks that do not take into account the tasks that these systems evolved to solve risk missing important insights.

Christiansen & Chater (C&C) propose that memory limitations force language comprehenders to compress perceptual data immediately, forgetting lower-level information and maintaining only higher-level categories ("chunks"). Recent data from speech perception and sentence processing, however, demonstrate that comprehenders can maintain fine-grained lower-level perception information for substantial durations. These results directly contradict the central idea behind the Now-or-Never bottleneck. To the extent that the framework allows them, it risks becoming so flexible that it fails to make substantive claims. On the other hand, these results are predicted by existing frameworks, such as bounded rationality, which are thus more productive frameworks for future research. We illustrate this argument with recent developments in our understanding of a classic result in speech perception: categorical perception.

Initial results in speech perception suggested that listeners are insensitive to fine-grained within-category differences in voice onset time (VOT, the most important cue distinguishing voiced and voiceless stop consonants, e.g., "b" versus "p" in *bill* versus *pill*), encoding only whether a sound is "voiced" or "voiceless" (Liberman et al. 1957). Subsequent work demonstrated sensitivity to within-category differences (Carney et al. 1977; Pisoni & Tash 1974), with some findings interpreted as evidence this sensitivity rapidly decays (e.g., Pisoni & Lazarus 1974). Such a picture is very similar to the idea behind Chunk-and-Pass: Listeners rapidly chunk phonetic detail into a phoneme, forgetting the subcategorical information in the process.

Although it may perhaps be intuitive, given early evidence that perceptual memory is limited (Sperling 1960), such discarding of subcategorical information would be surprising from the perspective of bounded rationality: Information critical to the successful recognition of phonetic categories often occurs downstream in the speech signal (Bard et al. 1988; Grosjean 1985). Effective language understanding thus requires maintaining and integrating graded support for different phonetic categories provided by a sound's acoustics (its subcategorical information) with information present in the downstream signal. Indeed, more recent work suggests that comprehenders do this. For example, within-category differences in VOT are *not* immediately forgotten but are still available downstream at the end of a multisyllabic word (McMurray et al. 2009; see Dahan 2010, for further discussion of right-context effects).

Of particular relevance is a line of work initiated by Connine et al. (1991, Expt. 1). They manipulated VOT in the initial segment of target words (*dent/tent*) and embedded these words in utterances with downstream information about the word's identity (e.g., "The dent/tent in the fender" or "… forest"). They found that listeners can maintain subcategorical phonetic detail and integrate it with downstream information *even beyond word boundaries*.

Chunk-and-Pass does not predict these results. Recognizing this, C&C allow violations of Now-or-Never, as long as "such online 'right-context effects' [are] highly local, because raw perceptual input will be lost if it is not rapidly identified" (sect. 3.1, para. 7). This substantially weakens the predictive power of their proposal. On the other hand, Connine et al.'s results do seem to support this qualification. They reported that subcategorical phonetic detail (a) was maintained only 3 syllables downstream, but not 6–8, and (b) was maintained only for maximally ambiguous tokens.

Recent work, however, points to methodological issues that call both of these limitations into question (Bicknell et al. 2015). Regarding (a), Connine et al. allowed listeners to respond at any point in the sentence: On 84% of trials in the 6–8 syllable condition, listeners categorized the target word prior to hearing the relevant right-context (e.g., *fender* or *forest*). Therefore, these responses could not probe access to subcategorical information. In a replication that avoided this problem, we found that subcategorical detail decays more slowly than Connine et al.'s analysis would suggest: Subcategorical detail was maintained for at least 6–8 syllables (the longest range investigated). Regarding (b), Connine et al.'s analysis was based on proportions, rather than log-odds. Rational integration of downstream information with subcategorical information should lead to additive effects in log-odds space (which, in proportional space, then are largest around the maximally ambiguous tokens; Bicknell et al. 2015), This is indeed what we found: The effect of downstream information on the log-odds of hearing *dent* (or *tent*) was constant across the entire VOT range. In short, subcategorical information is maintained longer than previous studies suggested, not immediately discarded by chunking (see also Szostak & Pitt 2013). Moreover, maintenance is not limited to special cases; it is the default (Brown et al. 2014).

Clearly, language processing is subject to cognitive limitations; many – if not most – theories of language processing acknowledge this. In its general form, the Now-or-Never bottleneck thus embodies an idea as old as the cognitive sciences: that observable behavior and the cognitive representations and mechanisms underlying this behavior are primarily driven by a priori (static/fixed) cognitive *limitations*. This contrasts with another view: Cognitive and neural systems have evolved *efficient* solutions to the computational tasks agents face (Anderson 1990). Both views have been productive, providing explanations for perception, motor control, and cognition, including language (and C&C have contributed to both views). A number of proposals have tied together these insights. This includes the idea of bounded rationality, that is, rational use of limited resources given task constraints (Howes et al. 2009; Neumann et al. 2014; Simon 1982; for language: e.g., Bicknell & Levy 2010; Feldman et al. 2009; Kleinschmidt & Jaeger 2015; Kuperberg & Jaeger 2016; Lewis et al. 2013). Chunk-and-Pass is a step backward because it blurs the connection between these two principled dimensions of theory development. Consequently, it fails to predict systematic maintenance of subcategorical information, whereas bounded rationality predicts this property of language processing *and* offers an explanation for it.

The Now-or-Never bottleneck makes novel, testable predictions only insofar as it makes strong claims about comprehenders' (in)ability to maintain lower-level information beyond the "now." The studies we summarized above are inconsistent with this claim. Similarly inconsistent is evidence from research on reading suggesting that lower-level information survives long enough to influence incremental parsing (Levy 2011; Levy et al. 2009). Moreover, the history of research on categorical perception provides a word of caution: Rather than focusing too much on cognitive limitations, it is essential for researchers to equally consider the computational problems of language processing and how comprehender goals can be effectively achieved.

Linguistic representations and memory architectures: The devil is in the details

doi:10.1017/S0140525X15000746, e68

Dustin Alfonso Chacón, Shota Momma, and Colin Phillips
Department of Linguistics, University of Maryland, College Park, MD 20742.
dustin.alfonso@gmail.com shotam2@gmail.com colin@umd.edu
ling.umd.edu/colin

Abstract: Attempts to explain linguistic phenomena as consequences of memory constraints require detailed specification of linguistic representations and memory architectures alike. We discuss examples of supposed locality biases in language comprehension and production, and their link to memory constraints. Findings do not generally favor Christiansen & Chater's (C&C's) approach. We discuss connections to debates that stretch back to the nineteenth century.

It is important to understand how language is shaped by cognitive constraints, and limits on memory are natural culprits. In this regard, Christiansen & Chater (C&C) join a tradition in language research that has a long pedigree (Frazier & Fodor 1978; Wundt 1904) and to which we are sympathetic. C&C's model aims to integrate an impressive range of phenomena, but the authors play fast and loose with the details; they mischaracterize a number of phenomena; and key predictions depend on auxiliary assumptions that are independent of their model. An approach that takes the details of linguistic representations and memory architectures more seriously will ultimately be more fruitful. We illustrate using examples from comprehension and production.

C&C propose that comprehenders can maintain only a few low-level percepts at once and must therefore quickly encode higher-order, abstract representations. They argue that this explains the pervasive bias for shorter dependencies. However, memory representations are more than simple strings of words that quickly vanish. Sentences are encoded as richly articulated, connected representations that persist in memory, perhaps without explicit encoding of order, and memory access is similarly articulated (Lewis et al. 2006). As evidence of their model, C&C cite *agreement attraction* in sentences like *The key to the cabinets are on the table*. These errors are common in production and often go unnoticed in comprehension, and it is tempting to describe them in terms of "proximity concord" (Quirk et al. 1972). But this is inaccurate. Agreement attraction is widespread in cases where the distractor is further from the verb than the true subject, as in *The musicians who the reviewer praise so highly will win* (Bock & Miller 1991). Attraction is asymmetrical, yielding "illusions of grammaticality" but not "illusions of ungrammaticality" (Wagers et al. 2009), and depends on whether the distractor is syntactically "active" (Franck et al. 2010). These facts are surprising if attraction reflects simple recency, but they can be captured in a model that combines articulated linguistic representations with a content-addressable memory architecture (Dillon et al. 2013;

McElree et al. 2003). Hence, agreement attraction fits C&C's broadest objective, deriving attraction from memory constraints, but only if suitably detailed commitments are made.

C&C also endorse the appealing view that locality constraints in syntax ("island effects": Ross 1967) can be reduced to memory-driven locality biases in the processing of filler-gap dependencies (Kluender & Kutas 1993). Details matter here, too, and they suggest a different conclusion. When linear and structural locality diverge, as in head-final languages such as Japanese, it becomes clear that the bias for shorter filler-gap dependencies in processing is linear, whereas grammatical locality constraints are structural (Aoshima et al. 2004; Chacón et al., submitted; Omaki et al. 2014).

The moral that we draw from these examples is that each reductionist claim about language must be evaluated on its own merits (Phillips 2013).

Turning to production, C&C argue that incrementality and locality biases reflect severe memory constraints, suggesting that we speak "into the void." This amounts to what is sometimes called *radical incrementality* (Ferreira & Swets 2002). It implies that sentence production involves word-by-word planning that is tightly synchronized with articulation – for example, planning is *just-in-time*, leading to a bias for local dependencies between words. However, this view of production does not reflect memory constraints alone, and it is empirically unwarranted.

Radical incrementality carries a strong representational assumption whose problems were pointed out in the late nineteenth century. The philologist Hermann Paul, an opponent of Wilhelm Wundt, argued that a sentence is essentially an associative sum of clearly segmentable concepts, each of which can trigger articulation in isolation. Radical incrementality requires this assumption, as it presupposes the isolability of each word or phrase in a sentence at all levels of representation. Memory constraints alone do not require this assumption, and so there is a gap in C&C's argument that memory constraints entail radical incrementality. Indeed, Wundt was already aware of memory limitations, and yet he adopted the contrasting view that sentence planning involves a successive scanning (*apperception*) of a sentence that is simultaneously present in the background of consciousness during speech (Wundt 1904). The historical debate illustrates that radical incrementality turns on representational assumptions rather than directly following from memory limitations.

Empirically, radical incrementality has had limited success in accounting for production data. Three bodies of data that C&C cite turn out to not support their view. First, the scope of planning at higher levels (e.g., conceptual) can span a clause (Meyer 1996; Smith & Wheeldon 1999). Also, recent evidence suggests that linguistic dependencies can modulate the scope of planning (Lee et al. 2013; Momma et al. 2015, in press). Second, since Wundt's time, availability effects on word order have not led researchers to assume radical incrementality (see Levelt 2012 for an accessible introduction to Wundt's views). Bock (1987) emphasized that availability effects on order result from the tendency for accessible words to be assigned a higher grammatical function (e.g., subject). In languages where word order and the grammatical functional hierarchy dissociate, availability effects support the grammatical function explanation rather than radical incrementality (Christianson & Ferreira 2005). Third, contrary to C&C's claim, early observations about speech errors indicated that exchange errors readily cross phrasal and clausal boundaries (Garrett 1980).

C&C could argue that their view is compatible with many of these data; memory capacity at higher levels of representation is left as a free parameter. But this is precisely the limitation of their model: Specific predictions depend on specific commitments. Radical incrementality is certainly possible in some circumstances, but it is not required, and this is unexpected under C&C's view that speaking reduces to a chain of word productions that are constrained by severe memory limitations.

To conclude, we affirm the need to closely link language processes and cognitive constraints, and we suspect the rest of the field does too. However, the specifics of the memory system and linguistic representations are essential for an empirically informative theory, and they are often validated by the counterintuitive facts that they explain.

Gestalt-like representations hijack Chunk-and-Pass processing

doi:10.1017/S0140525X15000758, e69

Magda L. Dumitru
Department of Cognitive Science, Macquarie University, Sydney NSW 2109, Australia.
magda.dumitru@gmail.com

Abstract: Christiansen & Chater (C&C) make two related and somewhat contradictory claims, namely that the ever abstract language representations built during Chunk-and-Pass processing allow for ever greater interference from extra-linguistic information, and that it is nevertheless the language system that re-codes incoming information into abstract representations. I analyse these claims and discuss evidence suggesting that Gestalt-like representations hijack Chunk-and-Pass processing.

Christiansen & Chater (C&C) argue that higher-level chunks preserve information from lower-level chunks albeit in a much impoverished form. However, they also suggest that there is no obligatory relationship between low-level chunks and high-level chunks. To support their claim, they cite the case of SF (cf. Ericsson et al. 1980), who could accurately recall as many as 79 digits after grouping them in locally meaningful units (e.g., historical dates and human ages). Moreover, they argue that the Now-or-Never bottleneck forbids broad parallelism in language at the expense of avoiding ambiguities (e.g., "garden path" sentences). In brief, C&C propose that chunks are only locally coherent and that their gist, however contradictory, is being safely kept track of at higher levels. Unfortunately, the authors remain silent about the mechanisms underlying higher-level representation formation.

C&C also declare themselves agnostic about the nature of chunks. Indeed, although there is ample psychological evidence for the existence of chunks in various types of experimental data, from pause durations in reading to naive sentence diagramming, chunks remain notoriously difficult to define. However, we have reasons to reject the possibility, which follows naturally from the Chunk-and-Pass framework, that chunks are arbitrary and may depend exclusively on memory limitations. To wit, chunks correspond most closely to intonational phrases (IPs) (cf. Gee & Grosjean 1983), which, in turn, are hard to capture by grammatical rules. For example, the sentence "This is the cat / that chased the rat / that ate the cheese" contains three IPs (separated by slashes) that fail to correspond to syntactic constituents (noun phrases or verb phrases). Yet IPs are not entirely free of structure, as they must begin at the edge of a syntactic constituent and end before or at the point where a syntactic constituent ends (cf. Jackendoff 2007). Moreover, although a given utterance can be carved up in several ways (hence, contain a variable number of IPs), carvings are not arbitrary and license only certain IP combinations and not others. We may therefore conclude that IPs and corresponding chunks must be globally coherent (i.e., fit well with each other) and depend on the meaning conveyed.

Furthermore, I believe that chunking is driven not by memory limitations nor by language structures, but by an overall need for coherence or meaningfulness (cf. Dumitru 2014). Indeed, evidence from memory enhancement techniques suggests that chunking must rely on global coherence. So, for example, memory contest champions who use the so-called *mind palace* technique (e.g., Yates 1966) often achieve impressive results. The method requires them to commit to long-term memory a vivid image associated with each item to be remembered (e.g., faces, digits, and lists of words) as well as a familiar walk through the palace where these images are stored at precise locations. Whenever necessary, contestants can retrace the walk through the palace (i.e., rely on global coherence) to recall a huge number of unrelated facts.

I also claim that coherence is grounded in a model of reality that people instantly build when recalling items or understanding language based on their experience with frequent patterns of perception and action (Barsalou 2008). Indeed, as shown in Altmann (2002) and in Altmann and Kamide (2009), for instance, people use available lexical information at the earliest stages of processing to anticipate upcoming words. Furthermore, as reported in Kamide et al. (2003), people target a larger sentential context during online processing, hence would look more readily toward a glass of beer when hearing "The man will taste the …" but towards candy when hearing "The girl will taste the …," for instance, although the verb itself combines equally well with "beer" and with "candy." Subsequently, Dumitru and Taylor (2014) reported that disjunction words like "or" cue knowledge about expected argument structure and sense depending on grounded sentential context. More important, language processing may reflect knowledge of the world that goes beyond people's awareness and beyond language structures (cf. Dumitru et al. 2013). In particular, when understanding conjunction and disjunction expressions, people rapidly establish grounded connections between the two items mentioned (i.e., the concepts evoked by the nouns linked by conjunction or by disjunction) in the form of Gestalts.

Accordingly, people shifted their gaze faster from the picture of an ant to the picture of a cloud when hearing "Nancy examined an ant and a cloud" than when hearing "Nancy examined an ant or a cloud"; they could instantly evoke a single Gestalt in conjunction situations (where they usually select both items mentioned) and two Gestalts in disjunction situations (where they usually select one of the items). As expected, their attention shifted faster between two object parts (in this case, two representations belonging to the same Gestalt) than between two objects (two representations belonging to different Gestalts). Subsequent work by Dumitru (2014) confirmed that Gestalts generate perceptual compatibility effects such that visual groupings of a particular set of stimuli (e.g., two different-coloured lines) had complementary effects on validation scores for conjunction descriptions as opposed to disjunction descriptions. Importantly, language users need not be aware of the dynamics of the concept-grounding process (i.e., why they shift their gaze between stimuli at a certain speed), and there are no language-related constraints that might explain these differences in behaviour.

To summarise, I have questioned the proposal by C&C that Chunk-and-Pass processing is exclusively driven by memory constraints and obeys the rules of the language system. Instead, I discussed recent evidence suggesting that chunking is driven by a need for global coherence manifested as Gestalt-like structures, which in turn underlie memory organisation and mirror real-world phenomena. I further suggested that Gestalts hijack Chunk-and-Pass processing when they generate online effects that are not language-related.

Consequences of the Now-or-Never bottleneck for signed versus spoken languages

doi:10.1017/S0140525X1500076X, e70

Karen Emmorey
School of Speech, Language, and Hearing Sciences, San Diego State University, San Diego, CA 92120.
kemmorey@mail.sdsu.edu
http://emmoreylab.sdsu.edu/

Abstract: Signed and spoken languages emerge, change, are acquired, and are processed under distinct perceptual, motor, and memory constraints. Therefore, the Now-or-Never bottleneck has different ramifications for these languages, which are highlighted in this commentary. The extent to which typological differences in linguistic structure can be traced to processing differences provides unique evidence for the claim that structure is processing.

Christiansen & Chater (C&C) make it clear that the consequences of the Now-or-Never bottleneck for language are not speech-specific. This commentary highlights how and why signed and spoken languages respond differently to the limitations imposed by the bottleneck. C&C argue that the Now-or-Never bottleneck arises from general principles of perceptuo-motor processing and memory, and both have different properties for visual-manual and aural-oral languages, which lead to adaptations and preferences that are specific to each language type.

The vocal articulators (lips, tongue, larynx) are smaller and quicker than the hands and arms, and the auditory system is generally more adept at temporal processing than the visual system, which is better at spatial processing. These perceptual and motoric differences exert distinct pressures and affordances when solving the problems presented by the Now-or-Never bottleneck. As a consequence, signed and spoken languages prefer different chunking strategies for structuring linguistic information. At the phonological level, spoken languages prefer what could be called *serial* chunking, whereas signed languages prefer *spatial* chunking. For example, for spoken languages, single-segment words are rare and multisegment words are common, but the reverse pattern holds for sign languages (Brentari 1998). Oversimplifying here, consonants and vowels constitute segment types for speech, while locations and movements constitute segment types for sign (e.g., Sandler 1986). Single-segment spoken words are rare because they are extremely short and generally limited to the number of vowels in the language. Single consonants violate the Possible-Word Constraint (Norris et al. 1997), which also applies to sign language (Orfanidou et al. 2010). Single-segment signs are not problematic because other phonological information – for example, hand configuration – can be produced (and perceived) simultaneously with a large number of possible single location or movement segments. Multisegment (> three) and multisyllabic signs are rare in part because the hands are relatively large and slow articulators, and this limits the number of serial segments that can be quickly chunked and passed on to the lexical level of representation.

Distinct preferences for serial versus spatial chunking are also found at the morphological level. Spoken languages show a general preference for linear affixation (specifically, suffixation) over nonconcatenative processes such as reduplication or templatic morphology (Cutler 1985). In contrast, linear affixation (particularly for inflectional morphology) is rare across sign languages, and simultaneous, nonconcatenative morphology is the norm. Aronoff et al. (2005) attributed the paucity of linear morphology to the youth of sign languages but acknowledged that processing constraints imposed by modality also shape this preference (Emmorey 1995). Specifically, the ability of the visual system to process spatially distributed information in parallel, the slow articulation rate of the hands, and limits on working memory all conspire to induce sign languages to favor simultaneous over sequential morphological processes. In fact, when the linear morphology of a spoken language is implemented in the visual-manual modality, as in Manually Coded English (MCE), deaf children who have no exposure to a natural sign language spontaneously create simultaneous morphology to mark verb arguments (Supalla 1991). In addition, linear manual suffixes in MCE are often incorrectly analyzed by children as separate signs because prosodically and perceptually they do not pattern like bound morphemes (Supalla & McKee 2002).

Although the architecture of the memory system is parallel for signed and spoken languages (Wilson & Emmorey 1997; 1998), immediate memory for sequences of items has consistently been found to be superior for speech (Bellugi et al. 1975, inter alia). Hall and Bavelier (2009) demonstrated that the serial span discrepancy between speech and sign arises during perception and encoding, but not during recall, where sign actually shows an advantage (possibly because visual feedback during signing does not interfere with the memory store, unlike auditory feedback during speaking; Emmorey et al. 2009). The source of these differences is still unclear, but the short-term memory capacity for sign (4–5 items) is typical of a variety of types of memory (Cowan 2000), and thus what needs to be explained is why the memory capacity for speech is unusually high.

Because sign languages emerge, change, are acquired, and are processed under distinct memory and perceptuo-motor constraints, they provide an important testing ground for C&C's controversial proposals that learning to process is learning the grammar and that linguistic structure is processing history. Typological differences between the structure of signed and spoken languages may be particularly revealing. Can such structural differences be explained by distinct processing adaptations to the Now-or-Never bottleneck? For example, given the bottleneck pressures, one might expect duality of patterning to emerge quickly in a signed language, but recent evidence suggests that it may not (Sandler et al. 2011). Could this be because the visual-manual and auditory-oral systems are "lossy" in different ways or because chunking processes differ between modalities? Given C&C's claim that "there is no representation of grammatical structure separate from processing" (sect. 6.2, para. 6), it is critical to determine whether the differences – and the commonalities – between signed and spoken languages can be traced to features of processing.

Linguistics, cognitive psychology, and the Now-or-Never bottleneck

doi:10.1017/S0140525X15000953, e71

Ansgar D. Endress[a] and Roni Katzir[b]
[a]*Department of Psychology, City University London, London EC1V 0HB, United Kingdom;* [b]*Department of Linguistics and Sagol School of Neuroscience, Tel Aviv University, Ramat Aviv 69978, Israel.*
ansgar.endress.1@city.ac.uk **rkatzir@post.tau.ac.il**

Abstract: Christiansen & Chater (C&C)'s key premise is that "if linguistic information is not processed rapidly, that information is lost for good" (sect. 1, para. 1). From this "Now-or-Never bottleneck" (NNB), C&C derive "wide-reaching and fundamental implications for language processing, acquisition and change as well as for the structure of language itself" (sect. 2, para. 10). We question both the premise and the consequentiality of its purported implications.

Problematic premises. Christiansen & Chater (C&C) base the Now-or-Never bottleneck (NNB) on the observation that sensory memory disappears quickly in explicit memory tasks. We note, first, that not all forms of explicit memory are short-lived. For example, children remember words encountered once after a month (Carey & Bartlett 1978; Markson & Bloom 1997). More important, it is by no means clear that explicit memory is the (only) relevant form of memory for language processing and acquisition, nor how quickly other forms of memory decay. For example, the perceptual learning literature suggests that learning can occur even in the absence of awareness of the stimuli (Seitz & Watanabe 2003; Watanabe et al. 2001) and sometimes has long-lasting effects (Schwab et al. 1985). Similarly, visual memories that start decreasing over a few seconds can be stabilized by presenting items another time (Endress & Potter 2014). At a minimum, then, such memory traces are long-lasting enough for repeated exposure to have cumulative learning effects.

Information that is not even perceived is thus used for learning and processing, and some forms of memory do not disappear immediately. Hence, it is still an open empirical question whether poor performance in explicit recall tasks provides severe constraints on processing and learning.

We note, in passing, that even if relevant forms of memory were short-lived, this would not necessarily be a bottleneck. Mechanisms to make representations last longer – such as self-sustained activity – are well documented in many brain regions (Major & Tank 2004), and one might assume that memories can be longer-lived when this is adaptive. Short-lived memories might thus be an adaptation rather than a bottleneck (e.g., serving to reduce information load for various computations).

Problematic "implications." C&C use the NNB to advance the following view: Language is a skill (specifically, the skill of parsing predictively); this skill is what children acquire (rather than some theory-like knowledge); and there are few if any restrictions on linguistic diversity. C&C's conclusions do not follow from the NNB and are highly problematic. Below, we discuss some of the problematic inferences regarding processing, learning, and evolution.

Regarding processing, C&C claim that the NNB implies that knowledge of language is the skill of parsing predictively. There is indeed ample evidence for a central role for prediction in parsing (e.g., Levy 2008), but this is not a consequence of the NNB: The advantages of predictive processing are orthogonal to the NNB, and, even assuming the NNB, processing might still occur element by element without predictions. C&C also claim that the NNB implies a processor with no explicit representation of syntax (other than what can be read off the parsing process as a trace). It is unclear what they actually mean with this claim, though. First, if C&C mean that the parser does not construct full syntactic trees but rather produces a minimum that allows semantics and phonology to operate, they just echo a view discussed by Pulman (1986) and others. Although this view is an open possibility, we do not see how it follows from the NNB. Second, if C&C mean that the NNB implies that parsing does not use explicit syntactic knowledge, this view is incorrect: Many parsing algorithms (e.g., LR, Earley's algorithm, incremental CKY) respect the NNB by being incremental and not needing to refer back to raw data (they can all refer to the result of earlier processing instead) and yet make reference to explicit syntax. Finally, we note that prediction-based, parser-only models in the literature that do not incorporate explicit representations of syntactic structure (e.g., Elman 1990; McCauley & Christiansen 2011) fail to explain why we can recognize unpredictable sentences as grammatical (e.g., *Evil unicorns devour xylophones*).

Regarding learning, C&C claim that the NNB is incompatible with approaches to learning that involve elaborate linguistic knowledge. This, however, is incorrect: The only implication of the NNB for learning is that if memory is indeed fleeting, any learning mechanism must be online rather than batch, relying only on current information. But online learning does not rule out theory-based models of language in any way (e.g., Börschinger & Johnson 2011). In fact, some have argued that online variants of theory-based models provide particularly good approximations to empirically observed patterns of learning (e.g., Frank et al. 2010).

Regarding the evolution of language (which they conflate with the biological evolution of language), C&C claim that it is item-based and gradual, and that linguistic diversity is the norm, with few if any true universals. However, how these claims might follow from the NNB is unclear, and C&C are inconsistent with the relevant literature. For example, language change has been argued to be abrupt and nonlinear (see Niyogi & Berwick 2009), often involving what look like changes in abstract principles rather than concrete lexical items. As for linguistic diversity, C&C repeat claims from Christiansen and Chater (2008) and Evans and Levinson (2009), but those works ignore the strongest typological patterns revealed by generative linguistics. For example, no known language allows for a single conjunct to be displaced in a question (Ross 1967): We might know that Kim ate *peas and* *something* yesterday and wonder what that *something* is, but in no language can we use a question of the form **What did Kim eat peas and yesterday?* to inquire about it. Likewise, in *Why did John wonder who Bill hit?*, one can only ask about the cause of the wondering, not of the hitting (see Huang 1982; Rizzi 1990). Typological data thus reveal significant restrictions on linguistic diversity.

Conclusion. Language is complex. Our efforts to comprehend it are served better by detailed analysis of the cognitive mechanisms at our disposal than by grand theoretical proposals that ignore the relevant psychological, linguistic, and computational distinctions.

ACKNOWLEDGMENTS
We thank Leon Bergen, Bob Berwick, Tova Friedman, and Tim O'Donnell.

Is Now-or-Never language processing good enough?

doi:10.1017/S0140525X15000771, e72

Fernanda Ferreira[a] and Kiel Christianson[b]

[a]*Department of Psychology, University of California, Davis, Davis, CA 95616;* [b]*College of Education, University of Illinois, Urbana–Champaign, Champaign, IL 61820.*
fferreira@ucdavis.edu
http://psychology.ucdavis.edu/people/fferreir
kiel@illinois.edu
http://epl.beckman.illinois.edu/

Abstract: Christiansen & Chater's (C&C's) Now-or-Never bottleneck framework is similar to the Good-Enough Language Processing model (Ferreira et al. 2002), particularly in its emphasis on sparse representations. We discuss areas of overlap and review experimental findings that reinforce some of C&C's arguments, including evidence for underspecification and for parsing in "chunks." In contrast to Good-Enough, however, Now-or-Never does not appear to capture misinterpretations or task effects, both of which are important aspects of comprehension performance.

Christiansen & Chater (C&C) offer an intriguing proposal concerning the nature of language, intended to explain fundamental aspects of language comprehension, production, learning, and evolution. We agree with the basic framework, and indeed we have offered our own theoretical approach, Good-Enough (GE) Language Processing, to capture many of the phenomena discussed in the target article, particularly those relating to both online and offline comprehension. In this commentary, we hope to expand the discussion by pointing to some of these connections and highlighting additional phenomena that C&C did not discuss but that reinforce some of their points. In addition, however, we believe the GE model is better able to explain important aspects of language comprehension that C&C consider, as well as several they leave out. Of course, no single article could be comprehensive when it comes to a field as broad and active as this one, but we believe a complete theory of language must ultimately have something to say about these important phenomena, and particularly the content of people's interpretations.

We begin, then, with a brief review of the GE approach (Ferreira et al. 2002). The fundamental assumption is that interpretations are often shallow and sometimes inaccurate. This idea that interpretations are shallow and underspecified is similar to C&C's suggestion that the comprehension system creates chunks that might not be combined into a single, global representation. In their model, this tendency arises from memory constraints that lead the system to build chunks at increasingly abstract levels of representation. As evidence for this assumption regarding underspecified representations, C&C might have discussed our work demonstrating that ambiguous relative clauses are often not definitively attached into the matrix structure if a failure to attach has no interpretive consequences (Swets et al. 2008; cf. Payne et al. 2014). Very much in line with C&C, Swets et al. observed that people who are asked detailed comprehension questions probing their interpretation of the ambiguous relative clause make definitive attachments, but those asked only shallow questions about superficial features of the sentence seem to leave the relative clause unattached – that is, they underspecify. This finding fits neatly with C&C's discussion of "right context effects," where here "right context" can be broadly construed to mean the follow-on comprehension question that influences the interpretation constructed online. An important difference, however, emerges as well, and here we believe the GE framework has some advantages over Now-or-Never as a broad model of comprehension: Our framework predicts that the language user's task will have a strong effect on the composition of "chunks" and the interpretation created from them (cf. Christianson & Luke 2011; Lim & Christianson 2015). We have reported these results in production as well, demonstrating that the extent to which speaking is incremental depends on the processing demands of the speaking task (Ferreira & Swets 2002). Given the importance of task effects in a range of cognitive domains, any complete model of language processing must include mechanisms for explaining how they arise.

Moreover, the idea that language processing proceeds chunk-by-chunk is not novel. C&C consider some antecedents of their proposal, but several are overlooked. For example, they argue that memory places major constraints on language processing, essentially obligating the system to chunk and interpret as rapidly as possible (what they term "eager processing"). This was a key motivation for Lyn Frazier's original garden-path model (Frazier & Rayner 1982) and the parsing strategies known as minimal attachment and late closure: The parser's goal is to build an interpretation quickly and pursue the one that emerges first rather than waiting for and considering multiple alternatives. This, too, is part of C&C's proposal – that the parser cannot construct multiple representations at the same level in parallel – but the connections to the early garden-path model are not mentioned, and the incompatibility of this idea with parallel models of parsing is also not given adequate attention. Another example is work by Tyler and Warren (1987), who showed that listeners form unlinked local phrasal chunks during spoken language processing and who conclude that they could find no evidence for the formation of a global sentence representation. Thus, several of these ideas have been part of the literature for many years, and evidence for them can be found in research motivated from a broad range of theoretical perspectives.

Perhaps the most critical aspect of comprehension that C&C's approach does not capture is meaning and interpretation: C&C describe an architecture that can account for some aspects of processing, but their model seems silent on the matter of the content of people's interpretations. This is a serious shortcoming given the considerable evidence for systematic misinterpretation (e.g., Christianson et al. 2001; 2006; Patson et al. 2009; van Gompel et al. 2006). In our work, we demonstrated that people who read sentences such as *While Mary bathed the baby played in the crib* often derive the interpretation that Mary bathed the baby, and they also misinterpret simple passives such as *The dog was bitten by the man* (Ferreira 2003). These are not small tendencies; the effects are large, and they have been replicated in numerous studies across many different labs. For C&C, these omissions are a lost opportunity because these results are consistent with their proposed architecture. For example, misinterpretations of garden-path sentences arise in part because the parser processes sentences in thematic chunks and fails to reconcile the various meanings constructed online. Recently, we demonstrated that the misinterpretations are attributable to a failure to "clean up" the interpretive consequences of creating these chunks (Slattery et al. 2013), a finding compatible with C&C's

idea that chunks are quickly recoded into more abstract levels of representation and that it is difficult to re-access the less abstract representations.

C&C's framework is exciting, and we believe it will inspire significant research. Their creative synthesis is a major achievement, and we hope we have contributed constructively to the project by pointing to areas of connection and convergence as well as by highlighting important gaps.

Reservoir computing and the Sooner-is-Better bottleneck

doi:10.1017/S0140525X15000783, e73

Stefan L. Frank[a] and Hartmut Fitz[b]

[a]*Centre for Language Studies, Radboud University Nijmegen, 6500 HD Nijmegen, The Netherlands;* [b]*Max Planck Institute for Psycholinguistics, 6500 AH Nijmegen, The Netherlands.*

s.frank@let.ru.nl
www.stefanfrank.info
hartmut.fitz@mpi.nl
www.mpi.nl/people/fitz-hartmut

Abstract: Prior language input is not lost but integrated with the current input. This principle is demonstrated by "reservoir computing": Untrained recurrent neural networks project input sequences onto a random point in high-dimensional state space. Earlier inputs can be retrieved from this projection, albeit less reliably so as more input is received. The bottleneck is therefore not "Now-or-Never" but "Sooner-is-Better."

Christiansen & Chater (C&C)argue that the "Now-or-Never" bottleneck arises because input that is not immediately processed is forever lost when it is overwritten by new input entering the same neural substrate. However, the brain, like any recurrent network, is a state-dependent processor whose current state is a function of both the previous state and the latest input (Buonomano & Maass 2009). The incoming signal therefore does not wipe out previous input. Rather, the two are integrated into a new state that, in turn, will be integrated with the next input. In this way, an input stream "lives on" in processing memory. Because prior input is implicitly present in the system's current state, it can be faithfully recovered from the state, even after some time. Hence, there is no need to immediately "chunk" the latest input to protect it from interference. This does not mean that no part of the input is ever lost. As the integrated input stream grows in length, it becomes increasingly difficult to reliably make use of the earliest input. Therefore, the sooner the input can be used for further processing, the more successful this will be: There is a "Sooner-is-Better" rather than a "Now-or-Never" bottleneck.

So-called *reservoir computing* models (Lukoševičius & Jaeger 2009; Maass et al. 2002) exemplify this perspective on language processing. Reservoir computing applies untrained recurrent networks to project a temporal input stream into a random point in a very high-dimensional state space. A "read-out" network is then calibrated, either online through gradient descent or offline by linear regression, to transform this random mapping into a desired output, such as a prediction of the incoming input, a reconstruction of (part of) the previous input stream, or a semantic representation of the processed language. Crucially, the recurrent network itself is not trained, so the ability to retrieve earlier input from the random projection cannot be the result of learned chunking or other processes that have been acquired from language exposure. Indeed, Christiansen and Chater (1999) found that even before training, the random, initial representations in a simple recurrent network's hidden layer allow for better-than-chance classification of earlier inputs. Reservoir computing has been applied to simulations of human language learning and comprehension, and such models accounted for experimental findings from both behavioural (Fitz 2011; Frank & Bod 2011) and neurophysiological studies (Dominey et al. 2003; Hinaut & Dominey 2013). Moreover, it has been argued that reservoir computing shares important processing characteristics with cortical networks (Rabinovich et al. 2008; Rigotti et al. 2013; Singer 2013), making this framework particularly suitable to the computational study of cognitive functions.

To demonstrate the ability of reservoir models to memorize linguistic input over time, we exposed an echo-state network (Jaeger & Haas 2004) to a word sequence consisting of the first 1,000 words (roughly the length of this commentary) of the Scholarpedia entry on echo-state networks. Ten networks were randomly generated with 1,000 units and static, recurrent, sparse connectivity (20% inhibition). The read-outs were adapted such that the network had to recall the input sequence 10 and 100 words back. The 358 different words in the corpus were represented orthogonally, and the word corresponding to the most active output unit was taken as the recalled word. For a 10-word delay, the correct word was recalled with an average accuracy of 96% (SD=0.6%). After 100 words, accuracy remained at 96%, suggesting that the network had memorized the entire input sequence. This indicates that there was sufficient information in the system's state-space trajectory to reliably recover previous perceptual input even after very long delays. Sparseness and inhibition, two pervasive features of the neocortex and hippocampus, were critical: Without inhibition, average recall after a 10-word delay dropped to 51%, whereas fully connected networks correctly recalled only 9%, which equals the frequency of the most common word in the model's input. In short, the more brain-like the network, the better its capacity to memorize past input.

The modelling results should not be mistaken for a claim that people are able to perfectly remember words after 100 items of intervening input. To steer the language system towards an interpretation, earlier input need not be available to explicit recall and verbalization. Thus, it is also irrelevant to our echo-state network simulation whether or not such specialized read-outs exist in the human language system. The simulation merely serves to illustrate the concept of state-dependent processing where past perceptual input is implicitly represented in the current state of the network. A more realistic demonstration would take phonetic, or perhaps even auditory, features as input, rather than presegmented words. Because the dynamics in cortical networks is vastly more diverse than in our model, there is no principled reason such networks should not be able to cope with richer information sources. Downstream networks can then access this information when interpreting incoming utterances, without explicitly recalling previous words. Prior input encoded in the current state can be used for any context-sensitive operation the language system might be carrying out – for example, to predict the next phoneme or word in the unfolding utterance, to assign a thematic role to the current word, or to semantically integrate the current word with a partial interpretation that has already been constructed.

Because language is structured at different levels of granularity (ranging from phonetic features to discourse relations), the language system requires neuronal and synaptic mechanisms that operate at different timescales (from milliseconds to minutes) in order to retain relevant information in the system's state. Precisely how these memory mechanisms are implemented in biological networks of spiking neurons is currently not well-understood; proposals include a role for diverse, fast-changing neuronal dynamics (Gerstner et al. 2014) coupled with short-term synaptic plasticity (Mongillo et al. 2008) and more long-term adaptation through spike-timing dependent plasticity (Bi & Poo 2001). The nature of processing memory will be crucial in any neurobiologically viable theory of language processing (Petersson & Hagoort 2012), and we should therefore not lock ourselves into architectural commitments based on stipulated bottlenecks.

ACKNOWLEDGMENTS

We would like to thank Karl-Magnus Petersson for helpful discussions on these issues. SLF is funded by the European Union Seventh Framework Programme under grant no. 334028.

Natural language processing and the Now-or-Never bottleneck

doi:10.1017/S0140525X15000795, e74

Carlos Gómez-Rodríguez
LyS (Language and Information Society) Research Group, Departamento de Computación, Universidade da Coruña, Campus de Elviña, 15071, A Coruña, Spain.
cgomezr@udc.es
http://www.grupolys.org/~cgomezr

Abstract: Researchers, motivated by the need to improve the efficiency of natural language processing tools to handle web-scale data, have recently arrived at models that remarkably match the expected features of human language processing under the Now-or-Never bottleneck framework. This provides additional support for said framework and highlights the research potential in the interaction between applied computational linguistics and cognitive science.

Christiansen & Chater (C&C) describe how the brain's limitations to retain language input (the Now-or-Never bottleneck) constrain and shape human language processing and acquisition.

Interestingly, there is a very strong coincidence between the characteristics of processing and learning under the Now-or-Never bottleneck and recent computational models used in the field of natural language processing (NLP), especially in syntactic parsing. C&C provide some comparison with classic cognitively inspired models of parsing, noting that they are in contradiction with the constraints of the Now-or-Never bottleneck. However, a close look at the recent NLP and computational linguistics literature (rather than the cognitive science literature) shows a clear trend toward systems and models that fit remarkably well with C&C's framework.

It is worth noting that most NLP research is driven by purely pragmatic, engineering-oriented requirements: The primary goal is not to find models that provide plausible explanations of the properties of language and its processing by humans, but rather to design systems that can parse text and utterances as accurately and efficiently as possible for practical applications like opinion mining, machine translation, or information extraction, among others.

In recent years, the need to develop faster parsers that can work on web-scale data has led to much research interest in incremental, data-driven parsers; mainly under the so-called *transition-based* (or shift-reduce) framework (Nivre 2008). This family of parsers has been implemented in systems such as MaltParser (Nivre et al. 2007), ZPar (Zhang & Clark 2011), ClearParser (Choi & McCallum 2013), or Stanford CoreNLP (Chen & Manning 2014), and it is increasingly popular because they are easy to train from annotated data and provide a very good trade-off between speed and accuracy.

Strikingly, these parsing models present practically all of the characteristics of processing and acquisition that C&C describe as originating from the Now-or-Never bottleneck in human processing:

Incremental processing (sect. 3.1): A defining feature of transition-based parsers is that they build syntactic analyses incrementally as they receive the input, from left to right. These systems can build analyses even under severe working memory constraints: Although the issue of "stacking up" with right-branching languages mentioned by C&C exists for so-called *arc-standard* parsers (Nivre 2004), parsers based on the arc-eager model (e.g., Gómez-Rodríguez & Nivre 2013; Nivre 2003) do not accumulate right-branching structures in their stack; as they build dependency links as soon as possible. In these parsers, we only need to keep a word in the stack while we wait for its head or its direct dependents, so the time that linguistic units need to be retained in memory is kept to the bare minimum.

Multiple levels of linguistic structure (sect. 3.2): As C&C mention, the organization of linguistic representation in multiple levels is "typically assumed in the language sciences"; this includes computational linguistics and transition-based parsing models. Traditionally, each of these levels was processed sequentially in a pipeline, contrasting with the parallelism of the Chunk-and-Pass framework. However, the appearance of general incremental processing frameworks spanning various levels, from segmentation to parsing (Zhang & Clark 2011), has led to recent research on joint processing where the processing of several levels takes place simultaneously and in parallel, passing information between levels (Bohnet & Nivre 2012; Hatori et al. 2012). These models, which improve accuracy over pipeline models, are very close to the Chunk-and-Pass framework.

Predictive language processing (sect. 3.3): The joint processing models just mentioned are hypothesized to provide accuracy improvements precisely because they allow for a degree of predictive processing. Contrary to pipeline approaches where information only flows in a bottom-up way, these systems allow top-down information from higher levels "to constrain the processing of the input at lower levels," just as C&C describe.

Acquisition as learning to process (sect. 4): Transition-based parsers learn a sequence of processing actions (transitions), rather than a grammar (Gómez-Rodríguez et al. 2014; Nivre 2008), making the learning process simple and flexible.

Local learning (sect. 4.2): This is also a general characteristic of all transition-based parsers. Because they do not learn grammar rules but processing actions to take in specific situations, adding a new example to the training data will create only local changes to the inherent language model. At the implementation level, this typically corresponds to small weight changes in the underlying machine learning model – be it a support vector machine (SVM) classifier (Nivre et al. 2007), perceptron (Zhang & Clark 2011), or neural network (Chen & Manning 2014), among other possibilities.

Online learning and learning to predict (sect. 4.1 and 4.3): Evaluation of NLP systems usually takes place in standard, fixed corpora, and so recent NLP literature has not placed much emphasis on online learning. However, some systems and frameworks do use online learning models with error-driven learning, like the perceptron (Zhang & Clark 2011). The recent surge of interest in parsing with neural networks (e.g., Chen & Manning 2014; Dyer et al. 2015) also seems to point future research in this direction.

Putting it all together, we can see that researchers whose motivating goal was not psycholinguistic modeling, but only raw computational efficiency, have nevertheless arrived at models that conform to the description in the target article. This fact provides further support for the views C&C express.

A natural question arises about the extent to which this coincidence is attributable to similarities between the efficiency requirements of human and automated processing – or rather to the fact that because evolution shapes natural languages to be easy to process by humans (constrained by the Now-or-Never bottleneck), computational models that mirror human processing will naturally work well on them. Relevant differences between the brain and computers, such as in short-term memory capacity, seem to suggest the latter. Either way, there is clearly much to be gained from cross-fertilization between cognitive science and computational linguistics: For example, computational linguists can find inspiration in cognitive models for designing NLP tools that work efficiently with limited resources, and cognitive scientists can use computational tools as models to test their hypotheses. Bridging the gap between these areas of research is essential to further our understanding of language.

ACKNOWLEDGMENTS
This work was funded by the Spanish Ministry of Economy and Competitiveness/ERDF (grant FFI2014-51978-C2-2-R) and Xunta de Galicia (grant R2014/034). I thank Ramon Ferrer i Cancho for helpful comments on an early version of this commentary.

Realizing the Now-or-Never bottleneck and Chunk-and-Pass processing with Item-Order-Rank working memories and masking field chunking networks

doi:10.1017/S0140525X15000801, e75

Stephen Grossberg
Center for Adaptive Systems, Boston University, Boston, MA 02215.
steve@bu.edu
http://cns.bu.edu/~steve

Abstract: Christiansen & Chater's (C&C's) key goals for a language system have been realized by neural models for short-term storage of linguistic items in an Item-Order-Rank working memory, which inputs to Masking Fields that rapidly learn to categorize, or chunk, variable-length linguistic sequences, and choose the contextually most predictive list chunks while linguistic inputs are stored in the working memory.

Key goals that Christiansen & Chater (C&C) propose for language processing have already been realized by real-time neural models of speech and language learning and performance, notably:

1. C&C write in their abstract about a Now-or-Never bottleneck whereby the brain compresses and recodes linguistic input as rapidly as possible; a multilevel linguistic representation; and a predictive system, which ensures that local linguistic ambiguities are dealt with Right-First-Time using Chunk-and-Pass processing.
2. At the beginning of paragraph 2 of section 3.3, C&C note that predictions for higher-level chunks may "run ahead" of those for lower-level chunks, as when listeners answer "two" in response to the question "How many animals of each kind did Moses take on the Ark?"

Neural models of speech and language embody design principles and mechanisms that automatically satisfy such properties. Introduced in Grossberg (1978a; 1978b), they have progressively developed to the present time. Two key contributions are as follows: (a) a model for short-term storage of sequences of language items that can include repeated items, called an Item-Order-Rank (IOR) working memory. The working memory inputs via an adaptive filter to (b) a model for learned unitization, categorization, or chunking of variable-length sequences of items that are stored in working memory, called a Masking Field (MF). An MF clarifies how the brain rapidly learns to categorize, or chunk, variable-length linguistic sequences, and uses recurrent competitive interactions to choose the most predictive sequence chunk, or list chunk, as linguistic inputs are processed in real time by the working memory. The MF, in turn, sends predictive top-down matching signals to the working memory to attentively select item sequences that the winning chunks represent.

Both IOR and MF networks are realized by recurrent on-center, off-surround networks whose cells obey the membrane equations of neurophysiology; that is, shunting dynamics (Grossberg 1973). These working memory and chunking networks have been used to explain and simulate many challenging properties of variable-rate variable-speaker speech and language data: for example, Ames and Grossberg (2008), Boardman et al. (1999), Cohen and Grossberg (1986), Grossberg (1986; 2003), Grossberg et al. (1997), Grossberg and Myers (2000), and Grossberg and Pearson (2008). Most recently, such working memories and chunking networks have been incorporated into the cARTWORD hierarchical laminar cortical model of speech learning and recognition (Grossberg & Kazerounian 2011; Kazerounian & Grossberg 2014).

The Item-Order-Rank working memory clarifies data such as in Goal 2 cited above because sufficient subsets of working memory items can choose a predictive chunk, even if there are incongruent sequence elements. A Masking Field clarifies data such as in Goal 1, above, because the most predictive list chunks are chosen in real time as linguistic data are processed in working memory.

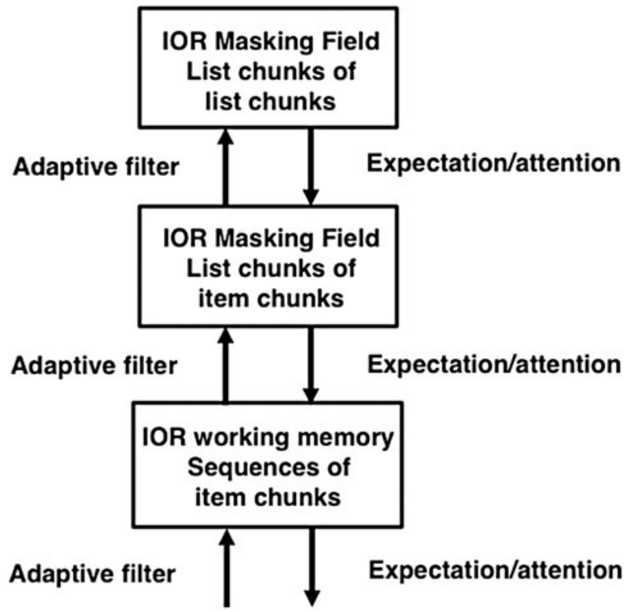

Figure 1 (Grossberg). Speech hierarchy. Each processing level in this hierarchy is an Item-Order-Rank (IOR) working memory that can store sequences with repeated items in short-term memory. The second and third IOR working memories are, in addition, multiple-scale Masking Fields (MF) that can chunk input sequences of variable length, and choose the sequence, or sequences, for storage that receive the most evidence from its inputs. Each level receives its bottom-up inputs from an adaptive filter and reads-out top-down expectations that focus attention on the feature patterns in their learned prototypes at the previous level. The first level stores sequences of item chunks. The second level stores sequences of list chunks. The individual list chunks of the third level thus represent sequences of list chunks at the second level, including sequences with repeated words, like "DOG EATS DOG." During rehearsal, each chunk at a higher level can read out its learned sequence through its top-down expectation with the support of a volitional signal that converts the top-down modulatory signals into signals that are capable of fully activating their target cells.

An IOR working memory (WM) stores a temporal stream of inputs through time as an evolving spatial pattern of content-addressable item representations. This WM model is called an IOR model because its nodes, or cell populations, represent list *items*, the temporal *order* in which the items are presented is stored by an *activity gradient* across nodes, and the same item can be repeated in different list positions, or *ranks*. A *primacy gradient* stores items in WM in the correct temporal order, with the first item having the highest activity. Recall occurs when a basal ganglia rehearsal wave activates WM read-out. The node with the highest activity reads out fastest and self-inhibits its WM representation. Such *inhibition-of-return* prevents perseveration of performance. Both psychophysical and neurophysiological data support this coding scheme; see Grossberg (2013), Grossberg and Pearson (2008), and Silver et al. (2011).

These circuits were derived by analyzing how a WM could be designed to enable list chunks of variable length to be rapidly learned and stably remembered through time, leading to postulates which imply that *all* working memories – linguistic, motor, and spatial – have a similar design, with similar data patterns across modalities, and that there exists an intimate link between list chunking and WM storage. Grossberg (2013) reviews supportive data.

An MF is a specialized IOR WM. The "items" stored in an MF are list chunks that are selectively activated, via a bottom-up adaptive filter, by subsequences of items that are stored in an item

WM. As items are stored in item WM, an adaptive filter activates the learned list chunks that represent the most predictive item groupings at any time, while suppressing less-predictive chunks. In order for an MF list chunk to represent lists (e.g., syllables or words) of multiple lengths, its cells interact within and between multiple spatial scales, with the cells of larger scales capable of selectively representing item sequences of greater length, and of inhibiting other MF cells that represent item sequences of lesser length ("self-similarity").

MFs solve the *temporal chunking problem*, which asks how a chunk for an unfamiliar list of familiar speech units – for example, a novel word composed of familiar subwords – can be learned under unsupervised learning conditions. What mechanisms prevent familiarity of subwords (e.g., MY, ELF, and SELF), which have already learned to activate their own list chunks, from forcing the novel longer list (e.g., MYSELF) to always be processed as a sequence of these smaller familiar chunks, rather than as a newly learned, unitized whole? How does a not-yet-established word representation overcome the salience of already well-established phoneme, syllable, or word representations to enable learning of the novel word to occur? This solution implies the properties of Goal 1, as well as psychophysical data like the Magical Number Seven and word superiority effects.

Lists with repeated words can be coded by a three-level network (Fig. 1): The first level contains item chunks that input to a Masking Field through an adaptive filter. This MF inputs to a second MF through an adaptive filter that compresses "items" from the first MF into list chunks. These "items" are, however, list chunks. Thus, the second MF's list chunks represent sequences of list chunks. Because it is also an IOR working memory, it can store sequences with repeated list chunks, including sequences with repeated words – for example, "DOG EATS DOG" – thereby providing the kind of multilevel, multiscale, predictive hierarchy that the authors seek.

Better late than Now-or-Never: The case of interactive repair phenomena

doi:10.1017/S0140525X15000813, e76

Patrick G. T. Healey,[a] Christine Howes,[b] Julian Hough,[c] and Matthew Purver[a]

[a]Cognitive Science Research Group, School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, United Kingdom; [b]Department of Philosophy, Linguistics and Theory of Science, University of Gothenburg, 405 30 Gothenburg, Sweden; [c]Fak. LiLi, Universität Bielefeld, D-33501 Bielefeld, Germany.

p.healey@qmul.ac.uk christine.howes@gu.se julian.hough@uni-bielefeld.de m.purver@qmul.ac.uk

Abstract: Empirical evidence from dialogue, both corpus and experimental, highlights the importance of interaction in language use – and this raises some questions for Christiansen & Chater's (C&C's) proposals. We endorse C&C's call for an integrated framework but argue that their emphasis on local, individual production and comprehension makes it difficult to accommodate the ubiquitous, interactive, and defeasible processes of clarification and repair in conversation.

Language is first encountered and deployed in interaction. A characteristic feature of natural interaction is that people often need to address problems with mutual understanding in which elements of what has been said need to be reworked or redone in some way. These processes raise some important questions for Christiansen & Chater's (C&C's) proposals. We support C&C's approach and agree that an integrated framework for the language sciences is a desirable goal for language researchers. However, we argue that C&C's emphasis on local, individual production and comprehension misses some of the key challenges posed by the processes of clarification and repair in conversation.

The thrust of C&C's approach is that language processing is a "Now-or-Never" process that involves rapid, local, lossy chunking of linguistic representations and facilitates a form of autonomous prediction of both our own and each other's utterances. This leads to the proposal that "Chunk-and-Pass processing implies that there is practically no possibility for going back once a chunk is created" (sect. 3.3, para. 2).

The phenomena of clarification and repair seem to present an important counterexample to this picture of language use. Dynamic revisions to utterances, or *repairs*, are ubiquitous in dialogue. In natural conversations it is rare for even a single utterance to be produced without some form of online revision, with these occurring approximately once every 25 words in conversational speech (Hough & Purver 2013), with the rate of repairs adjusted to task demands (Colman & Healey 2011) and to individual differences such as clinical conditions (Howes et al. 2012; Lake et al. 2011). Repair contagion, whereby the probability of another repair occurring increases after an initial one, is also common (Hough & Purver 2013).

Although many of these repairs are syntactically or lexically local in C&C's sense – for example, words or word fragments that are restarted – some involve more-substantial revisions, and some occur after a turn is apparently complete (Schegloff et al. 1977). Conversation analysts claim that the (minimum) space in which direct repairs or revisions to a speaker's utterance can be made is the four subsequent turns in the conversation (Schegloff 1995). This highlights the operation of significant, nonlocal mechanisms that can make use of prior phonetic, lexical, syntactic, and semantic information over relatively long intervals.

Even self-repairs, the most common and most local form of backtracking in conversation, are often nonlocal in a different sense, as they are produced in response to concurrent feedback from an interlocutor, which works against the idea of encapsulated local processing (e.g., Bavelas & Gerwing 2007; Goodwin 1979). The more strongly people are committed to the predictions of their own language processor, the less able they must be to deal with these real-time adjustments or reversals of decisions – potentially of phonetic, lexical, syntactic, or semantic information – in response to feedback from others. However, it seems that in conversation such revisions are the norm, not the exception. People can take advantage of each other's repair behavior, too: In a visual world paradigm, when experimental subjects hear repaired referring expressions compared to fluent ones, participants can use repaired material to speed up reference resolution (Brennan & Schober 2001). Additionally, experiments in interruptive clarification (Healey et al. 2011) show that participants often restart the interrupted turn after responding to a clarification request, again showing that people must, at least in some cases, have access to the previously produced material.

Ambiguities can emerge late in a dialogue, and people routinely deal with them. Although C&C do acknowledge the availability of mechanisms to "repair the communication by requesting clarification from the dialogue partner" (sect. 3.1, para. 8), they do not discuss how and whether these repair phenomena are consistent with the Chunk-and-Pass model. Similarly, C&C argue that early commitment to predictions about what is coming next should lead to frequent reuse of our own and each other's lexical and syntactic representations; however, the evidence for this in natural conversation is controversial. We have found that syntactic reuse is actually less common than would be expected by chance (Healey et al. 2014). The need to respond constructively to a conversational partner seems to overwhelm some of the processes observed in individual language processing.

These observations reinforce C&C's emphasis on the highly time-critical and piecemeal, incremental nature of language processing, but they also suggest that the demands of engaging with a live conversational partner requires more flexible, defeasible, and interactive mechanisms. Their proposal currently captures a type of incrementality that is essential for efficient working memory, what Levelt (1993) calls "Wundt's Principle," whereby a consuming module can begin operating with a minimal amount of characteristic input. However, repair phenomena entail other kinds of incrementality as desiderata for a psychological model: namely, recoverability and repairability of increments from the interactive context.

One existing formal and computational model capable of capturing the different facets of incrementality needed for repair mechanisms is Dynamic Syntax (DS, Purver et al. 2006; 2011). DS models language as a set of mechanisms for incrementally building up interpretations in context, and is therefore broadly commensurate with the C&C program; these mechanisms can also be induced (acquired) from the data available to a child learner (Eshghi et al. 2013), with the learning process being piecemeal, incremental, and process-driven as required by C&C. However, DS can also account for repair phenomena by using explicit recoverability mechanisms through backtracking over stored graphs of incrementally constructed semantic content (Eshghi et al. 2015; Hough & Purver 2012). We take this approach to be complementary to the C&C model, showing that many of their insights can be practically implemented, while also addressing the significant challenges posed by interactive repair phenomena in dialogue. In sum, we propose a model that is compatible with the "Now" aspect of their approach, but not with the "Never."

How long is now? The multiple timescales of language processing

doi:10.1017/S0140525X15000825, e77

Christopher J. Honey,[a] Janice Chen,[b] Kathrin Müsch,[a] and Uri Hasson[b]

[a]*Department of Psychology, Collaborative Program in Neuroscience, University of Toronto, Toronto, ON M5S 3G3, Canada;* [b]*Department of Psychology and the Neuroscience Institute, Princeton University, Princeton, NJ 08540.*

honey@psych.utoronto.ca **janice@princeton.edu**
kathrin.muesch@utoronto.ca **hasson@princeton.edu**
http://www.honeylab.org
http://hlab.princeton.edu

Abstract: Christiansen & Chater (C&C) envision language function as a hierarchical chain of transformations, enabling rapid, continuous processing of input. Their notion of a "Now-or-Never" bottleneck may be elaborated by recognizing that timescales become longer at successive levels of the sensory processing hierarchy – that is, the window of "Now" expands. We propose that a hierarchical "process memory" is intrinsic to language processing.

Meaningful interactions between linguistic units occur on many timescales. After listening to 10 minutes of a typical English narrative, a listener will have heard ~1,000 words composing ~100 sentences grouped into ~25 paragraphs. When the 1,001st word in the narrative arrives, it enters a rich syntactic and semantic context that spans multiple timescales and levels of abstraction. Christiansen & Chater (C&C) rightfully emphasize the constraints imposed by the rapidity of language input. Here we highlight the importance of a related class of constraints: those imposed by the need to integrate incoming information with prior information over multiple timescales.

C&C motivate the "Now-or-Never bottleneck" with the observations that "memory is fleeting" and "new material rapidly obliterates previous material" (abstract). These statements tend to hold true in low-level auditory masking (Elliott 1962) and in short-term memory experiments involving unrelated auditory items (Warren et al. 1969). However, in real-life language processing, memory cannot all be fleeting. This is because new stimuli have a prior relationship to, and must be actively integrated with, the stimuli that were just encountered. Thus, in real-life contexts, previous material exerts a powerful influence on the processing of new material.

Consider the difference between hearing the sequence of words "friend-mulch-key" and hearing the sequence of words "friend-ship-pact." In the first sequence, the representation of the word *friend* is degraded by interference with *mulch* and *key*. In the second sequence, by contrast, the word *friend* interacts meaningfully with *ship* and *pact*. This simple example reflects a general and ubiquitous phenomenon in real-life language: New material does not necessarily obliterate previous material. Instead, past and present information interact to produce understanding, and the memory of past events continually shapes the present (Nieuwland & Van Berkum 2006).

It seems the processing bottleneck that C&C describe applies best to early processing areas (e.g., primary sensory cortex), where sensory traces may have a very short lifetime (<200 ms). At higher levels of the language hierarchy, however, neural circuits must retain a longer history of past input to enable the integration of information over time. Temporal integration is necessary for higher-order regions to support the understanding of a new word in relation to a prior sentence or a new sentence in relation to the larger discourse. We have found that temporal integration occurs over longer timescales in higher-order regions (Hasson et al. 2008; Lerner et al. 2011), and that the intrinsic neural dynamics become slower across consecutive stages of the cortical hierarchy (Honey et al. 2012; Stephens et al. 2013). Thus, the temporal bottleneck appears to gradually widen across the consecutive stages of the language processing hierarchy, as increasingly abstract linguistic structures are processed over longer timescales.

Influenced by the ideas of Macdonald and Christiansen (1996), as well as Fuster (1997), concerning the memory that is intrinsic to ongoing information processing, and supported by recent single-unit, electrocorticography, and functional imaging data, we have developed a brain-based framework for such a functional organization (Hasson et al. 2015). In this framework, (a) virtually all cortical circuits can accumulate information over time, and (b) the timescale of accumulation varies hierarchically, from early sensory areas with short processing timescales (tens to hundreds of milliseconds) to higher-order areas with long processing timescales (many seconds to minutes). In this hierarchical systems perspective, memory is not restricted to a few localized stores and it is not transient; instead memory is intrinsic to information processing that unfolds throughout the brain on timescales from milliseconds to minutes. We have suggested the term "process memory" to refer to active traces of past information that are used by a local neural circuit to process incoming information in the present moment; this is in distinction to the more traditional notion of "working memory," which is a more functionally encapsulated memory store.

Process memory may support the Chunk-and-Pass mechanism that C&C propose for organizing inter-regional information flow. As they note: "incremental processing in comprehension and production takes place in parallel across multiple levels of linguistic representation, each with a characteristic temporal window" (sect. 3.2, para. 5). In our view, the Now-or-Never bottleneck can be made compatible with contextual language processing by allowing the "Now" (i.e., the local circuit memory of prior events) to have a variable duration. For example, the "Now" could be understood to have a short (e.g., milliseconds) timescale in sensory areas, where representations are fleeting, and then to gradually expand in duration in higher-order areas, where chunking is required over longer (e.g., seconds) and longer (e.g., minutes) windows of input. Thus, the "Now" may be understood as a time window around the present moment, in which information can be integrated, and the duration of the "Now" may lengthen as one moves from sensory areas toward higher-order language circuits.

In summary, we share the vision of C&C in which language function arises from a chain of transformations across a hierarchy of circuits, and that language learning is a kind of "learning to process." At the same time, we suggest that this hierarchical processing framework could be refined to account for the process memory that is intrinsic to language processing and is needed for comprehending incoming input within multiple timescales of prior context.

Neural constraints and flexibility in language processing

doi:10.1017/S0140525X15000837, e78

Christian R. Huyck
Department of Computer Science, Middlesex University, London NW4 4BT, United Kingdom.
c.huyck@mdx.ac.uk
http://www.cwa.mdx.ac.uk/chris/

Abstract: Humans process language with their neurons. Memory in neurons is supported by neural firing and by short- and long-term synaptic weight change; the emergent behaviour of neurons, synchronous firing, and cell assembly dynamics is also a form of memory. As the language signal moves to later stages, it is processed with different mechanisms that are slower but more persistent.

The Now-or-Never bottleneck in language processing that Christiansen & Chater (C&C) propose has a great deal of evidence to support it. Like all cognitive processes, language processing must be implemented in neurons, and the bottleneck is a neural one. Signals from the environment must be processed by neurons, and those neurons must keep a memory trace of those signals or they will be lost. Moreover, any processing mechanism must not only be implemented by the behaviour of neurons, but in the case of language, the process must be learned by those neurons.

Neural memory comes in several forms. Neurons spike propagating signals across their synapses to post-synaptic neurons taking tens of milliseconds. Neurons can be wired into cell assemblies (Hebb 1949) that can persistently fire for seconds. Synaptic weights can be modified for seconds to minutes by means of short-term potentiation (STP), or for days, months, or longer, through long-term potentiation (LTP). The formation of a cell assembly, by potentiation, can form a circuit that can last indefinitely. When that long-term memory is activated by a cascade of neural firing in the cell assembly, the long-term memory is also an active short-term memory.

When a sentence is parsed, either in speech or in text, the parsing is generally done in one pass. This single pass can be seen in eye-tracking evidence, especially when repairs are needed (Just & Carpenter 1980). One pass parsing is typically simulated with a stack, but a memory-based mechanism (Lewis & Vasishth 2005) can eliminate the need for a stack. A memory-based parsing mechanism has been implemented in a neural parsing model (Huyck 2009), with the persistence of the cell assembly showing the strength and duration of the memory. I am unaware of any existing simulated neural mechanism for backtracking in parsing.

One important aspect of eliminating the stack in parsing is that it reduces the need for binding. Binding is another type of neural memory mechanism that, although needed in standard computational models, is typically overlooked. In a standard program, if a variable is assigned a value, the two are bound. Binding is usually a primitive operation so it is ignored. Binding in a neural system is more difficult because it is not primitive. There are various binding mechanisms; synchronous firing is most widely used in the literature (Fuster & Alexander 1971). Two bound assemblies fire in roughly the same firing pattern, while another pair (or more) can be bound in a different pattern. Synchronous binding requires the neurons to continue firing. Moreover, only a small number of patterns can be supported simultaneously, so there are a limited number of bindings; the bound neurons do not all fire at the exact same time, so separate patterns must be quite distinct. Another option is to bind with STP. This method has neither of these limits, with a much larger number of bindings supported and the duration being up to minutes; it does, however, take longer to form. Binding can also be done with LTP, but this shades into permanent associative memory.

When people or computer systems process language, it is faster and safer to avoid binding. When binding is necessary, lower-level processing is likely to use synchrony. Higher-level processing is likely to use STP. So the speech signal uses synchrony; neurons representing the prime formants fire synchronously in the auditory cortex (Eggermont 2001). The simulated neural parser (Huyck 2009) uses STP for binding the slots in the neural implementation of verb frames associated with sentences. These can be used immediately after sentence processing to retrieve the meaning of the sentence, but they are gradually erased by the STP fading. The neurons that support the binding are reused later for processing other sentences.

Finite-state automata (FSA) do not require binding. The evidence from text engineering to support the bottleneck is that the Message Understanding Competitions for Text Extraction (Appelt et al. 1993) converged on an FSA cascade to solve the problem of processing text. One automaton separated words, a second categorised them lexically, a third did simple phrase parsing, and a fourth combined phrases. These could be run in a cascade, and perhaps a cascade is the basic mechanism that the brain uses.

As C&C note, the bottleneck also has ramifications for learning. First, the whole language cascade (whatever that may be) is being learned simultaneously. Initially, low-level phenomena, such as morphemes, are learned. Later, larger systems such as simple phrase grammars begin to be learned, but the lower-level systems are still being developed. We do not know how these biological neural systems work, much less how they are learned. One mechanism may be that things are being learned and cell assemblies (CAs) are formed; CAs can be connected to form FSA. Binding may be involved initially, and the synapse can then be modified to combine CAs into FSA; STP can support reverberation, which can then lead to LTP. Although one finite-state automaton in the cascade is being learned, both FSA above and below it can be learned so that the whole system continues to improve.

At the highest level, dialogue and above, the bottleneck begins to disappear. Rich cognitive maps support this kind of processing, and memory is formed mostly through LTP and CA circuit dynamics. Since these CAs can persistently fire, and the circuits can be reactivated using associative memory, it is possible to remember large amounts of things. (For example, I can still remember some of the dialogue from the movie I saw this weekend, and the plot.)

There is solid support for the Now-or-Never bottleneck in language processing, although the bottleneck's duration is reduced as the signal passes through stages of language processing. The distributed nature of neural processing supports multiple stages in processing, and the simultaneous learning of these stages. Processing and learning is implemented in neurons, although CA dynamics and binding issues are often not considered by researchers. By expanding understanding and modelling at the neural level, we can better understand language processing, and we can construct more robust language processing systems.

ACKNOWLEDGMENT
This work was supported by the Human Brain Project Grant 604102, Neuromorphic Embodied Agents that Learn.

Mechanisms for interaction: Syntax as procedures for online interactive meaning building

doi:10.1017/S0140525X15000849, e79

Ruth Kempson,[a] Stergios Chatzikyriakidis,[b] and Ronnie Cann[c]

[a]*Philosophy Department, King's College London, Strand, London WC2R 2LS, United Kingdom;* [b]*Department of Philosophy, Linguistics and Theory of Science, University of Gothenburg, 41124 Gothenburg, Sweden;* [c]*Linguistics and English Language, School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Edinburgh EH8 9AD, Scotland.*

ruth.kempson@kcl.ac.uk **stergios.chatzikyriakidis@gu.se**
r.cann@ed.ac.uk
http://www.kcl.ac.uk/artshums/depts/philosophy/people/staff/associates/emeritus/kempson/index.aspx
http://www.stergioschatzikyriakidis.com/contact.html
http://www.lel.ed.ac.uk/~ronnie/

Abstract: We argue that to reflect participant interactivity in conversational dialogue, the Christiansen & Chater (C&C) perspective needs a formal grammar framework capturing word-by-word incrementality, as in Dynamic Syntax, in which syntax is the incremental building of semantic representations reflecting real-time parsing dynamics. We demonstrate that, with such formulation, syntactic, semantic, and morpho-syntactic dependencies are all analysable as grounded in their potential for interaction.

Following their observation of a Now-or Never bottleneck on cognitive processing and a Chunk-and-Pass constraint to overcome this hurdle, Christiansen & Chater (C&C) set the challenge that existing grammars be evaluated in terms of commensurability with their claim that language itself should be seen in processing terms. Directly in line with their perspective is Dynamic Syntax (DS), in which syntax is a set of mechanisms for online building of semantic representations used in both production and perception (Cann et al. 2005; 2007; Kempson et al. 2001; 2011). These mechanisms involve anticipatory specifications of structure relative to some other structure as context, with the need for subsequent update, thus achieving the desired tightly time-constrained interpretation process. As co-developers of DS, we suggest three points of comparison between DS and the construction-grammar (CoG) perspective which C&C envisage: (1) incrementality; (2) the parsing-production interface; (3) lack of structural universals specific to language.

Though C&C stress the importance of incrementality of both parsing and production, given that CoG defines syntax as stored construction-types, somehow learned as wholes, it is not clear what basis this provides for the word-by-word incrementality displayed in conversation. In informal dialogue, participants can interrupt one another at any point, effortlessly switching roles. These switches can split any syntactic and semantic dependencies distributing them across more than one participant: In the following examples, number 1 involves a syntactic split between preposition and noun, and between infinitive and controlling subject; and number 2 involves a morpho-syntactic dependency split (*have* plus past participle) and a syntactic/semantic dependency split (reflexive and local antecedent).

(1) A: We're going to –
B: Burbage to see Granny.
(2) A (seeing B emerging from a smoke-filled kitchen): Are you OK? Have you –
B (interrupting): burnt myself? No fortunately not.

Such data, despite being widespread in conversation, pose severe challenges to conventional syntactic assumptions, including CoG, because the fragments are characteristically not induced as independently licensed by the grammar and even the sequence may not be well-formed, as in example number 2. Furthermore, it is hard to see how C&C's account of such interactions, given a Levelt-like characterisation of production as the inverse of parsing, can match the required level of granularity.

In contrast, such data follow as an immediate consequence of the DS view of syntax. Speakers and hearers both use the defined tree-growth mechanisms to construct a representation of what is being said, taking the immediate context as input: The only difference between them is the additional requirement on speakers that the construction process has to be commensurate with some more richly annotated (possibly incomplete) structure corresponding to what they have in mind. This dynamic predicts that switching from parsing to production, and the converse, will be seamless, yielding the effect of in-tandem construction without needing to invoke higher levels of inference (Poesio & Rieser 2011) or superimposed duplication of the one type of activity upon the other (Pickering & Garrod 2013b). Each individual will simply be constructing the emergent structure relative to the context he or she has just constructed in his or her other capacity (Gregoromichelaki et al. 2011; 2013). Despite the DS commitment to word-by-word incrementality, interpretation can be built up with apparent delays, because language input invariably encodes no more than partial content specifications, allowing subsequent enrichment.

The result is, as C&C say, that there will be no encapsulated linguistic universals specific to the language faculty as universals will be grounded in general constraints on online cognitive processing. Yet this should not be taken to deny the existence of such universals: to the contrary, robust structural universals are predicted as dictated by limits imposed by logical and processing constraints in combination.

Consider the syntactic puzzle precluding multiple long-distance dependencies. Within DS, semantic representations as trees are defined as sets of nodes each of which is uniquely identified in terms of its position relative to other nodes in the tree (Blackburn & Meyer-Viol 1994). This definition restricts emergent tree growth to transitions which meet this characterisation. The effect is to freely license multiply building any one node, while ensuring that no such multiple actions give rise to distinguishable output. In the case of left-periphery effects, where on the DS account, nodes can be constructed as not yet fixed ("unfixed") within the current domain, nothing precludes such an action being repeated. However, such multiple applications of this strategy will invariably give rise to one and the same node, yielding a well-formed result as long as attendant attributes are compatible: hence, the restriction precluding multiple long-distance dependency. Verb-final languages, with their as-yet unfixed arguments, might seem apparent counterexamples; but here, the Chunk-and-Pass constraint provides an answer: Case specifications on an unfixed node are taken to induce an immediate update of that node to a locally fixed relation, allowing another construction of an unfixed node again with potential from its case specifications for update in anticipation of the following verb. The supposed counterexample of NP NP NP V sequences in verb-final languages thus merely demonstrates the interaction of logic-based and processing-based constraints, in turn accounting for typological observations such that verb-final languages are typically case-marking (Kempson & Kiaer 2010).

This constraint extends to language change, further bolstering the overall perspective (Bouzouita & Chatzikyriakis 2009). As C&C observe, language change commonly involves prosodic reduction of adjacent items leading to composite grammaticalised forms. On the DS view, such novel creations would reflect what had earlier been discretely triggered sequences of update actions, now with the novel composite form triggering this sequence of update actions as a single macro induced by that form. Accordingly, we expect such grammaticalised forms to reflect whatever general limits are imposed by intersections of logic and processing constraints (see Chatzikyriakidis & Kempson [2011] for arguments that weak [clitic] pronoun clusters in Greek constitute such a case). In short, DS buttresses C&C's claims about language as a mechanism for progressive construction of information-bearing units. Despite much variation across languages, synchronic and diachronic, the C&C program promises to enable formally characterisable perspectives on language directly matching the dynamics of language behaviour in interaction.

On the generalizability of the Chunk-and-Pass processing approach: Perspectives from language acquisition and music

doi:10.1017/S0140525X15000850, e80

Usha Lakshmanan and Robert E. Graham

Department of Psychology, Southern Illinois University Carbondale, Carbondale, IL 62901.

usha@siu.edu r.graham@siu.edu

Abstract: Christiansen & Chater (C&C) offer the Chunk-and-Pass strategy as a language processing approach allowing humans to make sense of incoming language in the face of cognitive and perceptual constraints. We propose that the Chunk-and-Pass strategy is not adequate to extend universally across languages (accounting for typologically diverse languages), nor is it sufficient to generalize to other auditory modalities such as music.

Christiansen & Chater (C&C) claim universality and primacy for their Chunk-and-Pass processing approach in language acquisition and suggest that music provides an example of another complex acoustic signal with multilayered structuring, to which one could apply the Chunk-and-Pass strategy as well. However, fundamental issues that C&C leave unaddressed suggest that this strategy may not be generalizable to typologically diverse languages and to domains beyond language. We discuss two such issues: (1) cross-linguistic differences (e.g., morphology and word-order) and (2) domain-specific differences (e.g., language versus music).

It is unclear how the Chunk-and-Pass strategy would work in the acquisition of synthetic languages, with complex inflectional morphology (e.g., Tamil, Turkish, Navajo, Quechua, Cree, Swahili). Because there is extensive suffixation (through agglutination or fusion), the morpheme-to-word ratio in such languages is high, resulting in lengthy words. A single multimorphemic word expresses meanings that in a language with limited or no inflection would require a multiword clause or sentence to express. Although C&C suggest that chunking of complex multimorphemic words, by means of local mechanisms (e.g., formulaicity), also applies to agglutinative languages, they mainly consider evidence based on English, whose impoverished inflection and low morpheme-to-word ratio (particularly in its verb forms), facilitates chunking using the word (as opposed to its subparts) as a basic unit of analysis.

In C&C's framework, frequency and perceptual salience (rather than innate grammatical mechanisms) drive the chunking process. Existing studies on the acquisition of morphologically complex languages indicate that mechanisms proposed for English do not readily extend to synthetic language types (Kelly et al. 2014). Crucially, lexicon-building does not take place through storage of frequently encountered (and frequently used) exemplars in memory; instead, the chunking strategy may be only a first step in the process, in preparation for the next stage, namely, grammatical decomposition of stored units and the acquisition of the combinatorial principles determining their subparts (see Rose & Brittain 2011 for evidence from Northern East Cree). However, even a minor role for the chunking strategy in relation to morphologically complex languages may be problematic. A single verb root/stem, for example, is manifested through numerous surface realizations rendering the frequency factor unreliable. Additionally, evidence from children acquiring Quechua and Navajo, two morphologically rich languages, indicates that regardless of the perceptual salience of the verb root/stem (i.e., word initial in Quechua and word final in Navajo), the children's earliest verb forms were root/bare stems, not permitted in the adult grammar; however, they never produced isolated affixes, contrary to what would be predicted if they were using a simple chunking procedure (Courtney & Saville-Troike 2002). Interestingly, Tamil children use bare stems in imperative contexts, similar to adults. In contrast, their earliest indicative (nonimperative) verb forms are non-adult-like and consist predominantly of verbal participles (derived or inflected nonfinite stems) with the auxiliary, tense, and agreement suffixes stripped away (Lakshmanan 2006). The mismatch between the children's early verbs and the adult input (consisting of complex multimorphemic words) emphasizes the role of innate knowledge of fundamental grammatical concepts (e.g., verb root/stem, inflected stem, and affixes).

A Chunk-and-Pass strategy alone (without independent grammatical mechanisms), cannot explain children's success with "free word order" found in many morphologically complex languages. In Tamil, an SOV (Subject-Object-Verb) language, sentential constituents (NPs, PPs, and CPs) may appear in noncanonical sentential positions through rightward and leftward scrambling. Tamil is a null argument language, and sentences with overt realization of all arguments are rare. Tamil children between the ages of 17 months and 42 months, exhibit sensitivity to Case restrictions and movement constraints on scrambling and successfully use adult-like word order permutations to signal interpretive differences (Focus versus Topic) (Sarma 2003).

A Chunk-and-Pass strategy would predict that shorter sentences are easier for children to process and produce than longer sentences. However, this cannot explain scenarios where the reverse situation holds. For example, Tamil children (below age 5) produce significantly fewer participial relatives than older children. They also strongly prefer tag relatives to the participial relative, although the former are longer and less frequent than the latter. Crucially, the participial relative, though shorter (and more frequent), is structurally more complex because it involves movement (Lakshmanan 2000).

Let us now examine the generalizability of the Chunk-and-Pass approach to other complex acoustic input, as in the case of music. Some argue that music contains some semantic information, such as meaning that emerges from sound patterns resembling qualities of objects and suggesting emotional content, or sometimes as a result of symbolic connections with related but external material (Koelsch 2005). For example, Wagner was known to have short musical melodies (*leitmotif*) that represented characters in his operas, such that interactions between characters could be inferred or interpreted from musical composition. However, these more concrete occurrences are outliers among musical works, and other interpretations of musical semantics remain much weaker than in the context of language. Thus, although it is possible there is structural chunking, music lacks the semantic information to inform something like an "interactionist" approach (McClelland 1987) to parsing.

Another way in which music differs from language is in the context of anticipation. C&C discuss anticipation as a predictive perceptual strategy that helps streamline the process of organizing incoming speech signals. Although music perception involves anticipation, music provides clues of a different nature regarding what will follow in a phrase. Anticipation based on hierarchical phrase structure might be similar across language and music, but listeners also use rhythm, meter, and phrase symmetry to predict how a musical phrase will end. C&C also discuss anticipation in discourse; however, anticipation works differently in music. In ensemble performances, musicians often simultaneously produce and perceive (their own and others') music, which is different from linguistic turn-taking.

In sum, it is unclear why the "mini-linguist" theory of language acquisition and processing theory need to be mutually exclusive – why can't the child be acquiring grammar as a framework for processing and chunking? What also needs explanation is the question: Why would having only domain-general mechanisms for processing different types of complex acoustic input be advantageous?

"Process and perish" or multiple buffers with push-down stacks?

doi:10.1017/S0140525X15000862, e81

Stephen C. Levinson

Max Planck Institute for Psycholinguistics, Nijmegen & Donders Institute for Brain, Cognition and Behaviour, Radboud University, 6500 AH Nijmegen, The Netherlands.

stephen.levinson@mpi.nl

Abstract: This commentary raises two issues: (1) Language processing is hastened not only by internal pressures but also externally by turn-taking in language use; (2) the theory requires nested levels of processing, but linguistic levels do not fully nest; further, it would seem to require multiple memory buffers, otherwise there's no obvious treatment for discontinuous structures, or for verbatim recall.

Christiansen & Chater (C&C) have tried to convert a truism of psycholinguistics (essentially, Miller's 1956 short-term memory limitation) into a general theory of everything in language, in which representations are mere traces of processing, while hierarchy, patterns of change and the design features of language all follow from processing limitations. But like most general theories, this one seems underspecified, and it is hard to know exactly what would falsify it.

In this commentary I make two points. First, I suggest that the pressure for speed of processing comes not only from the effect of an evanescent signal on internal processing constraints, but also from outside, from facts about how language is used. Second, I would like to gently question the truism of the "process and perish" theory of linguistic signals.

Language comes, for the most part, as an acoustic signal that is delivered remarkably fast – as C&C note, faster than comparable nonlinguistic signals can be decoded. But why? One might try, speculatively, to relate this to the natural processing tempo of the auditory cortex (Hickok & Poeppel 2000) or to some general drive to efficiency. In fact, there are more obvious reasons for haste – namely, the turn-taking system of language use (Sacks et al. 1974). The turn-taking system operates with short units (usually a clause with prosodic closure), and after one speaker's such unit, any other speaker may respond, the first speaker gaining rights to that turn – thus ensuring communication proceeds apace. Turn transitions on average have a gap of only c. 200 ms, or the duration of a single syllable. Speakers are hastened on by the fact that delayed responses carry unwelcome semiotics (Kendrick & Torreira 2015). Now, the consequences of this system for language processing are severe: It takes c. 600 ms for preparation to speak a single word (Indefrey & Levelt 2004) and c. 1,500 ms to plan a single clause (Griffin & Bock 2000), so to achieve a gap of 200 ms requires that midway during an incoming turn, a responder is predicting the rest of it *and* planning his or her response well in advance of the end. To guard against prediction error, comprehension of the incoming turn must proceed even during preparation of the response – so guaranteeing overlap of comprehension and production processes (Levinson & Torreira 2015). This system pushes processing to the limit.

Let's now turn to the psycholinguistic "truism," namely that, given the short-term memory bottleneck and the problems of competition for lexical access, processing for both comprehension and production must proceed in "chunks" – the "increments" of incremental processing. Miller's (1956) short-term memory bottleneck is often married to Baddeley's (1987) auditory loop with a capacity of c. 2 seconds unless refreshed, rapidly overwritten by incoming stimuli. On these or similar foundations the current theory is built.

Assuming Miller's bottleneck, and chunking as a way of mitigating it, I see at least two points in the current theory that are either problematic or need further explication:

1. **How many buffers?** Chunking involves recoding longer lower-level strings into shorter, higher-level strings with "lossy" compression of the lower level. In Miller's theory, the higher-level chunks replace the lower ones, using that same short-term memory buffer. But in C&C's theory, the higher-level chunks will need to be retained in another buffer, as the next low-level increment is processed – otherwise, for example, discontinuous syntactic elements will get overwritten by new acoustic detail. Because there is a whole hierarchy of levels (acoustic, phonetic, phonological, morphological, syntactic, discourse, etc.), the "passing the buck upward" strategy will only allow calculation of coherence if there are just as many memory buffers as there are levels.

2. **Mismatching chunks across levels.** C&C's theory seems to presume nesting of chunks as one proceeds upward in comprehension from acoustics to meaning. A longstanding linguistic observation is that the levels do not in fact coincide. A well-known example is the mismatch between phonological and syntactic words (Dixon & Aikhenvald 2002): Consider resyllabification, as in the pronunciation of *my bike is small* as *mai.bai.kismall* (Vroomen & de Gelder 1999) – here, the lower-level units don't match the higher ones. Similarly, syntactic structure and semantic structure do not match: *All men* looks like *Tall men* in surface structure, but has a quite different underlying semantics. Jackendoff's (2002) theory of grammar, with interface rules handling the mismatch between levels, is an attempt to handle this lack of nesting across levels.

Another fly in the ointment is that, despite the hand-waving in sect. 6.1.2, nonlocal dependencies are not exceptional. Particle verbs, conditionals, parentheticals, wh-movement, center-embedding, topicalization, extraposition, and so forth, have been central to linguistic theorizing, and together such discontinuous constructions are frequent. Now, it is true that English – despite these constructions – generally likes to keep together the bits that belong together. But other languages (like the Australian ones) are much freer in word order – like classical Latin with c. 12% of NPs discontinuous, as in the three-way split (parts in bold, Pinkster 2005; Snijders 2012) in Figure 1.

Likewise, the preference for strictly local chunking runs into difficulties at other linguistic levels. Consider the phonological rule that, according to the grammar books, requires the French possessive pronoun *ma* to become *mon* before a vowel (*ma femme vs. mon épouse*, "my wife"); in fact, *mon* is governed by the properties of the head noun from which it may be separated, as in *Marie sera soit mon soit ton épouse* ("Marie will become either my or your wife"; Schlenker 2010). Morphology isn't necessarily well behaved either, some languages even randomizing affixes (Bickel et al. 2007). So we need to know how the local-

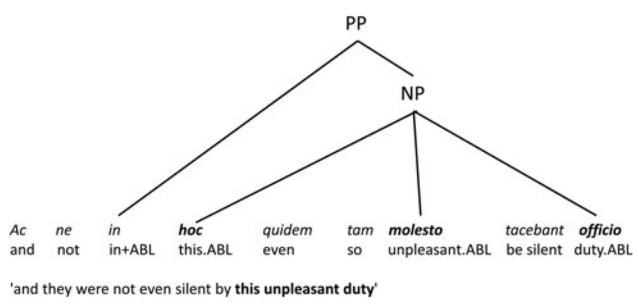

Figure 1 (Levinson). A discontinuous noun phrase (NP) in Latin wrapped around verb and adverb.

processing preference fails to outlaw all of the discontinuous structures in language, and where our push-down stack capacities actually reside.

Finally, C&C's Now-or-Never bottleneck theory suggests that details of an utterance cannot be retained in memory when following material overwrites it – only the gist of what was said may persist. But the practice of "other-initiated repair" suggests otherwise – in the following excerpt Sig repeats verbatim what he earlier said, just with extra stress on *shoot* even though three conversational turns intervene (Schegloff 2007, p. 109):

(1) Sig: Conservatives like to shoot people (and liberals don't?)
(2.0)
Dad: Conservatives like wha:t?
(0.8)
Sig: Wha:?
Dad: Whadyu say about conservatives? ((mouth full))
(0.3)
Sig: Conservatives like ta shoot people en (hh) liberals don't?

The fact that we can rerun the phonetics (? = rising intonation, underlining = stress) of utterances shows the existence of other buffers that escape the proposed bottleneck.

Linguistic structure emerges through the interaction of memory constraints and communicative pressures

doi:10.1017/S0140525X15000874, e82

Molly L. Lewis and Michael C. Frank
Department of Psychology, Stanford University, Stanford, CA 94305.
mll@stanford.edu **mcfrank@stanford.edu**
http://web.stanford.edu/~mll/
http://web.stanford.edu/~mcfrank/

Abstract: If memory constraints were the only limitation on language processing, the best possible language would be one with only one word. But to explain the rich structure of language, we need to posit a second constraint: the pressure to communicate informatively. Many aspects of linguistic structure can be accounted for by appealing to equilibria that result from these two pressures.

Christiansen & Chater (C&C) claim that memory limitations force the cognitive system to process the transient linguistic signal by compressing it. They suggest that this processing pressure influences the ultimate structure of language over the course of language evolution. Taken at face value, this proposal would lead to a degenerate linguistic structure, however. If memory constraints were the only pressure on language, languages would evolve to compress meaning into the simplest possible form – a single word (Horn 1984). But, as the authors point out, natural languages are not of this sort; they are richly structured into lexical and phrasal units of varying length. To account for this variability, we highlight the need to consider the communicative function of language. Communication serves as an important counter-pressure against compression in language processing, not just as a caveat.

Interlocutors use language with the goal of communicating information, but they also aim to minimize energetic cost (Zipf 1949). For the speaker, this goal implies minimizing production cost, and for the listener it implies minimizing comprehension cost. Importantly, these processing constraints have opposing cost functions (Horn 1984; Zipf 1949). For a producer, processing is minimized when a form is easy to say, and thus highly compressible. For the comprehender, however, processing is minimized when a form is minimally ambiguous and thus verbose. Compressing information is a useful strategy for a speaker who faces memory constraints, but it is useful only to the extent that the listener can still recover the intended meaning. This view of language use as rational action – minimizing costs while maximizing information transfer – is supported by a rich body of theoretical and empirical work (Clark 1996; Frank & Goodman 2012; Goodman & Stuhlmüller 2013; Grice 1975).

Although C&C argue that compression is the key factor in the emergence of structure, evidence at both the acquisition and evolution timescales suggests language is the product of the interaction between both compression and informativity. At the timescale of acquisition, experimental work suggests the resolution of reference in word learning is the product of communicative inferences (e.g., Baldwin 1991; 1993; Frank et al. 2009; Frank & Goodman 2014). And at the timescale of language evolution, a growing body of work suggests that the forms of words are also equilibria between these two pressures (Lewis & Frank 2014; Mahowald et al. 2012; Piantadosi et al. 2011; Zipf 1936). For example, Piantadosi et al. (2011) found that words that are less predictable in their linguistic context are longer, suggesting that speakers may lengthen words that are surprising in order to increase time for the listener to process.

In addition to linguistic form, these pressures influence the mapping between form and meaning. An equilibrium in the structure of form-meaning mappings is one in which the listener is able to recover the intended meaning, but the speaker does not exert additional effort over-describing. A range of semantic domains reflect this equilibrium (Baddeley & Attewell 2009; Kemp & Regier 2012; Regier et al. 2007), and ambiguity, more generally, has been argued to reflect this communicative tradeoff (Piantadosi et al. 2012). Ambiguity is an equilibrium in cases where the listener can recover the intended meaning from the communicative context. One example is the word "some," which has a literal meaning of "at least one and possibly all" but can be strengthened pragmatically to mean "at least one but not all" (Horn 1972). Because its meaning is determined through communicative context, its literal semantics can overlap those of its competitor, "all."

The key challenge associated with this broader proposal – that processing pressures influence linguistic structure – is providing direct evidence for a causal link between these two timescales. This problem is difficult to study in the laboratory because the proposed mechanism takes place over a long timescale and over multiple individual speakers. Furthermore, the presence of a causal link does not entail that phenomena in processing are directly reflected in linguistic structure – rather, entirely new properties may emerge at higher levels of abstraction from the interactions of more fundamental phenomena (Anderson 1972). It may, therefore, not be possible to directly extrapolate from brief communicative interactions observed in the laboratory to properties of linguistic structure.

Several recent pieces of experimental data begin to address this challenge, however. In one study, Fedzechkina et al. (2012) asked speakers to learn an artificial language that arbitrarily distinguished nouns through case-marking. Over learning sessions, speakers developed a system for marking in contexts where meanings were least predictable – a pattern reflected in the case-marking systems of natural language. Other work has used a similar paradigm to reveal the emergence of typologically prevalent patterns in the domains of word order (Culbertson et al. 2012; Culbertson & Newport 2015) and phonology (Wilson 2008).

A particularly promising approach for exploring this causal link is through transmission chains (Kirby et al. 2008; Reali & Griffiths 2009). In a transmission chain, a participant learns and recalls a language, and then the recalled language becomes the learning input for a new learner. By iterating over learners, we can observe how languages change across transmission of learners over the course of language evolution. Kirby et al. (2015) have compared the emergence of linguistic structure in a regime that iterates over different partners of learners versus a regime where the same two partners repeatedly interact with each other. They find that linguistic structure emerges only by iterating over different partners, demonstrating the unique contribution of cross-generational learning to the emergence of structure. Others have begun to use this paradigm to link the interaction of processing pressures to the emergence of communicative regularities in semantic structure (Carstensen et al. 2015; Lewis & Frank 2015).

In sum, the consequences of memory constraints are likely a critical factor in shaping language structure. But an additional important constraint is the pressure to communicate informatively, and this constraint should not be overlooked in accounting for linguistic structure.

The bottleneck may be the solution, not the problem

doi:10.1017/S0140525X15000886, e83

Arnon Lotem,[a] Oren Kolodny,[b] Joseph Y. Halpern,[c] Luca Onnis,[d] and Shimon Edelman[e]

[a]*Department of Zoology, Tel Aviv University, Tel Aviv 6997801, Israel;* [b]*Department of Biology, Stanford University, Stanford, CA 94305;* [c]*Department of Computer Science, Cornell University, Ithaca, NY 14853;* [d]*Division of Linguistics and Multilingual Studies, Nanyang Technological University, Singapore 637332;* [e]*Department of Psychology, Cornell University, Ithaca, NY 14853.*

lotem@post.tau.ac.il **okolodny@stanford.edu**
halpern@cs.cornell.edu **lucao@ntu.edu.sg** **se37@cornell.edu**

Abstract: As a highly consequential biological trait, a memory "bottleneck" cannot escape selection pressures. It must therefore coevolve with other cognitive mechanisms rather than act as an independent constraint. Recent theory and an implemented model of language acquisition suggest that a limit on working memory may evolve to help learning. Furthermore, it need not hamper the use of language for communication.

The target article by Christiansen & Chater (C&C) makes many useful and valid observations about language that we happily endorse. Indeed, several of C&C's major points appear in our own papers, including the following: (a) the inability of non-chunked, "analog" approaches to language to compete with "digital" combinatorics over chunks (Edelman, 2008b); (b) the centrality of chunking to modeling incremental, memory-constrained language acquisition and generation (Goldstein et al. 2010; Kolodny et al. 2015b) and the possible evolutionary roots of these features of language (Kolodny et al. 2014; 2015a; Lotem & Halpern 2012); (c) the realization that language experience has the form of a graph (Solan et al. 2005; cf. Edelman 2008a, p. 274), corresponding to C&C's "forest tracks" analogy; and (d) a proposed set of general principles for language acquisition and processing (Goldstein et al. 2010), one of which is essentially identical to C&C's "Now-or-Never bottleneck." However, our theory is critically different in its causality structure. Rather than assuming that the memory limit is a fixed constraint to which all other traits must adapt, we view it as an adaptation that evolved to cope with computational challenges. Doing so brings theory in line with standard practice in evolutionary biology, is more consistent with research findings, and raises numerous important research issues. We expand on these points in the following paragraphs.

No biological trait can be simply assumed as a "constraint." Viewing the Now-or-Never bottleneck as an evolutionary constraint to which language adapts – C&C's central idea – is unwarranted. In evolutionary theory, biological constraints – as opposed to constraints imposed by physics and chemistry, which are not subject to biological evolution – cannot simply be assumed; they must be understood in terms of trade-offs among selective pressures. Clearly, birds' wings evolved under aerodynamic constraints rather than vice versa. However, biological traits such as memory are not exempt from evolving. In proposing a bottleneck to which everything else in the system must adapt while the bottleneck itself remains fixed and independent (Fig. 1 in the target article), C&C implicitly assume that it cannot evolve.

To justify this assumption, C&C should have offered evidence of stabilizing selection pressures that act against genetic variants coding for a broader or narrower bottleneck, and thereby affecting cognition and, ultimately, fitness. Alternatively, they might have assumed that the biological mechanisms underlying the memory bottleneck cannot be genetically variable – an odd assumption, which runs counter to substantial evidence in humans of (a) a range of verbal memory decay rates (Mueller & Krawitz 2009), including in particular the longer verbal working memory span in individuals with Asperger's (Cui et al. 2010); (b) heritable variation in language and in word memory (Stromswold 2001; van Soelen et al. 2011) and in working memory (Blokland et al. 2011; Vogler et al. 2014); and (c) variation in perceptual memory across species (Lind et al. 2015; Mery et al. 2007). Given that heritable variation in a trait means that it can respond to selection (e.g., Falconer 1981), it is likely that the bottleneck *can* evolve, and that it is what it is because individuals with longer or shorter verbal working memory had lower biological fitness.[1]

If language is supported by domain-general mechanisms, verbal memory is even less immune to evolution. If the emergence of language constitutes a recent and radical departure from other cognitive phenomena, it is in principle possible that working memory evolved and stabilized prior to and separately from the "increasingly abstract levels of linguistic representation" (sect. 3.2, para. 2) posited by C&C. However, there are good arguments in support of a domain-general view of language (e.g., Chater & Christiansen 2010). In particular, linguistic representations and processes are hardly as modular as C&C assume (Onnis & Spivey 2012). Furthermore, theories of neural reuse (Anderson 2010) point to the massive redeployment of existing mechanisms for new functions, resulting in brain regions coming to be involved in diverse cognitive functions. If circuits that support language continue contributing to nonlinguistic functions (including working memory), a memory bottleneck is not a prior and independent constraint on language, but rather a trait that continues to evolve under multiple selective pressures, which include language.

The bottleneck may be the solution, not the problem. As we have suggested (Goldstein et al. 2010; Lotem & Halpern 2008; 2012; Onnis et al. 2008), a limited working memory may be an adaptation for coping with the computational challenges involved in segmentation and network construction. (Importantly, regardless of whether this specific hypothesis is correct, entertaining such hypotheses is the only way of distinguishing a function from a constraint; cf. Stephens & Krebs 1986, Ch. 10.) A recently implemented model that includes this hypothesis has been tested on tasks involving language, birdsong, and foraging (Kolodny et al. 2014; 2015a; 2015b; Menyhart et al. 2015) The model includes a time window during which natural and meaningful patterns are likely to recur and thus to pass a test for statistical significance, while spurious patterns decay and are forgotten. We stress that rather than acting as a constraint, the duration of the window must co-evolve with the mechanisms influencing the distribution of data so as to increase the effectiveness of memory representations (Lotem & Halpern 2012).

We do agree with C&C regarding some of the consequences of the memory bottleneck, such as the need for online incremental construction of hierarchical representation. Indeed, our model effectively implements what C&C call "Chunk-and-Pass" (Kolodny et al. 2015b).[2] We believe, however, that the ultimate constraint on learning structure (such as that of language) in time and space is not the memory bottleneck in itself, but rather the computational challenges of chunking the data and of building hierarchies.

Biological communication is about affecting behavior, not pumping bits. Our final point focuses on the communicative function of language. Viewing a memory window as a communication "bottleneck" suggests that massive amounts of information must flow through the channel in question. However, the real function of a message is to influence the rich network of connotations and interconnections already present in the listener's brain (cf. Edelman 2015, sect. 2.3). Communication is about generating adaptive behavioral changes (Burghardt 1970; Green & Marler 1979) – the listener gleans from it cues relevant to decision-making. For this, a signal must be informative and reliable in the given context (Leger 1993); the amount of information is not the main issue (except as a signal of quality, as in complex courtship songs; Lachmann et al. 2001). This implies that evolutionary selection in language is for how messages fit into the information already represented by their recipient; a bottleneck may not impose significant constraints here.

NOTES

1. If verbal memory indeed evolves, language is the niche in which it does so. The target article seems to gloss over the intimate connection between cultural evolution and niche construction (Odling-Smee et al. 2003). In focusing on how "linguistic patterns, which can be processed through that bottleneck, will be strongly selected" (sect. 5, para. 3), C&C ignore the possibility of there being also selection for individuals who can better process linguistic patterns.

2. As C&C note, correctly, regarding Chunk-and-Pass, "it is entirely possible that linguistic input can simultaneously, and perhaps redundantly, be chunked in more than one way" (sect. 3.2, para. 4). This point suggests that chunking on its own, especially when carried out recursively/hierarchically, is likely to severely exacerbate the combinatorial problem faced by the learner, rather than resolve the bottleneck issue.

Memory limitations and chunking are variable and cannot explain language structure

doi:10.1017/S0140525X15000898, e84

Maryellen C. MacDonald

Department of Psychology, University of Wisconsin-Madison, Madison, WI 53706.

mcmacdonald@wisc.edu

http://lcnl.wisc.edu/people/mcm/

Abstract: Both the Now-or-Never bottleneck and the chunking mechanisms hypothesized to cope with it are more variable than Christiansen & Chater (C&C) suggest. These constructs are, therefore, too weak to support C&C's claims for the nature of language. Key aspects of the hierarchical nature of language instead arise from the nature of sequencing of subgoals during utterance planning in language production.

Christiansen & Chater (C&C) overstate both the limitations of the Now-or-Never bottleneck and the lossy character of chunking, and they are overly optimistic that memory limitations can explain the nature of language. C&C correctly note that memory limitations during planning for language production promote *incremental planning* (where planning of the utterance and its execution of action are interleaved), but the memory limitations are not as strict as they suggest. Whereas "radical incrementality" – very minimal advance planning owing to a severe memory bottleneck – once had its proponents in language production, recent studies argue for looser constraints, with more tolerance for higher memory loads and more extensive advance planning (Ferreira & Swets 2002). The extent of advance planning may even be under some degree of implicit strategic control (Ferreira & Swets 2002; Wagner et al. 2010), suggesting that, rather than the memory bottleneck controlling us, we instead can exert some control over our own memory loads during language production. The bottleneck also isn't always so severe in comprehension, and chunking isn't as uniformly eager as C&C portray. Downstream linguistic input affects interpretation of earlier material (MacDonald 1994; Warren & Sherman 1974), which shouldn't occur if chunking greedily passes off the early information to the next level. Variability in the tolerance of memory loads suggests that the Now-or-Never bottleneck is really more of a wide-mouth jar, or perhaps more of an adjustable drawstring closure, and the consequences for the nature of language will therefore need adjustment as well.

Similarly, C&C view the lossy nature of Chunk-and-Pass processing as essential to explaining the nature of language processing, but chunking is neither as lossy nor as bottom-up as they suggest. C&C argue that in speech perception, sounds are rapidly chunked into words, leaving the sounds behind, so that the just-perceived sounds do not interfere with upcoming ones. These claims create several puzzles: First, this very bottom-up characterization of chunking is inconsistent with evidence for top-down influences in perception. C&C's focus on using context only for predicting the future is misplaced, because top-down processes also allow higher-level information to elaborate earlier percepts. Examples include the word superiority effect (Cattell 1886) and the phoneme restoration effect (Warren 1970), in which word representations affect perception of their parts (letters, phonemes). If chunking is so eager and lossy, it's not clear how higher-level word information could refine the lower-level percepts that should have already been discarded by lossy chunking. Second, if the memory bottleneck is so narrow, how is there room for interference, which by definition depends on several elements being in memory at the same time? There are numerous examples of semantic and sound overlap creating memory interference over fairly long distances during both comprehension (Acheson & MacDonald 2011; Van Dyke & Johns 2012), and production (Hsiao et al. 2014; Smith & Wheeldon 2004), again suggesting that the bottleneck can't be as strict at C&C describe. Third, if lossy chunking is the solution to memory interference, why is it so easy to find interference effects? The existence of memory interference suggests that chunking may not always be so lossy after all. In at least some circumstances, there appears to be real value in non-lossy processing, such as the Levy et al. (2009) example that C&C note as well as use of prosodic information over long distances (Morrill et al. 2014). These and other examples call into question the essence of lossy, greedy, bottom-up chunking as a design feature for language.

C&C note some variability in memory limits and chunking, but they do not discuss the consequences of variability for their account. They illustrate their ideas with an individual identified as SF, who can recall vast strings of meaningless digits by chunking them into meaningful units such as dates, and using the chunks to guide production. The analogy to language is unfortunate, because SF's chunking strategies are both conscious and idiosyncratic, inviting the inference that language users' chunking units are similarly variable. In sum, if memory limitations and the lossy and eager characteristics of chunking have notable exceptions and are subject to individual differences, then it is difficult to make them the foundation of claims for the nature of human language.

More seriously, no matter how we conceive the memory bottleneck, it can explain neither the existence of a hierarchy in language representations, nor why the hierarchy has certain levels of representation across individuals and not others. Consider a nonlinguistic analogy: the visual processes necessary to recognize a cup. Let's assume that these processes, also constrained by memory bottlenecks, have multiple stages of chunking and passing from low-level visual processing up to object recognition. From these perceptual stages, however, we would not want to conclude that the percept itself, the cup, has a hierarchical structure. Similarly, the memory-constrained chunking and passing for language perception, even if it works exactly as C&C describe, does not give the percept – language – its hierarchical structure.

Rather than trying to wring structure out of memory limitations, I suggest that key aspects of hierarchical structure emerge from how goals are realized in action (MacDonald 2013). Like all actions, language production must unfold over time, meaning that the various subgoals of the action must be planned and ordered in some way (Lashley 1951). For both nonlinguistic and linguistic actions, the nature of the hierarchy is constrained by the need to make decisions for some subgoals in order to plan others. To reach for a cup, the choice of which hand to use determines and must precede planning the reach. Similarly, a speaker must choose words (*cup* or *mug*?) before programming their articulation, naturally creating a hierarchy of lexical and sublexical plans. Although language and nonlinguistic action are not identical, important aspects of the hierarchical nature of language emerges from the staging of language production planning processes over time. Furthermore, although action plans are held in memory and are affected by the nature of that memory, memory limitations themselves cannot bear the explanatory burden that C&C ascribe to them.

Exploring some edges: Chunk-and-Pass processing at the very beginning, across representations, and on to action

doi:10.1017/S0140525X15000904, e85

Rose Maier and Dare Baldwin

Department of Psychology, University of Oregon, Eugene, OR 97403.
rosem@uoregon.edu baldwin@uoregon.edu

Abstract: We identify three "working edges" for fruitful elaboration of the Chunk-and-Pass proposal: (a) accounting for the earliest phases of language acquisition, (b) explaining diversity in the stability and plasticity of different representational types, and (c) propelling investigation of action processing.

Experience is dynamic and ephemeral, yet humans routinely generate abstract representations of their individualized experience that simultaneously achieve enough stability, plasticity, and interindividual parity to radically facilitate social and cognitive functioning. Christiansen & Chater's (C&C's) ambitious Chunk-and-Pass processing (CPP) proposal offers hope of a comprehensive and elegant account of how this can be. CPP has impressive explanatory breadth, neatly tying language acquisition to language change and language evolution, while also offering promise of a unified account of perception and cognition more generally. By C&C's own acknowledgment, however, many facets of the CPP account cry out for elaboration. In our view, three "working edges" will be (a) accounting for the earliest inception of language acquisition, (b) explaining stability and plasticity differences in learning profiles across knowledge systems (within language as well as across domains), and (c) elaborating CPP on the action processing front.

Regarding the first issue, C&C provide a workable framework for describing language acquisition once basic acoustic units have been discovered (e.g., phonemes, syllables), but do not describe how utter novices initially break into the system. Of course, there is a sizable literature investigating how infants initiate analysis of streaming speech (e.g., Vouloumanos & Werker 2007; Werker et al. 2012). One litmus test of the viability of CPP will be its ability to account for the phenomena documented in this literature within a unified Chunk-and-Pass framework. Among the complexities to be confronted here include findings indicating that infants' identification/construction of basic acoustic units may still be taking place at the same time that they are beginning to chunk longer strings of sounds together into words or morphemes. For example, infants remain quite sensitive to phonetic distributions until well into the first year; at 6 to 8 months, just 2–3 minutes of focused exposure to new distributions may be enough to temporarily rearrange infants' phonetic categories (Maye et al. 2002). And yet, by this same age, infants typically recognize at least a handful of words, including "mommy" and "daddy" (Tincoff & Jusczyk 1999), their own name (Bortfeld et al. 2005; Mandel et al. 1995), and several body part terms, such as "feet" and "hand" (Tincoff & Jusczyk 2012). Does CPP somehow build linguistic structure even without clear basic units over which to operate (in contradiction to hypotheses C&C articulate on this matter; e.g., sect. 3.2, para. 1)? Alternatively, does CPP operate on units only as they reach some criterion of availability, so that words composed of early-identified phonemes would potentially be available for chunking, whereas words with more difficult-to-identify phonemes are not? Or do processes other than Chunk-and-Pass need to be brought in to account for the earliest phases of language acquisition?

The second working edge we identify relates to stability and plasticity of representations. C&C note that stability and plasticity trade off: Learning depends on representations being updated to incorporate new content, but at the same time, some degree of stability is needed to avoid new information overwhelming previously acquired information. They argue that stability is a natural product of the compression that occurs during Chunk-and-Pass processing. The processing of linguistic content is "lossy" – the only features retained are those that are captured by a learner's current model of the language, making it difficult to dramatically alter that model since the features necessary to do so are likely the very ones lost in compression. This seems persuasive on the face of it, but leaves unclear how CPP can account for a different stability/plasticity issue: namely, the observation that representations of different types display distinct stability/plasticity profiles. In language, acquired representations of some kinds (e.g., phonetic and syntactic representations) display a strong propensity to stabilize and become markedly resistant to change (e.g., Johnson & Newport 1989; Kuhl 2004; Lenneberg 1964; Yoshida et al. 2010), whereas a variety of evidence suggests that other representational types (e.g., open-class lexical items) seem to display considerably more plasticity (e.g., Curtiss 1977; Newport 1990; Talmy 2000; Weber-Fox & Neville 1996). In question is whether these different plasticity profiles across representational types arise naturally from CPP. Are there differences in the information to be encoded across various types of representations such that the model would predict an emphasis on stability in some cases versus ongoing plasticity in others? Alternatively, will it be necessary to look to mechanisms beyond CPP to account for such differences, such as diverse neural commitment timetables?

Our third "working edge" focuses on action processing as a particularly fruitful target for broadening the scope of CPP-related investigation. Intuitively, language and action processing seem closely linked. Language can be regarded as one form of action, after all, and both language and action are subject to the Now-or-Never bottleneck, making them amenable to a CPP account, as C&C themselves note. Strikingly, however, investigation regarding action processing lags considerably behind language. One glaring example is the lack of a generally accepted inventory of basic actions, comparable to inventories of phonemes or syllables in language (cf. interesting but small-scale efforts along these lines, such as *therblig*, Gilbreth & Gilbreth 1919). Another example concerns hierarchical structure, which seems to be a fundamental organizing principle of both action and linguistic representations. To illustrate in the action context, observers typically note that an action such as getting a cup of coffee comprises embedded subgoals, such as getting a mug from a cupboard, placing it on a counter, pouring coffee into the mug, and so on. At the same time, relevant levels of that hierarchy seem not to be as crisp or well-defined as they are in language. A "learning to process" account may provide welcome guidance for continuing attempts to gain purchase on the representation of structure in action, and perhaps also will ultimately help to explain cross-domain differences in representational structure. All in all, as an explicitly domain-general approach, CPP holds promise for accelerating understanding in the action domain in a way that promotes interdisciplinary convergence with theorizing about language.

Many important language universals are not reducible to processing or cognition

doi:10.1017/S0140525X15000722, e86

David P. Medeiros, Massimo Piattelli-Palmarini, and Thomas G. Bever

Department of Linguistics, University of Arizona, Tucson, AZ 85721-0028.
medeiros@email.arizona.edu massimo@email.arizona.edu
tgb@email.arizona.edu
http://dingo.sbs.arizona.edu/~massimo/
http://dingo.sbs.arizona.edu/~tgb/

Abstract: Christiansen & Chater (C&C) ignore the many linguistic universals that cannot be reduced to processing or cognitive constraints, some of which we present. Their claim that grammar is merely acquired language processing skill cannot account for such universals. Their claim that all other universal properties are historically and culturally based is a nonsequitur about language evolution, lacking data.

In this latest attempt to reduce language to other mental systems, Christiansen & Chater (C&C) present two main points, each with two subpoints: (1a) Working memory constraints account for many features of sentence processing during comprehension; (1b) these features in turn can account for a variety of universal properties of language. (2a) Thus, learning a language is actually learning a set of rapidly deployable recoding templates and processes; (2b) what appear to be other kinds of psychologically or biologically determined structures of language are actually culturally and historically determined. Such attempts have a long history, with a considerable modern literature on the issue started in the 1970s (e.g., Bates & MacWhinney 1982; Hawkins 1983; Rumelhart & McClelland 1988; notable recent examples include Arbib 2012; Bybee 2007; Christiansen & Chater 2008; Perfors et al. 2011; Reali & Christiansen 2005; Rizzolatti & Arbib 1998; Tomasello 2003; 2006. All of these attempts have been quickly and persuasively countered: Berwick et al. 2013; Crain et al. 2009; Gualmini & Crain 2005; Kam & Fodor 2013; Piattelli-Palmarini et al. 2008; Pietroski 2008; Wexler 2002.)

Irreducible language universals. Many linguistic systems are irreducible to processing or cognitive explanations. We highlight several that seem particularly challenging to C&C's views.

(a) The Verb+Object Constraint (VOC) (Baker 2008; 2013). In our conceptualization of the world, actions are more intimately connected with their agent than with the object, but not syntactically so. Verb+Complement forms a syntactic constituent (a chunk) but Subject+Verb does not. This abstract structural relationship explains the fact that in all languages of the world idioms are formed by a verb and its object (In English, for example, *kick the bucket*, *sell the farm*, *hits the fan*, etc.). This fact is particularly surprising for VSO languages, on the "Chunk-and-Pass" perspective: Surface adjacency ought to lead to V+S idioms being more readily chunked and learned in such languages, while V … O idioms are, in simple clauses, discontinuous.

(b) There is a universal hierarchy of syntactic and semantic dominance relations (Belletti 2004; Cinque 1999; 2013): for example, evidential (*allegedly*) > epistemic (*probably*) > necessity (*necessarily*) > continuative (*still*) > durative (*briefly*) > obligation (*obligatorily*) > completive (*partially*). (The > indicates dominance in the ordering of modal modifications of a sentence, a transitive relation.) For example, in English we have:

(1) Jim is allegedly probably unable to frequently deliver assignments on time.
(2) *Jim is frequently unable to probably deliver allegedly his assignments on time.

There is a large literature on many languages suggesting that this ordering is universal. Explanations based on statistical regularity, general cognition, pure logic, or social conventions appear utterly implausible.

(c) Conceptually possible but linguistically impossible word ordering.
"[M]any potential orders are never found … which poses a puzzle for any culturally based account" (Cinque 2013, p. 17). Consider, for example, the relative ordering of the categories demonstrative, numeral, adjective, and noun, the topic of Greenberg's Universal 20 (Greenberg 1963; see also Hawkins 1983; Dryer 1992; 2009; Cinque 1996; 2005; 2013). All descriptions agree that some orders are never found: Whereas (3) and (4) are common orders, no language is reported to have as a basic noun phrase order (5) *Num Adj Dem N or (6) *Adj Dem N Num.

(3)	These three blind mice	Dem Num Adj N
(4)	Mice blind three these	N Adj Num Dem
(5)	*Three blind these mice	*Num Adj Dem N
(6)	*Blind these mice three	*Adj Dem N Num

The observed restrictions on nominal ordering are particularly interesting in light of experimental work by Culbertson et al. (e.g., Culbertson & Adger 2014; Culbertson et al. 2012). Briefly, they find their adult subjects, in a series of artificial grammar learning experiments, to reproduce typological word ordering patterns, apparently drawing on innate cognitive biases. This is a strong piece of evidence that the distribution of word order patterns is not historical bricolage; subjects discriminate novel typologically favored patterns from disfavored patterns, with no obvious basis in their native language.

Grammar learning is "merely" process and pattern learning. C&C argue that in learning to comprehend (and, we presume, talk), the child perforce must be learning a range of statistically valid local patterns so that the system can proceed rapidly. The heart of the idea is that learning patterns from repeated stimulus similarities is endemic to many aspects of maturation, hence not specific to language. In this, they agree with a variety of learned pattern accounts (e.g., Bever 1970; Townsend & Bever 2001). However, there are severe empirical problems. Their account says nothing, for instance, about which chunks may relate to each other; as far as C&C are concerned, anything goes. But there is considerable evidence for richly nuanced, universal principles governing many kinds of grammatical relations (subjacency, case, theta relations, etc.). It also makes long-distance dependencies mysterious. If learners look first for local associations in blindly segmenting their language, subject to a crippling limit on short-term memory, it is unclear how long-distance dependencies could be stable in any lineage, much less universal.

The "rest" of apparent linguistic structures (i.e., those that are not explained by immediate processing or by cognitive or statistical facts) are culturally and historically determined.

We do not belabor a response to this point because it is irrelevant to the major substantive claims by C&C, and they offer very little independent or new evidence for it. It is a claim about how the structures evolved that we see in today's languages that cannot be immediately accounted for in their interpretation of processing and cognitive constraints.

To us it seems like a very far-fetched claim about how things worked in the forest primeval. We do know from contemporary facts that (most) languages live in families suggesting some historical devolution; and there are clusters of shared properties among neighboring languages that do not share families, also suggesting historical influences. But these facts presuppose the existence of fully fledged languages, ready to differentiate and to be influenced by neighbors.

Processing cost and its consequences

doi:10.1017/S0140525X15000916, e87

William O'Grady

Department of Linguistics, University of Hawaii at Manoa, Honolulu, HI 96822.
ogrady@hawaii.edu

Abstract: I focus on two challenges that processing-based theories of language must confront: the need to explain why language has the particular properties that it does, and the need to explain why processing pressures are manifested in the particular way that they are. I discuss these matters with reference to two illustrative phenomena: proximity effects in word order and a constraint on contraction.

Christiansen & Chater's (C&C's) proposal has much to recommend it: Processing resources are severely limited, and Chunk-and-Pass is a promising strategy for accommodating those limitations. The hope and promise of this type of work is that in addition to shedding light on the nature of incremental processing, it can help explain specific properties of linguistic systems. C&C focus their attention on very general features of language, such as duality of patterning, the bounded nature of linguistic units, and the existence of multiple levels of representation. But many properties at a finer level of granularity also call for attention. Why, for example, do we find certain systems of agreement and case marking, but not others? Why are some languages ergative? Why are filler-gap dependencies subject to certain types of locality constraints? Traditionally, the answers to such questions invoke principles of grammar, not processing. However, a wave of recent research by C&C and others (e.g., Hawkins 2004; 2014; O'Grady 2005; 2013; 2015a) proposes a very different approach: Languages are the way they are because of their need to adapt to processing pressures.

At least two challenges immediately arise. On the one hand, it is necessary to demonstrate that processing pressures can help resolve the baffling puzzles that spring up everywhere in the phonology, morphology, and syntax of natural languages. On the other hand, it is necessary to develop a theory to explain why the effects of the processing bottleneck are felt when and where they are. Two examples help illustrate this point.

As C&C note (sect. 6.1.2), items that enter into a relationship with each other should occur in close proximity, for obvious processing reasons. But how close? In Thai, not even a determiner can intervene between a verb and the head of its direct object (one says "I read book that"). But the picture is complicated by data from other languages.

(1) a. A determiner intervenes: (English, French, Mandarin)
read [**that** book]

b. A possessor NP intervenes: (English, Mandarin)
read [**a good friend's** book]

c. A relative clause intervenes (Mandarin):
read [**that I just bought**] books
(compare English: read books **[that I just bought]**)

Hawkins (2004, p. 123ff) offers a key insight: All other things being equal, if a language permits a more costly implementation of a particular relationship, it will also permit a less costly implementation. For example, Mandarin allows a relative clause to appear between the verb and the head of its direct object, as in (1c) – a costly option in terms of working memory; as predicted, however, Mandarin also allows a less complex possessor phrase and a simple determiner to occur in that position. English sets the bar lower, allowing only possessor phrases and determiners to intervene – as in (1a,b) – but not a relative clause. The cut-off point for French is still lower: A determiner can intervene, as in (1a), but not a possessor or a relative clause. Most restrictive of all is Thai, in which even determiners cannot intervene. The processing bottleneck, it seems, is not absolute; it is manifested in different ways in different languages.

Another example of systematic variation in processing effects involves the notorious constraint on *want to* contraction illustrated below.

(2) a. Contraction allowed:
Ask whether they want to stay there. (cf. They want to stay there.)
wanna

b. Contraction prohibited:
Ask who they want to stay there. (cf., They want Mary to stay there.)
*wanna

Jaeggli (1980) proposed that contraction is blocked in (2b) by the presence of an invisible Case-marked trace between *want* and *to* – a classic example of grammatical analysis. In contrast, O'Grady (2005) outlined a processing-based alternative that turns on the interplay between two pressures: (a) for reasons related to working memory, filler-gap dependencies are best resolved at the first opportunity; (b) for articulatory reasons, contraction is most natural when *want* and *to* combine with each other without delay. Matters are straightforward in (2a), where the articulatory system moves seamlessly from *want* to *to*, producing a contracted pronunciation.

(3) Ask whether they **want-to** stay there.
↓
wanna

The situation is very different in (3) than in (2b), in which the transition from *want* to *to* is interrupted by the need to promptly resolve the filler-gap dependency by associating the *wh* word with *want*, which is transitive here (cf. *We want* ***her*** *to stay*). The resulting delay, often accompanied by prosodic reflexes such as lengthening of *want* (Warren et al. 2003), compromises the naturalness of contraction.

(4) Ask who they **want # to** stay there.

Here too, though, there is evidently room for variation. Ito (2005) reported that 5 of the 41 English speakers who she studied allowed *wanna* in patterns like (2b). Crucially, however, they also permitted contraction in the less-demanding (2a). The reverse is, of course, not true: Many speakers permit contraction in the easy pattern but not the difficult one.

In sum, case studies such as these help confirm that processing pressures (C&C's Now-or-Never bottleneck) shape the way language works, creating an explanatory narrative that is fundamentally different from traditional grammar-based accounts. At the same time, we gain insight into the nature of processing itself, for which an intriguing story is beginning to emerge. Because processing cost can never be reduced to zero, there is no perfect language and no single way to manage processing costs. What we find instead is systematic variation in what languages (and speakers) tolerate, with a preference for less-costly options over more-demanding alternatives. The end result is an array of effects in phenomena ranging from typological variation to developmental order (O'Grady 2013; 2015b).

Processing cost offers an important idea on which to build. The next step requires further close-range attention to the details of how languages work, how they differ from each other, and how they are acquired. Here, in the traditional data fields of linguistics, lie the clues needed to settle the disputes that define the contemporary study of language.

Conceptual short-term memory (CSTM) supports core claims of Christiansen and Chater

doi:10.1017/S0140525X15000928, e88

Mary C. Potter
Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139.
mpotter@mit.edu
http://mollylab-1.mit.edu/lab/

Abstract: Rapid serial visual presentation (RSVP) of words or pictured scenes provides evidence for a large-capacity conceptual short-term memory (CSTM) that momentarily provides rich associated material from long-term memory, permitting rapid chunking (Potter 1993; 2009; 2012). In perception of scenes as well as language comprehension, we make use of knowledge that briefly exceeds the supposed limits of working memory.

Christiansen & Chater (C&C) focus on cognitive limitations in language understanding and production that force immediate decisions at multiple levels. Our experiments using rapid serial visual presentation (RSVP) of written words and of pictured scenes show that a large-capacity but short-lasting conceptual short-term memory (CSTM), consisting of associations from long-term memory, is retrieved in response to currently active stimuli and thoughts (Potter 1993; 2012). We "understand" when some structural connections are found between the current stimuli and CSTM. In visual perception of scenes and objects, as well as in language comprehension, we make quick use of knowledge that briefly exceeds the supposed limits of short-term memory. Consistent with C&C's core ideas, rich but unselective associations arise quickly but last only long enough for selective pattern recognition – chunking, in C&C's terms. Irrelevant associations never become conscious (or are immediately forgotten).

Three interrelated characteristics of CSTM support key ideas in C&C's target article. Demos of some of these effects can be seen on Scholarpedia (Potter 2009).

1. There is rapid access to conceptual (semantic) information about a stimulus and its associations. Conceptual information about a word or a picture is available within 100–300 ms, as shown by experiments using semantic priming (Neely 1991), including masked priming (Forster & Davis 1984); eye tracking when reading (Rayner 1983; 1992) or looking at pictures (Loftus 1983); measurement of event-related potentials during reading (Kutas & Hillyard 1980; Luck et al. 1996); and target detection in RSVP with letters and digits (Chun & Potter 1995; Sperling et al. 1971), with pictures (Intraub 1981; Meng & Potter 2008; Potter 1976; Potter et al. 2010), or with words (Davenport & Potter 2005; Lawrence 1971b; Meng & Potter 2011; Potter et al. 2002). Conceptually defined targets can be detected in a stream of nontargets presented at rates of 8–10 items per second or faster (Potter et al. 2014), showing that categorical information about a written word or picture is activated and then selected extremely rapidly. The converging evidence shows that semantic or conceptual characteristics of a stimulus have an effect on performance as early as 100 ms after its onset. This time course is too rapid for slower cognitive processes, such as intentional encoding, deliberation, or serial comparison in working memory.

2. New structures can be discovered or built out of the momentarily activated conceptual information, influenced by the observer's task or goal. Evidence for this claim comes from comparing responses to RSVP sentences, scrambled sentences, and lists of unrelated words. It is possible to process the syntactic and conceptual structure in a sentence and, hence, subsequently to recall it, when reading at a rate such as 12 words per second (Forster 1970; Potter 1984; 1993; Potter et al. 1980; 1986). In contrast, when short lists of unrelated words are presented at that rate, only two or three words can be recalled (see also Lawrence 1971a). For sentences, the meaning and plausibility of the sentence, as well as the syntactic structure, are recovered as the sentence is processed. Words that do not fit the syntax or meaning are systematically misperceived (Potter et al. 1993). Syntactic and semantic choices are made online (Potter et al. 1998). Memory for the sentence may be reconstructed from meaning, rather than recalled word for word (Lombardi & Potter 1992; Potter & Lombardi 1990; 1998). Because almost all of the sentences one normally encounters (and all of the experimental sentences) include new combinations of ideas, structure-building is not simply a matter of locating a previously encountered pattern in long-term memory: It involves the creation of a new relationship among existing concepts.

As with words, so with a new pictured scene: Not only must critical objects and the setting be identified, but also the relations among them – the gist of the picture (e.g., Davenport & Potter 2004). Associated long-term memory of visual scenes must be activated to recognize that one is looking at a picnic, or a bride and groom, or a ball game. As C&C suggest, structure-building presumably takes advantage of as much old structure as possible, using any preexisting associations and chunks of information to bind elements.

3. There is rapid forgetting of information that is not structured or that is not selected for further processing. Conceptual information is activated rapidly, but the initial activation is highly unstable and will be deactivated and forgotten within a few hundred milliseconds if it is not incorporated into a structure, consistent with C&C's proposal. As a structure is built – for example, as a sentence is being parsed and interpreted – the resulting interpretation can be held in memory and ultimately stabilized or consolidated in working or long-term memory as a unit, whereas only a small part of an unstructured sequence such as a string of unrelated words or an incoherent picture can be consolidated in the same time period.

Because similar principles seem to apply to language comprehension and to nonlinguistic visual understanding, I have proposed that understanding in both cases is abstractly conceptual rather than fundamentally language-based. For example, pictured objects and their names give equivalent and equally rapid information about meaning (Potter & Faulconer 1975; Potter et al. 1977). Other perceptual senses such as audition and touch also have rapid access to the same conceptual level.

If the CSTM hypothesis is correct, then the Now-or-Never bottleneck occurs after a rich set of associations from long-term memory has enabled conceptual chunking of incoming linguistic or visual information. At that point, the information can be passed through the bottleneck to a more abstract level of discourse or scene understanding. Moreover, the severe limitations of working memory seen for arbitrary lists of letters, numbers, or geometric figures are largely overcome when proactive interference from reuse of a small set of stimuli is eliminated (Endress & Potter 2014a). The desperate speed of processing noted by C&C is not due solely to the limitations of short-term memory, but more generally reflects the pressure to think, see, understand, and act as fast as possible, in order to survive in a predatory world.

Language acquisition is model-based rather than model-free

doi:10.1017/S0140525X1500093X, e89

Felix Hao Wang and Toben H. Mintz
Department of Psychology University of Southern California, 3620 McClintock Ave, Los Angeles, CA 90089-1061.
wang970@usc.edu tmintz@usc.edu
http://dornsife.usc.edu/tobenmintz

Abstract: Christiansen & Chater (C&C) propose that learning language is learning to process language. However, we believe that the general-purpose prediction mechanism they propose is insufficient to account for many phenomena in language acquisition. We argue from theoretical considerations and empirical evidence that many acquisition tasks are model-based, and that different acquisition tasks require different, specialized models.

Given the Chunk-and-Pass processing necessitated by the Now-or-Never bottleneck, Christiansen & Chater (C&C) propose that learning language is learning to *process* language. In C&C's conceptualization, the learning and prediction processes are general, (henceforth, *model-free*), and knowledge used in prediction arises gradually. In discussing the consequences of this scenario, C&C impose a dichotomy between these prediction-based models that are the outcome of learning to process, and learning based on more specialized constraints on how linguistic information is processed (the "child as linguist" approach, henceforth, *model-based*). In this commentary, we leave aside discussion of the Now-or-Never bottleneck per se and focus on C&C's claims about its theoretical consequences for language acquisition.

C&C's perspective provides an interesting framework for guiding research and developing theories. However, we argue that it does not provide significant constraints on the broader theoretical debates with which the field is engaged: in particular, debates about the nature of constraints on learning. Our argument is based on theoretical necessity and empirical evidence. Theoretically, the model-free approach is destined to be misled by surface-level information. Specifically, the general-purpose learning procedure is underspecified with respect to the level of analysis given different problems: Information for particular problems may exist at different levels, and using the wrong level may lead the learner astray. Empirically, when the model-based and model-free approaches are computationally equivalent, the model-free approach simply may not coincide with human performance. To support these claims we cite two cases: one from syntax, and another from word learning.

Many arguments for model-based learning come from phenomena that require a specific level of analysis. An oft-cited example is the constraint on structure-dependence, which specifies that grammatical operations apply to abstract phrasal structures, not linear sequences. It accounts for the fact that the yes/no question in 1(b), following, is the correct form that is related to the declarative 1(a), but question in 1(c) is not.

1. a. The girl who is smiling is happy.
   b. Is the girl who is smiling happy?
   c. *Is the girl who smiling is happy?

The distinction hinges superficially on which *is* is moved to the beginning of the sentence in the question. The grammatical principle that governs this operation is subject-auxiliary inversion; in 1(a), the subject is the complex noun phase [*the girl who is smiling*], so the entire structure inverts with *is*. The model-based argument is that young children's input lacks the positive examples of the complex embedded questions as in 1(b), but rather consists of simpler utterances such as 2(a) and 2(b); without the notion that syntactic operations operate over phrasal structures, why would a learner not conclude from 2(a) and 2(b) to simply front the first *is*?

(2) a. The girl is happy.
    b. Is the girl happy?

Reali and Christiansen's (2005) model-free approach addresses this question. They demonstrated that a model-free learner who is sensitive to local bigram patterns could make the correct predictions about the structure of yes/no questions with complex noun phrases. This demonstration showed how attending to local sequential patterns could achieve the appropriate behavior despite not representing linguistic material at the level of syntactic hierarchies, as called for by model-based accounts. However, it turned out that the success of the model-free mechanism was an artifact of idiosyncrasies in English that had nothing to do with the syntactic structures in question (Kam et al. 2008). This does not rule out the possibility that a different model-free mechanism would succeed at learning the right generalizations, but adopting the view that learning language is learning to process language does not get around the fundamental challenges.

We now turn to an example from our own work in cross-situational word-learning, where model-based and model-free versions of learning mechanisms can both work in principle (Yu et al. 2007). Cross-situational word learning refers to naturalistic situations where learners encounter words under referential ambiguity, and learn the correct word-to-referent mappings via the accumulation of cross situational statistics (Yu & Smith 2007, among others). The associative learning account for how cross-situational statistics are used proposes that learning is model-free, in that passive accumulation of the co-occurrence statistics between words and their possible referents suffices for learning word-referent mappings. In contrast, model-based word-learning accounts posit that, like a mini-linguist, learners have the overarching assumption that words are referential, and learners actively evaluate possible word-referent mappings (e.g, Trueswell et al. 2013; Waxman & Gelman 2009). Although computationally, both accounts are plausible (Yu et al. 2007), we recently carried out an experiment showing the importance of learners' knowledge that words are referential – a model-based, top-down constraint (Wang & Mintz, under revision). We created a cross-situational learning experiment in which there was referential ambiguity within trials, but reliable cross-situational statistical information as to the word-referent mappings. In two different conditions, we held word and referent co-occurrence statistics constant but gave each group of participants different instructions. Both groups were instructed to perform a distractor task, and only one group was also told to learn word meanings. Only the latter group successfully learned the mappings, even though both groups were exposed to the same word-to-referent co-occurrence patterns. Thus, although a model-free learner could succeed in the task, human learners required the notion that words refer for word learning. We take this as evidence that model-based hypothesis testing is required for word learning empirically, even though the model-free version could have worked in principle.

In sum, although the Now-or-Never bottleneck presents interesting challenges for theories of language acquisition, the perspective C&C espouse does not solve problems that model-based approaches do, and empirically, model-free mechanisms do not apply to certain learning situations. Thus, casting acquisition as learning to process across levels of linguistic abstraction does not avoid the theoretical controversies and debates that inhabit the field. It simply shifts the debate from the nature of the constraints on linguistic knowledge acquisition to the nature of the constraints on "learning to process." We do not believe that this shift has substantial theoretical consequences for understanding the nature of the constraints on language learning.

What gets passed in "Chunk-and-Pass" processing? A predictive processing solution to the Now-or-Never bottleneck

doi:10.1017/S0140525X15000941, e90

Sam Wilkinson
Department of Philosophy, Durham University, Durham DH1 3HN, United Kingdom.
sam.wilkinson@durham.ac.uk

Abstract: I agree with the existence, and importance, of the "Now-or-Never" bottleneck. However, there is a far simpler and more parsimonious solution to it. This solution is predictive processing, and the failure to view the solution that this provides fundamentally boils

down to viewing prediction as one aspect of cognition, rather than as its central principle.

The "Now-or-Never" bottleneck presents a real challenge. Rather than the solution presented by Christiansen and Chater (C&C), however, an alternative – one that is both simpler and more economical – is possible. They do allude to the solution I want to present, but they apply it locally rather than globally. The solution in question is prediction. One explanation for why this globally applied solution is not presented is that C&C adopt a traditional view of cognition, according to which inputs come in, get processed, and passed on. Adherence to this view is evidenced in talk of "Chunk-and-Pass" processing. Inputs "come in" and get "chunked" and "passed." Within a predictive processing framework, on the other hand, the direction, if anything, is reversed. "Processing" constitutes inputs having been successfully predicted from the top down. What gets "passed" is prediction error, not some honed incoming product. What does get honed, in light of incoming prediction error, is predictions. Indeed,

> an expected event does not need to be explicitly represented or communicated to higher cortical areas which have processed all of its relevant features prior to its occurrence. (Bubic et al. 2010, p. 10, quoted in Clark 2013)

If we adopt a wholesale predictive processing approach, according to which prediction is not an *aid* to processing as traditionally construed but is rather its fundamental principle, then we overcome the Now-or-Never bottleneck in an evolutionarily, biologically, and computationally plausible way, and end up with all of the same consequences for how to understand language that the authors are at pains to point out.

Firstly, all of the presented solutions to the Now-or-Never bottleneck need not be seen, as C&C present them, as separate, but rather may be viewed as different facets of predictive processing. In other words, (a) "eager recoding and compression" of input, and (b) hierarchical levels of "representation" become consequences of, and not additions to, a need to (c) "deploy all available information predictively." Let me explain why.

Predictive processing is a concrete implementation of a Bayesian strategy. Incoming signals are noisy and ambiguous, and so the brain uses Bayesian inference (it takes into account not only the "fit" of the hypothesis with the input, but also its "prior probability") to settle on one hypothesis rather than another. Thus, a hypothesis can have a really good fit but such a low prior probability that it isn't selected (or it can have a poor fit, but such a high prior probability that it is selected).

This Bayesian strategy gets implemented in the brain as follows. The selection of a hypothesis determines a set of predictions about subsequent inputs, namely, inputs that are compatible with the hypothesis. If the hypothesis does a bad job of predicting inputs, it will be tweaked or abandoned altogether in favour of another hypothesis. These hypotheses are hierarchically arranged, with the hypotheses of one level providing the inputs (prediction error) for the next. "Higher" parts of the hierarchy are, roughly, those parts that are further away from the sensory stimulus. These tend to operate at longer timescales, and at higher levels of abstraction. "Lower" parts of the hierarchy are closer to the sensory stimulus. These tend to be at shorter timescales, and at low levels of abstraction. These, for example, correspond to early stages of visual processing: your brain's early statistically driven attempts to make sense of (predict) noisy inputs.

Predictive processing is time and energy efficient, and it involves compression (more or less "lossy" depending on the occasion). You save on bandwidth by passing on only what is newsworthy. What counts as "newsworthy" is simply what the receiver of the message hasn't already predicted, namely, prediction error. To sum up, then, predictive processing in the brain always involves (1) compression and (2) hierarchical arrangement of hypotheses (which, to use C&C's terminology, can be thought of as "representations").

Now I'd like to gesture towards some of the consequences that predictive processing has for how we think about language. It has all of the same nine consequences that C&C enumerate, but, again, they are (at least for the most part) facets of each other rather than separate consequences. Let me illustrate this with two seemingly distant consequences: the multilevel organization of language (Consequence 1), the nature of what is learned during language acquisition (Consequence 5).

The hierarchical arrangement of hypotheses in predictive processing clearly suggests that language processing has a "multilevel organization." Of course, our processing of other, nonlinguistic, stimuli has a similar organization, but that structure is not so clearly delineated since nonlinguistic worldly items *themselves* lack that structure. As theorists, we tend to use rough-and-ready descriptions in natural language (e.g., "light comes from above" or "This is a face") when talking about neurally encoded hypotheses, but there is nothing intrinsically linguistic about the hypotheses themselves. The same applies when that which is being processed is linguistic (e.g., a communicative utterance or a written sentence). Very schematically put, from the "bottom" to the "top" it goes like this: One's brain can be initially "uncertain" about the shapes seen, or the sounds heard. Having resolved that, it can be uncertain about the letters or phonemes, and then the words used, and then what they mean, and then what is meant by them, or by the whole utterance, and so on. Within this picture, the way in which hypotheses are hierarchically arranged, and priors are updated and can become entrenched, developing a sensitivity to deep, below-the-surface structured statistical regularities in the world, suggests (in line with C&C's suggestion) that acquisition is indeed learning to process. Gone is the need for innate linguistic knowledge (although some of our priors – or our propensity to form them – may be, in some sense, innate). However, learning to process is learning to predict, where this involves being attuned to the dynamic statistical structure of the world of which language, and language users, are an important part.

In conclusion, although I am sympathetic to the spirit of what C&C present, a more wholesale predictive processing account yields very similar consequences but casts things in a different (and arguably more plausible) light.

ACKNOWLEDGMENT
This research was supported by a Wellcome Trust Strategic Award (WT098455MA).

Authors' Response

Squeezing through the Now-or-Never bottleneck: Reconnecting language processing, acquisition, change, and structure

doi:10.1017/S0140525X15001235, e91

Nick Chater[a] and Morten H. Christiansen[b,c,d]
[a]*Behavioural Science Group, Warwick Business School, University of Warwick, Coventry, CV4 7AL, United Kingdom;* [b]*Department of Psychology, Cornell University, Ithaca, NY 14853;* [c]*The Interacting Minds Centre, Aarhus University, 8000 Aarhus C, Denmark;* [d]*Haskins Laboratories, New Haven, CT 06511.*
nick.chater@wbs.ac.uk
christiansen@cornell.edu

Abstract: If human language must be squeezed through a narrow cognitive bottleneck, what are the implications for language processing, acquisition, change, and structure? In our target article, we suggested that the implications are far-reaching and

form the basis of an integrated account of many apparently unconnected aspects of language and language processing, as well as suggesting revision of many existing theoretical accounts. With some exceptions, commentators were generally supportive both of the existence of the bottleneck and its potential implications. Many commentators suggested additional theoretical and linguistic nuances and extensions, links with prior work, and relevant computational and neuroscientific considerations; some argued for related but distinct viewpoints; a few, though, felt traditional perspectives were being abandoned too readily. Our response attempts to build on the many suggestions raised by the commentators and to engage constructively with challenges to our approach.

R1. Introduction

In our target article, we argued that a powerful and general cognitive constraint, the Now-or-Never bottleneck, has far-reaching consequences for both language comprehension and production. This perspective implies that language acquisition and language change proceed construction-by-construction, rather than involving more abrupt, system-wide shifts. We argued, moreover, that the picture that arises from the Now-or-Never bottleneck has implications for the structure of language itself: Syntactic structure is viewed as processing history, thus enforcing a tight link between the psychology of language processing and linguistic theory.

The Now-or-Never bottleneck is a general cognitive constraint that, we suggest, applies to perception, motor control, reasoning, and memory: Unless information is recoded and/or used rapidly, it is subject to severe interference from an onslaught of further information. Our article explores possible implications of the Now-or-Never bottleneck for language: how it is processed and acquired, how languages change, and the structure of language itself. The argument is that the Now-or-Never bottleneck has profound implications in each of these domains: For example, it requires that processing is incremental and predictive, using a Chunk-and-Pass mechanism; that acquisition is item-based; that languages change construction-by-construction; and that there may be an intimate relationship between language structure and processing.

The commentators on our article have provided a rich variety of perspectives and challenges with which to evaluate and potentially to further develop this account. We have grouped our responses to commentators according to key themes that emerged.

The first set of issues, discussed in section R2, concerns the evidence for, and nature of, the bottleneck. Key questions include: Does the psychological and linguistic evidence support (**Ferreira & Christianson**; **Kempson, Chatzikyriakidis, & Cann** [**Kempson et al.**]; **Potter**) or contradict (**Baggio & Vicario**; **Chacón, Momma, & Phillips** [**Chacón et al.**]; **Endress & Katzir**) the existence of the bottleneck? Have we overstated its scope (**Levinson; MacDonald**)? What is its neural basis (**Frank & Fitz**; **Grossberg**; **Honey, Chen, Müsch, & Hasson** [**Honey et al.**]; **Huyck**)? How can the hypothesis be elaborated (**Dumitru**; Potter)? And, if we accept the existence of the Now-or-Never bottleneck, should it be treated as basic, or as arising from more fundamental principles (e.g., **Badets**; **Bicknell, Jaeger, & Tanenhaus** [**Bicknell et al.**]; **Lotem, Kolodny, Halpern, Onnis, & Edelman** [**Lotem et al.**]; **Wilkinson**)?

A second set of issues, which we discuss in section R3, focuses on the empirical and computational viability of the framework for language processing that we derive from the Now-or-Never bottleneck. According to the Chunk-and-Pass framework, language comprehension requires a succession of increasingly abstract chunking operations, and, at each level, chunking must occur as rapidly as possible and the resulting chunks immediately passed to higher levels. The reverse process, where the speaker converts an abstract message into articulatory instructions, is proposed to involve what we term Just-in-Time language production. Key questions include the following: How does how the Chunk-and-Pass framework relate to existing theories of language processing, both in psycholinguistics (**Bicknell et al**.; **Chacón et al.**; **Ferreira & Christianson**; **MacDonald; O'Grady**) and computational linguistics (**Gómez-Rodríguez** and **Huyck**)? How do these proposals relate to experimental data (**Baggio & Vicario**; **Healey, Howes, Hough, & Purver** [**Healey et al.**]), including effects of top-down processing (**Dumitru**; **Healey et al.**; **MacDonald**; **Potter**)? Can our account meet the challenges of interactive dialogue (**Badets**; **Baggio & Vicario**; **Healey et al**.; **Kempson et al**.; **Levinson**)? How far does the Chunk-and-Pass approach apply to sign language (**Emmorey**), and to nonlinguistic domains such as music and action (**Lakshmanan & Graham; Maier & Baldwin**)?

A third set of issues, which we address in section R4, concerns the implications of the Now-or-Never bottleneck and Chunk-and-Pass processing for language acquisition, evolution, and structure. In our target article, we argued that the bottleneck has far-reaching implications for language across multiple timescales, ranging from the duality of patterning observed across languages (roughly, having distinct phonological and lexical levels), the locality of most linguistic regularities, and what we take to be the instance-based nature of language acquisition and language change. Key questions include whether our account provides sufficient constraints to explain language acquisition (**Endress & Katzir**; **Lakshmanan & Graham; Wang & Mintz**) and how it may be developed further (**Lewis & Frank**; **Maier & Baldwin**); and how far the account can explain language change and evolution (**Behme**; **Bergmann, Dale, & Lupyan [Bergmann et al.]**; **Endress & Katzir**; **Lewis & Frank**; **Lotem et al**.). Some commentators explore how this approach can be a productive framework for understanding regularities within and across languages (**Kempson et al.**; **O'Grady**); others believe that further constraints are required (**Chacón et al.**; **Endress & Katzir**; **Medeiros, Piatelli-Palmarini, & Bever** [**Medeiros et al.**]; **Wang & Mintz**).

In the remainder of this response to commentators, we will discuss these three sets of issues in turn before drawing general conclusions and considering directions for future work.

R2. The nature of the Now-or-Never bottleneck

Memory is fleeting: Sensory and linguistic information is subject to severe interference from the continual onslaught of new material. If the input is not used or recoded right

https://doi.org/10.1017/S0140525X1500031X

away, it will be lost forever: This is the Now-or-Never bottleneck.

The fleeting nature of memory is illustrated by the finding that our ability to recall arbitrary sequences of sounds is extraordinarily limited (Warren et al. 1969). Yet we are able to process highly complex, nonarbitrary sequences of linguistic input (and, similarly, musical input and action sequences). We proposed that these observations imply that sensory and linguistic information must be used or recoded into higher-level representations right away, to avoid being lost forever.

What is the origin of the Now-or-Never bottleneck? In our target article, we stressed the importance of interference – new input interferes with existing input, particularly between elements that overlap phonologically or semantically. Such interference has been observed at a wide variety of representational levels in studies of memory for serial order (Brown et al. 2007). Likewise, as noted by **MacDonald,** sentences containing words with overlapping phonological forms and meaning create processing problems (e.g., "The baker that the banker sought bought the house" vs. "The runner that the banker feared bought the house," Acheson & MacDonald 2011; see also Van Dyke & Johns 2012 for a review).

Another possible origin of the bottleneck stems not from interference, but from one or more capacity-limited buffers (discussed by **Levinson**). So, for example, Miller (1956) famously suggested a capacity of 7±2 chunks in short-term memory, an approach enriched and updated by Baddeley and colleagues (e.g., Baddeley 1992; Baddeley & Hitch 1974). More recently, Cowan (2000) argued for a capacity limit of 4±1 items. Our reading of the recent memory literature is that many, and perhaps all, aspects of memory limitations may best be understood in terms of interference rather than capacity-limited buffers, because the same patterns of forgetting and memory errors are observed over many timescales (e.g., Brown et al. 2007). From this perspective, apparent capacity limitations are a side effect of interference, rather than stemming from, for example, a fixed number of "slots" in memory (see also Van Dyke & Johns 2012).

From the point of view we expressed in the target article, the key issue is the limited nature of the bottleneck, whether it stems primarily from interference, capacity limitations, or a combination of the two. Note, in particular, that memory performance depends on the number of chunks involved, and what counts as a chunk depends on prior experience with relevant material. Hence, the same sequence of phonemes may, over experience, be chunked into a series of syllables or words, or into a single multiword chunk (Jones 2012). We stress, too, that interference effects will operate between chunks – that is, chunks are not merely encapsulated units – so that some of the internal structure of chunks will be retained. This is evident, for example, in phonological interference effects in memory for serial order (Burgess & Hitch 1999). Thus, although some commentators (e.g., **Bicknell et al.; MacDonald**) seem to have taken our notion of "lossy compression" as indicating a near total loss of information, we use the term in the standard computer science sense as indicating that not all information is retained**.** More generally, we are able to outline the consequences of the Now-or-Never bottleneck without taking a stand on the exact nature of the underlying memory representations – although, of course, within the general framework developed here, more detailed memory models will allow for more fine-grained predictions about language processing.

Indeed, we suggest that one fruitful direction for research is to explore cognitive models in which processing and memory are not distinct mechanisms. As **Honey et al.** point out, it may be appropriate to see memory as arising from ongoing neural processing activity, rather than as located in distinct stores (see, e.g., Crowder 1993; Kolers & Roediger 1984). From this viewpoint, processing and memory operations should be located in the same brain regions (Hasson et al. 2015). This perspective has also been applied to accounts of individual differences in language processing, modeled using simple recurrent networks (Elman 1990), in which the same connections and weights encode and process linguistic input (MacDonald & Christiansen 2002). This type of model captures the relationship between language processing and short-term memory performance, without any functionally distinct working memory (by contrast with, for example, production system models such as Just and Carpenter's [1992] CC-READER). As we shall discuss further, in this integrated perspective on memory and processing it is not possible to modify memory capacity independently of processing operations (Christiansen & Chater 2015; 2016; MacDonald & Christiansen 2002). Thus, memory capacity is not a free parameter that can be independently selected for by natural selection (see our discussion of **Lotem et al.**).

Honey et al. underscore our claim that the Now-or-Never bottleneck implies longer integration timescales for more abstract levels of representation. They substantiate this view with evidence from functional magnetic resonance imaging (fMRI) and intracranial recordings, countering **Vicario & Baggio**'s concern that our multilevel representational approach lacks neural foundations. According to **Honey et al.**, incoming information is continually integrated with prior information – yet once integration has occurred, the resulting interpretation and knowledge updating becomes entrenched and difficult to revise (**Ferreira & Christianson**). Consistent with such interpretative entrenchment, Tylén et al. (2015) found that when a narrative had a coherent storyline, then incidental facts tended to be forgotten if they were not central to the plot. However, when the storyline was jumbled, there was a greater recall of incidental semantic facts, presumably because integration was not possible. Importantly, an fMRI version of the same experiment yielded activation of the same cortical hierarchies, from lower-level sensory circuits to higher-level cognitive areas, as noted by **Honey et al.** (and discussed in the target article).

R2.1. Challenges to the Now-or-Never bottleneck

Several commentators question the severity of the Now-or-Never bottleneck. Some of these concerns, however, focus on consequences that do not follow from the bottleneck. For example, as illustrated by SF's spectacular memory for sequences of numbers chunked by running times, chunking low-level material facilitates memory for that material. More broadly, low-level information is remembered only to the extent that it has been processed. So the Now-or-Never bottleneck does not imply complete amnesia for past low-level sensory or linguistic information – people can, after all, remember tunes and poems by heart. What they cannot do is recall unprocessed sequences of noises or letters, which they are unable to chunk in light of prior experience. So, although we can remember new words in our language, recalling a complex sound-pattern from a foreign language (e.g., for speakers of English, a word or phrase in Khoisan) will be very difficult. Hence, **Endress & Katzir**'s claim that children can learn a word from a single encounter does not challenge the Now-or-Never bottleneck (the notion of fast-mapping has, though, been questioned in some recent studies, e.g., Horst & Samuelson 2008; McMurray et al. 2012).

We stress also (pace **Endress & Katzir**) that the bottleneck applies equally to explicit and so-called implicit memory (i.e., with or without awareness), if indeed such a distinction can be defended (e.g., Shanks & St. John 1994). Our claim is that memory is dependent on processing, and this remains true irrespective of whether memory is assessed through explicit or implicit measures. For example, many psychology undergraduates will have been exposed to the hard-to-see "Dalmatian" (see, e.g., Gregory 2005). Famously, once one can see the pattern as a Dalmatian, the Dalmatian interpretation is typically available many years later (e.g., to help segment the image, an implicit measure of memory) – and the image will immediately be recognized as familiar and as a Dalmatian (explicit measures). But, of course, people who have *not* successfully found the Dalmatian "gestalt" will, of course, not remember that they have seen this specific pattern of black-and-white marks on a piece of paper or a computer screen many years before. In short, an image is memorable only to the extent that it has been successfully processed. This explains why prior exposure to an image will assist the processing of later copies of the same image, because such exposure helps create a "gist" that can be reused, allowing for cumulative learning effects over multiple exposures (see, for example, Endress & Potter 2014a).

Similarly, **Bicknell et al.** stress that perceptual data are not necessarily immediately forgotten – and we agree. The Now-or-Never bottleneck implies that perceptual or linguistic data that cannot be successfully processed into higher-level representations will suffer severe interference from subsequent material. But where that data can be recoded successfully, more low-level details may be retained because they are embedded within a richer memory structure, thus countering interference from subsequent material to some extent. Nonetheless, we would anticipate that recalling such low-level details is likely to be cognitively effortful, although some details may be retained when crucial to the task at hand.

The influence of task constraints is illustrated by a study that **Bicknell et al.** describe, by Connine et al. (1991), employing a phoneme labeling task. Participants indicate which of two sounds they heard at the beginning of the third word in a sentence, and are instructed to use any available information from the sentence to make their response. The stimuli were ambiguous between a voiced and unvoiced initial consonant, yielding a blend of *dent* and *tent*, followed by a disambiguating context: "When the __ in the fender/forest …" Therefore, while encoding the word, participants are explicitly instructed to pay attention to the details of the first phoneme. Accordingly, some low-level information is likely to be retained over a short period. **Bicknell et al.** report their own study indicating slightly longer periods of retention of phonemic information, over six syllables, when participants are refrained from responding until the end of the sentence. But this hardly changes the broad message that the "raw" sensory input is rapidly lost, presumably through interference, although some limited information can, as we would predict, be retained through being encoded in larger units (e.g., through retaining a memory of the degree of "ambiguousness" of the word *dent* or *tent*).

Note that Connine et al. (1991) highlighted task-specific effects as a possible driver of their results: "One major issue left unresolved by the present research is the degree to which delayed commitment is subject to strategic factors introduced by task specific demands" (p. 246). With this in mind, we can only agree with **Bicknell et al.** (and also **Ferreira & Christianson**), that memory (including memory for low-level information encoded into higher-level units) can be used strategically in the service of task goals (e.g., Anderson & Milson 1989; Anderson & Schooler 1991). Indeed, as noted by **Potter**, our framework seems naturally compatible with allowing newly built structures to be "influenced by the observer's task or goal" (para. 4). Moreover, it is possible that such strategic task-related effects may appropriately be modeled by bounded rational analysis, as **Bicknell et al.** suggest. Similarly, we suggest that this approach to modeling task-specific effects is compatible with the "good enough" processing model described by **Ferreira & Christianson** (Ferreira & Swets 2002). We see the Now-or-Never viewpoint as providing a framework within which "boundedly rational" and "good enough" models may fruitfully be integrated.

Whereas **Endress & Katzir** and **Bicknell et al.** stress, and we agree, that not all low-level information is lost immediately (though it will be lost if it cannot be processed into higher-level units), **Baggio & Vicario** argue that the processing of sequential material such as language should not be viewed as a race against time at all. They do not deny the existence of the Now-or-Never bottleneck, but suggest that the brain has a number of mechanisms through which the effects of the bottleneck can be countered, including inference, pragmatics, and skills associated with literacy.

Yet we are not sure that **Baggio & Vicario**'s suggestions change the picture substantially. Focusing for now on reading, even though we can always refixate a word that we have missed or misread while reading, becoming a *fluent* reader requires overcoming a reading-based analogue of the Now-or-Never bottleneck for three reasons: (1) memory for visual information is short-lived (60–70 ms; Pashler 1998); (2) visual input is taken in at a fast rate during normal reading (about 200 words per minute; Legge et al. 1985); and (3) memory for visual sequences is limited (to about four items; Luck & Vogel 1997). Because memory for what has just been read is short-lived and subject to rapid interference, we suggest that readers must perform chunking operations on text input as quickly as possible in order to read fluently. Indeed, individual differences in chunking ability predict self-paced reading performance (McCauley & Christiansen 2015b).

R2.2. *Is the Now-or-Never bottleneck a side effect of a deeper constraint?*

In our target article, we argued that the Now-or-Never bottleneck provides a powerful motivation for online

https://doi.org/10.1017/S0140525X1500031X

prediction in language processing, and in cognition more broadly. Given the underspecified nature of the sensory and linguistic input, predictive information is required to analyze new input as rapidly as possible, before it is obliterated by the onslaught of further material. Similarly, prediction is required for online learning, in which the disparity between predictions and sensory data can immediately be used to drive learning. According to the Now-or-Never bottleneck, unless the disparity between predictions and input is computed and exploited right away, the sensory information will be lost, and with it, the opportunity for learning.

By contrast, **Wilkinson** and **Badets** argue, from different perspectives, that online prediction should not be seen as helping to deal with the Now-or-Never bottleneck, but as the central engine of cognition. There might not be substantial disagreement here, however. A cognitive theory based on prediction still has to specify at which point the error between prediction and sensory or linguistic input is assessed, to guide action and shape learning. The Now-or-Never bottleneck requires that prediction error is calculated and used to drive learning *online*: If the disparity between prediction and sensory input is not calculated right away, then sensory input will be lost. Notice that, by contrast, many prediction-based learning methods do not learn online. For example, the parameters in connectionist networks or Bayesian models are often adapted to provide the best fit to the whole "batch" of available data, which typically involves storing and resampling these data throughout learning. Indeed, the requirement for learning to be online is very strong: Online learning algorithms face the danger of so-called "catastrophic interference" where learning new items damages memories of old items (e.g., French 1999).

Such catastrophic interference can, as we note, be avoided by using item-based learning models, so that learning from experience involves not refitting the parameters of a model (e.g., a stochastic-phrase structure grammar, or the like), but continually adding to, and then generalizing from, a database of stored exemplars (e.g., an inventory of constructions). Needless to say, sensory experience must be encoded in an abstract form (rather than purely as "raw" acoustic or visual input) to reduce interference with other stored items. In our target article, we argued that item-based learning is a plausible model for language acquisition (Tomasello 2003); and the need for online predictive learning, imposed by the Now-or-Never bottleneck, may favor item-based learning throughout perception and cognition more broadly (e.g., Kolodner 1993; Poggio & Edelman 1990).

From a different theoretical viewpoint, **Lotem et al.** raise the possibility that the Now-or-Never bottleneck should not necessarily be viewed as a fixed constraint on cognitive machinery, but may instead itself be an adaptation of our learning mechanisms, driven by natural selection (see also **Endress & Katzir**'s discussion of Major & Tank 2004). The argument of our target article focused on the nature and implications of the Now-or-Never bottleneck, but the question of the origins of the bottleneck is, of course, of great interest. Lotem et al. argue that the bottleneck has been adapted through natural selection to optimize the brain's ability to learn. They note that a wide variety of evidence shows that memory performance varies between individuals, and is to some extent heritable. They interpret this variation to suggest that the size of the bottleneck is itself variable – and that this size can potentially be selected for. This viewpoint would, for example, be compatible with theories of memory, mentioned earlier, in which memory consists of one or more capacity-limited buffers (e.g., Baddeley 1992) – and hence where the capacity limit can be adjusted (as is appropriate, for example, in thinking about computer RAM memory or hard disk capacity).

We suggest, by contrast, that human memory and processing are fundamentally integrated and that the Now-or-Never bottleneck arises from interference effects that are unavoidable, given that the same neural and computational machinery is used for successive, and potentially strongly overlapping, and hence interfering, inputs (e.g., Brown et al. 2007; Hintzman 1988; Murdock 1983). From this standpoint, the Now-or-Never bottleneck is not usefully characterized as having a variable size, which is subject to independent variation and selection. Rather, the bottleneck emerges from the computational architecture of the brain; and variation in memory performance depends on the effectiveness of Chunk-and-Pass mechanisms to mitigate its impact. So SF's ability to encode streams of digits as running times indicates not a particularly wide "bottleneck" but rather a particularly efficient recoding strategy (Ericsson et al. 1980). Expert chess players are able to recall positions of real chess games by encoding them using a rich set of "chunks" from prior games (yet even top chess players have no memory advantage for "nonsense" chess positions and neither do they have significantly above-average general visuospatial abilities; Simon & Chase 1973; Waters et al. 2002). Similarly, we suggest that individual differences in the efficacy of language processing operations will depend on being able to draw on a rich set of prior linguistic experiences to efficiently recode linguistic input (Christiansen & Chater 2016; Jones 2012; MacDonald & Christiansen 2002).

From this standpoint, it is not appropriate to see the size of the Now-or-Never bottleneck as a free parameter that can be optimized through selection and variation, as embodied in **Lotem et al.**'s variable "time-window" in their computer simulations (e.g., Kolodny et al. 2014; 2015a; 2015b). Note, too, that the "window" in this model is large (e.g., 50–300 items) compared with buffers typically postulated in the study of human memory (Baddeley 1992), so its psychological status is not clear either.

In any case, to the extent that **Lotem et al.** see the Now-or-Never bottleneck for language as shaped specifically to the linguistic environment, their approach appears to depend on the structure of language being exogenously fixed, to provide a stable target for adaption of the Now-or-Never bottleneck. But language is not given exogenously; it is shaped by generations of rapid cultural evolution to fit with, among other things, the learning and processing biases of the brain, including the Now-or-Never bottleneck. We have suggested elsewhere that language is shaped by the brain, rather than the brain being shaped by language (Christiansen & Chater 2008). So linguistic regularities will arise from, among other things, the Now-or-Never bottleneck; and hence the Now-or-Never bottleneck is prior to, rather than an adaptation for, the structure of language.

R2.3. Neural plausibility?

How might the Now-or-Never bottleneck be implemented neurally? **Grossberg** argues that many key aspects of our approach are already embodied in existing computational models of neural function created by his research team, and, in particular, in the notions of Item-Order-Rank (IOR) working memory and by a learning and chunking mechanism called the Masking Field (MF) (for a less detailed discussion along somewhat similar lines, see **Huyck**). We are sympathetic with the proposal that Chunk-and-Pass processing, and, more broadly, the serial character of high-level thought (e.g., Pashler 1998), derive from the basic operating principles of the brain, as carrying out a sequence of parallel constraint satisfaction processes. The data outlined by **Honey et al.** suggest that each computational step (e.g., chunking and recoding linguistic input) may work in parallel across large areas of the brain, so that multiple processes at the same representational level cannot be carried out simultaneously, and hence language processing, and high-level thought more generally, is sequential (e.g., Rumelhart et al. 1986b). If this is right, then the Now-or-Never bottleneck may be a side effect of the basic principles of neural computation, rather than a free parameter that can be readily modified by natural selection (contra **Lotem et al.**).

Frank & Fitz offer a very different perspective on brain function inspired by the processing properties of the cerebellum (Fitz 2011). They question the severity of the bottleneck in light of computational results from what they term "reservoir computing," in which an untrained neural network projects a temporal input stream into a high dimensional space; a second network is trained to read off information from the "reservoir." They report simulations that they take to show that the network can reliably recover complex sequential input after long delays. Interesting as these results are, they seem to provide a poor fit with the large literatures on both human memory limitations and restrictions on language processing. It is thus unclear whether such networks would predict the aspects of language processing discussed in our target article, and by other commentators (e.g., **Ferreira & Christianson**; **Grossberg**; **Kempson et al**.).

R3. The case for Chunk-and-Pass language processing

The Now-or-Never bottleneck is a fundamental constraint on memory that the language system deals with by Chunk-and-Pass comprehension and Just-in-Time production. The very phrase *Chunk-and-Pass* has, to some commentators, suggested a link with the Sausage Machine parsing model of Frazier and Fodor (1978). This has led some commentators to level concerns at the Chunk-and-Pass approach that are more appropriately directed at the Sausage Machine (**Bicknell et al.**; **Chacón et al**.; **Ferreira & Christianson**; **Healy et al.**; **MacDonald**). According to the Sausage Machine model, a preliminary syntactic analysis is created within a window of about six words and then shunted off as a packet (like successive sausages coming out of a real sausage machine) to a second stage that completes the syntactic parsing. But although the Sausage Machine has a packet-by-packet character, it differs fundamentally from the Chunk-and-Pass model along at least three key dimensions. First, the Chunk-and-Pass account operates at a variety of representational levels, using units that have been acquired by item-based learning – so Chunk-and-Pass processing is not restricted to the syntactic units used in parsing. Second, while the operation of the Sausage Machine is informationally encapsulated from semantic and pragmatic factors, the Chunk-and-Pass model assumes that all sources of information, from low-level sensory input to pragmatics and world knowledge are brought to bear online to create and recode chunks at all levels of analysis. Thus, we stress that the Chunk-and-Pass view includes top-down influences (see **Dumitru**), rather than operating purely bottom-up in a modular fashion (a concern raised by **Healey et al.**, **Lotem et al.**, and **MacDonald**).

The third difference to note is that, unlike the Sausage Machine, which postulates cognitively decisive breakpoints at the boundaries between "sausages" (i.e., phrase structure created by the parser), the Chunk-and-Pass viewpoint allows links (and interference) between items that are not grouped within the same chunk (e.g., words which are not in the same phrase or clause). But the strength of such links will reduce rapidly, in a graded fashion, as the "distance" between items increases, as would be predicted by the memory interference processes that we take to underlie the Now-or-Never bottleneck. Chunk-and-Pass processing implies a strong bias toward local structure in language, but is entirely compatible with the existence of some nonlocal dependencies (see **Medeiros et al.**; **Healey et al.**; **Levinson**). We emphasize that the Now-or-Never bottleneck explains the remarkably, though not completely, local structure of language (as noted by **Kempson et al.**), with its hierarchy of levels of representations largely corresponding to local sequences of linguistic material. As we outlined in our target article, this contrasts with the batch-coded communication signals used in engineering and computer science and which are optimal within an information theory framework (Cover & Thomas 2006).

Turning to production, we argued that the Now-or-Never bottleneck implies that once detailed low-level production instructions have been assembled, they must be executed right away, or they will be obliterated by interference from the oncoming stream of later instructions: This is Just-in-Time production. Some commentators (**Chacón et al**.; **Ferreira & Christianson**; **MacDonald**) have taken Just-in-Time production to imply so-called *radical incrementality*, in which phonological words are articulated immediately in the absence of any planning ahead. They have rightly noted that such radical incrementality is inconsistent with evidence of task-related effects on production. For example, Ferreira and Swets (2002) showed that participants plan ahead when producing utterances involving the results of arithmetic calculations. Indeed, speakers appear to plan beyond the immediate phonological word, but likely no more than a clause in advance (e.g., Bock & Cutting 1992). We want to stress, though, that just as comprehension at the discourse level takes place over a relatively long timescale, so does planning at the discourse or conceptual level in production. This is because chunks at the discourse level have a longer duration than articulatory chunks (see **Honey et al**.). Whereas planning at the level of the phonological word may be quite short in temporal scope, planning will extend further ahead at the level of multi-word combinations (what might traditionally be called the "grammatical level"), and even longer at the conceptual/discourse level (e.g., Smith & Wheeldon 2004). Thus, the evidence that **Chacón et al**. discuss in this regard (e.g., Lee et al. 2013; Smith & Wheeldon 1999) is not inconsistent with Just-in-Time production.

Nonetheless, it is important to note that people do interleave planning and articulation processes when producing utterances under time pressure (Ferreira & Swets 2002). Given the speed of turn-taking (e.g., as noted by **Levinson**), such time pressures may be the norm in normal conversations, limiting the amount of advance planning possible. This is reflected by the patterns of disfluencies observed in production, indicative of brief planning ahead at the clausal level (e.g., Ferreira 1993; Holmes 1988). We see this limited planning ahead as compatible with Just-in-Time production, whereby production is limited to just a few chunks ahead for a given level of representation. Crucially, as noted in the target article, such chunks may involve multi-word sequences, which are articulated as units rather than as a chain of individual words (Arnon & Cohen Priva 2013; Bybee & Scheibman 1999). This allows speakers to plan ahead to some degree when this is required by task demands, though our account suggests that such planning would be limited to a few chunks within a given level of linguistic representation.[1] Future work is needed to further develop this perspective on production in more detail.

The Now-or-Never bottleneck, and the processing consequence that follows from it, applies across modalities. Just-in-Time mechanisms of motor planning will be used whether the language output is speech or sign. Similarly, Chunk-and-Pass processing will be required to deal with the onslaught of linguistic material, whether that material is spoken or signed. However, as **Emmorey** points out, the detailed implications of the Now-or-Never bottleneck may differ between modalities. She notes that the speed of the speech articulators, in contrast to manual gestures, contributes to a rapid serial information transmission strategy being adopted for speech, while greater parallelism is used in signed communication. So, for example, she points out that while spoken words consist of a sequence of phonemes, signed words typically correspond to multiple sign elements (spatial locations and temporally defined movements). Similarly, **Emmorey** notes that spoken languages deploy affixes temporally before or after the modified item, whereas morphology is usually signaled simultaneously in signed languages. We suggest that differences in sequential learning abilities in the auditory and visual domains may also be important: The perceptual system readily finds sequential structure in auditory material in comparison with visual material (Conway & Christiansen 2005; 2009; Frost et al. 2015); conversely, the visual modality readily creates visual gestalts to encode simultaneously presented movements in one or more effectors (compare Bregman 1990; Wagemans et al. 2012 – see also **Dumitru**).

We have presented Chunk-and-Pass processing as a general solution to the constraints imposed by the Now-or-Never bottleneck. We appreciate the call for proposals concerning how such a framework might be elaborated, for example, with respect to the nature of discourse representations (**Chacón et al**.), developmental underpinnings (**Maier & Baldwin**), and the nature of processing and representational levels used in Chunk-and-Pass processing (**Levinson**). In this regard, we are encouraged by the detailed examples provided by **O'Grady**, illustrating how an account of this kind can be elaborated to deal with linguistically complex phenomena such as the *wanna* contraction, and his more detailed processing-based explanations of central linguistic phenomena including binding and quantification across languages (O'Grady 2013; 2015a).

R3.1. Chunk-and-Pass processing and semantic interpretation

Several commentators (e.g., **Chacón et al.**; **Ferreira & Christianson**; **Frank & Fitz**; **Honey et al.**) rightly stressed that a Chunk-and-Pass model of comprehension must integrate current input with past input to produce a semantic interpretation that can interface with general knowledge. The final stages of such interpretation, therefore, has to do more than merely chunk linguistic input: Inferential processes will be required to resolve anaphora and other referring expressions (Garnham & Oakhill 1985), first, to bridge between current input and prior linguistic and nonlinguistic context (Clark 1975) and, second, to update beliefs about the speaker's intentions (e.g., Levinson 2000) and about the environment (Gärdenfors & Rott 1995). We argue, though, that the Now-or-Never bottleneck implies that processes of semantic and pragmatic interpretation and belief revision must occur right away, or the opportunity for such interpretation is lost – that is, belief updating, as well as semantic interpretation narrowly construed, is incremental.

The phenomenon of rapid semantic analysis and belief updating is exemplified, for example, in the celebrated demonstration that so-called "close" shadowers (i.e., people able to repeat speech input at a latency of 250–300 ms or even less) are sensitive not only to syntactic structure, but also to semantic interpretation (Marslen-Wilson 1987). Or consider a very different paradigm, in which a potentially baffling paragraph of text is read either with or without an explanatory title or context (Bransford & Johnson 1972). In the absence of the explanatory context, memory for the passage is poor. This means that, even if the clarifying context is provided later, the cognitive system is unable to make much sense of the passage in retrospect. Unless it is understood at the time, the details will be too poorly remembered to be reinterpreted successfully. **Potter** offers a possible framework for such interpretations in terms of what she calls *conceptual short term memory* (CSTM): activations of long-term memory associated with active stimuli and thoughts (Potter 2012). Importantly, she notes that "rich but unselective associations arise quickly but last only long enough for selective pattern recognition – chunking, in C&C's terms." Thus, CSTM may allow the rapid integration of conceptual information, influenced by task demands and goals, which will facilitate incremental interpretation through Chunk-and-Pass processing. It also enables the building of the kinds of online semantic and discourse-related representations called for by **Chacón et al**. CSTM may further provide a nonsyntactic basis for the successful processing of nonlocal dependencies (an issue raised by **Medeiros et al.**; **Healy et al.**; **Levinson**).

As noted by **Ferreira & Christianson**, however, the resulting interpretations may often be rather shallow and underspecified (e.g., Ferreira et al. 2002), with the depth and focus of such "good-enough" representations being affected by task demands (Swets et al. 2008). This can lead to systematic misinterpretations, such as when participants in a study by Christianson et al. (2001) tended to derive the incorrect interpretation that Mary bathed the baby from the temporally ambiguous sentence *While Mary bathed the baby played in the crib*. The difficulty of backtracking appears to be a key contributing factor in such misinterpretations because the language system has limited opportunity for going back to correctly reinterpret previous input (Slattery et al. 2013).

Our formulation of Chunk and-Pass processing emphasized the importance both of bottom-up and top-down processes. Indeed, we stressed that the pressure rapidly to chunk locally ambiguous speech input provides a powerful reason to harness the full range of relevant informational sources as rapidly as possible. Integrating these sources of information will best predict what input is likely, so that it can be chunked and passed to higher levels of representation as quickly as possible. Parallel models of word recognition (e.g., Marslen-Wilson 1987; McClelland & Elman 1986) nicely exemplify this viewpoint: Acoustic, lexical, semantic, and pragmatic information is brought to bear in real time in order to identify words rapidly, and indeed, the "recognition point" for a word is thereby often reached well before the end of the word. We are therefore highly sympathetic to the call from some commentators to highlight the importance of top-down processing (**Dumitru**; **Healey et al.**; **MacDonald**; **Potter**). Note, though, that top-down expectations from prior context or world knowledge may in some cases also produce misinterpretations, as when study participants misinterpret the sentence *The man bit the dog* as if it was the dog that did the biting (Ferreira 2003; see also Potter). In such cases, higher-level expectations can run ahead of the linguistic input (as emphasized by Dumitru), potentially leading unanticipated linguistic input to be misinterpreted.

R3.2. The importance of dialogue

Since the turn of the millennium, researchers have become increasingly aware of how viewing language processing in the context of dialogue, rather than considering the isolated production and comprehension of utterances, can have profound implications for the psychology of language (e.g., Pickering & Garrod 2004). We therefore agree with the various commentators who emphasize the centrality of dialogue in assessing the implications of Chunk-and-Pass processing (**Badets**; **Baggio & Vicario**; **Healey et al.**; **Kempson et al.**; **Levinson**)

Kempson et al. and **Levinson** note the theoretical challenges arising from real-world dialogue in which there is often rapid turn-taking, in which partners may, for example, complete each other's sentences. This possibility seems compatible with the idea that production and comprehension processes are closely intertwined (see Pickering & Garrod 2007; 2013a for reviews). For example, if comprehension involves an "analysis-by-synthesis" reconstruction of the process by which the utterance was produced, then the comprehension process itself creates a representation that can be used to continue the sentence. This is particularly natural within the present framework: The same inventory of chunks can be deployed both by comprehension and production processes. Indeed, a single-system model for processing and producing language in the spirit of the Chunk-and-Pass framework is exemplified in a recent computational model (Chater et al. 2016; McCauley & Christiansen 2013).

The remarkable speed and rapid turn-taking of interactive dialogue (**Levinson**) presents a considerable challenge to the cognitive system – although the fact that we are able to understand time-compressed speech, which is several times faster than the normal speech-rate, strongly suggests that the limiting factor is rate of articulation rather than comprehension (Pallier et al. 1998). As Levinson points out, the ability of participants to turn-take with latencies of a fraction of a second implies that significant speech planning has occurred before the other speaker has finished; and prediction is, of course, required to plan an appropriate response before a partner's utterance is completed. We suggest, though, that online prediction and incremental interpretation are required, even when such constraints are relaxed: Unless the speech signal is not recoded right away, it will be obliterated by interference from later material. Thus, the ability to anticipate later material at higher levels of representation (e.g., at a discourse level) requires rapid online analysis at lower levels of representations, so that the output of such analysis can be fed into predictive mechanisms.

Levinson points out that dialogue can include repetitions that span fairly long stretches of material – for example, repeating a query after some intervening comments. Note that this does not provide a problem for the present approach, as long as that material has been recoded into a more abstract form. The Now-or-Never bottleneck implies that maintaining a representation of an acoustic stream, a string of arbitrary acoustic instructions, or a string of phonemes, will be impossible, but such information can be recalled, at least with much greater accuracy, when encoded into a hierarchy of larger units. Hence, this type of example is entirely compatible with the Now-or-Never bottleneck.

Rapid interactive dialogue often goes wrong: We continually self-correct, or correct each other (**Healey et al**.). The ability to do this provides further evidence for an incremental Chunk-and-Pass model of language comprehension – the incrementally created chunked representation can then be revised and reformulated, as appropriate, by making fairly local modifications to the chunk structure. We agree with Healey et al., **Kempson et al.**, and **Baggio & Vicario** that the ability to switch and repair is a source of strong constraints on theories of processing and, to the extent that the structure of processing matches linguistic structure (Pulman 1985), by extension to syntactic theory, arguably favoring approaches such as dynamic syntax (Healey et al.; Kempson et al.) and construction grammar (**O'Grady**).

R3.3. Meeting the computational challenges of natural dialogue

Huyck and **Gómez-Rodríguez** highlight the parallel between our framework and the computational solutions used by engineers to implement real-time natural language processing (NLP). Of particular interest is the observation that such artificial systems are not subject to whatever hardware limitations the human brain may be working under but nonetheless end up employing the same solution.

One possibility is that the limitations of the brain are actually shaped by the problem, as **Lotem et al.** suggest. Another possibility is that NLP systems are dealing with human language, which is adapted to the Now-or-Never bottleneck (as discussed further subsequently), and therefore has a very local structure. Artificial NLP systems must process language that embodies these human constraints – and, to replicate natural human conversational interaction successfully, they may need to embody those very same constraints. Importantly, Gómez-Rodríguez argues in favor of the latter, because computers are not limited by memory to the same degree as humans. But these systems face the same problems as humans when interacting with another person: Language needs to be processed here-and-now so that responses can be made within a reasonably short amount of time (e.g., there are about 200 ms between turns in human conversation; Stivers et al. 2009). For example, so-called *chatbots* (e.g., Wallace 2005) receive human language input (in text or voice) and produce language output (in text or synthesized voice) in real time. Because no one is willing to wait even a few seconds for a response, and because we expect responses even to our half-formed, fragmentary utterances, these chatbots need to process language in the here-and-now, just like people. The strategies they employ to do this are revealing. As Gómez-Rodríguez notes, these artificial systems essentially implement the same Chunk-and-Pass processing solutions that we discussed in our target article: incremental processing, multiple levels of linguistic structure, predictive language processing, acquisition as learning to process, local learning, and online learning to predict. We see this convergence as further evidence in favor of the feasibility of Chunk-and-Pass processing as a solution to the pressures from the Now-or-Never bottleneck.

R3.4. Chunk-and-Pass in nonlinguistic domains?

If, as we have argued, the Now-or-Never bottleneck is a domain-general constraint on memory, then we should expect Chunk-and-Pass processing to apply not just to language comprehension, but also to a wide range of perceptual domains. Similarly, it seems likely that the principles of Just-in-Time production may be extended beyond speech production to action planning and motor control in general (**MacDonald**; **Maier & Baldwin**). Indeed, as we noted in our target article, planning one's own actions and perceiving the actions of others appear to involve the creation of multilevel representational hierarchies, and we conjecture that Chunk-and-Pass and Just-in-Time processes will operate in these domains (Botvinick 2008; MacKay 1987).

In the target article, we speculated that music might be a domain in which Chunk-and-Pass and Just-in-Time mechanisms might be required to process a highly complex and hierarchically structured auditory sequence, of comparable complexity to human language. **Lakshmanan & Graham** appear skeptical, apparently on the grounds that music and language differ in a number of regards (e.g., music does not have a semantics; or music does not involve turn-taking – although improvised styles of music including jazz and Indian classical music do frequently involve rapid turn-taking between players). But these concerns seem beside the point when considering the key question at issue: Music and language appear to share a hierarchical organization, and both can be processed highly effectively despite the severe pressure of the Now-or-Never bottleneck, and far better than humans can process unstructured sequences of sounds (Warren et al. 1969). We therefore believe that the Chunk-and-Pass framework might fruitfully be applied in future studies of music and other aspects of perception and action.

R4. Implications for language acquisition, evolution, and structure

The Now-or-Never bottleneck and its processing consequences (Chunk-and-Pass comprehension and Just-in-Time production) have, we argue, implications for how language is acquired, how it changes and evolves over time, and for how we should think about the structure of language. The commentators have raised important issues in each of these domains.

R4.1. Implications for language acquisition

In our target article, we argued that the Now-or-Never bottleneck implies that language learning is online: Learning must occur as processing unfolds, or the linguistic material will be obliterated by later input, and learning will not be possible. For parameter-based models of language, this can be difficult – learning seems to require surveying a large corpus of linguistic input to "check" the appropriateness of parameter settings. But if learning must occur online, without the ability to retain, and review, a large verbatim corpus, then parameter setting is difficult (witness the difficulties of making "trigger" models in the principles and parameters tradition [Gibson & Wexler 1994] learn successfully). An item-based model of language acquisition provides an alternative conception of online learning – new constructions can be added to the learner's model of the language one-by-one. Such item-based models also fit well with empirical evidence on child language acquisition (e.g., Tomasello 2003), as well as with item-based models of linguistic structure, such as construction grammar (e.g., Goldberg 2006).

Wang & Mintz characterize this view as a "model-free" approach to language, contrasting it with a "model-based" perspective incorporating linguistic constraints. They suggest that because our approach involves domain-general learning, it is unable to capture many of the constraints on linguistic structure (such as the apparent sensitivity to structure, rather than linear order, in question formation[2]). This suggestion incorrectly presupposes that domain-general learning necessarily has to be constraint-free. All too often it is implicitly assumed that either language acquisition is guided by (presumably innate) linguistic constraints or that there can be no constraints at all. But this is, of course, a false dichotomy. Indeed, we have argued elsewhere (e.g., Christiansen & Chater 2008; 2016) that there are substantial constraints on language, deriving from a wide variety of perceptual, communicative, and cognitive factors (we discuss this point subsequently).

Of these constraints, the Now-or-Never bottleneck is of particular importance, but it is not the only one – and so we agree with **Wang & Mintz** that many additional constraints will shape both language itself and our ability to acquire it.

The stronger the confluence of multiple cognitive and other biases that shape language, the *easier* language will be to learn, because each generation of learners simply have to "follow in the footsteps" of past learners. Language has been shaped by many generations of cultural evolution to fit with our learning and processing biases as well as possible. Thus, considering language as culturally evolved to be easy to learn and process helps explain why language is learned so readily (e.g., Chater & Christiansen 2010). This viewpoint fits nicely with the iterative learning studies (Kirby et al. 2008; Reali & Griffiths 2009) described by **Lewis & Frank**, and their emphasis on language as emerging from the interaction of cognitive and communicative pressures.

Compatible with this viewpoint, and in contrast to **Lakshmanan & Graham**'s suggestion of acquisition guided by (unspecified) "innate grammatical mechanisms," Kelly et al. (2014), in their survey of the acquisition of polysynthetic languages, highlighted the importance of several properties of the input in explaining children's patterns of acquisition. For example, children learning Quiché Mayan and Mohawk initially produce the most perceptually prominent units of speech, and such perceptual salience also appears to play a role in the acquisition of Navajo, Inuktitut, Quechua, and Tzeltal (Lakshmanan & Graham's stipulations notwithstanding). Another property of the input – frequency – has been shown by Xanthos et al. (2012) to be key to the acquisition of complex morphology across a typologically diverse set of languages: French, Dutch, German (weakly inflecting languages); Russian, Croatian, and Greek (strongly inflecting languages); and Turkish, Finnish, and Yucatec Maya (agglutinating languages). Using corpus analyses, Xanthos et al. (2012) found that the frequency of different morphological patterns predicted the speed of acquisition of morphology, consistent with usage-based suggestions regarding the importance of variation in the input for learning complex patterns in language (e.g., Brodsky et al. 2007) as well as for distributional learning more generally (e.g., Gómez 2002).

Lakshmanan & Graham suggest that "without independent grammatical mechanisms" Chunk-and-Pass processing cannot explain children's acquisition of "free word order" languages such as Tamil. However, a recent computational model by McCauley and Christiansen (2014; 2015c) casts doubt on this claim. This chunk-based learner (CBL) implements Chunk-and-Pass processing at the word level, using simple statistical computations to build up an inventory of chunks consisting of one or more words, when exposed to child-directed speech from a typologically broad set of 29 Old World languages, including Tamil. Importantly, the model works entirely incrementally using online learning, as required by the Now-or-Never bottleneck. Following our idea, expressed in the target article, that acquisition involves learning how to process language, CBL gradually learns simplified versions of both comprehension and production. "Comprehension" consists of the chunking of natural child-directed speech, presented to the model word-by-word (essentially, a variation of "shallow parsing," in line with evidence for the relatively underspecified nature of child and adult language comprehension; e.g., Frank & Bod 2011; Gertner & Fisher 2012; Sanford & Sturt 2002 – see also **Ferreira & Christianson**).

In "production," the task of CBL is to recreate the child utterances encountered in the corpus, given the inventory of chunks learned thus far in the acquisition process. When exposed to a corpus of Tamil child-directed speech, the model was able to use its inventory of chunks to successfully produce a large proportion of the child utterances in that corpus in the absence of "independent grammatical mechanisms." Indeed, CBL performed as well on Tamil as it did on Mandarin and English. Although not a definitive proof, the CBL simulations do suggest that Chunk-and-Pass processing may be more powerful than a priori speculations might suggest.[3] This underscores the importance of implementing theoretical accounts computationally – whether these accounts are usage-based or rely on innate grammatical mechanisms – in order to determine the degree to which they account for actual linguistic behavior.

Maier & Baldwin raise important questions about how item-based acquisition gets off the ground: For example, what principles can the learner use to establish the basic units from which structures can be built? One possible answer is that information-theoretic properties of the sequence (e.g., points of unusually low predictability) may provide clues to chunk boundaries. A simplified version of this approach is employed by the CBL model (McCauley & Christiansen 2014; 2015c), which uses dips in backward transitional probabilities (which infants track; cf. Pelucchi et al. 2009) to chunk words together. Another approach might discover chunks by way of undersegmentation, essentially treating intonational units as preliminary chunks. The PUDDLE (Phonotactics from Utterances Determine Distributional Lexical Elements) model of word segmentation (Monaghan & Christiansen 2010) adopts this method and is able to build a vocabulary by using shorter chunks to split up larger chunks. For example, the model is able to use the frequent occurrence of a child's name in isolation to segment larger utterances in which that name also appears, mirroring the kind of developmental data (e.g., Bortfeld et al. 2005) that Maier & Baldwin mention. As discussed in McCauley et al. (2015), the two ways of discovering chunks in CBL and PUDDLE likely occur side-by-side in development, possibly alongside other mechanisms. Future research is needed to fully understand the interplay between these different mechanisms and their specific characteristics. Fortunately, as Maier & Baldwin point out, there is considerable empirical evidence that can potentially help constrain models of initial chunk formation (for reviews, see, e.g., Arnon & Christiansen, submitted; Werker et al. 2012).

R4.2. Implications for language change and language evolution

Item-based models of language processing and acquisition imply an item-based model of language change. So, assuming that items can be identified with constructions at various levels of abstraction (e.g., from individual lexical items, all the way to constructions determining, for example, canonical word order), then the structure of the language, both within a person, and across individuals, will change construction-by-construction, rather than through the flipping of an abstract parameter, which may have diverse and widespread implications (e.g., Lightfoot 1991). Note, though, that more abstract constructions may be relevant to a large number of sentences of the language. So the fact that the language changes one construction at a time does not imply, for example, that it changes one sentence at a time.

We see language change within a given language community as an accumulation of changes within the set of constructions acquired by the members of that community. And we view the evolution of language as nothing more than language change writ large. In particular, this implies that we see the evolution of language as a result of processes of cultural evolution over long periods of human history constrained by communicative goals, as well as our cognitive and neural machinery, rather than resulting from the biological evolution of a language faculty through processes of natural selection or some other mechanism. In short, language is shaped by the brain, rather than the brain being shaped by language (Chater & Christiansen 2010; Chater et al. 2009; Christiansen & Chater 2008). In the target article, we aimed to expand on this perspective by exploring how specific properties of language, such as its highly local structure, the existence of duality of patterning, and so on, might arise given the powerful constraint imposed by the Now-or-Never bottleneck.

In this light, **Endress & Katzir**'s concern that we may be conflating the cultural and biological evolution of language can be set aside; we explicitly reject the idea that there is any substantive biological evolution of language (any more than there has been substantive *biological* evolution of any other cultural form, whether writing, mathematics, music, or chess) although, of course, there will be an interesting biological evolutionary story to tell about the cognitive and neural precursors upon which language has been built. Similarly, **Lotem et al.**'s worry that we have forgotten about biological evolution is also misplaced. The "fit" between language and language users arises because language is a cultural product that is shaped around us (and our memory limitations), rather than a fixed and exogenously given system to which the brain must adapt. Indeed, our perspective aligns with Charles Darwin's suggestion that the cultural evolution of language can be viewed as analogous to biological evolution through natural selection. As early as in *The Descent of Man*, Darwin discussed the cultural evolution of linguistic forms in light of biological adaptation: "The formation of different languages and of distinct species, and the proofs that both have been developed through a gradual process, are curiously the same" (Darwin 1871, p. 59).

One of the great challenges of evolution by natural selection is to explain how biological organisms can increase in complexity. Darwin's answer was that such complexity may be favored if it increases the number of offspring at the next generation – that is, if it improves "fitness." A parallel challenge arises for explaining the presumed increase in complexity of human languages, from, we may assume, initially limited systems of signed or vocal communication, to the huge richness in phonology, morphology, vocabulary, and syntax, of contemporary natural languages, an issue raised by **Behme**. Indeed, gradual increases in complexity can happen relatively quickly, as indicated by the fact that children can "outperform" the adults from whom they learn language (Singleton & Newport 2004), and the incremental incorporation of new linguistic structures into emergent languages such as the Nicaraguan Sign Language (Senghas et al. 2004) or the Al-Sayyid Bedouin Sign Language (Sandler 2012). The pressure for such increases in complexity seems clear: the drive to communicate. While some theorists have argued that the language is not primarily "designed" for communication, but rather for thought (e.g., Chomsky 2010), we suggest that the social importance of communication underlies the continual generation of new linguistic items, and the recombination of existing items in creative new ways. Of course, such forms are then subject to the forces of simplification and erosion when they are transmitted across generations of speakers – the forces described by theories of grammaticalization. The picture of language as attempting to maximize communication richness, in the face of memory constraints, is elegantly outlined by **Lewis & Frank**.

Bergmann et al. note that language change can be affected by the nature of the language community. For example, the presence of a large number of second-language speakers (and the properties of their first language) will affect how the new language is processed and transmitted. After all, the Chunk-and-Pass machinery built for a first language will typically be recruited to process a second language, resulting in non-native patterns of chunking. Preliminary support for this perspective comes from analyses of the productions of first (L1) and second (L2) language learners using the earlier mentioned CBL model (McCauley & Christiansen 2014a; 2015c). McCauley and Christiansen (2015a) used the CBL model to compare the "chunkedness" of productions by native Italian speakers learning English or German, when compared with either child or adult native speakers of English and German. The results showed that, compared to those of the L2 speakers, the productions of the native speakers – whether children or adults – were considerably more chunked as measured by repeated multiword sequences. The inability of L2 speakers to chunk incoming input in a native-like way is likely to negatively influence their mastery of fundamental regularities such as morphology and case (Arnon & Christiansen, submitted). In languages with a preponderance of non-native speakers, the L2 learners may exert a greater pressure to regularize and otherwise simplify the language, as **Bergmann et al.** point out. Thus, the impact of the Now-or-Never bottleneck and the specifics of Chunk-and-Pass processing will vary to some degree based on individual experiences with particular languages.

Viewing language change as operating construction-by-construction does not necessarily rule out the possibility of abrupt change, as we noted earlier – modifying a single abstract construction (e.g., a ditransitive construction, Subj V Obj_1 Obj_2) may have far-reaching consequences. Hence, we can disregard **Endress & Katzir**'s contention that our approach is inconsistent with a part of the literature, which they suggest reports that language change is abrupt and substantial.

The question of whether language change provides evidence for modifications of "deep" linguistic principles or operates construction-by-construction is by no means settled in the literature, as is the question of whether macroscopic linguistic change in a community over historical time is actually abrupt at the level of individual speakers (e.g., Hopper & Traugott 1993; Lightfoot 1991; Wang 1977 – an issue parallel to the gradualist vs. punctate equilibrium controversy in biological evolutionary theory, e.g., Dawkins 1986; Eldredge & Gould 1972). If compelling evidence could be found suggesting that language change involves modification of highly abstract linguistic principles not embedded in a single construction, then this would contradict the item-based model that we see as following from the Now-or-Never bottleneck. But we do not believe that the literature provides such evidence.

R4.3. Implications for language structure

Commentators provide a wide range of viewpoints concerning the relationship between the present account and language structure. A particular concern is how far the Now-or-Never bottleneck is able to capture so-called language universals. We stress that we see broad patterns across languages, whether exception-less or merely statistical universals, as arising from the interaction of a multitude of constraints, including perceptual and cognitive factors, communicative pressures, the structure of thought, and so on (Christiansen & Chater 2008). Moreover, the trajectory of change observed for a particular language will also be determined by a range of cultural and historical forces, including sociolinguistic factors, language contact, and so on. In view of the interaction of this broad range of factors, it may be unlikely that many aspects of language are strictly universal, and indeed, human languages do seem to exhibit spectacular variety, including on such basic matters as the nature and number of syntactic categories. Yet even if strict language universals are, to some degree at least, a myth (Evans & Levinson 2009), we should nonetheless expect that language will be shaped, in part, by cognitive constraints, such as the Now-or-Never bottleneck.

In this light, concerns that the Now-or-Never bottleneck does not provide an account of all putatively universal features of language (**Medeiros et al**.; **Chacón et al.**; **Endress & Katzir**) can be set aside. Explaining the cross-linguistic patterns they mention using the aforementioned multiple constraints is likely to be a valuable direction for future research. Indeed, we would argue that the Now-or-Never bottleneck is a specific and concrete example of the type of cognitive constraint that Medeiros et al. believe to underlie universal or near-universal features of language.

Kempson et al. argue that the Now-or Never Bottleneck and its implications have interesting links with formal theories of grammar, such as dynamic syntax, in which there is a close relationship between grammatical structure and processing operations. Similarly, **O'Grady** suggests that the processing bottleneck is manifested differently in different languages. We agree insofar as memory limitations arise from the interaction of the cognitive system with the statistical structure of the language being learned. O'Grady's specific proposals here and elsewhere (2013; 2015a) provide a promising direction for the development of a detailed – and cross-linguistically valid – analysis, linking structure and processing in a way that is consistent with the Now-or-Never bottleneck.

One property mentioned by several commentators as being a widespread (**Chacón et al.; Levinson; MacDonald**) if not universal property of language (**Medeiros et al.**) is the existence of nonlocal dependencies. We have provided a broad account of complex recursive structures incorporating long-distance dependencies elsewhere (Christiansen & Chater 2015; 2016). Here, we briefly discuss an often-cited example of long-distance dependencies in the form of center-embedding as exemplified in (1) and (2), where the subscripts indicate subject-noun/verb relationships:

1. *The chef$_1$ who the waiter$_2$ appreciated$_2$ admired$_1$ the musicians*.
2. *The chef$_1$ who the waiter$_2$ who the busboy$_3$ offended$_3$ appreciated$_2$ admired$_1$ the musicians*.

Whereas (1) is easy to comprehend, (2) creates problems for most people (e.g., Blaubergs & Braine 1974; Hakes et al. 1976; Hamilton & Deese 1971; Wang 1970). This problem with multiple long-distance dependencies is not unique to English but has also been observed for center-embedded constructions in French (Peterfalvi & Locatelli 1971), German (Bach et al. 1986), Spanish (Hoover 1992), Hebrew (Schlesinger 1975), Japanese (Uehara & Bradley 1996), and Korean (Hagstrom & Rhee 1997). Indeed, corpus analyses of Danish, English, Finnish, French, German, Latin, and Swedish (Karlsson 2007) indicate that doubly center-embedded sentences such as (2) are practically absent from spoken language. Evidence from sequence learning suggests that the problems with multiple center-embeddings do not derive from semantic or referential complications but rather are due to basic memory limitations for sequential information (de Vries et al. 2012), as discussed in the target article. These memory limitations may even result in the kind of "illusion of grammaticality" noted by **Chacón et al.**, as when the second verb in (2) is removed to yield the sentence in (3), which to many people seems quite acceptable and even comprehensible (e.g., Christiansen & MacDonald 2009; Gibson & Thomas 1999; Vasishth et al. 2010):

3. *The chef$_1$ who the waiter$_2$ who the busboy$_3$ offended$_3$ admired$_1$ the musicians*.

However, these memory limitations interact with the statistics of the language being used (as discussed previously) such that the above "missing verb" effect can be observed in French (Gimenes et al. 2009) but not in German (Vasishth et al. 2010) or Dutch (Frank et al. 2016). Because verb-final constructions are common in German and Dutch, requiring the listener to track dependency relations over a relatively long distance, substantial prior experience with these constructions likely has resulted in language-specific processing improvements (see also Engelmann & Vasishth 2009; Frank et al. 2016, for similar perspectives). Nonetheless, in some cases the missing verb effect may appear even in German, under conditions of high processing load (Trotzke et al. 2013). We would expect that other nonlocal dependencies (e.g., as noted by **Medeiros et al.**, **Chacón et al.**, **Levinson**, **MacDonald**) would be amenable to similar types of explanation within the framework of Chunk-and-Pass processing (as also noted by **Kempson et al.** and **O'Grady**).

R5. Conclusions and future directions

Our target article highlights a fundamental constraint imposed by memory interference on the processing and production of sequential material, and in particular, on language. Dealing with this Now-or-Never bottleneck requires, we argue, the chunking and recoding of incoming material as rapidly as possible, across a hierarchy of representational levels (this is Chunk-and-Pass processing). Similarly, it requires specifying the representations involved in producing language just before they are used (this is Just-in-Time production). These proposals themselves have, we suggest, a variety of implications for language structure (e.g., that such structure is typically highly local), for acquisition, and for language change and evolution (e.g., that language changes construction-by-construction both within individuals during learning, and over generations within entire language communities).

The commentaries on our article have raised important issues of clarification (e.g., differentiating the present proposals from bottom-up, syntax-driven models such as the Sausage Machine, Frazier & Fodor 1978); have clarified important links with prior models and empirical results (e.g., the link with "good enough" parsing, **Ferreira & Christianson**); and have outlined supporting evidence (e.g., from the time-course of neural activity involved in language processing, e.g., **Honey et al.**) and pointed out ways in which the approach can be deepened and made more linguistically concrete (**O'Grady**). One commentator fears that our proposals may be unfalsifiable (**Levinson**); others suspect that our approach may actually be falsified by known features of language structure (**Medeiros et al.**), processing (**MacDonald**), acquisition (**Wang & Mintz**), or language change (**Endress & Katzir**). We hope that our target article will persuade readers that memory constraints have substantial implications for understanding many aspects of language, and that our response to commentators makes the case that the many claims flowing from the Now-or-Never bottleneck are compatible with what is known about language (although not always with what is presumed to be the case by prior theories). Most important, we encourage interested readers to continue the work of the many commentators who provide constructive directions to further explore the nature of the Now-or-Never bottleneck, further elaborate and test the Chunk-and-Pass and Just-in-Time perspectives on language processing, and help integrate the study of these performance constraints into our understanding of key aspects of language structure, acquisition, and evolution (for some steps in this direction, see Christiansen & Chater 2016).

NOTES

1. **Chacón et al.** contend that "early observations about speech errors indicated that exchange errors readily cross phrasal and clausal boundaries (Garrett 1980)" (para. 7). A careful reading of Garrett, however, shows that most exchange errors tend to occur *within* phrases, as would be expected from our perspective.

2. **Wang & Mintz** seem to have misunderstood the aim of the modeling by Reali and Christiansen (2005). Their point was not to provide a full-fledged model of so-called auxiliary fronting in complex yes/no questions (such as *Is the dog that is on the chair black?*) but rather to demonstrate that the input to young children provided sufficient statistical information for them to distinguish between grammatical and ungrammatical forms of such sentences. Kam et al. (2008) noted some limitations of the simplest bigram model used by Reali and Christiansen, but failed to address the fact that not only did the model fit the results from the classic study by Crain and Nakayama (1987) but also correctly predicted that children should make fewer errors involving high-frequency word chunks compared to low-frequency chunks in a subsequent question elicitation study (Ambridge et al. 2008; see Reali & Christiansen 2009). For example, higher rates of auxiliary-doubling errors occur for questions where such errors involved high-frequency word category combinations (e.g., more errors such as **Is the boy who is washing the elephant is tired?* than **Are the boys who are washing the elephant are tired?*). Most important for current purposes is the fact that Reali and Christiansen – in line with our account of Chunk-and-Pass processing – do not assume that distributional information is all there is to language acquisition: "Young learners are likely to rely on many additional sources of information (e.g., semantic, phonological, prosodic) to be able to infer different aspects of the structure of the target language" (Reali & Christiansen 2009, p. 1024).

3. **Endress & Katzir** (see also **Wang & Mintz**) raise a common concern relating to usage-based models: that the sparseness of the input will prevent them from being able to process novel word sequences that are grammatical but not predictable (such as *Evil unicorns devour xylophones*). Reali et al. (2005) addressed this challenge head-on, showing in a statistical learning experiment that human participants become sufficiently sensitive to the regularities of training examples to recognize novel sequences whose bigram transitions are absent in training. They subsequently showed that a simple recurrent network (Elman 1990) could correctly process sequences that contain null-probability bigram information by relying on distributional regularities in the training corpus. Thus, in contrast to the claims of **Endress & Katzir**, distributional learning appears to be sufficiently powerful to deal with unpredictable but grammatical sequences such as Chomsky's (1957) famous sentence *Colorless green ideas sleep furiously* (see also Allen & Seidenberg 1999).

References

[The letters "a" and "r" before author's initials stand for target article and response references, respectively]

Acheson, D. J. & MacDonald, M. C. (2011) The rhymes that the reader perused confused the meaning: Phonological effects on on-line sentence comprehension. *Journal of Memory and Language* 65:193–207. [MCM, rNC]

Adams, J. A. (1971) A closed-loop theory of motor learning. *Journal of Motor Behavior* 3:111–50. [AB]

Allen, J. & Seidenberg, M. S. (1999) The emergence of grammaticality in connectionist networks. In: *The emergence of language*, ed. B. MacWhinney, pp. 115–51. Erlbaum. [rNC]

Allopenna, P. D., Magnuson, J. S. & Tanenhaus, M. K. (1998) Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language* 38:419–39. [aMHC]

Altmann, G. T. M. (2002) Learning and development in neural networks: The importance of prior experience. *Cognition* 85:43–50. [aMHC, MLD]

Altmann, G. T. M. (2004) Language-mediated eye movements in the absence of a visual world: The "blank screen paradigm." *Cognition* 93:79–87. [aMHC]

Altmann, G. T. M. & Kamide, Y. (1999) Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition* 73:247–64. [aMHC]

Altmann, G. T. M. & Kamide, Y. (2009) Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Cognition* 111:55–71. [MLD, aMHC]

Altmann, G. T. M. & Mirkovic, J. (2009) Incrementality and prediction in human sentence processing. *Cognitive Science* 33:583–609. [aMHC]

Altmann, G. T. M. & Steedman, M. J. (1988) Interaction with context during human sentence processing. *Cognition* 30:191–38. [aMHC]

Ambridge, B., Rowland, C. & Pine, J. (2008) Is structure dependence an innate constraint? New experimental evidence from children's complex question production. *Cognitive Science* 32:222–55. [rNC]

Ames, H. & Grossberg, S. (2008) Speaker normalization using cortical strip maps: A neural model for steady state vowel categorization. *Journal of the Acoustical Society of America* 124:3918–36. [SG]

Anderson, J. R. (1990) *The adaptive character of thought*. Erlbaum. [aMHC, KB]

Anderson, J. R. & Milson, R. (1989) Human memory: An adaptive perspective. *Psychological Review* 96:703. [rNC]

Anderson, J. R. & Schooler, L. J. (1991) Reflections of the environment in memory. *Psychological Science* 2:396–408. [rNC]

Anderson, M. L. (2010) Neural reuse: A fundamental organizational principle of the brain. *Behavioral and Brain Sciences* 34:245–66. [AB, AL]

Anderson, P. W. (1972) More is different. *Science* 177:393–96. [MLL]

Aoshima, S., Phillips, C. & Weinberg, A. (2004) Processing filler-gap dependencies in a head-final language. *Journal of Memory and Language* 51:23–54. [DAC]

Appelt, D., Hobbs, J., Bear, J., Israel, D. & Tyson, M. (1993) FASTUS: A finite-state processor for information extraction from real-world text. *International Joint Conference on Artificial Intelligence, Chambery, France, August 1993.* ed. A. Ralescu, pp. 1172–78. [CRH]

Arbib, M. A. (2005) From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences* 28:105–24. [aMHC]

Arbib, M. A. (2012) *How the brain got language: The mirror system hypothesis*. Oxford University Press. [DPM]
Arnon, I. & Christiansen, M. H. (submitted) The role of multiword building blocks in explaining L1-L2 differences [aMHC, rNC]
Arnon, I. & Clark, E. V. (2011) Why brush your teeth is better than teeth – Children's word production is facilitated in familiar sentence-frames. *Language Learning and Development* 7:107–29. [aMHC]
Arnon, I. & Cohen Priva, U. (2013) More than words: The effect of multi-word frequency and constituency on phonetic duration. *Language and Speech* 56:349–73. [aMHC, rNC]
Arnon, I. & Snider, N. (2010) More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language* 62:67–82. [aMHC]
Aronoff, M., Meir, I. & Sandler, W. (2005) The paradox of sign language morphology. *Language* 81(2):301–44. [KE]
Austin, J. L. (1962) *How to do things with words*. Harvard University Press. [aMHC]
Bach, E., Brown, C. & Marslen-Wilson, W. (1986) Crossed and nested dependencies in German and Dutch: A psycholinguistic study. *Language and Cognitive Processes* 1:249–62. [rNC]
Baddeley, A. (1987) *Working memory*. Oxford University Press. [SCL]
Baddeley, A. (1992) Working memory. *Science* 255:556–59. [rNC]
Baddeley, A. D. & Hitch, G. J. (1974) Working memory. *The Psychology of Learning and Motivation* 8:47–89. [rNC]
Baddeley, R. & Attewell, D. (2009) The relationship between language and the environment: Information theory shows why we have only three lightness terms. *Psychological Science* 20:1100–107. [MLL]
Badets, A., Koch, I. & Philipp, A. M. (2016) A review of ideomotor approaches to perception, cognition, action, and language: Advancing a cultural recycling hypothesis. *Psychological Research* 80:1–15. doi: 10.1007/s00426-014-0643-8. [AB]
Badets, A. & Osiurak, F. (2015) A goal-based mechanism for delayed motor intention: Considerations from motor skills, tool use and action memory. *Psychological Research* 79:345–60. [AB]
Badets, A. & Rensonnet, C. (2015) Une approche idéomotrice de la cognition. *L'Année psychologique* 115:591–635. [AB]
Baker, M. (2001) *The atoms of language*. Basic Books. [aMHC]
Baker, M. C. (2008) *The syntax of agreement and concord*. Cambridge University Press. [DPM]
Baker, M. C. (2013) On agreement and its relationship to case: Some generative ideas and results. *Lingua* 130(June):14–32. [DPM]
Baldwin, D. (1991) Infants' contribution to the achievement of joint reference. *Child Development* 62:874–90. [MLL]
Baldwin, D. A. (1993) Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language* 20:395–18. [MLL]
Bannard, C. & Matthews, D. (2008) Stored word sequences in language learning. *Psychological Science* 19:241–48. [aMHC]
Bar, M. (2004) Visual objects in context. *Nature Reviews Neuroscience* 5:617–29. [aMHC]
Bar, M. (2007) The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences* 11:280–89. [aMHC]
Bard, E. G., Shillcock, R. C. & Altmann, G. T. M. (1988) The recognition of words after their acoustic offsets in spontaneous speech: Effects of subsequent context. *Perception and Psychophysics* 44:395–408. [KB]
Barsalou, L. W. (2008) Grounded cognition. *Annual Review of Psychology* 59:617–45. [MLD]
Bates, E. & MacWhinney, B. (1982) Functionalist approaches to grammar. In: *Language acquisition: The state of the art*, ed. E. Wanner & L. Gleitman, pp. 173–218. Cambridge University Press. [DPM]
Baumann, T. & Schlangen, D. (2012) INPRO_iSS: A component for just-in-time incremental speech synthesis. In: *Proceedings of the ACL 2012 System Demonstrations*, pp. 103–108. Association for Computational Linguistics. [aMHC]
Bavelas, J. B. & Gerwing, J. (2007) Conversational hand gestures and facial displays in face-to-face dialogue. In: *Social communication*, ed. K. Fiedler, pp. 283–308. Psychology Press. [PGTH]
Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N., Holland, J., Ke, J., Larsen- Freeman, D. & Schoenemann, T. (2009) Language is a complex adaptive system: Position paper. *Language Learning* 59 (Suppl. 1):1–27. [aMHC]
Behme, C. (2014a) Assessing direct and indirect evidence in linguistic research. *Topoi* 33:373–83. [CB]
Behme, C. (2014b) *Evaluating Cartesian linguistics: From historic antecedents to computational modeling*. Peter Lang. [CB]
Belletti, A., ed. (2004) *Structures and beyond. The cartography of syntactic structures, vol. 3.* Oxford University Press. [DPM]
Bellugi, U. & Fisher, S. (1972) A comparison of sign language and spoken language. *Cognition* 1:173–200. [aMHC]
Bellugi, U., Klima, E. S. & Siple, P. (1975) Remembering in signs. *Cognition* 3:93–25. [KE]
Bentz, C. & Winter, B. (2013) Languages with more second language learners tend to lose nominal case. *Language Dynamics and Change* 3:1–27. [TB]
Berwick, R. C. (1985) *The acquisition of syntactic knowledge*. MIT Press. [aMHC]
Berwick, R. C., Chomsky, N. & Piattelli-Palmarini, M. (2013) Poverty of the stimulus stands: Why recent challenges fail. In: *Rich languages from poor inputs*, ed. M. Piattelli-Palmarini & R. C. Berwick, pp. 19–42. Oxford University Press. [DPM]
Berwick, R. C., Friederici, A. D., Chomsky, N. & Bolhuis, J. J. (2013) Evolution, brain, and the nature of language. *Trends in Cognitive Sciences* 17:91–100. [aMHC]
Berwick, R. C. & Weinberg, A. S. (1984) *The grammatical basis of linguistic performance*. MIT Press. [aMHC]
Bever, T. (1970) The cognitive basis for linguistic structures. In: *Cognition and the development of language*, ed. J. R. Hayes, pp. 279–362. Wiley. [aMHC, DPM]
Bever, T. G. (1975) Cerebral asymmetries in humans are due to the differentiation of two incompatible processes: Holistic and analytic. *Annals of the New York Academy of Sciences* 263(1):251–62. [DPM]
Bi, G. & Poo, M. (2001) Synaptic modification of correlated activity: Hebb's postulate revisited. *Annual Review of Neuroscience* 24:139–66. [SLF]
Bickel, B., Banjade, G., Gaenszle, M., Lieven, E., Paudyal, N., Rai, I. P., Rai, M., Rai, N. K. & Stoll, S. (2007) Free prefix ordering in Chintang. *Language* 83(1):43–73. [SCL]
Bickerton, D. (1984) The language bioprogram hypothesis. *Behavioral and Brain Sciences* 7:173–221. [aMHC]
Bicknell, K. & Levy, R. (2010) A rational model of eye movement control in reading. In: *Proceedings of the 48th annual meeting of the Association for Computational Linguistics*, ed. J. Hajič, S. Carberry, S. Clark, & J. Nivre, pp. 1168–78. Association for Computational Linguistics. [KB]
Bicknell, K., Tanenhaus, M. K. & Jaeger, T. F. (2015) Listeners can maintain and rationally update uncertainty about prior words. Manuscript submitted for publication. [KB]
Blackburn, S. & Meyer-Viol, W. (1994) Linguistics, logic and finite trees. *Bulletin of Interest Group in Pure and Applied Logics* 2:3–29. [RK]
Blaubergs, M. S. & Braine, M. D. S. (1974) Short-term memory limitations on decoding self- embedded sentences. *Journal of Experimental Psychology* 102:745–48. [rNC]
Blokland, G. A. M., McMahon, K. L., Thompson, P. M., Martin, N. G., de Zubicaray, G. I. & Wright, M. J. (2011) Heritability of working memory brain activation. *The Journal of Neuroscience* 31:10882–90. [AL]
Boardman, I., Grossberg, S., Myers, C. & Cohen, M. (1999) Neural dynamics of perceptual order and context effects for variable-rate speech syllables. *Perception and Psychophysics* 6:1477–500. [SG]
Bock, J. K. (1982) Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review* 89:1–47. [aMHC]
Bock, J. K. (1986) Meaning, sound, and syntax: Lexical priming in sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 12:575–86. [aMHC]
Bock, J. K. & Loebell, H. (1990) Framing sentences. *Cognition* 35:1–39. [aMHC]
Bock, J. K. & Miller, C. A. (1991) Broken agreement. *Cognitive Psychology* 23:45–93. [aMHC]
Bock, K. (1987) Exploring levels of processing in sentence production. In: *Natural language generation*, ed. G. Kempen, pp. 351–63. Springer. [DAC]
Bock, K. & Cutting, J. C. (1992) Regulating mental energy: Performance units in language production. *Journal of Memory and Language* 31:99–127. [rNC]
Bock, K. & Miller, C. A. (1991) Broken agreement. *Cognitive Psychology* 23:45–93. [DAC]
Bod, R. (2009) From exemplar to grammar: A probabilistic analogy–based model of language learning. *Cognitive Science* 33:752–93. [aMHC]
Boeckx, C. & Leivada, E. (2013) Entangled parametric hierarchies: Problems for an overspecified universal grammar. *PLOS ONE* 8(9):e72357. [aMHC]
Boemio, A., Fromm, S., Braun, A. & Poeppel, D. (2005) Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience* 8(3):389–95. [GB]
Bohnet, B. & Nivre, J. (2012) A transition-based system for joint part-of-speech tagging and labeled non-projective dependency parsing. In: *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Jeju Island, Korea, July 12–14, 2012*, ed. J. Tsujii, J. Henderson & M. Pasca, pp. 1455–65. Association for Computational Linguistics. [CG-R]
Borovsky, A., Elman, J. L. & Fernald, A. (2012) Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology* 112:417–36. [aMHC]
Börschinger, B. & Johnson, M. (2011) A particle filter algorithm for Bayesian word segmentation. In: *Proceedings of the Australasian Language Technology*

Association Workshop, Canberra, Australia, ed. D. Mollá & D. Martinez, pp. 10–18. [ADE]
Bortfeld, H., Morgan, J. L., Golinkoff, R. M. & Rathbun, K. (2005) Mommy and me: Familiar names help launch babies into speech-stream segmentation. *Psychological Science* 16(4):298–304. [RM, rNC]
Botvinick, M. M. (2008) Hierarchical models of behavior and prefrontal function. *Trends in Cognitive Sciences* 12:201–208. [rNC]
Bouzouita, M. & Chatzikyriakidis, S. (2009) Clitics as calcified processing strategies. In: *Proceedings of LFG09, Cambridge, UK, July 2009*, ed. M. Butt & T. Holloway-King, pp. 189–207. CSLI Press. Available at: http://www.stergioschatzikyriakidis.com/uploads/1/0/3/6/10363759/lfg09bouzouitachatzikyriakidis_2.pdf. [RK]
Boyd, R. & Richerson, P. J. (1987) The evolution of ethnic markers. *Cultural Anthropology* 2:65–79. [aMHC]
Branigan, H., Pickering, M. & Cleland, A. (2000) Syntactic co-ordination in dialogue. *Cognition* 75:13–25. [aMHC]
Bransford, J. D. & Johnson, M. K. (1972) Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior* 11:717–26. [rNC]
Bregman, A. S. (1990) *Auditory scene analysis: The perceptual organization of sound*. MIT Press. [aMHC, rNC]
Brennan, S. E. & Schober, M. F. (2001) How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language* 44(2):274–96. [PGTH]
Brentari, D. (1998) *A prosodic model of sign language phonology.* The MIT Press. [KE]
Brinton, B., Fujiki, M., Loeb, D. F. & Winkler, E. (1986) Development of conversational repair strategies in response to requests for clarification. *Journal of Speech, Language, and Hearing Research* 29(1):75–81. [GB]
Broadbent, D. (1958) *Perception and communication*. Pergamon Press. [aMHC]
Brodsky, P., Waterfall, H. & Edelman, S. (2007) Characterizing motherese: On the computational structure of child-directed language. In: *Proceedings of the 29th Meeting of the Cognitive Science Society,* August 2007, pp. 833–38, ed. D. S. McNamara & J. G. Trafton. Cognitive Science Society. [rNC]
Brown, G. D. A., Neath, I. & Chater, N. (2007) A temporal ratio model of memory. *Psychological Review* 114:539–76. [aMHC, rNC]
Brown, M., Dilley, L. C. & Tanenhaus, M. K. (2014) Probabilistic prosody: Effects of relative speech rate on perception of (a) word(s) several syllables earlier. In: *Proceedings of the 7th International Conference on Speech Prosody, Dublin, Ireland, May 20–23, 2014*, ed. N. Campbell, D. Gibbon & D. Hirst. pp. 1154–58. Dublin. [KB]
Brown-Schmidt, S. & Konopka, A. E. (2011) Experimental approaches to referential domains and the on-line processing of referring expressions in unscripted conversation. *Information* 2:302–26. [aMHC]
Brown-Schmidt, S. & Konopka, A. (2015) Processes of incremental message planning during conversation. *Psychonomic Bulletin and Review* 22:833–43. [aMHC]
Brown-Schmidt, S. & Tanenhaus, M. K. (2008) Real-time investigation of referential domains in unscripted conversation: A targeted language game approach. *Cognitive Science* 32:643–84. [aMHC]
Bubic, A., von Cramon, D. Y. & Schubotz, R. I. (2010) Prediction, cognition and the brain. *Frontiers in Human Neuroscience* 4(25):1–15. [SW]
Buonomano, D. V. & Maass, W. (2009) State-dependent computations: Spatiotemporal processing in cortical networks. *Nature Reviews Neuroscience* 10:113–25. [SLF]
Burgess, N. & Hitch, G. J. (1999) Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review* 106:551. [rNC]
Burghardt, G. M. (1970) Defining "communication." In: *Communication by chemical signals*, ed. J. W. Johnston Jr., D. G. Moulton & A. Turk, pp. 5–18. Appleton-Century-Crofts. [AL]
Bybee, J. (2002) Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* 14:261–90. [aMHC]
Bybee, J. (2006) From usage to grammar: The mind's response to repetition. *Language* 82:711–33. [aMHC]
Bybee, J. (2007) *Frequency of use and the organization of language*. Oxford University Press. [aMHC, DPM]
Bybee, J. (2009) Language universals and usage-based theory. In: *Language universals*, ed. M. H. Christiansen, C. Collins & S. Edelman, pp. 17–39. Oxford University Press. [aMHC]
Bybee, J. & Hopper, P., eds. (2001) *Frequency and the emergence of linguistic structure*. John Benjamins. [aMHC]
Bybee, J. & McClelland, J. L. (2005) Alternatives to the combinatorial paradigm of linguistic theory based on general principles of human cognition. *The Linguistic Review* 22:381–410. [aMHC]
Bybee, J., Perkins, R. D. & Pagliuca, W. (1994) *The evolution of grammar: Tense, aspect and modality in the languages of the world*. University of Chicago Press. [aMHC]
Bybee, J. & Scheibman, J. (1999) The effect of usage on degrees of constituency: The reduction of don't in English. *Linguistics* 37:575–96. [aMHC, rNC]
Bybee, J. L. & Slobin, D. I. (1982) Rules and schemas in the development and use of the English past tense. *Language* 58:265–89. [aMHC]
Campbell, L. (2000) What's wrong with grammaticalization? *Language Sciences* 23:113–61. [aMHC]
Cann, R. & Kempson, R. (2008) Production pressures, syntactic change and the emergence of clitic pronouns. In: *Language in flux: Dialogue coordination, language variation, change and evolution*, ed. R. Cooper & R. Kempson, pp. 221–63. College Publications. [aMHC]
Cann, R., Kempson, R. & Marten, L. (2005) *The dynamics of language: An introduction*. Elsevier (now published by Emerald Insight Publishers). [RK]
Cann, R., Kempson, R. & Wedgwood, D. (2012) Representationalism and linguistic knowledge. In: *Philosophy of linguistics*, ed. R. Kempson, T. Fernando & N. Asher, pp. 357–402. Elsevier. [aMHC]
Cann, R., Purver, M. & Kempson, R. (2007) Context and wellformedness: The dynamics of ellipsis. *Research on Language and Computation* 5:333–58. [RK]
Carey, S. & Bartlett, E. (1978) Acquiring a single word. *Papers and Reports on Child Language Development* 15:17–29. [ADE]
Carney, A. E., Widin, G. P. & Viemeister, N. F. (1977) Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America* 62:961–70. [KB]
Carr, M. F., Jadhav, S. P. & Frank, L. M. (2011) Hippocampal replay in the awake state: A potential substrate for memory consolidation and retrieval. *Nature Neuroscience* 14:147–53. [aMHC]
Carstairs-McCarthy, A. (1992) *Current morphology*. Routledge. [aMHC]
Carstensen, A., Xu, J., Smith, C. & Regier, T. (2015) Language evolution in the lab tends toward informative communication. In: *Proceedings of the 37th Annual Meeting of the Cognitive Science Society, Austin, TX, July 2015*, ed. D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings & P. P. Maglio, pp. 303–308. Cognitive Science Society. [MLL]
Catani, M., Jones, D. K. & ffytche, D. H. (2005) Perisylvian language networks of the human brain. *Annals of Neurology* 57(1):8–16. [GB]
Cattell, J. M. (1886) The time it takes to see and name objects. *Mind* 11:63–65. [MCM]
Chacón, D., Imtiaz, M., Dasgupta, S., Murshed, S., Dan, M. & Phillips, C. (submitted) Locality in the processing of filler-gap dependencies in Bangla. [DAC]
Chang, F., Dell, G. S. & Bock, K. (2006) Becoming syntactic. *Psychological Review* 113:234–72. [aMHC]
Chater, N. & Christiansen, M. H. (2010) Language acquisition meets language evolution. *Cognitive Science* 34:1131–57. [AL, rNC]
Chater, N., Crocker, M. J. & Pickering, M. J. (1998) The rational analysis of inquiry: The case of parsing. In: *Rational models of cognition*, ed. M. Oaksford & N. Chater, pp. 441–68. Oxford University Press. [aMHC]
Chater, N., McCauley, S. M. & Christiansen, M. H. (2106) Language as skill: Intertwining comprehension and production. *Journal of Memory and Language* 89:244–54. [aMHC, rNC]
Chater, N., Reali, F. & Christiansen, M. H. (2009) Restrictions on biological adaptation in language evolution. *Proceedings of the National Academy of Sciences* 106:1015–20. [rNC]
Chater, N., Tenenbaum, J. B. & Yuille, A. (2006) Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Sciences* 10:287–91. [aMHC]
Chatzikyriakids, S. & Kempson, R. (2011) Standard modern and Pontic Greek person restrictions: Feature-free dynamic account. *Journal of Greek Linguistics* 10:127–66. Available at: http://www.kcl.ac.uk/innovation/groups/ds/publications/assets/chatzikyriakidis-kempson-jgl-draft.pdf. [RK]
Chen, D. & Manning, C. D. (2014) A fast and accurate dependency parser using neural networks. In: *Proceedings of the 2014 Conference on Empirical Methods on Natural Language Processing, Doha, Qatar, October 25–29, 2014*, ed. A. Moschitti, B. Pang & W. Daelemans, pp. 740–50. Association for Computational Linguistics. [CG-R]
Cherry, E. C. (1953) Some experiments on the recognition of speech with one and with two ears. *Journal of the Acoustical Society of America* 25:975–79. [aMHC]
Choi, J. D. & McCallum, A. (2013) Transition-based dependency parsing with selectional branching. In: *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Sofia, Bulgaria, August 4–9, 2013*, ed. P. Fung & M. Poesio, pp. 1052–62. Association for Computational Linguistics. [CG-R]
Chomsky, N. (1957) *Syntactic structures*. Mouton. [aMHC, rNC]
Chomsky, N. (1965) *Aspects of the theory of syntax*. MIT Press. [aMHC]
Chomsky, N. (1975) *Reflections on language*. Pantheon Books. [CB]
Chomsky, N. (1980) The linguistic approach. In: *Language and learning*, ed. M. Piattelli-Palmerini, pp. 107–30. Harvard University Press. [CB]
Chomsky, N. (1981) *Lectures on government and binding*. Mouton de Gruyter. [aMHC]
Chomsky, N. (1986) *Knowledge of language*. Praeger Publishing. [CB]
Chomsky, N. (2000) *The architecture of language*. Oxford University Press. [CB]
Chomsky, N. (2002) *On nature and language*. Cambridge University Press. [CB]

Chomsky, N. (2010) Some simple evo-devo theses: How true might they be for language? In: *Approaches to the evolution of language*, ed. R. K. Larson, V. M. Deprez & H. Yamakido. Cambridge University Press. [rNC]
Chomsky, N. (2012) *Noam Chomsky, the science of language – interviews with James McGilvray.* Cambridge University Press. [CB]
Christiansen, M. H. & Chater, N. (1999) Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science* 23:157–205. [aMHC, SLF]
Christiansen, M. H. & Chater, N. (2008) Language as shaped by the brain. *Behavioral & Brain Sciences* 31(05):489–58. [aMHC, ADE, DPM, rNC]
Christiansen, M. H. & Chater, N. (2015) The language faculty that wasn't: A usage-based account of natural language recursion. *Frontiers in Psychology* 6:1182. doi: 10.3389/fpsyg.2015.01182. [rNC]
Christiansen, M. H. & Chater, N. (2016) *Creating language: Integrating evolution, acquisition, and processing*. MIT Press. [rNC]
Christiansen, M. H. & MacDonald, M. C. (2009) A usage-based approach to recursion in sentence processing. *Language Learning* 59(Suppl. 1):126–61. [aMHC, rNC]
Christianson, K. & Ferreira, F. (2005) Conceptual accessibility and sentence production in a free word order language (Odawa). *Cognition* 98:105–35. [DAC]
Christianson, K., Hollingworth, A., Halliwell, J. F. & Ferreira, F. (2001) Thematic-roles assigned along the garden path linger. *Cognitive Psychology* 42:368–407. [FF, rNC]
Christianson, K. & Luke, S. G. (2011) Context strengthens initial misinterpretations of text. *Scientific Studies of Reading* 15:136–66. [FF]
Christianson, K., Williams, C. C., Zacks, R. T. & Ferreira, F. (2006) Misinterpretations of garden-path sentences by older and younger adults. *Discourse Processes* 42:205–38. [FF]
Chun, M. M. & Potter, M. C. (1995) A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception and Performance* 21:109–27. [MCP]
Cinque, G. (1996) The 'antisymmetric' program: Theoretical and typological implications. *Journal of Linguistics* 32(2):447–64. [DPM]
Cinque, G. (1999) *Adverbs and functional heads: A cross-linguistic perspective*. Oxford University Press. [DPM]
Cinque, G. (2005) Deriving Greenberg's universal 20 and its exceptions. *Linguistic Inquiry* 36(3):315–32. [DPM]
Cinque, G. (2013) Cognition, universal grammar, and typological generalizations. *Lingua* 130:50–65. doi: 10.1016/j.lingua.2012.10.007. [DPM]
Cisek, P. & Kalaska, J. F. (2001) Common codes for situated interaction. *Behavioral and Brain Sciences* 24:883–84. [AB]
Clark, A. (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences* 36(3):181–253. [aMHC, SW]
Clark, H. H. (1975) Bridging. In: *Proceedings of the 1975 Workshop on Theoretical Issues in Natural Language Processing, Cambridge, MA, June 1975*, ed. B. L. Nash-Webber & R. Shank, pp. 169–74. Association for Computational Linguistics. [rNC]
Clark, H. H. (1996) *Using language*. Cambridge University Press. [aMHC, GB, MLL]
Clark, J., Yallop, C. & Fletcher, J. (2007) *An introduction to phonetics and phonology, third edition*. Wiley-Blackwell. [aMHC]
Clément, S., Demany, L. & Semal, C. (1999) Memory for pitch versus memory for loudness. *Journal of the Acoustical Society of America* 106:2805–11. [aMHC]
Cohen, M. A. & Grossberg, S. (1986) Neural dynamics of speech and language coding: Developmental programs, perceptual grouping, and competition for short-term memory. *Human Neurobiology* 5:1–22. [SG]
Colman, M. & Healey, P. G. T. (2011) The distribution of repair in dialogue. In: *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society, Boston, MA, July 2011*, ed. L. Carlson & T. F. Shipley, pp. 1563–68. [PGTH]
Coltheart, M. (1980) Iconic memory and visible persistence. *Perception & Psychophysics* 27:183–228. [aMHC]
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* 30:234–50. [KB, rNC]
Conway, C. & Christiansen, M. H. (2005) Modality constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31:24–39. [rNC]
Conway, C. M. & Christiansen, M. H. (2009) Seeing and hearing in space and time: Effects of modality and presentation rate on implicit statistical learning. *European Journal of Cognitive Psychology* 21:561–80. [rNC]
Cooper, R. P. & Shallice, T. (2006) Hierarchical schemas and goals in the control of sequential behavior. *Psychological Review* 113:887–916. [aMHC]
Corballis, M. C. (2009) Mental time travel and the shaping of language. *Experimental Brain Research* 192:553–60. [AB]
Courtney, E. H. & Saville-Troike, M. (2002) Learning to construct verbs in Navajo and Quechua. *Journal of Child Language* 29:623–54. [UL]
Cover, T. M. & Thomas, J. A. (2006) *Elements of information theory, second edition*. Wiley. [rNC]
Cowan, N. (2000) The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences* 24:87–185. [aMHC, KE, rNC]
Crain, S. (1991) Language acquisition in the absence of experience. *Behavioral and Brain Sciences* 14:597–50. [DPM]
Crain, S. & Nakayama, M. (1987) Structure dependence in grammar formation. *Language* 63:522–43. [rNC]
Crain, S., Thornton, R. & Khlentzos, D. (2009) The case of the missing generalizations. *Cognitive Linguistics* 20(1):145–55. [DPM]
Crick, F. & Mitchison, G. (1983) The function of dream sleep. *Nature* 304:111–14. [aMHC]
Croft, W. (2001) *Radical construction grammar*. Oxford University Press. [aMHC]
Crowder, R. G. (1993) Systems and principles in memory theory: Another critique of pure memory. In: *Theories of memory*, ed. A. Collins, S. Gathercole, M. Conway & P. Morris. Erlbaum. [rNC]
Crowder, R. G. & Neath, I. (1991) The microscope metaphor in human memory. In: *Relating theory and data: Essays on human memory in honour of Bennet B. Murdock, Jr.*, ed. W. E. Hockley & S. Lewandowsky. Erlbaum. [aMHC]
Cui, J., Gao, D., Chen, Y., Zou, X. & Wang, Y. (2010) Working memory in early-school-age children with Asperger's syndrome. *Journal of Autism and Developmental Disorders* 40:958–67. [AL]
Culbertson, J. & Adger, D. (2014) Language learners privilege structured meaning over surface frequency. *Proceedings of the National Academy of Sciences* 111 (16):5842–47. [DPM]
Culbertson, J. & Newport, E. L. (2015) Harmonic biases in child learners: In support of language universals. *Cognition* 139:71–82. [MLL]
Culbertson, J., Smolensky, P. & Legendre, G. (2012) Learning biases predict a word order universal. *Cognition* 122(3):306–29. [MLL, DPM]
Culicover, P. W. (1999) *Syntactic nuts: Hard cases, syntactic theory, and language acquisition*. Oxford University Press. [aMHC]
Culicover, P. W. (2013) The role of linear order in the computation of referential dependencies. *Lingua* 136:125–44. [aMHC]
Curtiss, S. (1977) *Genie: A psycholinguistic study of a modern-day "wild child."* Academic Press. [RM]
Cutler, A., ed. (1982) *Slips of the tongue and language production*. De Gruyter Mouton. [aMHC]
Cutler, A., Hawkins, J. A. & Gilligan, G. (1985) The suffixing preference: A processing explanation. *Linguistics* 23:723–58. [KE]
Dahan, D. (2010) The time course of interpretation in speech comprehension. *Current Directions in Psychological Science* 19:121–26. [aMHC]
Dahl, Ö. (2004) *The growth and maintenance of linguistic complexity*. Benjamins. [TB]
Dale, R. & Lupyan, G. (2012) Understanding the origins of morphological diversity: The linguistic niche hypothesis. *Advances in Complex Systems* 15(3–4):1–16. Available at: http://doi.org/10.1142/S0219525911500172. [TB]
Darwin, C. (1871) *The descent of man, and selection in relation to sex, vol. 1*. John Murray. [rNC]
Davenport, J. L. & Potter, M. C. (2004) Scene consistency in object and background perception. *Psychological Science* 15:559–64. [MCP]
Davenport, J. L. & Potter, M. C. (2005) The locus of semantic priming in RSVP target search. *Memory and Cognition* 33:241–48. [MCP]
Dawkins, R. (1986) *The blind watchmaker: Why the evidence of evolution reveals a universe without design*. Norton. [rNC]
de Vries, M. H., Christiansen, M. H. & Petersson, K. M. (2011) Learning recursion: Multiple nested and crossed dependencies. *Biolinguistics* 5:10–35. [aMHC]
de Vries, M. H., Geukes, S., Zwitserlood, P., Petersson, K. M. & Christiansen, M. H. (2012) Processing multiple non-adjacent dependencies: Evidence from sequence learning. *Philosophical Transactions of the Royal Society B: Biological Sciences* 367:2065–76. [rNC]
Dediu, D., Cysouw, M., Levinson, S. C., Baronchelli, A., Christiansen, M. H., Croft, W., Evans, N., Garrod, S., Gray, R., Kandler, A. & Lieven, E. (2013) Cultural evolution of language. In: *Cultural evolution: Society, technology, language and religion*, ed. P. J. Richerson & M. H. Christiansen, pp. 303–32. MIT Press. [aMHC]
Dehaene, S. (2009) *Reading in the brain: The new science of how we read*. Viking. [GB]
Dehaene-Lambertz, G., Dehaene, S., Anton, J. L., Campagne, A., Ciuciu, P., Dehaene, G. P., Denghien, I., Jobert, A., Lebihan, D., Sigman, M., Pallier, C. & Poline, J.-P. (2006a) Functional segregation of cortical language areas by sentence repetition. *Human Brain Mapping* 27:360–71. [aMHC]
Dehaene-Lambertz, G., Hertz-Pannier, L., Dubois, J., Meriaux, S., Roche, A., Sigman, M. & Dehaene, S. (2006b) Functional organization of perisylvian activation during presentation of sentences in preverbal infants. *Proceedings of the National Academy of Sciences* 103:14240–45. [aMHC]
Dell, G. S., Burger, L. K. & Svec, W. R. (1997) Language production and serial order: A functional analysis and a model. *Psychological Review* 104:123–47. [aMHC]
Dell, G. S. & Chang, F. (2014) The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions*

of the Royal Society B: Biological Sciences 369(1634):20120394. [aMHC]

DeLong, K. A., Urbach, T. P. & Kutas, M. (2005) Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience* 8:1117–21. [aMHC]

Dikker, S., Rabagliati, H., Farmer, T. A. & Pylkkänen, L. (2010) Early occipital sensitivity to syntactic category is based on form typicality. *Psychological Science* 21:629–34. [aMHC]

Dillon, B., Mishler, A., Sloggett, S. & Phillips, C. (2013) Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language* 69:85–103. [DAC]

Ding, N., Melloni, L., Zhang, H., Tian, X. & Poeppel, D. (2016) Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience* 19:159–64. [GB]

Dixon, R. & Aikhenvald, A. (2002) Word: A typological framework. In: *Word: A cross-linguistic typology*, ed. R. M. W. Dixon & A. Y. Aikhenvald, pp. 1–41. Cambridge University Press. [SCL]

Dominey, P. F., Hoen, M., Blanc, J.-M. & Lelekov-Boissard, T. (2003) Neurological basis of language and sequential cognition: Evidence from simulation, aphasia and ERP studies. *Brain and Language* 86:207–25. [SLF]

Dumitru, M. L. (2014) Moving stimuli guide retrieval and (in)validation of coordination simulations. *Cognitive Processing* 15(3):397–403. [MLD]

Dumitru, M. L., Joergensen, G. H., Cruickshank, A. G. & Altmann, G. T. M. (2013) Language-guided visual processing affects reasoning: The role of referential and spatial anchoring. *Consciousness and Cognition* 22(2):562–71. [MLD]

Dumitru, M. L. & Taylor, A. J. (2014) *Or* cues knowledge of alternatives: Evidence from priming. *Scandinavian Journal of Psychology* 55(2):97–101. [MLD]

Dunn, M., Greenhill, S. J., Levinson, S. C. & Gray, R. D. (2011) Evolved structure of language shows lineage-specific trends in word-order universals. *Nature* 473:79–82. [aMHC]

Durrant, P. (2013) Formulaicity in an agglutinating language: The case of Turkish. *Corpus Linguistics and Linguistic Theory* 9:1–38. [aMHC]

Dryer, M. (1992) The Greenbergian word order correlations. *Language* 68:81–138. [DPM]

Dryer, M. S. (2009) The branching direction theory of word order correlations revisited. In: *Universals of language today*, ed. S. Scalise, E. Magni & A. Bisetto, pp. 185–207. Springer Netherlands. [DPM]

Dyer, C., Ballesteros, M., Ling, W., Matthews, A. & Smith, N. (2015) Transition-based dependency parsing with stack long short-term memory. In: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Beijing, China, July 26–31, 2015*, ed. C. Zong & M. Strube, pp. 334–43. Association for Computational Linguistics. [CG-R]

Edelman, S. (2008a) *Computing the mind: How the mind really works*. Oxford University Press. [AL]

Edelman, S. (2008b) On the nature of minds, or: Truth and consequences. *Journal of Experimental and Theoretical AI* 20:181–96. [AL]

Edelman, S. (2015) The minority report: Some common assumptions to reconsider in the modeling of the brain and behavior. *Journal of Experimental and Theoretical Artificial Intelligence* 27, doi 10.1080/0952813X.2015.1042534. [AL]

Eggermont, J. (2001) Between sound and perception: Reviewing the search for a neural code. *Hearing Research* 157(1): 1–42. [CRH]

Eldredge, N. & S. J. Gould (1972) Punctuated equilibria: An alternative to phyletic gradualism. In: *Models in paleobiology*, ed. T. J. M. Schopf, pp. 82–115. Freeman Cooper. [rNC]

Elliott, L. L. (1962) Backward and forward masking of probe tones of different frequencies. *Journal of the Acoustical Society of America* 34:1116–17. [aMHC, CJH]

Ellis, N. C. (2002) Frequency effects in language processing. *Studies in Second Language Acquisition* 24:143–88. [aMHC]

Elman, J. L. (1990) Finding structure in time. *Cognitive Science* 14(2):179–211. [aMHC, ADE, rNC]

Emmorey, K. (1995) Processing the dynamic visual-spatial morphology of signed languages. In: *Morphological aspects of language processing: Cross-linguistic perspectives*, ed. L. M. Feldman, pp. 29–54. Erlbaum. [KE]

Emmorey, K., Bosworth, R. & Kraljic, T. (2009) Visual feedback and self-monitoring of sign language. *Journal of Memory and Language* 61:398–11. [KE]

Endress, A. D. & Potter, M. C. (2014a) Large capacity temporary visual memory. *Journal of Experimental Psychology: General*. 143(1):548–65. doi: 10.1037/a0033934. [MCP, rNC]

Endress, A. D. & Potter, M. C. (2014b) Something from (almost) nothing: Buildup of object memory from forgettable single fixations. *Attention, Perception, and Psychophysics* 76:2413–23. [ADE]

Enfield, N. J. (2013) *Relationship thinking: Enchrony, agency, and human sociality*. Oxford University Press. [aMHC]

Engelmann, F. & Vasishth, S. (2009) Processing grammatical and ungrammatical center embeddings in English and German: A computational model. In: *Proceedings of the Ninth International Conference on Cognitive Modeling, Manchester, UK, July 2009*, pp. 240–45. ed. A. Howes, D. Peebles & R. Cooper. [rNC]

Erickson, T. A. & Matteson, M. E. (1981) From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior* 20:540–52. [aMHC]

Ericsson, K. A., Chase, W. G. & Faloon, S. (1980) Acquisition of a memory skill. *Science* 208:1181–82. [aMHC, MLD, rNC]

Eshghi, A., Hough, J. & Purver, M. (2013) Incremental grammar induction from child-directed dialogue utterances. In: *Proceedings of the 4th Annual ACL Workshop on Cognitive Modeling and Computational Linguistics (CMCL)*, Sofia, Bulgaria, August 8 2013, pp. 94–103, ed. V. Demberg & R. Levy. Association for Computational Linguistics. [PGTH]

Eshghi, A., Howes, C., Gregoromichelaki, E., Hough, J. & Purver, M. (2015) Feedback in conversation as incremental semantic update. In: *Proceedings of the 11th International Conference on Computational Semantics (IWCS)*, London, UK, April 15–17 2015, pp. 261–71, ed. M. Purver, M. Sadrzadeh & M. Stone. Association for Computational Linguistics. [PGTH]

Evans, N. & Levinson, S. (2009) The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 32:429–92. [aMHC, ADE, rNC]

Falconer, D. S. (1981) *Introduction to quantitative genetics*. Longman. [AL]

Farmer, T. A., Christiansen, M. H. & Monaghan, P. (2006) Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences* 103:12203–208. [aMHC]

Farmer, T. A., Monaghan, P., Misyak, J. B. & Christiansen, M. H. (2011) Phonological typicality influences sentence processing in predictive contexts: A reply to Staub et al. (2009) *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37:1318–25. [aMHC]

Federmeier, K. D. (2007) Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology* 44:491–505. [aMHC]

Fedzechkina, M., Jaeger, T. F. & Newport, E. L. (2012) Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences* 109:17897–17902. [MLL]

Feldman, N. H., Griffiths, T. L. & Morgan, J. L. (2009) The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review* 116:752–82. [KB]

Ferreira, F. (1993) Creation of prosody during sentence production. *Psychological Review* 100:233. [rNC]

Ferreira, F. (2003) The misinterpretation of noncanonical sentences. *Cognitive Psychology* 47(2):164–203. [FF, rNC]

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

Ferreira, F. & Swets, B. (2002) How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language* 46(1):57–84. [DAC, FF, MCM, rNC]

Ferreira, V. (2008) Ambiguity, accessibility, and a division of labor for communicative success. *Psychology of Learning and Motivation* 49:209–46. [aMHC]

Ferrer i Cancho, R. (2004) The Euclidean distance between syntactically linked words. *Physical Review E* 70:056135. [aMHC]

Ferrer i Cancho, R. & Liu, H. (2014) The risks of mixing dependency lengths from sequences of different length. *Glottotheory* 5:143–55. [aMHC]

Fine, A. B., Jaeger, T. F., Farmer, T. A. & Qian, T. (2013) Rapid expectation adaptation during syntactic comprehension. *PLOS ONE* 8(10):e77661. [aMHC]

Fiske, S. T. & Taylor, S. E. (1984) *Social cognition*. Addison-Wesley. [aMHC]

Fitneva, S. A., Christiansen, M. H. & Monaghan, P. (2009) From sound to syntax: Phonological constraints on children's lexical categorization of new words. *Journal of Child Language* 36:967–97. [aMHC]

Fitz, H. (2011) A liquid-state model of variability effects in learning nonadjacent dependencies. In: *Proceedings of the 33rd Annual Conference of the Cognitive Science Society, Boston, MA, July 2011*, ed. L. Carlson, C. Hölscher & T. Shipley, pp. 897–902. Cognitive Science Society. [SLF, rNC]

Flanagan, J. R. & Wing, A. M. (1997) The role of internal models in motor planning and control: Evidence from grip force adjustments during movements of hand-held loads. *Journal of Neuroscience* 17:1519–28. [aMHC]

Fleischman, S. (1982) *The future in thought and language: Diachronic evidence from Romance*. Cambridge University Press. [aMHC]

Fodor, J. D. (1998) Unambiguous triggers. *Linguistic Inquiry* 29:1–36. [aMHC]

Ford, M., Bresnan, J. W. & Kaplan, R. M. (1982) A competence-based theory of syntactic closure. In: *The mental representation of grammatical relations*, ed. J. W. Bresnan, pp. 727–96. MIT Press. [aMHC]

Forster, K. I. (1970) Visual perception of rapidly presented word sequences of varying complexity. *Perception & Psychophysics* 8:215–21. [MCP]

Forster, K. I. & Davis, C. (1984) Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 10:680–698. [MCP]

Franck, J., Soare, G., Frauenfelder, U. H. & Rizzi, L. (2010) Object interference: The role of intermediate traces of movement. *Journal of Memory and Language* 62:166–82. [DAC]

Frank, M., Goldwater, S., Griffiths, T. & Tenenbaum, J. (2010) Modeling human performance in statistical word segmentation. *Cognition* 117:107–25. [ADE]

Frank, M. C. & Goodman, N. (2012) Predicting pragmatic reasoning in language games. *Science* 336:998–98. [MLL]

Frank, M. C. & Goodman, N. D. (2014) Inferring word meanings by assuming that speakers are informative. *Cognitive Psychology* 75:90–96. [MLL]

Frank, M. C., Goodman, N. D. & Tenenbaum, J. B. (2009) Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science* 20:579–85. [MLL]

Frank, S. L. & Bod, R. (2011) Insensitivity of the human sentence-processing system to hierarchical structure. *Psychological Science* 22:829–34. [SLF, rNC]

Frank, S. L., Bod, R. & Christiansen, M. H. (2012) How hierarchical is language use? *Proceedings of the Royal Society B: Biological Sciences* 279:4522–31. [aMHC]

Frank, S. L., Trompenaars, T. & Vasishth, S. (2016) Cross-linguistic differences in processing double-embedded relative clauses: Working-memory constraints or language statistics? *Cognitive Science*. 40:554–78. [rNC]

Frazier, L. & Fodor, J. D. (1978) The sausage machine: A new two-stage parsing model. *Cognition* 6:291–25. [aMHC, DAC, rNC]

Frazier, L. & Rayner, K. (1982) Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology* 14(2):178–10. [FF]

French, R. M. (1999) Catastrophic forgetting in connectionist networks. *Trends in Cognitive Sciences* 3:128–35. [aMHC, rNC]

Frost, R., Armstrong, B. C., Siegelman, N. & Christiansen, M. H. (2015) Domain generality versus modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences* 19:117–25. [rNC]

Fuster, J. & Alexander, G. (1971) Neuron activity related to short-term memory. *Science* 173:652–54. [CRH]

Fuster, J. M. (1995) Temporal processing. *Annals of the New York Academy of Science* 769:173–82. [GB]

Fuster, J. M. (1997) Network memory. *Trends in Neuroscience* 20:451–59. [CJH]

Gahl, S. & Garnsey, S. M. (2004) Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. *Language* 80:748–75. [aMHC]

Gallace, A., Tan, H. Z. & Spence, C. (2006) The failure to detect tactile change: A tactile analogue of visual change blindness. *Psychonomic Bulletin & Review* 13:300–303. [aMHC]

Gärdenfors, P. (2004) Cooperation and the evolution of symbolic communication. In: *Evolution of communication systems*, ed. D. K. Oller & U. Griebel, pp. 237–56. MIT Press. [AB]

Gärdenfors, P. & Rott, H. (1995) Belief revision. In: *Handbook of logic in artificial intelligence and logic programming, vol. 4*, ed. D. Gabbay, pp. 35–132. Oxford University Press. [rNC]

Garnham, A. & Oakhill, J. (1985) On-line resolution of anaphoric pronouns: Effects of inference making and verb semantics. *British Journal of Psychology* 76:385–93. [rNC]

Garrett, M. F. (1980) Levels of processing in sentence production. In: *Language production: Vol. 1. Speech and talk*, ed. B. Butterworth. pp. 177–221. Academic Press. [DAC, rNC]

Gee, J. & Grosjean, F. (1983) Performance structures: A psycholinguistics and linguistic appraisal. *Cognitive Psychology* 15:411–58. [MLD]

Gerstner, W., Kistler, W. M., Naud, R. & Paninski, L. (2014) *Neuronal dynamics: From single neurons to networks and models of cognition*. Cambridge University Press. [SLF]

Gertner, Y. & Fisher, C. (2012) Predicted errors in children's early sentence comprehension. *Cognition* 124:85–94. [rNC]

Gibson, E. (1998) Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68:1–76. [aMHC]

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* 110:8051–56. [aMHC]

Gibson, E. & Thomas, J. (1999) Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes* 14:225–48. [rNC]

Gibson, E. & Wexler, K. (1994) Triggers. *Linguistic Inquiry* 25:407–54. [aMHC, rNC]

Gil, D. (2009) How much grammar does it take to sail a boat? In: Language complexity as an evolving variable, ed. G. Sampson, D. Gil & P. Trudgill, pp. 19–33. Oxford University Press. [CB]

Gilbreth, F. B. & Gilbreth, L. M. (1919) *Applied motion study*. Macmillan. [RM]

Gildea, D. & Temperley, D. (2010) Do grammars minimize dependency length? *Cognitive Science* 34:286–310. [aMHC]

Gimenes, M., Rigalleau, F. & Gaonac'h, D. (2009) When a missing verb makes a French sentence more acceptable. *Language and Cognitive Processes* 24:440–49. [rNC]

Giraud, A. L., Kleinschmidt, A., Poeppel, D., Lund, T. E., Frackowiak, R. S. J. & Laufs, H. (2007) Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron* 56(6):1127–34. [GB]

Giraud, A. L. & Poeppel, D. (2012) Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience* 15(4):511–17. [GB]

Givón, T. (1971) Historical syntax and synchronic morphology: An archaeologist's field trip. *Chicago Linguistic Society* 7:394–415. [aMHC]

Givón, T. (1979) *On understanding grammar*. Academic Press. [aMHC]

Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C. H., Jones, G., Oliver, I. & Pine, J. M. (2001) Chunking mechanisms in human learning. *Trends in Cognitive Sciences* 5:236–43. [aMHC]

Gold, E. M. (1967) Language identification in the limit. *Information and Control* 10:447–74. [aMHC]

Goldberg, A. (2006) *Constructions at work.* Oxford University Press. [aMHC, rNC]

Goldinger, S. D. (1998) Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105:251–79. [aMHC]

Goldstein, M. H., Waterfall, H. R., Lotem, A., Halpern, J., Schwade, J., Onnis, L. & Edelman, S. (2010) General cognitive principles for learning structure in time and space. *Trends in Cognitive Sciences* 14:249–58. [AL]

Golinkoff, R. M. (1986) "I beg your pardon?": The preverbal negotiation of failed messages. *Journal of Child Language* 13(3):455–76. [GB]

Golinkoff, R. M., Hirsh-Pasek, K., Bloom, L., Smith, L., Woodward, A., Akhtar, N., Tomasello, M. & Hollich, G., eds. (2000) *Becoming a word learner: A debate on lexical acquisition*. Oxford University Press. [aMHC]

Gómez, R. L. (2002) Variability and detection of invariant structure. *Psychological Science* 13:431–36. [rNC]

Gómez-Rodríguez, C. & Nivre, J. (2013) Divisible transition systems and multiplanar dependency parsing. *Computational Linguistics* 39(4):799–45. [CG-R]

Gómez-Rodríguez, C., Sartorio, F. & Satta, G. (2014) A polynomial-time dynamic oracle for non-projective dependency parsing. In: *Proceedings of the 2014 Conference on Empirical Methods on Natural Language Processing, Doha, Qatar, October 25–29, 2014*, ed. A. Moschitti, B. Pang & W. Daelemans, pp. 917–27. Association for Computational Linguistics. [CG-R]

Goodman, N. D. & Stuhlmüller, A. (2013) Knowledge and implicature: Modeling language understanding as social cognition. *Topics in Cognitive Science* 5:173–84. [MLL]

Goodwin, C. (1979) The interactive construction of a sentence in natural conversation. In: *Everyday language: Studies in ethnomethodology*, ed. G. Psathas, pp. 97–121. Irvington Publishers. [PGTH]

Gordon, R. G. (2005) *Ethnologue: Languages of the world, 15th edition*. SIL International. [TB]

Gorrell, P. (1995) *Syntax and parsing*. Cambridge University Press. [aMHC]

Goswami, U. (2015) Sensory theories of developmental dyslexia: Three challenges for research. *Nature Reviews Neuroscience* 16:43–54. [GB]

Gould, S. J. & Vrba, E. S. (1982) Exaptation: A missing term in the science of form. *Paleobiology* 8:4–15. [AB]

Graesser, A. C., Singer, M. & Trabasso, T. (1994) Constructing inferences during narrative text comprehension. *Psychological Review* 101(3):371–95. [GB]

Grandin, T. (2005) *Animals in translation*. Scribner. [CB]

Graybiel, A. M. (1998) The basal ganglia and chunking of action repertoires. *Neurobiology of Learning and Memory* 70:119–36. [aMHC]

Green, S. & Marler, P. (1979) The analysis of animal communication. In: *Handbook of behavioral neurobiology: Vol. 3. Social behavior and communication*, ed. P. Marler & J. G. Vandenbergh, pp. 73–158. Plenum Press. [AL]

Greenberg, J. (1963) Some universals of grammar with particular reference to the order of meaningful elements. In: *Universals of language*, ed. J. Greenberg, pp. 73–113. MIT Press. [DPM]

Greenwald, A. G. (1970) Sensory feedback mechanisms in performance control: With special reference to the ideo-motor mechanism. *Psychological Review* 77:73–99. [AB]

Greenwald, A. G. (1972) On doing two things at once: Time sharing as a function of ideomotor compatibility. *Journal of Experimental Psychology* 94:52–57. [AB]

Gregoromichelaki, E., Kempson, R., Howes, C. & Eshghi, A. (2013) On making syntax dynamic: The challenge of compound utterances and the architecture of the grammar. In: *Alignment in communication: Towards a new theory of communication*, ed. I. Wachsmuth, J. de Ruiter, P. Jaecks & S. Kopp, pp. 57–86. John Benjamins. [RK]

Gregoromichelaki, E., Kempson, R., Purver, M., Mills, G., Cann, R., Meyer-Viol, W. & Healey, P. (2011) Incrementality and intention-recognition in utterance processing. *Dialogue and Discourse* 2:199–33. Available at: http://dad.uni-bielefeld.de/index.php/dad/article/view/363/1460. [RK]

Gregory, R. L. (2005) The Medawar lecture 2001 knowledge for vision: Vision for knowledge. *Philosophical Transactions of the Royal Society B: Biological Sciences* 360:1231–51. [rNC]

Grice, H. P. (1967) *Logic and conversation*. William James Lectures. Manuscript, Harvard University. [aMHC]

Grice, H. P. (1975) Logic and conversation. In: *Syntax and Semantics*, ed. P. Cole & J. Morgan, pp. 41–58. Academic Press. [MLL]

Griffin, Z. M. & Bock, K. (2000) What the eyes say about speaking. *Psychological Science* 4:274–79. [SCL]

Griffiths, T. L. & Tenenbaum, J. B. (2009) Theory-based causal induction. *Psychological Review* 116:661–716. [aMHC]

Grosjean, F. (1985) The recognition of words after their acoustic offset: Evidence and implications. *Perception and Psychophysics* 38:299–10. [KB]

Grossberg, S. (1973) Contour enhancement, short-term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics* 52:213–57. [SG]

Grossberg, S. (1978a) A theory of human memory: Self-organization and performance of sensory- motor codes, maps, and plans. In: *Progress in theoretical biology, volume* 5, ed. R. Rosen & F. Snell, pp. 233–74. Academic Press. [SG]

Grossberg, S. (1978b) Behavioral contrast in short-term memory: Serial binary memory models or parallel continuous memory models? *Journal of Mathematical Psychology* 3:199–19. [SG]

Grossberg, S. (1986) The adaptive self-organization of serial order in behavior: Speech, language, and motor control. In: *Pattern recognition by humans and machines, vol. 1: Speech perception*, ed. E. C. Schwab & H. C. Nusbaum, pp. 187–94. Academic Press. [SG]

Grossberg, S. (2003) Resonant neural dynamics of speech perception. *Journal of Phonetics* 31:423–45. [SG]

Grossberg, S. (2013) Adaptive resonance theory: How a brain learns to consciously attend, learn, and recognize a changing world. *Neural Networks* 37:1–47. [SG]

Grossberg, S., Boardman, I. & Cohen, C. (1997) Neural dynamics of variable-rate speech categorization. *Journal of Experimental Psychology: Human Perception and Performance* 23:418–503. [SG]

Grossberg, S. & Kazerounian, S. (2011) Laminar cortical dynamics of conscious speech perception: A neural model of phonemic restoration using subsequent context in noise. *Journal of the Acoustical Society of America* 130:440–60. [SG]

Grossberg, S. & Myers, C. W. (2000) The resonant dynamics of speech perception: Interword integration and duration-dependent backward effects. *Psychological Review* 107:735–67. [SG]

Grossberg, S. & Pearson, L. (2008) Laminar cortical dynamics of cognitive and motor working memory, sequence learning and performance: Toward a unified theory of how the cerebral cortex works. *Psychological Review* 115:677–32. [SG]

Gualmini, A. & Crain, S. (2005) The structure of children's linguistic knowledge. *Linguistic Inquiry* 36(3):463–74. [DPM]

Gurevich, O., Johnson, M. A. & Goldberg, A. E. (2010) Incidental verbatim memory for language. *Language and Cognition* 2:45–78. [aMHC]

Haber, R. N. (1983) The impending demise of the icon: The role of iconic processes in information processing theories of perception. *Behavioral and Brain Sciences* 6:1–11. [aMHC]

Hagoort, P. (2009) Reflections on the neurobiology of syntax. In: *Biological foundations and origin of syntax. Strüngmann Forum Reports, volume 3*, ed. D. Bickerton & E. Szathmáry, pp. 279–96. MIT Press. [aMHC]

Hagstrom, P. & Rhee, R. (1997) The dependency locality theory in Korean. *Journal of Psycholinguistic Research* 26:189–206. [rNC]

Hahn, U. & Nakisa, R. C. (2000) German inflection: Single route or dual route? *Cognitive Psychology* 41:313–60. [aMHC]

Hakes, D. T., Evans, J. S. & Brannon, L. L (1976) Understanding sentences with relative clauses. *Memory and Cognition* 4:283–90. [rNC]

Hale, J. (2001) A probabilistic Earley parser as a psycholinguistic model. *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics, Pittsburgh, PA, June 2–7, 2001*, pp. 159–66. Association for Computational Linguistics. [aMHC]

Hale, J. (2006) Uncertainty about the rest of the sentence. *Cognitive Science* 30:609–42. [aMHC]

Hall, M. & Bavelier, D. (2009) Short-term memory stages in sign versus speech: The source of the serial span discrepancy. *Cognition* 120:54–66. [KE]

Hamilton, H. W. & Deese, J. (1971) Comprehensibility and subject-verb relations in complex sentences. *Journal of Verbal Learning and Verbal Behavior* 10:163–70. [rNC]

Haspelmath, M. (1999) Why is grammaticalization irreversible? *Linguistics* 37:1043–68. [aMHC]

Hasson, U., Chen, J. & Honey, C. J. (2015) Hierarchical process memory: Memory as an integral component of information processing. *Trends in Cognitive Sciences* 19(6):304–13. doi:10.1016/j.tics.2015.04.006. [CJH, rNC]

Hasson, U., Yang, E., Vallines, I., Heeger, D. J. & Rubin, N. (2008) A hierarchy of temporal receptive windows in human cortex. *The Journal of Neuroscience* 28 (10):2539–50. [aMHC, CJH]

Hatori, J., Matsuzaki, T., Miyao, Y. & Tsujii, J. (2012) Incremental joint approach to word segmentation, POS tagging, and dependency parsing in Chinese. In: *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Jeju Island, Korea, July 8–14, 2012*, ed. H. Li, C-Y. Lin, M. Osborne, G. G. Lee & J. C. Park, pp. 1216–24. Association for Computational Linguistics. [CG-R]

Hauser, M. D., Chomsky, N. & Fitch, W. T. (2002) The faculty of language: What is it, who has it, and how did it evolve? *Science* 298:1569–79. [aMHC]

Hauser, M. D., Yang, C., Berwick, R. C., Tattersall, I., Ryan, M., Watumull, J., Chomsky, N. & Lewontin, R. (2014) The mystery of language evolution. *Frontiers of Psychology* 5:401. doi:10.3389/fpsyg.2014.00401. [CB]

Hawkins, J. A. (1983) *Word order universals.* Academic Press. [DPM]

Hawkins, J. A. (1990) A parsing theory of word order universals. *Linguistic Inquiry* 21:223–61. [DPM]

Hawkins, J. A. (2004) *Efficiency and complexity in grammars*. Oxford University Press. [aMHC, DPM, WO]

Hawkins, J. A. (2009) Language universals and the performance-grammar correspondence hypothesis. In: *Language universals*, ed. M. H. Christiansen, C. Collins & S. Edelman, pp. 54–78. Oxford University Press. [aMHC]

Hawkins, J. A. (2014) *Cross-linguistic variation and efficiency*. Oxford University Press. [WO]

Hay, J. (2000) Morphological adjacency constraints: A synthesis. *Proceedings of Student Conference in Linguistics 9. MIT Working Papers in Linguistics* 36:17–29. [aMHC]

Healey, P. G. T., Eshghi, A., Howes, C. & Purver, M. (2011) Making a contribution: Processing clarification requests in dialogue. In: *Proceedings of the 21st Annual Meeting of the Society for Text and Discourse*. Poitiers, France, July 11–13 2011, [PGTH]

Healey, P. G. T., Purver, M. & Howes, C. (2014) Divergence in dialogue. *PLOS ONE* 9(6):e98598. doi: 10.1371/journal.pone.0098598. [PGTH]

Heathcote, A., Brown, S. & Mewhort, D. J. K. (2000) The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin and Review* 7:185–207. [aMHC]

Hebb, D. (1949) *The organization of behavior*. Wiley. [CRH]

Heider, P., Dery, J. & Roland, D. (2014) The processing of *it* object relative clauses: Evidence against a fine-grained frequency account. *Journal of Memory and Language* 75:58–76. [aMHC]

Heine, B. & Kuteva, T. (2002) *World lexicon of grammaticalization*. Cambridge University Press. [aMHC]

Heine, B. & Kuteva, T. (2007) *The genesis of grammar*. Oxford University Press. [aMHC]

Hickok, G. & Poeppel, D. (2000) Towards a functional anatomy of speech perception. *Trends in Cognitive Sciences* 4:131–38. [SCL]

Hinaut, X. & Dominey, P. F. (2013) Real-time parallel processing of grammatical structure in the fronto-striatal system: A recurrent network simulation study using reservoir computing. *PLOS ONE* 8(2):e52946. [SLF]

Hinojosa, J. A., Moreno, E. M., Casado, P., Munõz, F. & Pozo, M. A. (2005) Syntactic expectancy: An event-related potentials study. *Neuroscience Letters* 378:34–39. [aMHC]

Hintzman, D. L. (1988) Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review* 95:528–51. [aMHC, rNC]

Hinton, G. E. & Sejnowski, T. J. (1986) *Learning and relearning in Boltzmann machines*. In: *Graphical models: Foundations of neural computation*, ed. M. J. Irwin & T. J. Sejnowski, pp. 45–76. MIT Press. [aMHC]

Hoey, M. (2005) *Lexical priming: A new theory of words and language*. Psychology Press. [aMHC]

Hofmeister, P. & Sag, I. A. (2010) Cognitive constraints and island effects. *Language* 86:366–415. [aMHC]

Holmes, V. M. (1988) Hesitations and sentence planning. *Language and Cognitive Processes* 3:323–61. [rNC]

Hommel, B., Müsseler, J., Aschersleben, G. & Prinz, W. (2001) The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences* 24:849–78. [AB]

Honey, C. J., Thesen, T., Donner, T. H., Silbert, L. J., Carlson, C. E., Devinsky, O., Doyle, W. K., Rubin, N., Heeger, D. J. & Hasson, U. (2012) Slow cortical dynamics and the accumulation of information over long timescales. *Neuron* 76(2):423–34. [aMHC, CJH]

Hoover, M. L. (1992) Sentence processing strategies in Spanish and English. *Journal of Psycholinguistic Research* 21:275–99. [rNC]

Hopper, P. J. & Traugott, E. C. (1993) *Grammaticalization*. Cambridge University Press. [aMHC, rNC]

Horn, L. (1984) Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In: *Meaning, form, and use in context*, ed. D. Schiffrin, pp. 11–42. Georgetown University Press. [MLL]

Horn, L. R. (1972) On the semantic properties of logical operators in English. Unpublished doctoral dissertation, UCLA. [MLL]

Horst, J. S. & Samuelson, L. K. (2008) Fast mapping but poor retention in 24-month-old infants. *Infancy* 13:128–57. [rNC]

Hough, J. & Purver, M. (2012) Processing self-repairs in an incremental type-theoretic dialogue system. In: *Proceedings of the 16th SemDial Workshop on the Semantics and Pragmatics of Dialogue (SeineDial),* Paris, France, September 19–21, 2012, pp. 136–44, ed. S. Brown-Schmidt, J. Ginzburg & S. Larsson. SemDial [PGTH]

Hough, J. & Purver, M. (2013) Modelling expectation in the self-repair processing of annotat-, um, listeners. In: *Proceedings of the 17th SemDial Workshop on the Semantics and Pragmatics of Dialogue (DialDam),* Amsterdam, Netherlands, December 16–18, 2013, pp. 92–101, ed. R. Fernández & A. Isard. SemDial. [PGTH]

Howes, A., Lewis, R. L. & Vera, A. (2009) Rational adaptation under task and processing constraints: Implications for testing theories of cognition and action. *Psychological Review* 116:717–51. [KB]

Howes, C., Purver, M., McCabe, R., Healey, P. G. & Lavelle, M. (2012) Helping the medicine go down: Repair and adherence in patient-clinician dialogues. In: *Proceedings of the 16th SemDial Workshop on the Semantics and Pragmatics of Dialogue (SeineDial),* Paris, France, September 19–21 2012, pp. 155–56, ed. S. Brown-Schmidt, J. Ginzburg & S. Larsson. SemDial. [PGTH]

Hruschka, D., Christiansen, M. H., Blythe, R. A., Croft, W., Heggarty, P., Mufwene, S. S., Pierrehumbert, J. H. & Poplack, S. (2009) Building social cognitive models of language change. *Trends in Cognitive Sciences* 13:464–69. [aMHC]

Hsiao, Y., Gao, Y. & MacDonald, M. C. (2014) Agent-patient similarity affects sentence structure in language production: Evidence from subject omissions in Mandarin. *Frontiers in Psychology* 5:1015. [MCM]

Hsu, A., Chater, N. & Vitányi, P. (2011) The probabilistic analysis of language acquisition: Theoretical, computational, and experimental analysis. *Cognition* 120:380–90. [aMHC]

Huang, C.-T. J. (1982) Logical relations in Chinese and the theory of grammar. Unpublished doctoral dissertation. Departments of Linguistics and Philosophy, Massachusetts Institute of Technology. [ADE]

Hudson Kam, C. L. & Newport, E. L. (2005) Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development* 1:151–95. [aMHC]

Hurford, J. (1999) The evolution of language and languages. In: *The evolution of culture*, ed. R. Dunbar, C. Knight & C. Power, pp. 173–93. Edinburgh University Press. [aMHC]

Huyck, C. (2009) A psycholinguistic model of natural language parsing implemented in simulated neurons. *Cognitive Neurodynamics* 3(4):316–30. [CRH]

Indefrey, P. & Levelt, W. J. M. (2004) The spatial and temporal signatures of word production components. *Cognition* 92(1–2):101–44. [SCL]

Intraub, H. (1981) Rapid conceptual identification of sequentially presented pictures. *Journal of Experimental Psychology: Human Perception and Performance* 7:604–10. [MCP]

Ito, Y. (2005) A psycholinguistic approach to *wanna* contraction in second language acquisition. Unpublished doctoral dissertation. University of Hawaii at Manoa. Department of Second Language Studies. [WO]

Jackendoff, R. (2002) *Foundations of language*. Oxford. [SCL]

Jackendoff, R. (2007) A parallel architecture perspective on language processing. *Brain Research* 1146:2–22. [aMHC, MLD]

Jacoby, L. L., Baker, J. G. & Brooks, L. R. (1989) The priority of the specific: Episodic effects in picture identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15:275–81. [aMHC]

Jaeger, H. & Haas, H. (2004) Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science* 304:78–80. [SLF]

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

Jaeger, T. F. & Tily, H. (2011) On language "utility": Processing complexity and communicative efficiency. *Wiley Interdisciplinary Reviews: Cognitive Science* 2:323–35. [aMHC]

Jaeggli, O. (1980) Remarks on *to*-contraction. *Linguistic Inquiry* 11:239–45. [WO]

Jensen, M. S., Yao, R., Street, W. N. & Simons, D. J. (2011) Change blindness and inattentional blindness. *Wiley Interdisciplinary Reviews: Cognitive Science* 2:529–46. [aMHC]

Johnson, J. S. & Newport, E. L. (1989) Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*. http://doi.org/10.1016/0010-0285(89)90003-0. [RM]

Johnson-Laird, P. N. (1983) *Mental models: Towards a cognitive science of language, inference, and consciousness*. Harvard University Press. [aMHC]

Jolsvai, H., McCauley, S. M. & Christiansen, M. H. (2013) Meaning overrides frequency in idiomatic and compositional multiword chunks. In: *Proceedings of the 35th Annual Conference of the Cognitive Science Society, ,Berlin, Germany, July 31–August 3, 2013,*, ed. M. Knauff, M. Pauen, N. Sebanz & I. Wachsmuth, pp. 692–97. Cognitive Science Society. [aMHC]

Jones, G. (2012) Why chunking should be considered as an explanation for developmental change before short-term memory capacity and processing speed. *Frontiers in Psychology* 3:167. doi: 10.3389/fpsyg.2012.00167. [aMHC, rNC]

Jurafsky, D. (1996) A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science* 20:137–94. [aMHC]

Jurafsky, D., Bell, A., Gregory, M. L. & Raymond, W. D. (2001) Probabilistic relations between words: Evidence from reduction in lexical production. In: *Frequency and the emergence of linguistic structure*, ed. J. L. Bybee & P. Hopper, pp. 229–54. John Benjamins. [aMHC]

Jurafsky, D., Martin, J. H., Kehler, A., Vander Linden, K. & Ward, N. (2000) *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Prentice Hall. [aMHC]

Just, M. & Carpenter, P. (1980) A theory of reading: From eye fixations to comprehension. *Psychological Review* 87(4):123–54. [CRH]

Just, M. A. & Carpenter, P. A. (1992) A capacity theory of comprehension: Individual differences in working memory. *Psychological Review* 99:122. [rNC]

Kam, X.-N. C. & Fodor, J. D. (2013) Children's acquisition of syntax: Simple models are too simple. In: *Rich languages from poor inputs*, ed. M. Piattelli-Palmarini & R. C. Berwick, pp. 43–60. Oxford University Press. [DPM]

Kam, X. N. C., Stoyneshka, I., Tornyova, L., Fodor, J. D. & Sakas, W. G. (2008) Bigrams and the richness of the stimulus. *Cognitive Science* 32(4):771–87. [FHW, rNC]

Kamide, Y. (2008) Anticipatory processes in sentence processing. *Language and Linguistic Compass* 2:647–70. [aMHC]

Kamide, Y., Altmann, G. T. M. & Haywood, S. L. (2003) The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language* 49:133–56. [MLD]

Kamin, L. J. (1969) Predictability, surprise, attention and conditioning. In: *Punishment and aversive behavior*, ed. B. A. Campbell, R. M. Church, pp. 279–96. Appleton-Century-Crofts. [aMHC]

Karlsson, F. (2007) Constraints on multiple center-embedding of clauses. *Journal of Linguistics* 43:365–92. [aMHC, rNC]

Kashima, Y., Bekkering, H. & Kashima, E. S. (2013) Communicative intentions can modulate the linguistic perception-action link. *Behavioral and Brain Sciences* 36:33–34. [AB]

Katz, J. J. (1984) An outline of Platonist grammar. In: *Talking minds: The study of language in cognitive science*, ed. T. G. Bever, J. M. Carroll & L. A. Miller, pp. 17–48. MIT Press. [CB]

Katz, J. J. (1996) The unfinished Chomskyan revolution. *Mind and Language* 11:270–94. [CB]

Katz, J. J. (1998) *Realistic rationalism*. MIT Press. [CB]

Katz, J. J. (2004) *Sense, reference, and philosophy*. Oxford University Press. [CB]

Katz, J. J. & Postal, P. M. (1991) Realism vs. conceptualism in linguistics. *Linguistics and Philosophy* 14:515–54. [CB]

Kazerounian, S. & Grossberg, S. (2014) Real-time learning of predictive recognition categories that chunk sequences of items stored in working memory. *Frontiers in Psychology: Language Sciences.* doi: 10.3389/fpsyg.2014.01053. [SG]

Kelly, B., Wigglesworth, G., Nordlinger, R. & Blythe, J. (2014) The acquisition of polysynthetic languages. *Language and Linguistics Compass* 8(2):51–64. [UL, rNC]

Kemp, C. & Regier, T. (2012) Kinship categories across languages reflect general communicative principles. *Science* 336:1049–54. [MLL]

Kempson, R., Gregoromichelaki, E. & Howes, C. (2011) *The dynamics of lexical interfaces*. CSLI Press. [RK]

Kempson, R. & Kiaer, J. (2010) Multiple long-distance scrambling: Syntax as reflections of processing. *Journal of Linguistics* 6:127–92. [RK]

Kempson, R., Meyer-Viol W. & Gabbay, D. (2001) *Dynamic syntax: The flow of language understanding*. Blackwell. [aMHC, RK]

Kendrick, K. H. & Torreira, F. (2015) The timing and construction of preference: A quantitative study. *Discourse Processes* 52(4):255–89. [SCL]

Kimball, J. (1973) Seven principles of surface structure parsing in natural language. *Cognition* 2:15–47. [aMHC]

Kirby, S., Cornish, H. & Smith, K. (2008) Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences* 105:10681–85. [aMHC, MLL, rNC]

Kirby, S., Tamariz, M., Cornish, H. & Smith, K. (2015) Compression and communication drive the evolution of language. *Cognition* 141:87–102. [MLL]

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

Kluender, R. & Kutas, M. (1993) Subjacency as a processing phenomenon. *Language & Cognitive Processes* 8:573–633. [aMHC, DAC]

Koelsch, S. (2005) Neural substrates of processing syntax and semantics in music. *Current Opinion in Neurobiology* 15(2):207–12. [UL]

Kolers, P. A. & Roediger, H. L. (1984) Procedures of mind. *Journal of Verbal Learning and Verbal Behavior* 23:425–49. [rNC]

Kolodner, J. (1993) *Case-based reasoning*. Morgan Kaufmann. [rNC]

Kolodny, O., Edelman, S. & Lotem, A. (2014) The evolution of continuous learning of the structure of the environment. *Journal of the Royal Society Interface* 11:20131091. [AL, rNC]

Kolodny, O., Edelman, S. & Lotem, A. (2015a) Evolution of protolinguistic abilities as a by-product of learning to forage in structured environments. *Proceedings of the Royal Society of London B* 282(1811):20150353. [AL, rNC]

Kolodny, O., Lotem, A. & Edelman, S. (2015b) Learning a generative probabilistic grammar of experience: A process-level model of language acquisition. *Cognitive Science* 39:227–67. [AL, rNC]

Konopka, A. E. (2012) Planning ahead: How recent experience with structures and words changes the scope of linguistic planning. *Journal of Memory and Language* 66:143–62. [aMHC]

Kraljic, T. & Brennan, S. (2005) Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? *Cognitive Psychology* 50:194–231. [aMHC]

Kuhl, P. K. (2004) Early language acquisition: Cracking the speech code. *Nature Reviews. Neuroscience* 5(11):831–43. http://doi.org/10.1038/nrn1533. [RM]

Kuperberg, G. & Jaeger, T. F. (2016) What do we mean by prediction in language comprehension? *Language, Cognition, and Neuroscience* 31:32–59. [KB]

Kutas, M., Federmeier, K. D. & Urbach, T. P. (2014) The "negatives" and "positives" of prediction in language. In: *The Cognitive Neurosciences V*, ed. M. S. Gazzaniga & G. R. Mangun, pp. 649–56. MIT Press. [aMHC]

Kutas, M. & Hillyard, S. A. (1980) Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science* 207:203–205. [MCP]

Lachmann, M., Számadó, S. & Bergstrom, C. T. (2001) Cost and conflict in animal signals and human language. *Proceedings of the National Academy of Science* 98:13189–94. [AL]

Lake, J. K., Humphreys, K. R. & Cardy, S. (2011) Listener vs. speaker-oriented aspects of speech: Studying the disfluencies of individuals with autism spectrum disorders. *Psychonomic Bulletin and Review* 18(1):135–40. [PGTH]

Lakshmanan, U. (2000) The acquisition of relative clauses by Tamil children. *Journal of Child Language* 21:587–17. [UL]

Lakshmanan, U. (2006) Assessing linguistic competence: Verbal inflection in child Tamil. *Language Assessment Quarterly* 3(2):171–205. [UL]

Lashley, K. S. (1951) The problem of serial order in behavior. In: *Cerebral mechanisms in behavior: The Hixon Symposium*, ed. L. A. Jeffress, pp. 112–46. Wiley. [aMHC, MCM]

Lawrence, D. H. (1971a) Temporal numerosity estimates for word lists. *Perception and Psychophysics* 10:75–78. [MCP]

Lawrence, D. H. (1971b) Two studies of visual search for word targets with controlled rates of presentation. *Perception and Psychophysics* 10:85–89. [MCP]

Lee, E. K., Brown-Schmidt, S. & Watson, D. G. (2013) Ways of looking ahead: Hierarchical planning in language production. *Cognition* 129:544–62. [DAC, rNC]

Leger, D. W. (1993) Contextual sources of information and responses to animal communication signals. *Psychological Bulletin* 113:295–304. [AL]

Legge, G. E., Pelli, D. G., Rubin, G. S. & Schleske, M. M. (1985) Psychophysics of reading – I. Normal vision. *Vision Research* 25:239–52. [rNC]

Lenneberg, E. H. (1964) A biological perspective of language. In: *New directions in the study of language*, ed. E. Lenneberg, pp. 65–88. MIT Press. [RM]

Lerdahl, F. & Jackendoff, R. (1983) *A generative theory of tonal music*. MIT Press. [aMHC]

Lerner, Y., Honey, C. J., Silbert, L. J. & Hasson, U. (2011) Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *Journal of Neuroscience* 31:2906–15. [aMHC, CJH]

Levelt, W. (2012) *A history of psycholinguistics: The pre-Chomskyan era*. Oxford University Press. [DAC]

Levelt, W. J. (1993) *Speaking: From intention to articulation, vol. 1*. MIT Press. [PGTH]

Levelt, W. J. M. (2001) Spoken word production: A theory of lexical access. *Proceedings of the National Academy of Sciences* 98:13464–71. [aMHC]

Levinson, S. C. (2000) *Presumptive meanings: The theory of generalized conversational implicature*. MIT Press. [aMHC, rNC]

Levinson, S. C. (2013) Recursion in pragmatics. *Language* 89:149–62. [aMHC]

Levinson, S. C. & Torreira, F. (2015) Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology* 6:731. doi:10.3389/fpsyg.2015.00731. [SCL]

Levy, R. (2008) Expectation-based syntactic comprehension. *Cognition* 106:1126–77. [aMHC, ADE]

Levy, R. (2011) Integrating surprisal and uncertain-input models in online sentence comprehension: formal techniques and empirical results. In: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, Portland, OR, June 19–24, 2011*, pp. 1055–65, ed. Y. Matsumoto & R. Mihalcea.. Association for Computational Linguistics. [KB]

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* 106:21086–90. [aMHC, KB, MCM]

Lewis, M. & Frank, M. C. (2015) Conceptual complexity and the evolution of the lexicon. In: *Proceedings of the 37th Annual Meeting of the Cognitive Science Society, Austin, TX, July 2015*, ed. D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings & P. P. Magli, pp. 1138–343. Cognitive Science Society. [MLL]

Lewis, M., Sugarman, E. & Frank, M. C. (2014) The structure of the lexicon reflects principles of communication. In: *Proceedings of the 36th Annual Meeting of the Cognitive Science Society, Austin, TX, July 2014*, pp. 845–50, Cognitive Science Society. [MLL]

Lewis, R. & Vasishth, S. (2005) An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science* 29(3):375–419. [CRH]

Lewis, R. L., Shvartsman, M. & Singh, S. (2013) The adaptive nature of eye movements in linguistic tasks: How payoff and architecture shape speed-accuracy trade-offs. *Topics in Cognitive Science* 5:1–30. [KB]

Lewis, R. L., Vasishth, S. & Van Dyke, J. A. (2006) Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences* 10:447–54. [DAC]

Liberman, A. M., Harris, K. S., Hoffman, H. S. & Griffith, B. C. (1957) The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology* 54:358–68. [KB]

Lightfoot, D. (1989) The child's trigger experience: Degree-0 learnability. *Behavioral and Brain Sciences* 12:321–34. [aMHC]

Lightfoot, D. (1991) *How to set parameters: Arguments from language change*. MIT Press. [rNC]

Lim, J.-H. & Christianson, K. (2015) L2 Sensitivity to agreement errors: Evidence from eye movements during comprehension and translation. *Applied Psycholinguistics* 36:1283–315. [FF]

Lind, J., Enquist, M. & Ghirlanda, S. (2015) Animal memory: A review of delayed matching-to-sample data. *Behavioural Processes* 117:52–58. [AL]

Loftus, G. R. (1983) Eye fixations on text and scenes. In: *Eye movements in reading: Perceptual and language processes*, ed. K. Rayner, pp. 359–76. Academic Press. [MCP]

Logan, G. D. (1988) Toward an instance theory of automatization. *Psychological Review* 95:492–527. [aMHC]

Lombardi, L. & Potter, M. C. (1992) The regeneration of syntax in short term memory. *Journal of Memory and Language* 31:713–33. [MCP]

Lotem, A. & Halpern, J. Y. (2008) A data-acquisition model for learning and cognitive development and its implications for autism. Computing and information science technical reports, Cornell University. Available at: http://hdl.handle.net/1813/10178. [AL]

Lotem, A. & Halpern, J. Y. (2012) Coevolution of learning and data-acquisition mechanisms: A model for cognitive evolution. *Philosophical Transactions of the Royal Society B* 367:2686–94. [AL]

Louwerse, M. M., Dale, R., Bard, E. G. & Jeuniaux, P. (2012) Behavior matching in multimodal communication is synchronized. *Cognitive Science* 36:1404–26. [aMHC]

Luck, S. J. & Vogel, E. K. (1997) The capacity of visual working memory for features and conjunctions. *Nature* 390:279–81. [rNC]

Luck, S. J., Vogel, E. K. & Shapiro, K. L. (1996) Word meanings can be accessed but not reported during the attentional blink. *Nature* 383:616–18. [MCP]

Lukoševičius, M. & Jaeger, H. (2009) Reservoir computing approaches to recurrent neural network training. *Computer Science Review* 3:127–49. [SLF]

Lupyan, G. & Christiansen, M. H. (2002) Case, word order, and language learnability: Insights from connectionist modeling. In: *Proceedings of the 24th Annual Conference of the Cognitive Science Society, Fairfax, VA, August 2002*, pp. 596–601, ed. W. D. Gray & C. Schunn. Erlbaum. [aMHC]

Lupyan, G. & Dale, R. (2010) Language structure is partly determined by social structure. *PLOS ONE* 5(1):1–10. http://doi.org/10.1371/journal.pone.0008559. [TB]

Lupyan, G. & Dale, R. A. (2015) The role of adaptation in understanding linguistic diversity. In: *Language structure and environment: Social, cultural, and natural factors*, ed. R. de Busser & R. J. LaPolla, pp. 287–16. John Benjamins. [TB]

Maass, W., Natschläger, T. & Markram, H. (2002) Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Computation* 14:2531–60. [SLF]

MacDonald, M. C. (1994) Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes* 9:157–201. [aMHC, MCM]

MacDonald, M. C. (2013) How language production shapes language form and comprehension. *Frontiers in Psychology* 4:226. doi: 10.3389/fpsyg.2013.00226. [aMHC, MCM]

MacDonald, M. C. & Christiansen, M. H. (2002) Reassessing working memory: A comment on Just & Carpenter (1992) and Waters & Caplan (1996). *Psychological Review* 109:35–54. [aMHC, rNC]

MacDonald, M. C., Pearlmutter, M. & Seidenberg, M. (1994) The lexical nature of ambiguity resolution. *Psychological Review* 101:676–703. [aMHC]

MacKay, D. G. (1987) *The organization of perception and action: A theory for language and other cognitive skills*. Springer. [rNC]

MacKay, D. J. (2003) *Information theory, inference and learning algorithms*. Cambridge University Press. [aMHC]

Magyari, L. & de Ruiter, J. P. (2012) Prediction of turn-ends based on anticipation of upcoming words. *Frontiers in Psychology* 3:376. doi: 10.3389/fpsyg.2012.00376 [aMHC]

Mahowald, K., Fedorenko, E., Piantadosi, S. & Gibson, E. (2012) Info/information theory: Speakers actively choose shorter words in predictable contexts. *Cognition* 126:313–18.

Major, G. & Tank, D. (2004) Persistent neural activity: Prevalence and mechanisms. *Current Opinion in Neurobiology* 14(6):675–84. [ADE, rNC]

Mandel, D. R., Jusczyk, P. W. & Pisoni, D. B. (1995) Infants' recognition of the sound patterns of their own names. *Psychological Science* 6(5):314–17. [RM]

Mani, N. & Huettig, F. (2012) Prediction during language processing is a piece of cake – but only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance* 38:843–47. [aMHC]

Manning, C. D. & Schütze, H. (1999) *Foundations of statistical natural language processing*. MIT press. [aMHC]

Marchman, V. A. & Fernald, A. (2008) Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science* 11:F9–16. [aMHC]

Marcus, M. P. (1980) *Theory of syntactic recognition for natural languages*. MIT Press. [aMHC]

Markson, L. & Bloom, P. (1997) Evidence against a dedicated system for word learning in children. *Nature* 385:813–15. [ADE]

Marr, D. (1976) Early processing of visual information. *Philosophical Transactions of the Royal Society of London, B: Biological Sciences* 275:483–519. [aMHC]

Marr, D. (1982) *Vision*. W. H. Freeman. [aMHC]

Marslen-Wilson, W. (1987) Functional parallelism in spoken word recognition. *Cognition* 25:71–102. [rNC]

Marslen-Wilson, W. D. (1975) Sentence perception as an interactive parallel process. *Science* 189:226–28. [aMHC]

Marslen-Wilson, W. D., Tyler, L. K. & Koster, C. (1993) Integrative processes in utterance resolution. *Journal of Memory and Language* 32:647–66. [aMHC]

Martin, R. C. & He, T. (2004) Semantic STM and its role in sentence processing: A replication. *Brain and Language* 89:76–82. [aMHC]

Martin, R. C., Shelton, J. R. & Yaffee, L. S. (1994) Language processing and working memory: Neuropsychological evidence for separate phonological and semantic capacities. *Journal of Memory and Language* 33:83–111. [aMHC]

Maye, J., Werker, J. F. & Gerken, L. (2002) Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82:101–111. [RM]

McCauley, S. M. & Christiansen, M. H. (2011) Learning simple statistics for language comprehension and production: The CAPPUCCINO model. In: *Proceedings of the 33rd Annual Conference of the Cognitive Science Society, Boston, MA, July 2011.* pp. 1619–24, ed. L. A. Carlson, C. Hölscher & T. F. Shipley. Cognitive Science Society. [aMHC, ADE]

McCauley, S. M. & Christiansen, M. H. (2013) Toward a unified account of comprehension and production in language development. *Behavioral and Brain Sciences* 36:366–67. [rNC]

McCauley, S. M. & Christiansen, M. H. (2014a) Acquiring formulaic language: A computational model. *Mental Lexicon* 9:419–36. [rNC]

McCauley, S. M. & Christiansen, M. H. (2014b) Reappraising lexical specificity in children's early syntactic combinations. In: *Proceedings of the 36th Annual Conference of the Cognitive Science Society.* pp. 1000–1005. ed. P. Bello, M. Guarini, M. McShane, & B. Scassellati. Cognitive Science Society. [aMHC]

McCauley, S. M. & Christiansen, M. H. (2015a) *Computational investigations of multiword chunks in language learning.* Submitted. [rNC]

McCauley, S. M. & Christiansen, M. H. (2015b) Individual differences in chunking ability predict on-line sentence processing. In: *Proceedings of the 37th Annual Conference of the Cognitive Science Society, Pasadena, CA, July 2015*, pp. 1553–58, ed. D. C. Noelle & R. Dale. Cognitive Science Society. [rNC]

McCauley, S. M. & Christiansen, M. H. (2015c) *Language learning as language use: A cross-linguistic model of child language development*. Manuscript in preparation. [rNC]

McCauley, S. M., Monaghan, P. & Christiansen, M. H. (2015) Language emergence in development: A computational perspective. In: *The handbook of language emergence*, ed. B. MacWhinney & W. O'Grady, pp. 415–36. Wiley-Blackwell. [rNC]

McClelland, J. L. (1987) The case for interactionism in language processing. In: *Attention & performance XII: The psychology of reading*, ed. M. Coltheart, pp. 3–35. Erlbaum. [UL]

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

McCloskey, M. & Cohen, N. J. (1989) Catastrophic interference in connectionist networks: The sequential learning problem. *Psychology of Learning and Motivation* 24:109–65. [aMHC]

McElree, B., Foraker, S. & Dyer, L. (2003) Memory structures that subserve sentence comprehension. *Journal of Memory and Language* 48:67–91. [DAC]

McMurray, B., Horst, J. S. & Samuelson, L. K. (2012) Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review* 119:831–77. [rNC]

McMurray, B., Tanenhaus, M. K. & Aslin, R. N. (2009) Within-category VOT affects recovery from "lexical" garden-paths: Evidence against phoneme-level inhibition. *Journal of Memory and Language* 60:65–91. [KB]

McWhorter, J. (2001) The world's simplest grammars are creole grammars. *Linguistic Typology* 5(2):125–66. Available at: http://doi.org/10.1515/lity.2001.001. [TB]

Meng, M. & Potter, M. C. (2008) Detecting and remembering pictures with and without visual noise. *Journal of Vision* 8(9):7. Available at: http://journalofvision.org/8/9/7/ [MCP]

Meng, M. & Potter, M. C. (2011) An attentional blink for nontargets? *Attention, Perception, and Psychophysics* 73:440–46. [MCP]

Menyhart, O., Kolodny, O., Goldstein, M. H., Devoogd, T. & Edelman, S. (2015) Juvenile zebra finches learn the underlying statistical regularities in their father's song. *Frontiers in Psychology* 6:571. [AL]

Mermillod, M., Bugaïska, A. & Bonin, P. (2013) The stability-plasticity dilemma: Investigating the continuum from catastrophic forgetting to age-limited learning effects. *Frontiers in Psychology* 4:504. doi: 10.3389/fpsyg.2013.00504. [aMHC]

Mery, F., Belay, A. T., So, A. K., Sokolowski, M. B. & Kawecki, T. J. (2007) Natural polymorphism affecting learning and memory in Drosophila. *Proceedings of the National Academy of Science* 104:13051–55. [AL]

Mesulam, M. M. (1998) From sensation to cognition. *Brain* 121(6):1013–52. [GB]

Meyer, D. E. & Schvaneveldt, R. W. (1971) Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *Journal of Experimental Psychology* 90:227–34. [aMHC]

Meyer, A. S. (1996) Lexical access in phrase and sentence production: Results from picture-word interference experiments. *Journal of Memory and Language* 35:477–96. [aMHC, DAC]

Miller, G. A. (1956) The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63(2):81–97. [aMHC, SCL, rNC]

Miller, G. A., Galanter, E. & Pribram, K. H. (1960) *Plans and the structure of behavior*. Holt, Rinehart & Winston. [aMHC]

Miller, G. A. & Taylor, W. G. (1948) The perception of repeated bursts of noise. *Journal of the Acoustic Society of America* 20:171–82. [aMHC]

Mintz, T. H., Wang, F. H. & Li, J. (2014) Word categorization from distributional information: Frames confer more than the sum of their (Bigram) parts. *Cognitive Psychology* 75:1–27. [FHW]

Misyak, J. B. & Christiansen, M. H. (2010) When "more" in statistical learning means "less" in language: Individual differences in predictive processing of adjacent dependencies. In: *Proceedings of the 32nd Annual Cognitive Science Society Conference, Portland, OR, August 11–14, 2010*, ed. R. Catrambone & S. Ohlsson, pp. 2686–91. Cognitive Science Society. [aMHC]

Momma, S., Slevc, L. R. & Phillips, C. (2015) The timing of verb planning in active and passive sentence production. *Poster presented at the 28th annual CUNY Conference on Human Sentence Processing, Los Angeles, CA, March 19–21, 2015.* [DAC]

Momma, S., Slevc, L. R. & Phillips, C. (in press) The timing of verb planning in Japanese sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. [DAC]

Monaghan, P. & Christiansen, M. H. (2008) Integration of multiple probabilistic cues in syntax acquisition. In: *Trends in corpus research: Finding structure in data* (TILAR Series), ed. H. Behrens, pp. 139–63. John Benjamins. [aMHC]

Monaghan, P. & Christiansen, M. H. (2010) Words in puddles of sound: Modelling psycholinguistic effects in speech segmentation. *Journal of Child Language* 37:545–64. [rNC]

Mongillo, G., Barak, O. & Tsodyks, M. (2008) Synaptic theory of working memory. *Science* 319:1543–46. [SLF]

Morgan, J. L. & Demuth, K. (1996) *Signal to syntax: Bootstrapping from speech to grammar in early acquisition.* Erlbaum. [aMHC]

Morillon, B., Lehongre, K., Frackowiak, R. S. J., Ducorps, A., Kleinschmidt, A., Poeppel, D. & Giraud, A. L. (2010) Neurophysiological origin of human brain asymmetry for speech and language. *Proceedings of the National Academy of Sciences* 107(43):18688–93. [GB]

Morrill, G. (2010) *Categorial grammar: Logical syntax, semantics, and processing*. Oxford University Press. [aMHC]

Morrill, T. H., Dilley, L. C., McAuley, J. D. & Pitt, M. A. (2014) Distal rhythm influences whether or not listeners hear a word in continuous speech: Support for a perceptual grouping hypothesis. *Cognition* 131:69–74. [MCM]

Mueller, S. T. & Krawitz, A. (2009) Reconsidering the two-second decay hypothesis in verbal working memory. *Journal of Mathematical Psychology* 53:14–25. [AL]

Mufwene, S. (2008) *Language evolution: Contact, competition and change*. Continuum International Publishing Group. [aMHC]

Murdock B. B., Jr. (1968) Serial order effects in short-term memory. *Journal of Experimental Psychology Monograph Supplement* 1–15. [aMHC]

Murdock, B. B. (1983) A distributed memory model for serial-order information. *Psychological Review* 90:316–38. [rNC]

Nadig, A. S. & Sedivy, J. C. (2002) Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science* 13:329–36. [aMHC]

Navon, D. & Miller, J. O. (2002) Queuing or sharing? A critical evaluation of the single-bottleneck notion. *Cognitive Psychology* 44:193–251. [aMHC]

Neal, R. M. & Hinton, G. E. (1998) A view of the EM algorithm that justifies incremental, sparse, and other variants. In: *Learning in graphical models*, ed. M. I. Jordan, pp. 355–68. Kluwer. [aMHC]

Neef, M. (2014) Das nächste Paradigma: Realistische Linguistik. Eine Ergänzung zum Beitrag Wo stehen wir in der Grammatiktheorie? von Wolfgang Sternefeld und Frank Richter. *Muttersprache* 124:105–20. [CB]

Neely, J. H. (1991) Semantic priming effects in visual word recognition: A selective review of current findings and theories. In: *Basic processes in reading: Visual word recognition*, ed. D. Besner & G. W. Humphreys, pp. 264–36. Erlbaum. [MCP]

Neumann, R., Rafferty, A. N. & Griffiths, T. L. (2014) A bounded rationality account of wishful thinking. In: *Proceedings of the 36th Annual Conference of the Cognitive Science Society, Quebec City, Canada, July 2014, Quebec City, Quebec, Canada,* pp. 1210–15, ed. P. Bello, M. Guarini, M. McShane, & B. Scassellati. Cognitive Science Society. [KB]

Nevins, A. (2010) *Locality in vowel harmony*. MIT Press. [aMHC]

Newell, A. & Rosenbloom, P. S. (1981) Mechanisms of skill acquisition and the law of practice. In: *Cognitive skills and their acquisition*, ed. J. R. Anderson, pp. 1–55. Erlbaum. [aMHC]

Newport, E. L. (1990) Maturational constraints on language learning. *Cognitive Science* 14:11–28. [RM]

Nicol, J. L., Forster, K. I. & Veres, C. (1997) Subject-verb agreement processes in comprehension. *Journal of Memory and Language* 36:569–87. [aMHC]

Nieuwland, M. S. & Van Berkum, J. J. (2006) When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience* 18(7):1098–11. [CJH]

Nivre, J. (2003) An efficient algorithm for projective dependency parsing. In: *Proceedings of the 8th International Workshop on Parsing Technologies (IWPT 03), Nancy, France, April 23–25, 2003*, ed. H. Bunt & G. van Noord, pp. 149–60. Association for Computational Linguistics. [CG-R]

Nivre, J. (2004) Incrementality in deterministic dependency parsing. In: *Proceedings of the Workshop on Incremental Parsing: Bringing Engineering and Cognition Together*, ed. F. Keller, S. Clark, M. Crocker & M. Steedman, pp. 50–57. Association for Computational Linguistics. [CG-R]

Nivre, J. (2008) Algorithms for deterministic incremental dependency parsing. *Computational Linguistics* 34(4):513–53. [CG-R]

Nivre, J., Hall, J., Nilsson, J., Chanev, A., Eryigit, G., Kübler, S., Marinov, S. & Marsi, E. (2007) MaltParser: A language-independent system for data-driven dependency parsing. *Natural Language Engineering* 13(2):95–35. [CG-R]

Niyogi, P. & Berwick, R. C. (1996) A language learning model for finite parameter spaces. *Cognition* 61:161–93. [aMHC]

Niyogi, P. & Berwick, R. C. (2009) The proper treatment of language acquisition and change in a population setting. *Proceedings of the National Academy of Sciences* 106:10124–29. [ADE]

Norman, D. A. & Shallice, T. (1986) Attention to action: Willed and automatic control of behavior. In: *Consciousness and self-regulation*, ed. R. J. Davidson, G. E. Schwartz & D. Shapiro, pp. 1–18. Plenum Press. [aMHC]

Norris, D., McQueen, J. M., Cutler, A. & Butterfield, S. (1997) The possible word constraint in the segmentation of continuous speech. *Cognitive Psychology* 34:191–43. [KE]

Nosofsky, R. M. (1986) Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General* 115:39. [aMHC]

Noveck, I. A. & Reboul, A. (2008) Experimental pragmatics: A Gricean turn in the study of language. *Trends in Cognitive Sciences* 12:425–31. [aMHC]

Nygaard, L. C., Sommers, M. S. & Pisoni, D. B. (1994) Speech perception as a talker-contingent process. *Psychological Science* 5:42–46. [aMHC]

O'Grady, W. (2005) *Syntactic carpentry: An emergentist approach to syntax*. Erlbaum. [aMHC, WO]

O'Grady, W. (2013) The illusion of language acquisition. *Approaches to Bilingualism* 3:253–85. [aMHC, WO, rNC]

O'Grady, W. (2015a) Anaphora and the case for emergentism. In: *The handbook of language emergence*, ed. B. MacWhinney & W. O'Grady, pp. 100–22. Wiley-Blackwell. [aMHC, WO, rNC]

O'Grady, W. (2015b) Processing determinism. *Language Learning* 65:6–32. [WO]

Oaksford, M. & Chater, N., eds. (1998) *Rational models of cognition.*. Oxford University Press. [aMHC]

Oaksford, M. & Chater, N. (2007) *Bayesian rationality*. Oxford University Press. [aMHC]

Odling-Smee, F. J., Laland, K. N. & Feldman, M. W. (2003) *Niche construction: The neglected process in evolution, vol. MPB 37*. Princeton University Press. [AL]

Ohno, T. & Mito, S. (1988) *Just-in-time for today and tomorrow*. Productivity Press. [aMHC]

Omaki, A., Davidson-White, I., Goro, T., Lidz, J. & Phillips, C. (2014) No fear of commitment: Children's incremental interpretation in English and Japanese. *Language Learning and Development* 10:206–33. [DAC]

Onnis, L. & Spivey, M. J. (2012) Toward a new scientific visualization for the language sciences. *Information* 3:124–50. [AL]

Onnis, L., Waterfall, H. R. & Edelman, S. (2008) Learn locally, act globally: Learning language from variation set cues. *Cognition* 109:423–30. [AL]

Orfanidou, E., Morgan, G., Adam, R. & McQueen, J. (2010) Recognition of signed and spoken language: Different sensory inputs, the same segmentation procedure. *Journal of Memory and Language* 62:272–83. [KE]

Orr, D. B., Friedman, H. L. & Williams, J. C. C. (1965) Trainability of listening comprehension of speeded discourse. *Journal of Educational Psychology* 56:148–56. [aMHC]

Orwin, M., Howes, C. & Kempson, R. (2013) *Language, music and interaction*. College Publications. [aMHC]

Padó, U., Crocker, M. W. & Keller, F. (2009) A probabilistic model of semantic plausibility in sentence processing. *Cognitive Science* 33:794–838. [aMHC]

Pallier, C., Devauchelle, A. D. & Dehaene, S. (2011) Cortical representation of the constituent structure of sentences. *Proceedings of the National Academy of Sciences* 108(6):2522–27. [GB]

Pallier, C., Sebastian-Gallés, N., Dupoux, E., Christophe, A. & Mehler, J. (1998) Perceptual adjustment to time-compressed speech: A cross-linguistic study. *Memory and Cognition* 26(4):844–51. [rNC]

Pani, J. R. (2000) Cognitive description and change blindness. *Visual Cognition* 7:107–26. [aMHC]

Pashler, H. (1988) Familiarity and visual change detection. *Perception and Psychophysics* 44:369–78. [aMHC]

Pashler, H. (1998) *The psychology of attention*. MIT Press. [rNC]

Patson, N. D., Darowski, E. S., Moon, N. & Ferreira, F. (2009) Lingering misinterpretations in garden-path sentences: Evidence from a paraphrasing task. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 35:280–85. [FF]

Pavani, F. & Turatto, M. (2008) Change perception in complex auditory scenes. *Perception and Psychophysics* 70:619–29. [aMHC]

Payne, B. R., Grison, S., Gao, X., Christianson, K., Morrow, D. G. & Stine-Morrow, E. A. (2014) Aging and individual differences in binding during sentence understanding: Evidence from temporary and global syntactic attachment ambiguities. *Cognition* 130:157–73. [FF]

Pearlmutter, N. J., Garnsey, S. M. & Bock, K. (1999) Agreement processes in sentence comprehension. *Journal of Memory and Language* 41:427–56. [aMHC]

Pellegrino, F., Coupé, C. & Marsico, E. (2011) A cross-language perspective on speech information rate. *Language* 87:539–58. [aMHC]

Pelucchi, B., Hay, J. F. & Saffran, J. R. (2009) Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition* 113:244–47. [rNC]

Pereira, P. & Schabes, Y. (1992) Inside–outside reestimation from partially bracketed corpora. *Proceedings of the 30th Annual Meeting on Association for Computational Linguistics, Newark, DE, June 28–July 2, 1992*, pp. 128–35, ed. H. S. Thompson. Association for Computational Linguistics. [aMHC]

Perfors, A., Tenenbaum, J. B. & Regier, T. (2011) The learnability of abstract syntactic principles. *Cognition* 118(3):306–38. doi: 10.1016/j.cognition.2010.11.001. [DPM]

Peterfalvi, J. M. & Locatelli, F (1971) L'acceptabilité des phrases [The acceptability of sentences]. *L'Année Psychologique* 71(2):417–27. [rNC]

Petersson, K. M. & Hagoort, P. (2012) The neurobiology of syntax: Beyond string sets. *Philosophical Transactions of the Royal Society B* 367:1971–83. [SLF]

Petkov, C. I., O'Connor, K. N. & Sutter, M. L. (2007) Encoding of illusory continuity in primary auditory cortex. *Neuron* 54(1):153–65. [GB]

Phillips, B. S. (2006) *Word frequency and lexical diffusion*. Palgrave Macmillan. [aMHC]

Phillips, C. (1996) Merge right: An approach to constituency conflicts. In: *Proceedings of the Fourteenth West Coast Conference on Formal Linguistics, volume 15, Los Angeles, CA, March 1994*, pp. 381–95, ed. J. Camacho, L. Choueiri & M. Watanabe. University of Chicago Press. [aMHC]

Phillips, C. (2003) Linear order and constituency. *Linguistic Inquiry* 34:37–90. [aMHC]

Phillips, C. (2013) Some arguments and nonarguments for reductionist accounts of syntactic phenomena. *Language and Cognitive Processes* 28:156–87. [DAC]

Piantadosi, S., Tily, H. & Gibson, E. (2011) Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences* 108:3526–29. [aMHC, MLL]

Piantadosi, S., Tily, H. & Gibson, E. (2012) The communicative function of ambiguity in language. *Cognition* 122:280–91. [aMHC]

Piattelli-Palmarini, M., Hancock, R. & Bever, T. (2008) Language as ergonomic perfection [Peer Commentary] *Behavioral and Brain Sciences* 31(5):530–31. [DPM]

Pickering, M. J. & Branigan, H. P. (1998) The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language* 39:633–51. [aMHC]

Pickering, M. J. & Garrod, S. (2004) Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences* 27:169–226. [aMHC, rNC]

Pickering, M. J. & Garrod, S. (2007) Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences* 11:105–10. [aMHC, rNC]

Pickering, M. J. & Garrod, S. (2013a) An integrated theory of language production and comprehension. *Behavioral and Brain Sciences* 36: 329–47. [aMHC, AB, rNC]

Pickering, M. J. & Garrod, S. (2013b) How tightly are production and comprehension interwoven? *Frontiers in Psychology* 4:377–92. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3636456/. [RK]

Pierrehumbert, J. (2002) *Word-specific phonetics*. Laboratory Phonology VII. Mouton de Gruyter. [aMHC]

Pietroski, P. (2008) Minimalist meaning, internalist interpretation. *Biolinguistics* 2(4):317–41. [DPM]

Pine, J. M., Freudenthal, D., Krajewski, G. & Gobet, F. (2013) Do young children have adult-like syntactic categories? Zipf's law and the case of the determiner. *Cognition* 127:345–60. [aMHC]

Pinker, S. (1984) *Language learnability and language development*. Harvard University Press. [aMHC]

Pinker, S. (1994) *The language instinct: How the mind creates language*. William Morrow. [aMHC]

Pinker, S. & Bloom, P. (1990) Natural language and natural selection. *Behavioral and Brain Sciences* 13:707–27. [aMHC]

Pinker, S. & Prince, A. (1988) On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition* 28:73–193. [aMHC]

Pinkster, H. (2005) The language of Pliny the Elder. In: *The language of Latin Prose*, ed. T. Reinhardt, M. Lapidge & J. N. Adams, pp. 239–56. Oxford University Press. [SCL]

Pisoni, D. B. & Lazarus, J. H. (1974) Categorical and noncategorical modes of speech perception along the voicing continuum. *Journal of the Acoustical Society of America* 55:328–33. [KB]

Pisoni, D. B. & Tash, J. (1974) Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics* 15:285–90. [KB]

Poesio, M & Rieser, H. (2011) An incremental model of anaphora and reference resolution based on resource situations. *Dialogue and Discourse* 2:235–77. http://dad.uni-bielefeld.de/index.php/dad/article/view/373/1461. [RK]

Poggio, T. & Edelman, S. (1990) A network that learns to recognize 3D objects. *Nature* 343:263–66. [rNC]

Postal, P. M. (2003) Remarks on the foundations of linguistics. *The Philosophical Forum* 34:233–51. [CB]

Postal, P. M. (2009) The incoherence of Chomsky's "Biolinguistic" ontology. *Biolinguistics* 3:104–23. [CB]

Potter, M. C. (1976) Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory* 2:509–22. [MCP]

Potter, M. C. (1984) Rapid serial visual presentation (RSVP): A method for studying language processing. In: *New methods in reading comprehension research*, ed. D. Kieras & M. Just, pp. 91–18. Erlbaum. [MCP]

Potter, M. C. (1993) Very short-term conceptual memory. *Memory and Cognition* 21:156–61. [MCP]

Potter, M. C. (2009) Conceptual short term memory. *Scholarpedia* 5(2):3334. [MCP]

Potter, M. C. (2012) Conceptual short term memory in perception and thought. *Frontiers in Psychology* 3:113. doi: 10.3389/fpsyg.2012.00113. [MCP, rNC]

Potter, M. C. & Faulconer, B. A. (1975) Time to understand pictures and words. *Nature* 253:437–38. [MCP]

Potter, M. C. & Lombardi, L. (1990) Regeneration in the short-term recall of sentences. *Journal of Memory and Language* 29:633–54. [MCP]

Potter, M. C. & Lombardi, L. (1998) Syntactic priming in immediate recall of sentences. *Journal of Memory and Language* 38:265–82. [aMHC, MCP]

Potter, M. C., Kroll, J. F. & Harris, C. (1980) Comprehension and memory in rapid sequential reading. In: *Attention and Performance VIII*, ed. R. Nickerson, pp. 395–18. Erlbaum. [MCP]

Potter, M. C., Kroll, J. F., Yachzel, B., Carpenter, E. & Sherman, J. (1986) Pictures in sentences: Understanding without words. *Journal of Experimental Psychology: General* 115:281–94. [MCP]

Potter, M. C., Moryadas, A., Abrams, I. & Noel, A. (1993) Word perception and misperception in context. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19:3–22. [MCP]

Potter, M. C., Staub, A. & O'Connor, D. H. (2002) The time course of competition for attention: Attention is initially labile. *Journal of Experimental Psychology: Human Perception and Performance* 28:1149–62. [MCP]

Potter, M. C., Stiefbold, D. & Moryadas, A. (1998) Word selection in reading sentences: Preceding versus following contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24:68–100. [MCP]

Potter, M. C., Valian, V. V. & Faulconer, B. A. (1977) Representation of a sentence and its pragmatic implications: Verbal, imagistic, or abstract? *Journal of Verbal Learning and Verbal Behavior* 16:1–12. [MCP]

Potter, M. C., Wyble, B., Hagmann, C. E. & McCourt, E. S. (2014) Detecting meaning in RSVP at 13 ms per picture. *Attention, Perception, and Performance* 76(2):270–79. doi: 10.3758/s13414-013-0605-z. [MCP]

Potter, M. C., Wyble, B., Pandav, R. & Olejarczyk, J. (2010) Picture detection in RSVP: Features or identity? *Journal of Experimental Psychology: Human Perception and Performance* 36:1486–94. [MCP]

Poulet, J. F. A. & Hedwig, B. (2006) New insights into corollary discharges mediated by identified neural pathways. *Trends in Neurosciences* 30:14–21. [aMHC]

Princeton, N. J. & Stromswold, K. (2001) The heritability of language: A review and metaanalysis of twin, adoption, and linkage studies. *Language* 77:647–23. [AL]

Pulman, S. G. (1985) A parser that doesn't. *Proceedings of the Second European Meeting of the Association for Computational Linguistics, Geneva, Switzerland, March 1985*, pp. 128–35, Association for Computational Linguistics. [aMHC, rNC]

Pulman, S. G. (1986) Grammars, parsers, and memory limitations. *Language and Cognitive Processes* 1(3):197–225. [ADE]

Purver, M., Cann, R. & Kempson, R. (2006) Grammars as parsers: Meeting the dialogue challenge. *Research in Language and Computation* 4(2–3):289–326. [PGTH]

Purver, M., Eshghi, A. & Hough, J. (2011) Incremental semantic construction in a dialogue system. *Proceedings of the Ninth International Conference on Computational Semantics (IWCS),*, Oxford, UK, January 12–14, 2011, pp. 365–69, ed. J. Box & S. Pulman. Association for Computational Linguistics. [PGTH]

Quirk, R., Greenbaum, S., Leech, G. N. & Svartvik, J. (1972) *A grammar of contemporary English*. Longman. [DAC]

Rabinovich, M., Huerta, R. & Laurent, G. (2008) Transient dynamics for neural processing. *Science* 321:48–50. [SLF]

Ratcliff, R. (1990) Connectionist models of recognition memory: Constraints imposed by learning and forgetting functions. *Psychological Review* 97:285–308. [aMHC]

Rayner, K., ed. (1983) *Eye movements in reading: Perceptual and language processes*. Academic Press. [MCP]

Rayner, K., ed. (1992) *Eye movements and visual cognition: Scene perception and reading*. Springer. [MCP]

Reali, F. & Christiansen, M. H. (2005) Uncovering the richness of the stimulus: Structure dependence and indirect statistical evidence. *Cognitive Science* 29(6):1007–28. [DPM, FHW, rNC]

Reali, F. & Christiansen, M. H. (2007a) Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language* 57:1–23. [aMHC]

Reali, F. & Christiansen, M. H. (2007b) Word-chunk frequencies affect the processing of pronominal object-relative clauses. *Quarterly Journal of Experimental Psychology* 60:161–70. [aMHC]

Reali, F. & Christiansen, M. H. (2009) On the necessity of an interdisciplinary approach to language universals. In: *Language universals*, ed. M. H. Christiansen, C. Collins & S. Edelman, pp. 266–77. Oxford University Press. [rNC]

Reali, F., Dale, R. & Christiansen, M. H. (2005) *Colorless green ideas sleep furiously* revisited: A statistical perspective. In: *Proceedings of the 27th Annual Meeting of the Cognitive Science Society, Stresa, Italy, July 2005*, pp. 1821–26, ed. B. G. Bara, L. W. Barsalou & M. Bucciarelli. Erlbaum. [rNC]

Reali, F. & Griffiths, T. (2009) The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition* 111:317–28. [MLL, rNC]

Redington, M., Chater, N. & Finch, S. (1998) Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science* 22:425–469. [aMHC]

Regier, T., Kay, P. & Khetarpal, N. (2007) Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences* 104:1436–41. [MLL]

Remez, R. E., Fellowes, J. M. & Rubin, P. E. (1997) Talker identification based on phonetic information. *Journal of Experimental Psychology: Human Perception and Performance* 23:651–66. [aMHC]

Remez, R. E., Ferro, D. F., Dubowski, K. R., Meer, J., Broder, R. S. & Davids, M. L. (2010) Is desynchrony tolerance adaptable in the perceptual organization of speech? *Attention, Perception, and Psychophysics* 72:2054–58. [aMHC]

Richerson, P. J. & Christiansen, M. H., eds. (2013) *Cultural evolution: Society, technology, language and religion.* MIT Press. [aMHC]

Rigotti, M., Barak, O., Warden, M. R., Wang, X. -J., Daw, N. D., Miller, E. K. & Fusi, S. (2013) The importance of mixed selectivity in complex cognitive tasks. *Nature* 497:585–90. [SLF]

Rizzi, L. (1990) *Relativized minimality*. MIT Press. [ADE]

Rizzolatti, G. & Arbib, M. A. (1998) Language within our grasp. [Viewpoint] *Trends in Neurosciences* 21(5):188–94. [DPM]

Roland, D., Elman, J. & Ferreira, V. (2006) Why is that? Structural prediction and ambiguity resolution in a very large corpus of English sentences. *Cognition* 98:245–72. [aMHC]

Rose, Y. & Brittain, J. (2011) Grammar matters: Evidence from phonological and morphological development in Northern East Cree. In: *Selected proceedings of the 4th Conference on Generative Approaches to Language Acquisition North America (GALANA 2010), Somerville, MA*, pp. 193–208, ed. M. Pirvulescu, M. C. Cuervo, A. T. Pérez-Leroux, J. Steele & N. Strik. Cascadilla Proceedings Project. Available at: www.lingref.com, document #2596. [UL]

Ross, J. R. (1967) Constraints on variables in syntax. Unpublished doctoral dissertation. Department of Linguistics, MIT. [ADE]

Rumelhart, D. E. & McClelland, J. L. (1986) On learning the past tenses of English verbs. In: *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 2: Psychological and biological models*, ed. J. L. McClelland, D. E. Rumelhart & the PDP Research Group, pp. 216–71. MIT Press. [aMHC]

Rumelhart, D. E., McClelland, J. L. & the PDP Research Group (1986a) *Parallel distributed processing: Explorations in the microstructure of cognition, volumes 1 and 2*. MIT Press. [aMHC]

Rumelhart, D. E., McClelland, J. L. & the PDP Research Group (1988) *Parallel distributed processing, vol. 1*, pp. 354–62. IEEE. [DPM]

Rumelhart, D. E., Smolensky, P., McClelland, J. L. & Hinton, G. (1986b) Sequential thought processes in PDP models. In: *Parallel distributed processing: explorations in the microstructures of cognition, vol. 2*, ed, J. L. McClelland & D. E. Rumelhart, pp. 3–57. MIT Press. [rNC]

Saad, D., ed. (1998) *On-line learning in neural networks.* Cambridge University Press. [aMHC]

Sacks, H., Schegloff, E. & Jefferson, G. (1974) Simplest systematics for organization of turn-taking for conversation. *Language* 50(4):696–35. [SCL]

Sagarra, N. & Herschensohn, J. (2010) The role of proficiency and working memory in gender and number processing in L1 and L2 Spanish. *Lingua* 120:2022–39. [aMHC]

Sahin, N. T., Pinker, S., Cash, S. S., Schomer, D. & Halgren, E. (2009) Sequential processing of lexical, grammatical, and articulatory information within Broca's area. *Science* 326:445. [aMHC]

Sandler, W. (1986) The spreading hand autosegment of American Sign Language. *Sign Language Studies* 50:1–28. [KE]

Sandler, W. (2012) Dedicated gestures and the emergence of sign language. *Gesture* 12:265–307. [rNC]

Sandler, W., Aronoff, M., Meir, I. & Padden, C. (2011) The gradual emergence of phonological form in a new language. *Natural Language and Linguistic Theory* 29:503–43. [KE]

Sandler, W., Meir, I., Padden, C. & Aronoff, M. (2005) The emergence of grammar: Systematic structure in a new language. *Proceedings of the National Academy of Sciences* 102:2661–65. [aMHC]

Sanford, A. J. & Sturt, P. (2002) Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognitive Sciences* 6:382–86. [rNC]

Sarma, V. (2003) Noncanonical word order: Topic and focus in adult and child Tamil. In: *Word order and scrambling*, ed. S. Karimi, pp. 238–72. Blackwell. [UL]

Saxton, M., Houston-Price, C. & Dawson, N. (2005) The prompt hypothesis: Clarification requests as corrective input for grammatical errors. *Applied Psycholinguistics* 26(3):393–14. [GB]

Schegloff, E. (2007) *Sequence organization in interaction*. Cambridge University Press. [SCL]

Schegloff, E. A., Jefferson, G. & Sacks, H. (1977) The preference for self-correction in the organization of repair in conversation. *Language* 53(2):361–82. [PGTH]

Schlenker, P. (2010) A phonological condition that targets discontinuous syntactic units: Ma/mon suppletion in French. *Snippets* 22:11–13. [SCL]

Schlesinger, I. M. (1975) Why a sentence in which a sentence in which a sentence is embedded is embedded is difficult. *Linguistics* 153:53–66. [rNC]

Schmidt, R. A. & Wrisberg, C. A. (2004) *Motor learning and performance, third edition*. Human Kinetics. [aMHC]

Schultz, W., Dayan, P. & Montague, P. R. (1997) A neural substrate of prediction and reward. *Science* 275:1593–99. [aMHC]

Schwab, E. C., Nusbaum, H. C. & Pisoni, D. B. (1985) Some effects of training on the perception of synthetic speech. *Human Factors* 27:395–408. [ADE]

Seidenberg, M. S. (1997) Language acquisition and use: Learning and applying probabilistic constraints. *Science* 275:1599–603. [aMHC]

Seidenberg, M. S. & McClelland, J. L. (1989) A distributed, developmental model of word recognition and naming. *Psychological Review* 96:523–68. [aMHC]

Seitz, A. R. & Watanabe, T. (2003) Psychophysics: Is subliminal learning really passive? *Nature* 422:36. [ADE]

Senghas, A., Kita, S. & Özyürek, A. (2004) Children creating core properties of language: Evidence from an emerging sign language in Nicaragua. *Science* 305:1779–82. [rNC]

Shanks, D. R. & St. John, M. F. (1994) Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences* 17:367–95. [rNC]

Shannon, C. (1948) A mathematical theory of communication. *Bell System Technical Journal* 27:623–56. [aMHC]

Shin, Y. K., Proctor, R. W. & Capaldi, E. J. (2010) A review of contemporary ideomotor theory. *Psychological Bulletin* 136:943–74. [AB]

Siegel, D. (1978) The adjacency constraint and the theory of morphology. *North East Linguistics Society* 8:189–97. [aMHC]

Sigman, M. & Dehaene, S. (2005) Parsing a cognitive task: A characterization of the mind's bottleneck. *PLOS Biology* 3(2):e37. doi:10.1371/journal.pbio.0030037. [aMHC]

Silbert, L. J., Honey, C. J., Simony, E., Poeppel, D. & Hasson, U. (2014) Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proceedings of the National Academy of Sciences* 111:E4687–96. [aMHC]

Silver, M. R., Grossberg, S., Bullock, D., Histed, M. H. & Miller, E. K. (2011) A neural model of sequential movement planning and control of eye movements: Item-order-rank working memory and saccade selection by the supplementary eye fields. *Neural Networks* 26:29–58. [SG]

Simon, H. A. (1956) Rational choice and the structure of the environment. *Psychological Review* 63:129–38. [aMHC]

Simon, H. A. (1982) *Models of bounded rationality: Empirically grounded economic reason*. MIT Press. [aMHC, KB]

Simon, H. A. & Chase, W. G. (1973) Skill in chess. *American Scientist* 61:393–403. [rNC]

Simons, D. J. & Levin, D. T. (1998) Failure to detect changes to people during a real-world interaction. *Psychonomic Bulletin and Review* 5:644–49. [aMHC]

Singer, W. (2013) Cortical dynamics revisited. *Trends in Cognitive Sciences* 17:616–26. [SLF]

Singleton, J. L. & Newport, E. L. (2004) When learners surpass their models: The acquisition of American Sign Language from inconsistent input. *Cognitive Psychology* 49:370–407. [rNC]

Siyanova-Chanturia, A., Conklin, K. & Van Heuven, W. J. B. (2011) Seeing a phrase "time and again" matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37:776–784. [aMHC]

Slattery, T. J., Sturt, P., Christianson, K., Yoshida, M. & Ferreira, F. (2013) Lingering misinterpretations of garden path sentences arise from flawed semantic processing. *Journal of Memory and Language* 69:104–20. [FF, rNC]

Smith, K. & Kirby, S. (2008) Cultural evolution: Implications for understanding the human language faculty and its evolution. *Philosophical Transactions of the Royal Society B* 363:3591–603. [aMHC]

Smith, K. & Wonnacott, E. (2010) Eliminating unpredictable variation through iterated learning. *Cognition* 116:444–49. [aMHC]

Smith, M. & Wheeldon, L. (1999) High level processing scope in spoken sentence production. *Cognition* 73:205–46. [aMHC, DAC, rNC]

Smith, M. & Wheeldon, L. (2004) Horizontal information flow in spoken sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30:675–686. [MCM, rNC]

Snedeker, J. & Trueswell, J. (2003) Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language* 48:103–30. [aMHC]

Snijders, L. (2012) Issues concerning constraints on discontinuous NPs in Latin. In: *Proceedings of the LFG12 Conference, Bali, Indonesia, June 28–July 1, 2012*, pp. 565–81, ed. M. Butt & T. H. King. CSLI Publications. Available at: http://web.stanford.edu/group/cslipublications/cslipublications/LFG/17/lfg12.html [SCL]

Solan, Z., Horn, D., Ruppin, E. & Edelman, S. (2005) Unsupervised learning of natural languages. *Proceedings of the National Academy of Science* 102:11629–34. [AL]

Sperling, G. (1960) The information available in brief visual presentations. *Psychological Monographs: General and Applied* 74:1–29. [aMHC, KB]

Sperling, G., Budiansky, J. Spivak, J. G. & Johnson, M. C. (1971) Extremely rapid visual search: The maximum rate of scanning letters for the presence of a numeral. *Science* 174:307–11. [MCP]

Staub, A. & Clifton, C., Jr. (2006) Syntactic prediction in language comprehension: Evidence from either … or. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32:425–36. [aMHC]

Steedman, M. (1987) Combinatory grammars and parasitic gaps. *Natural Language and Linguistic Theory* 5:403–39. [aMHC]

Steedman, M. (2000) *The syntactic process*. MIT Press. [aMHC]

Stephens, D. W. & Krebs, J. R. (1986) *Foraging theory*. Princeton University Press. [AL]

Stephens, G. J., Honey, C. J. & Hasson, U. (2013) A place for time: The spatiotemporal structure of neural dynamics during natural audition. *Journal of Neurophysiology* 110(9):2019–26. [aMHC, CJH]

Stephens, G. J., Silbert, L. J. & Hasson, U. (2010) Speaker–listener neural coupling underlies successful communication. *PNAS* 107:14425–30. [aMHC]

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., de Ruiter, J. P., Yoon, K.-Y. & Levinson, S. C. (2009) Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences* 106:10587–92. [aMHC, rNC]

Studdert-Kennedy, M. (1986) Some developments in research on language behavior. In: *Behavioral and social science: Fifty years of discovery: In commemoration of the fiftieth anniversary of the "Ogburn Report," recent social trends in the United States*, ed. N. J. Smelser & D. R. Gerstein, pp. 208–48. National Academy Press. [aMHC]

Sturt, P. & Crocker, M. W. (1996) Monotonic syntactic processing: A cross-linguistic study of attachment and reanalysis. *Language and Cognitive Processes* 11:449–94. [aMHC]

Supalla, S. (1991) Manually coded English: The modality question in signed language development. In: *Theoretical issues in sign language research*, ed. P. Siple & S. D. Fischer, pp. 85–109. University of Chicago Press. [KE]

Supalla, S. & McKee, C. (2002) The role of manually coded English in language development of deaf children. In: *Modality and structure in signed and spoken languages*, ed. R. P. Meier, K. Cormier & D. Quinto-Pozos, pp. 143–65. Cambridge University Press. [KE]

Swaab, T., Brown, C. M. & Hagoort, P. (2003) Understanding words in sentence contexts: The time course of ambiguity resolution. *Brain and Language* 86:326–43. [aMHC]

Swets, B., Desmet, T., Clifton, C. & Ferreira, F. (2008) Underspecification of syntactic ambiguities: Evidence from self-paced reading. *Memory and Cognition* 36:201–16. [rNC]

Swinney, D. A. & Osterhout, L. (1990) Inference generation during auditory language comprehension. *Psychology of Learning and Motivation* 25:17–33. [GB]

Szostak, C. M. & Pitt, M. A. (2013) The prolonged influence of subsequent context on spoken word recognition. *Attention, Perception, and Psychophysics* 75:1533–1546. [KB]

Talmy, L. (2000) *Toward a cognitive semantics.* MIT Press. [RM]

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* 268:1632–34. [aMHC]

Tenenbaum, J. B., Kemp, C., Griffiths, T. L. & Goodman, N. D. (2011) How to grow a mind: Statistics, structure, and abstraction. *Science* 331:1279–85. [aMHC]

Thornton, R., MacDonald, M. C. & Gil, M. (1999) Pragmatic constraint on the interpretation of complex noun phrases in Spanish and English. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 25:1347–65. [aMHC]

Tincoff, R. & Jusczyk, P. W. (1999) Some beginnings of word comprehension in 6-month-olds. *Psychological Science* 10(2):172–75. Available at: http://doi.org/10.1111/1467-9280.00127. [RM]

Tincoff, R. & Jusczyk, P. W. (2012) Six-month-olds comprehend words that refer to parts of the body. *Infancy* 17(4):432–44. Available at: http://doi.org/10.1111/j.1532-7078.2011.00084.x. [RM]

Tomasello, M. (1992) *First verbs: A case study of early grammatical development*. Cambridge University Press. [aMHC]

Tomasello, M. (2003) *Constructing a language: A usage-based theory of language acquisition.* Harvard University Press. [aMHC, DPM, rNC]

Tomasello, M. (2006) Acquiring linguistic constructions. In: *Handbook of child psychology. 2. Cognition, perception, and language*, ed. W. Damon, R. Lerner, D. Kuhn & R. Siegler, pp. 255–98. Wiley. [DPM]

Tomasello, M., Conti-Ramsden, G. & Ewert, B. (1990) Young children's conversations with their mothers and fathers: Differences in breakdown and repair. *Journal of Child Language* 17(1):115–30. [GB]

Townsend, D. J. & Bever, T. G. (2001) *Sentence comprehension: The integration of habits and rules*. MIT Press. [aMHC]

Treisman, A. (1964) Selective attention in man. *British Medical Bulletin* 20:12–16. [aMHC]

Treisman, A. & Schmidt, H. (1982) Illusory conjunctions in the perception of objects. *Cognitive Psychology* 14:107–41. [aMHC]

Tremblay, A. & Baayen, H. (2010) Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In: *Perspectives on formulaic language*, ed. D. Wood, pp. 151–67. Continuum International Publishing. [aMHC]

Tremblay, A., Derwing, B., Libben, G. & Westbury, C. (2011) Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning* 61:569–613. [aMHC]

Trotzke, A., Bader, M. & Frazier, L. (2013) Third factors and the performance interface in language design. *Biolinguistics* 7:1–34. [rNC]

Trudgill, P. (2011) *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford University Press. [aMHC, TB]

Trueswell, J. C. & Tanenhaus, M. K. (1994) Towards a lexicalist framework of constraint-based syntactic ambiguity resolution. In: *Perspectives on sentence processing*, ed. C. Clifton, L. Frazier & K. Rayner, pp. 155–79. Erlbaum. [aMHC]

Trueswell, J. C., Medina, T. N., Hafri, A. & Gleitman, L. R. (2013) Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive Psychology* 66(1):126–56. [FHW]

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* 73:89–134. [aMHC]

Trueswell, J. C., Tanenhaus, M. K. & Garnsey, S. M. (1994) Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language* 33:285–318. [aMHC]

Trueswell, J. C., Tanenhaus, M. K. & Kello, C. (1993) Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19:528–53. [aMHC]

Tylén, K., Christensen, P., Roepstorff, A., Lund, T., Østergaard, S. & Donald, M. (2015) Brains striving for coherence: Long-term cumulative plot formation in the default mode network. *NeuroImage* 121:106–14. [rNC]

Tyler, L. K. & Warren, P. (1987) Local and global structure in spoken language comprehension. *Journal of Memory and Language* 26(6):638–57. [FF]

Uehara, K. & Bradley, D. (1996) The effect of *-ga* sequences on processing Japanese multiply center-embedded sentences. In: *11th Pacific-Asia conference on language, information, and computation,* pp. 187–196. Seoul: Kyung Hee University. [rNC]

Ullman, M. T. (2001) The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research* 30:37–69. [DPM]

Valian, V., Solt, S. & Stewart, J. (2009) Abstract categories or limited-scope formulae? The case of children's determiners. *Journal of Child Language* 36:743–78. [aMHC]

Van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V. & Hagoort, P. (2005) Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31:443–67. [aMHC]

van den Brink, D., Brown, C. M. & Hagoort, P. (2001) Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience* 13:967–85. [aMHC]

Van Dyke, J. A. & Johns, C. L. (2012) Memory Interference as a determinant of language comprehension. *Language and Linguistics Compass* 6:193–11. [MCM, rNC]

Van Everbroeck, E. (1999) Language type frequency and learnability: A connectionist appraisal. In: *Proceedings of the 21st Annual Conference of the Cognitive Science Society, Vancouver, British Columbia, Canada, August 1999*, pp. 755–60, ed. M. Hahn & S. C. Stoness. Erlbaum. [aMHC]

van Gompel, R. P. & Liversedge, S. P. (2003) The influence of morphological information on cataphoric pronoun assignment. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29:128–39. [aMHC]

van Gompel, R. P. G., Pickering, M. J., Pearson, J. & Jacob, G. (2006) The activation of inappropriate analyses in garden-path sentences: Evidence from structural priming. *Journal of Memory and Language* 55:335–62. [FF]

van Soelen, I. L. C., Brouwer, R. M., van Leeuwen, M., Kahn, R. S., Hulshoff Pol, H. E. & Boomsma, D. I. (2011) Heritability of verbal and performance intelligence in a pediatric longitudinal sample. *Twin Research and Human Genetics* 14:119–28. [AL]

Vasishth, S., Suckow, K., Lewis, R. L. & Kern, S. (2010) Short-term forgetting in sentence comprehension: Crosslinguistic evidence from verb-final structures. *Language and Cognitive Processes* 25:533–67. [rNC]

Vicario, C. M., Martino, D. & Koch, G. (2013) Temporal accuracy and variability in the left and right posterior parietal cortex. *Neuroscience* 245:121–28. [GB]

Vogler, C., Gschwind, L., Coyne, D., Freytag, V., Milnik, A., Egli, T., Heck, A., de Quervain, D. J. & Papassotiropoulos, A. (2014) Substantial SNP-based heritability estimates for working memory performance. *Translational Psychiatry* 4:e438. [AL]

Vouloumanos, A. & Werker, J. F. (2007) Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental Science* 10(2):159–64. Available at: http://doi.org/10.1111/j.1467-7687.2007.00549.x. [RM]

Vroomen, J. & de Gelder, B. (1999) Lexical access of resyllabified words: Evidence from phoneme monitoring. *Memory and Cognition* 27(3):413–21. [SCL]

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M. & von der Heydt, R. (2012) A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychological Bulletin* 138:1172–217. [rNC]

Wagers, M., Lau, E. & Phillips, C. (2009) Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language* 61:206–37. [DAC]

Wagner, V., Jescheniak, J. D. & Schriefers, H. (2010) On the flexibility of grammatical advance planning during sentence production: Effects of cognitive load on multiple lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 36:423–40. [MCM]

Wallace, R. S. (2005) *Be your own botmaster, second edition*, ALICE A.I. Foundation. [rNC]

Wang, F. H. & Mintz, T. H. (under revision) The limits of associative learning in cross-situational word learning. [FHW]

Wang, M. D. (1970) The role of syntactic complexity as a determiner of comprehensibility. *Journal of Verbal Learning and Verbal Behavior* 9:398–404. [rNC]

Wang, W. S.-Y. (1969) Competing changes as a cause of residue. *Language* 45:9–25. [aMHC]

Wang, W. S.-Y., ed. (1977) *The lexicon in phonological change*. Mouton. [aMHC, rNC]

Wang, W. S.-Y. & Cheng, C.-C. (1977) Implementation of phonological change: The Shaung-feng Chinese case. In: *The lexicon in phonological change*, ed. W. S.-Y. Wang, pp. 86–100. Mouton. [aMHC]

Warren, P. & Marslen-Wilson, W. (1987) Continuous uptake of acoustic cues in spoken word recognition. *Perception & Psychophysics* 41:262–75. [aMHC]

Warren, P., Speer, S. & Schafer, A. (2003) *Wanna*-contraction and prosodic disambiguation in US and NZ English. *Wellington Working Papers in Linguistics* 15:31–49. [WO]

Warren, R. M. (1970) Perceptual restoration of missing speech sounds. *Science* 167 (3917):392–93. [GB, MCM]

Warren, R. M., Obusek, C. J., Farmer, R. M. & Warren, R. P. (1969) Auditory sequence: Confusion of patterns other than speech or music. *Science* 164:586–87. [aMHC, CJH, rNC]

Warren, R. M. & Sherman, G. L. (1974) Phonemic restorations based on subsequent context. *Perception and Psychophysics* 16:150–56. [MCM]

Wasow, T. & Arnold, J. (2003) Post-verbal constituent ordering in English. In: *Determinants of grammatical variation in English*, ed. G. Rohdenburg & B. Mondorf, pp. 119–54. Mouton de Gruyter. [aMHC]

Watanabe, T., Náñez, J. E. & Sasaki, Y. (2001) Perceptual learning without perception. *Nature* 413:844–48. [ADE]

Waters, A. J., Gobet, F. & Leyden, G. (2002) Visuospatial abilities of chess players. *British Journal of Psychology* 93(4):557–65. [rNC]

Waxman, S. R. & Gelman, S. A. (2009) Early word-learning entails reference, not merely associations. *Trends in Cognitive Sciences* 13(6):258–63. [FHW]

Weber-Fox, C. M. & Neville, H. J. (1996) Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience* 8(3):231–56. [RM]

Weissenborn, J. & Höhle, B., eds. (2001) *Approaches to bootstrapping: Phonological, lexical, syntactic and neurophysiological aspects of early language acquisition*. John Benjamins. [aMHC]

Werker, J. F., Yeung, H. H. & Yoshida, K. A. (2012) How do infants become experts at native-speech perception? *Current Directions in Psychological Science* 21(4):221–26. Available at: http://doi.org/10.1177/0963721412449459. [RM, rNC]

Wexler, K. (2002) Lenneberg's dream: Learning, normal language development and specific language impairment. In: *Language competence across populations: Towards a definition of specific language impairment*, ed. J. Schaffer & Y. Levy, pp. 11–60. Erlbaum. [DPM]

Wicha, N. Y. Y., Moreno, E. M. & Kutas, M. (2004) Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience* 16:1272–88. [aMHC]

Wiener, M., Turkeltaub, P. & Coslett, H. B. (2010) The image of time: A voxel-wise meta-analysis. *NeuroImage* 49(2):1728–40. [GB]

Wilbur, R. B. & Nolkn, S. B. (1986) The duration of syllables in American Sign Language. *Language and Speech* 29:263–80. [aMHC]

Wilson, C. (2006) Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science* 30:945–82. [MLL]

Wilson, M. & Emmorey, K. (1997) A visual-spatial "phonological loop" in working memory: Evidence from American Sign Language. *Memory and Cognition* 25(3):313–20. [KE]

Wilson, M. & Emmorey, K. (1998) A "word length effect" for sign language: Further evidence on the role of language in structuring working memory. *Memory and Cognition* 26(3):584–90. [KE]

Wilson, M. & Emmorey, K. (2006) Comparing sign language and speech reveals a universal limit on short-term memory capacity. *Psychological Science* 17:682–83. [aMHC]

Winograd, T. (1972) Understanding natural language. *Cognitive Psychology* 3:1–191. [aMHC]

Wolpert, D. M., Diedrichsen, J. & Flanagan, J. R. (2011) Principles of sensorimotor learning. *Nature Reviews Neuroscience* 12:739–51. [aMHC, AB]

Wolpert, D. M., Ghahramani, Z. & Flanagan, J. R. (2001) Perspectives and problems in motor learning. *Trends in Cognitive Sciences* 5:487–94. [AB]

Wundt, W. (1904) The psychology of the sentence. In: *Language and psychology: Historical aspects of psycholinguistics*, ed. A. L. Blumenthal, pp. 9–32. Wiley. [DAC]

Xanthos, A., Lahaa, S., Gillis, S., Stefany, U., Aksu-Koc, A., Christofidou, A., Gagarina, N., Hrzica, G., Nihan Ketrez, F., Kilani-Schoch, M., Korecky-Kroll, K., Kovacevic, M., Laalo, K., Palmovic, M., Pfeiler, B., Voeikova, M. D. & Dressler, W. U. (2012) On the role of morphological richness in the early development of noun and verb inflection. *First Language* 31:461–79. [rNC]

Yang, C. (2002) *Knowledge and learning in natural language*. Oxford University Press. [aMHC]

Yang, C. (2013) Ontogeny and phylogeny of language. *Proceedings of the National Academy of Sciences* 110:6324–27. [aMHC]

Yates, F. (1966) *The art of memory.* Routledge & Kegan Paul. [MLD]

Yonata, L. (1999) Early metalinguistic competence: Speech monitoring and repair behavior. *Developmental Psychology* 35(3):822–34. [GB]

Yoshida, K. A., Pons, F., Maye, J. & Werker, J. F. (2010) Distributional phonetic learning at 10 months of age. *Infancy* 15(4):420–33. Available at: http://doi.org/10.1111/j.1532-7078.2009.00024.x. [RM]

Yu, C. & Smith, L. B. (2007) Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science* 18(5):414–20. [FHW]

Yu, C., Smith, L. B., Klein, K. & Shiffrin, R. M. (2007) Hypothesis testing and associative learning in cross-situational word learning: Are they one and the same? In: *Proceedings of the 29th Annual Conference of the Cognitive Science Society, Nashville, TN, August 2007*, pp. 737–42, ed. D. S. McNamara & J. G. Trafton. Cognitive Science Society. [FHW]

Zhang, Y. & Clark, S. (2011) Syntactic processing using the generalized perceptron and beam search. *Computational Linguistics* 37(1):105–51. [CG-R]

Zhu, L., Chen, Y., Torrable, A., Freeman, W. & Yuille, A. L. (2010) Part and appearance sharing: Recursive compositional models for multi-view multi-object detection. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPRW 2010), San Francisco, CA, June 2010*, Institute of Electrical and Electronics Engineers (IEEE). [aMHC]

Zipf, G. (1936) *The psychobiology of language*. Routledge. [MLL]

Zipf, G. K. (1949) *Human behavior and the principle of least effort*. Addison-Wesley. [aMHC, MLL]