Language as skill: Intertwining comprehension and production

Nick Chater [a,*], Stewart M. McCauley [b], Morten H. Christiansen [b,c,d]

[a] Behavioural Science Group, Warwick Business School, University of Warwick, Coventry CV4 7AL, UK
[b] Department of Psychology, Cornell University, Ithaca, NY 14853, USA
[c] Department of Language and Communication, University of Southern Denmark, 5230 Odense M, Denmark
[d] Haskins Laboratories, New Haven, CT 06511, USA

ARTICLE INFO

ABSTRACT

Are comprehension and production a single, integrated skill, or are they separate processes drawing on a shared abstract knowledge of language? We argue that a fundamental constraint on memory, the Now-or-Never bottleneck, implies that language processing is incremental and that language learning occurs on-line. These properties are difficult to reconcile with the 'abstract knowledge' viewpoint, and crucially suggest that language comprehension and production are facets of a unitary skill. This viewpoint is exemplified in the Chunk-Based Learner, a computational acquisition model that processes incrementally and learns on-line. The model both parses and produces language; and implements the idea that language acquisition is nothing more than learning to process. We suggest that the Now-or-Never bottleneck also provides a strong motivation for unified perception–production models in other domains of communication and cognition.

© 2016 Published by Elsevier Inc.

Language as knowledge; language as skill

The ability to comprehend and produce language requires spectacular levels of skill. But is it one skill, or two? Transfer between comprehension and production appears to suggest a unitary system; hearing a language is, it seems, crucially important for speaking that language. After all, children don't simultaneously learn to understand German while producing Mandarin. Indeed, across diverse theoretical perspectives in the language sciences, there is general agreement that there exists an important overlap between the knowledge and processing operations involved in comprehension and production. But there is considerable disagreement about the nature of this relationship.

One viewpoint (e.g., Chomsky, 1965), which has been dominant in many theoretical approaches to language,

starts by assuming a strong separation between linguistic competence (i.e., an abstract specification of the speaker/hearer's knowledge of the language) and linguistic performance (the processes by which this abstract competence is deployed in language processing). From this standpoint, the overlap between production and comprehension may reside purely in the shared abstract linguistic competence that is being drawn upon by both comprehension and production processes. But the processes of comprehension and production could, in principle, be completely unrelated. We call this the 'language as knowledge' view.

An opposing viewpoint suggests that no such abstract linguistic competence exists—rather, acquiring language is no more than acquiring the ability to process language. There is no separate representation of the abstract structure of the language (e.g., a grammar) distinct from the mechanisms of language production and comprehension; instead there are simply procedures for language processing (e.g., Kempson, Meyer-Viol, & Gabbay, 2001; O'Grady, 2013). From this point of view, the overlap

* Corresponding author.
E-mail address: Nick.Chater@wbs.ac.uk (N. Chater).

between comprehension and production requires that the same (or highly overlapping) processes underpin both the comprehension and production of utterances. We call this account the 'language as skill' perspective (see also Christiansen & Chater, 2016).

We will argue, in the next section, *Fundamental memory constraints on skill learning*, that basic limitations on memory strongly favor the language as skill perspective, and hence the assumption that the processing operations of comprehension and production are intimately related. This requires viewing language acquisition as a process of on-line skill learning, where processing operations lay down traces that facilitate further processing—there is no opportunity for inferring abstract general principles of language structure. In light of the limitations we outline, the challenge of building cognitively plausible models of language processing and acquisition is considerable. We take some initial steps toward addressing this challenge in Section 'A unified model of production and comprehension', describing a computational model incorporating incremental processing and on-line learning. Crucially, although comprehension and production are closely integrated in the model, we present new simulation results demonstrating that the model gives rise to the kind of comprehension–production asymmetry often observed in language acquisition (e.g., Fraser, Bellugi, & Brown, 1963). In Section 'Integrated production and comprehension', we then reflect on the broader theoretical issues raised by proposing a unitary model, before drawing brief conclusions.

Fundamental memory constraints on skill learning

On just about any measure, language processing is astonishingly fast. Speaking rates are typically 10–15 phonemes per second, which translates to as much as 150 words per minute (Studdert-Kennedy, 1986). This daunting speed implies that the comprehension system is faced with a relentless onslaught of new input. Similarly, the production system has to generate and execute a stream of articulatory instructions at a remarkable speed. Furthermore, the interleaving of comprehension and production processes is also very fast, as made evident by the rapid turn-taking observed across languages and cultures (e.g., the mean latency between 'turns' is typically about 200 ms, Stivers et al., 2009), the ability to 'shadow' speech within a 250 ms latency or less (Marslen-Wilson, 1973, 1985), and our ability to fluently complete each other's sentences (Clark & Wilkes-Gibbs, 1986).

Our impressive performance in processing language contrasts strikingly with our very limited ability to process sequences of arbitrary auditory or visual stimuli. For example, in a classic study, Warren, Obusek, Farmer, and Warren (1969) found that naïve listeners were unable correctly to recall the order of just four different non-speech sounds, even though each could easily be identified in isolation. More broadly, memory for the temporal order of arbitrary stimuli appears restricted to 4 ± 1 items (Cowan, 2000); and even the identities of the items in a sequence is typically rapidly lost, e.g., when measured by probed or free recall (e.g., Baddeley, 2007). In short, memory is fleeting:

unless information is recoded and/or used rapidly, it is subject to severe interference from an onslaught of new material. We call this the Now-or-Never bottleneck (Christiansen & Chater, in press).

To cope with the flow of speech input, it is crucial that phonemes are rapidly recoded into higher-level units—for example, into syllables, lexical items, phrases, and beyond (although the specific hierarchy of linguistic levels is, of course, controversial; and may vary from language to language): we call this Chunk-and-Pass language comprehension (Christiansen & Chater, in press). These more abstract levels correspond both to larger units of speech input (so that the same three- or four-item limit corresponds to a longer stretch of speech), and also will typically lead to less interference with subsequent material. This is because, on just about any theoretical account, the space of lexical items is very much larger than space of phonemes, so that confusability between phonemes will be much greater than confusability between lexical items.

Parallel issues arise in speech production. Here, the constraint is to maintain the shortest possible 'inventory' of material waiting to be generated, to avoid interference between items to be produced, at a given level of representation (e.g., Dell, Burger, & Svec, 1997). And as with comprehension, this, too, can only be done by decoding higher-level representations into lower-level representations in a piecemeal manner. If, for example, an entire message were decoded into a string of phonemes before even beginning to speak, that stream of phonemes would vastly exceed the few items we can accurately hold in memory, and hence would rapidly be lost. The production system must, therefore, decode a higher-level representation into a more detailed lower-level representation when that lower level representation will be used, and not substantially before. Thus, while our speech production system may look ahead several words in advance, those words will only be converted into, say, a phoneme representation, when the word is almost ready to spoken; and the phonemes will in turn only be translated into still more detailed articulatory instructions at the very last moment. The cascade of different levels of speech production thus obeys a principle of 'Just in Time' processing, analogous to the inventory management system pioneered in Japanese manufacturing (Ohno & Mito, 1988). A 'stock' of representations cannot be allowed to accumulate, because such stock is highly 'perishable'—specifically, it will be rapidly interfered with or overwritten by the continual arrival of new material (Christiansen & Chater, in press)[1].

The Now-or-Never Bottleneck has particularly striking implications for language acquisition. If linguistic input can only be retained very briefly before being overwritten, then learning must occur *as language is being processed.*

---

[1] We want to stress that we are not advocating for so-called 'radical incrementality' in production, in which words are articulated immediately without any planning ahead. Rather, we see production as involving planning a few chunks ahead at every level of linguistic abstraction. Importantly, though, 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 multiword combinations, and even longer at the conceptual/discourse level (see Chater & Christiansen, in press, for further discussion).

There is no opportunity for the learner to survey previous linguistic input in order to carry out, for example, distributional tests to determine linguistic categories (e.g., Redington, Chater, & Finch, 1998) or to search for phrase structure rules (Pereira & Schabes, 1992). Indeed, typically, the learner will not be able to take a synoptic view even of a single utterance: the early part of an utterance will typically have been recoded into higher level representations before the end of the utterance is reached. We call the constraint that learning occurs in-the-moment *on-line learning* (Christiansen & Chater, in press).

In this article, we propose that the constraint that language is *processed incrementally* and *learned on-line* imposes very strong restrictions on computational models of language processing, and, crucially for the present discussion, suggests an intimate relationship between comprehension and production.[2] Note, in particular, that from the language as skill perspective, it is natural to assume that learning is simply a matter of storing traces of past (incremental) language processing operations (chunks, in the model outlined below). For example, according to instance-based theories of skill learning (e.g., Logan, 1988), lexical access (Goldinger, 1998), and memory (Hintzman, 1986), the processing of new input draws on traces of past processing of previous items. But if this viewpoint is right, the existence of any *transfer* between comprehension and production must imply that the *same* processing traces are relevant to both—otherwise, traces from comprehension (e.g., hearing a new word or syntactic construction) could not be recruited in production (e.g., using that same new word or construction). The on-line nature of learning strongly suggests, we argue, a unitary model of comprehension and production. Language processing is a skill; and, to a large extent, comprehension and production are the *same* skill. Moreover, in this view, language acquisition is nothing more than the learning the skill of language processing. There is, in particular, no additional challenge of acquiring knowledge of the language, over and above the (unitary) ability to comprehend and produce.

By contrast, it is difficult to reconcile incremental processing and on-line learning with the "language as knowledge" viewpoint, according to which comprehension and production correspond potentially to non-overlapping skills drawing on a common body of abstract knowledge of the structure of the whole language (i.e., linguistic competence, Chomsky, 1986). The challenge of extracting a putative abstract linguistic competence most naturally and reliably operates 'off-line' by finding regularities over a large corpus of full sentences, in order to find the best overall match between the grammar and corpus (e.g., Klein & Manning, 2004; Pereira, 1992).

Yet while the Now-or-Never Bottleneck can be used to provide a theoretical argument for a unitary model of comprehension and production, this account appears to be challenged by an important empirical observation: the asymmetry often observed between comprehension and production in language acquisition. Although children in specific cases may exhibit adult-like production of sentence types that they do not appear to fully comprehend (cf. Grimm, Muller, Hamann, & Ruigendijk, 2011), their comprehension abilities generally appear to run ahead of their production skills (e.g., Fraser et al., 1963). But how can such asymmetries arise, if comprehension and production are as closely intertwined as we have proposed? In the next section, we consider a computational model compatible with the Now-or-Never bottleneck, and report new simulations that demonstrate how a comprehension–production asymmetry might arise within a system that unifies these two fundamental components of language use.[3]

A unified model of production and comprehension

Inspired by the restrictions imposed by the Now-or-Never bottleneck, McCauley and Christiansen (2011, 2014) implemented a computational model of language acquisition, the Chunk-Based Learner (CBL), which provides a unified approach to aspects of comprehension and production. The model aims to recreate the early language production of individual children given the linguistic input to which they have been exposed. Individual models are created for specific children in the CHILDES database (MacWhinney, 2000). The input to comprehension is the caregiver speech to the individual child, and production involves reproducing the utterances of that particular child. Thus, each CBL model simulates language acquisition in a single child, with child-directed speech (or speech spoken in the presence of the child) as input for comprehension and the child's own utterances as the target for the model's productions.

The CBL model acquires item-based knowledge of the language in a purely incremental fashion. It learns on-line, using peaks and dips in "backwards" transitional probabilities between words to chunk words together as they are encountered, incrementally building up a "shallow parse" as each incoming utterance unfolds. By storing the word sequences that it groups together, the model gradually builds up an inventory of chunks consisting of one or more words – a "chunkatory" – which forms the basis for both comprehension and production.

As it passes through the corpus, the CBL model attempts to reproduce each child utterance using only the chunks and distributional information it has acquired up to the point at which the child produced that particular utterance. Crucially, the very same chunks and distributional information used during production are also employed incrementally to build a "shallow parse" of utterances produced by caregivers. Thus, language acquisition in CBL

---

[2] The present paper extends the argument of Christiansen and Chater (in press) to suggest that the Now-or-Never bottleneck most naturally fits with a unified model of comprehension and production, and that language learning should be viewed as the acquisition of a single, unitary, skill. The computational simulations reported below extend previous work developing the Chunk-Based Learner (McCauley & Christiansen, 2011, 2014) to show in detail how a unitary model of parsing and production may nonetheless exhibit behavioral parsing-production asymmetries.

[3] As we discuss below, the comprehension–production asymmetry may be somewhat overdetermined. There are additional, plausible, auxiliary assumptions that can also yield the asymmetry. Our contribution here is to point out that an asymmetry can arise directly from the operation of a unitary model, without depending on such auxiliary assumptions.

involves the simultaneous acquisition of two skills, concerning parsing and producing language: the model learns to become better at both skills, using the same chunkatory and the same distributional information within a unitary system. Moreover, the model's emphasis on multiword chunks is consistent with evidence highlighting the importance of such units in both child and adult language comprehension and production (e.g., Arnon & Clark, 2011; Arnon & Snider, 2010; Bannard & Matthews, 2008; Janssen & Barber, 2012; Reali & Christiansen, 2007).

CBL, in its present form, is limited to purely distributional learning and processing, involving combinations of words in individual utterances, and does not extract meaning from the utterance—hence, we will speak of CBL embodying a unified model of *parsing* and production, rather than comprehension and production. However, given the broader theoretical framework of the Now-or-Never bottleneck (Christiansen & Chater, in press), we envisage that the kind of chunking the model performs (roughly akin to shallow parsing) will be supplemented by contextual top-down information (e.g., tied to semantic and pragmatic knowledge) to assist in full-blown language comprehension. This allows children to arrive at some rudimentary understanding of grammatical constructions they have not yet mastered in full (and therefore cannot use effectively in production). Through chunking, the model can arrive at an item-based shallow parse of an utterance; theoretically, a child might then use such a representation in conjunction with semantic and pragmatic information to arrive at a "good enough" interpretation of the utterance (Ferreira, Bailey, & Ferraro, 2002).

Just as CBL does not model the extraction of meaning from linguistic input in comprehension, production in CBL does not address the process of deciding on the meaning to be conveyed. Instead the focus is on retrieving and sequencing words and chunks in the appropriate order to reconstruct the target child's utterances. Thus, the model captures only some aspects of language comprehension and production. Yet this restricted model is sufficient to explore how behavioral asymmetries can arise through differing task demands, despite the use of the very same representations and cognitive mechanisms (i.e., the same chunks and distributional statistics) during both comprehension and production.[4] Here, we expand on previous CBL simulations (McCauley & Christiansen, 2011, 2014) to explore how asymmetries between comprehension and production may arise, even in an entirely unified model of parsing and production. We first briefly outline the operation of

the CBL model before reporting the outcome of a new set of simulations using analyses in which comprehension and production performance are scored according to the same metric. We then show, using natural performance measures, that the model exhibits better comprehension than production performance across sets of randomly selected test utterances, despite its unified architecture. Moreover, as CBL's experience with processing language increases, the difference between comprehension and production performance decreases, as is also observed in child language acquisition.

Method

The operation of the CBL model

The CBL model has been described in detail in previous work (McCauley & Christiansen, 2011, 2014). We therefore provide only a brief overview of the model architecture here (although the details should be sufficient to implement the model). The model processes the corpus word-by-word, tracking frequency information for words and pairs of words (bigrams). This distributional information is used on-line to compute the backwards transitional probability (BTP) between words (which 8-month-old infants appear to track; cf. Pelucchi, Hay, & Saffran, 2009). The CBL maintains a running average BTP value over previously encountered word pairs, and this value serves as a threshold. Specifically, each time the model calculates a BTP between words which reach or surpass the running average, the two words are grouped together to form part (or all) of a chunk. By contrast, whenever the BTP between words falls below the running average, a chunk boundary is inserted between the words and the chunk thus created (consisting of the word or words preceding the boundary) is placed in the model's chunkatory. Each time a chunk is encountered which already exists in the inventory, its frequency count is incremented by 1. Because there are no *a priori* limits on the number or size of the multiword building blocks that can be learned, the resulting chunkatory will contain a mix of words and multiword chunks.

The model also uses its chunkatory to make on-line "predictions" concerning which words should be chunked together, based on previously discovered chunks: when a word-pair is encountered, it is grouped together if it has occurred more than twice, either as a complete chunk or as part of a larger chunk in the chunkatory. If the word pair does not occur in the chunkatory, the BTP between the words is checked against the running average threshold with the same consequences as usual. For example, suppose the model encounters the phrase *the blue ball* for the first time and its chunkatory includes *the blue car* and *blue ball*. When processing *the* and *blue*, the model will not place a boundary between these two words because the word pair is already represented in the chunkatory (as in *the blue car*). Instead, it predicts that this bigram will form part of a chunk. Next, when processing *blue* and *ball*, the model reacts similarly, as this bigram is also represented in the chunkatory. The model thereby combines its knowledge of two chunks to discover a new, third chunk, *the blue ball*; this is added to the chunkatory, and the model then goes on to process the next word in the utterance.

---

[4] CBL was created to directly embody the twin constraints of incremental processing and on-line learning within a unified model of comprehension and production (McCauley & Christiansen, 2014). Other computational models have related properties. For example, Bod's (2009) Unsupervised Data-Oriented Parsing provides an integrated, though non-incremental, model of production and comprehension in which fragments of current input are learned on-line, and used to parse or produce later sentences. Another example is the dual-mechanism model of Chang, Dell, and Bock (2006), who use a unified connectionist neural network architecture to simulate production and comprehension (in the form of next word prediction), but without being fully online and using a small artificial input corpus. Importantly, CBL has not previously been used to explore the comprehension/production asymmetry in a quantitative way, as in the present paper.

A record of the model's on-line chunking of each utterance is maintained, and can be compared to a gold standard shallow parse (described below) to gain a measure of "parsing" performance. However, the model also simulates "production": this aspect of the model comes into play each time the model encounters an utterance produced by the "target child" (although in the present simulations we also model production for specific adult utterances as well, as described below). When the model encounters an utterance produced by the target child in the input corpus, it attempts to recreate the utterance using only building blocks and distributional statistics discovered in the parsing of the previous input. Following Chang, Lieven, and Tomasello (2008), we make the simplifying assumption that the overall message the child wishes to convey can be roughly approximated by treating the utterance as a "bag-of-words" (i.e., a randomly-ordered set of words). The model's task is to output these words in the correct order such that the output matches the child utterance. Consistent with usage-based approaches, the model uses chunks from the chunkatory during production. In order to model the retrieval of stored chunks, word combinations represented as multiword units in the chunkatory are placed into the bag-of-words. The model then seeks to reproduce the child's utterance using the chunks in the bag-of-words. This is achieved according to an incremental, chunk-to-chunk process: the model begins by removing from the bag the chunk which has the highest BTP, given the start of utterance marker, and placing it as the first chunk in a new utterance. Then, at each subsequent timestep, the model removes from the bag the chunk which has the highest BTP given the most recently placed chunk, producing it as the next part of the utterance. This process continues until the bag is empty. The resulting utterance can then be compared to the child's original utterance for scoring purposes, awarding a score of 1 if the model's utterance matches the child utterance perfectly, and 0 otherwise (McCauley & Christiansen, 2011). Below, however, we introduce a different scoring method that can be used to evaluate parsing and production on an equal footing.

Because we view comprehension and production as forming a unified system, the child's own productions are taken to reinforce statistics previously learned during comprehension. For this reason, after CBL has produced an utterance, it will immediately update its distributional statistics and chunkatory given a parse of that utterance. Thus, the child is taken to "hear" its own productions, consistent the hypothesized lack of separation between the processes involved in comprehension and production.

A performance "Gold standard." In order to provide a gold standard against which to evaluate the model's performance, a shallow parser is used to parse the input corpus. Shallow parsers are widely used in natural language processing tasks (e.g., Hammerton, Osborne, Armstrong, & Daelemans, 2002), and simply segment out the phrases in a text in a non-embedded, non-overlapping fashion (thus yielding a "shallow" rather than fully-articulated, "deep" parse). We chose shallow parsing as the gold standard: (1) for consistency with the Now-or-Never bottleneck; (2) according to evidence for the relatively

underspecified nature of adult and child comprehension (e.g., Ferreira et al., 2002; Frank & Bod, 2011; Gertner & Fisher, 2012; Sanford & Sturt, 2002); and (3) because it approximates children's item-based sentence processing as hypothesized by usage-based theories (e.g., O'Grady, 2013; Tomasello, 2003). After the entire corpus has been shallow-parsed, phrase labels (VP, NP, etc.) are removed and replaced by boundary markers of the sort produced by CBL. The model's parsing *and* production attempts are then scored by comparison to the gold standard parse of the corresponding utterances (using an automated procedure described below).

Input corpus and test utterances

For the present simulations, we use the two largest publicly available corpora of child speech and child-directed speech: (1) the Thomas corpus (Maslen, Theakston, Lieven, & Tomasello, 2004), an English corpus spanning 36 months of the target child's development from ages 2;0 to 5;0, consisting of 2.5 million words, and (2) the Leo corpus (Behrens, 2006), a German corpus spanning 36 months of the target child's development from ages 1;11 to 4;11, consisting of 1.7 million words. Importantly, each corpus consists of speech directed to (and produced by) a single target child. The corpora were stripped of all codes, tags, and punctuation, leaving only the speaker identifiers and the original sequence of words. A marker was added to the beginning of each utterance (to serve as the start-of-utterance marker).

To serve as test utterances, we held back a random selection of 100 adult utterances and 100 child utterances from each corpus, with the constraint that each utterance contained at least four words (to ensure the presence of at least two trigrams in each utterance; see the section on scoring below). For the resulting test utterances, the average length of the adult set in words was 7.93 for English and 5.73 for German, while the average length of the child set was 5.26 for English and 5.59 for German. Importantly, any utterance that perfectly matched one of the randomly selected test utterances was removed from the corpus prior to running the model.

Scoring and testing procedure

For the present simulations, we sought to evaluate both comprehension and production according to a single, established measure that could be used to compare model output to a gold standard parse. In other words, we aimed to provide a "level playing field" on which comprehension and production could be compared in a meaningful way, providing an advantage over scoring approaches that would evaluate each according to a separate metric. To achieve this, we used the Bilingual Evaluation Understudy (BLEU) algorithm (Papineni, Roukos, Ward, & Zhu, 2002). BLEU is typically used to compare the output of a machine translation algorithm to that of a human translator, and has yielded strong correlations with human assessment of translation quality (e.g., Coughlin, 2003). In the case of the present simulations, we used BLEU to score CBL's parsing and production performance by comparing model output to that of the gold standard parse for the target utterance. The Illinois Chunker (Punyakanok & Roth,

2001) was used to generate the gold standard parses for English, whereas TreeTagger (Schmid, 1995) was used for German (as described above).

BLEU operates by comparing the *n*-grams that appear in a candidate string to the *n*-grams that occur in a target string. *Precision* is defined as the number of *n*-grams (a fixed *n*-gram size is chosen) in the candidate string that appear in the target string, normalized by the total number of *n*-grams in the candidate string. *Recall* is defined as the number of *n*-grams in the target string that are in the candidate string, normalized by the total number of *n*-grams in the target string. Importantly, when calculating precision, a candidate *n*-gram cannot be counted more times than it actually occurs in the target string (to prevent score inflation). Thus, if comparing two strings according to bigrams, "*the cat*" in the candidate string "*the cat the cat the cat*" only counts once toward the precision score, rather than three times, when evaluated against the target string "*the dog chased the cat*," since "*the cat*" only occurs once in the target string.

Given that we have used trigram models as a baseline for evaluating CBL performance in past studies (McCauley & Christiansen, 2011; McCauley, Monaghan, & Christiansen, 2015), as well as the typically shorter length of child utterances, we focused on trigrams for calculating the BLEU scores. For this purpose, we treat chunk boundaries as part of the utterance. Precision is calculated as the number of trigrams in the model's attempted production, or attempted shallow parse in the case of comprehension, that are actually in the gold standard, normalized by the total number of trigrams in the attempt (with the constraint, mentioned above, that trigrams cannot count toward precision more times than they appear in the gold standard). Recall is the number of trigrams in the gold standard that are in the comprehension or production attempt, normalized by the total number of trigrams in the gold standard. After mean precision and recall have been calculated for the entire test set, we calculate the harmonic mean between precision and recall (the *F*-score; Van Rijsbergen, 1979) to gain an overall measure of performance.

As an example of BLEU scoring in action, consider the utterance (from the Thomas corpus) "*I'm just going to cut you up an apple.*" The shallow parser used to provide a gold standard for the model identifies the following set of chunks: *[I'm] [just] [going to cut] [you] [up] [an apple]*. During comprehension, CBL chunks the utterance into the following units: *[I'm] [just] [going to cut] [you up] [an apple]*. For scoring purposes, utterances are represented as a string of words, bisected by a boundary marker at chunk boundaries (boundary markers are then considered part of the utterance). Thus, in the case of the model's comprehension attempt, *I'm || just || going to cut || you up || an apple* is compared to the gold standard parse, *I'm || just || going to cut || you || up || an apple*, using BLEU. This yields an *F*-score of 0.78. However, when attempting to produce the same utterance, CBL outputs the following sequence of chunks: *I'm || going to cut || up || an apple || just || you*. When compared to the gold standard parse using BLEU, this yields an *F*-score of just 0.58.

To gain insight into the trajectory of model performance, we tested the model's ability to comprehend and produce the novel test utterances after exposure to each 10th of the input corpus (yielding ten separate timesteps for each test set). Crucially, the model was prevented from learning new distributional information or chunks during testing (thus, being tested at a previous timestep gave the model no advantage at subsequent timesteps; as mentioned above, all utterances identical to test utterances were removed from the corpus prior to evaluating the model).

Results

The overall trajectory of performance is depicted in Fig. 1, separately for each corpus and for the child and adult test utterances. Across all ten time points, the mean BLEU *F*-score for the English adult test utterances was .44 for comprehension and .37 for production. The mean *F*-score for the English child utterances was .44 for comprehension and .35 for production. The German results followed the same general pattern, with a child utterance mean *F*-score of .37 for comprehension and .33 for production, and an adult utterance mean *F*-score of .35 for comprehension and .31 for production.

We further analyzed the data using multiple linear regression, with *Timestep* (0–10), *Language* (English vs. German), *Task* (Comprehension vs. Production), and *Utterance Type* (Adult vs. Child) as predictors of the *F*-scores (which, representing proportional data, were logit-transformed prior to the analyses). This revealed (1) a significant effect of *Timestep* ($\beta = 0.017$, $t = 5.36$, $p < .0001$), indicating that performance improved over time, (2) a significant effect of *Language* ($\beta = -0.24$, $t = -13.1$, $p < .0001$), indicating that BLEU scores were lower for the German simulations overall, and (3) a significant effect of *Task* ($\beta = -0.26$, $t = -14.3$, $p < .0001$), indicating superior BLEU scores on the comprehension task. There was no significant effect of *Utterance Type*, indicating similar performance for child and adult utterances once the other variables were taken into account.

A key result of the simulations was that CBL exhibited a comprehension–production asymmetry in the expected direction, cross-linguistically, despite its unified architecture. For both child and adult test sets, the asymmetry between parsing and production was greatest at early timesteps. While the asymmetry between parsing and production arises in part from differing task demands (as discussed above), note that any such asymmetry diminishes as learning proceeds, and indeed disappears in the German child test set, (Fig. 1c). The model thus eventually attains similar parsing and production performance as its chunk-based knowledge becomes more adult-like with increased input exposure, despite the differing task demands of production and parsing. Our simulations therefore suggest that the differing task demands of comprehension and production, as they relate to distributional information, may play an important role in driving the comprehension/production asymmetry observed in children, but that the greater difficulty imposed by the distributional aspects
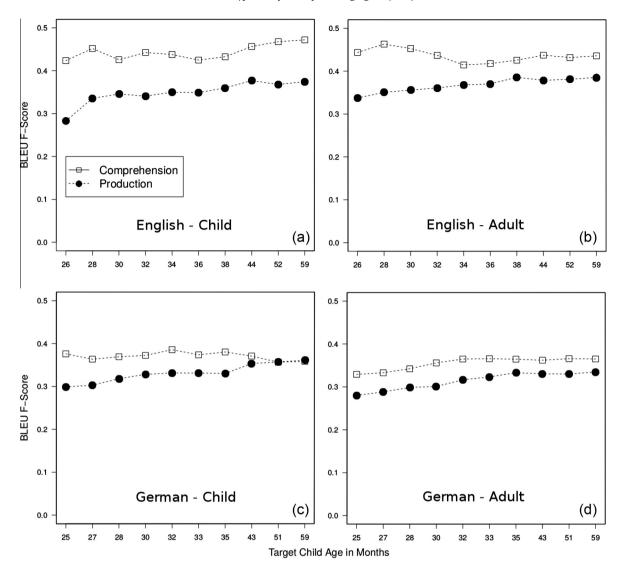
Fig. 1. *CBL performance in comprehension and production*. Mean BLEU *F*-scores for CBL comprehension and production performance after exposure to each tenth of the input corpus, shown separately for: (a) the English child test set, (b) the English adult test set, (c) the German child test set, and (d) the German adult test set.

of production (as opposed to comprehension) may be overcome with sufficient linguistic experience, consistent with the 'language as skill' perspective.[5]

It is important to note that the BLEU scores were calculated for comprehension and production of the *same utterances*. Moreover, for each language, the performance of the model was remarkably similar across the adult and child test utterances. Finally, it should be noted that the utterances in each test set are longer than typical utterances

throughout the corpus, particularly in the case of the child utterances, which are typically shorter than the average test utterance length of 5.26 for English and 5.59 for German (recall that a minimum utterance length of 4 words was chosen to ensure the presence of at least two trigrams in each utterance for BLEU scoring). Thus, performance scores would be expected to be far higher on the entire set of child utterances.

In sum, even in the absence of other constraints, such as the use of top-down information (discussed above), CBL exhibits the expected comprehension/production asymmetry through the interaction of purely distributional learning with the differing task demands of comprehension and production. This does not, of course, imply that other sources of information are unimportant (and, indeed, they are likely to magnify these effects); rather,

---

[5] That differing task demands play a role in our simulations is inevitable; it should be noted that task demands would also affect any behavioral assessment of a child's linguistic abilities, making it difficult to compare comprehension and production on an equal footing. We take a step toward such a comparison by using the same scoring metric for both the comprehension and production aspects of our simulations.

distributional learning and processing are likely to play an important role in driving the asymmetry between comprehension and production observed in children.

The comprehension–production asymmetry reconsidered

The results reported here indicate that a comprehension–production asymmetry can arise directly from the operation of a completely unified parsing-production mechanism, which draws on the very same information to accomplish both tasks. This does not rule out the possibility, of course, that there may be further additional factors that magnify the effect.[6] For example, one might reasonably conjecture that as human motor control develops slowly (Payne & Isaacs, 2012), it may lag behind perceptual processes in development. If so, the comprehension–production asymmetry might simply arise as a special case of this putative 'motor lag,' in the context of language. This may indeed be so, but then the question about the meaning of this observed general perception-motor asymmetry recurs at a more general level. Does such an asymmetry imply, for example, that perceptual and motor processes are distinct and develop at different speeds, or is it possible that a completely integrated system for perception and action, developing in unison, might nonetheless lead to observed asymmetries, for reasons parallel to those observed in our simulations? This question is of particular interest in light of considerable evidence from neuroscience and behavioral studies that many aspects of perception and action use a common representational code, and perhaps also a common set of computational processes (Prinz, 1990; Rizzolatti & Sinigaglia, 2007). Indeed, if the Now-or-Never bottleneck is a general cognitive constraint and applies not just to language but to perception and action in general (as argued by Christiansen & Chater, in press), then perhaps a unitary model of perception and action may be consistent with observed perception–action asymmetries during development.

Another interesting possibility is the effect of practice: language learners surely hear a great deal more than they themselves say—and hence one might anticipate that comprehension will run ahead of production purely because it is practiced more intensively. Yet, from the point of view of unitary model of comprehension and production, this argument may not necessarily apply, as the basic operations recruited for the two tasks may be identical. So, for example, in the CBL model practice in either parsing or production will enrich, and make it easier to access, the chunkatory, and, more broadly, in analysis-by-synthesis models—whether in speech processing or visual perception (Liberman & Mattingly, 1985; Yuille & Kersten, 2006)—production is "practiced" automatically as a side-effect of perception. So this alternative explanation is not quite as straightforward as may first appear.

In light of these observations concerning putative alternative explanations, it is perhaps particularly interesting that a parsing–production asymmetry can arise without

recourse to any additional assumptions; the disparity between parsing and production can be generated even in an entirely unitary system, using a uniform scoring method.

Integrated production and comprehension

The model that we have developed here is, of course, very limited compared to the remarkable richness of human language and language processing. Nonetheless, the CBL illustrates how processing and learning may operate under the severe challenge of the Now-or-Never bottleneck. Most importantly, from the point of view of this special issue, the CBL model embodies a completely integrated approach to production and comprehension: the very same representations and processing operations are involved. A chunk can be viewed as a computational procedure that maps, in either direction, between individual words and sequences of words (Christiansen & Chater, in press). The same viewpoint applies across linguistic levels. From this perspective, the chunkatory is not an abstract dictionary of words and phrases with their phonological and semantic properties. Instead, it is a set of bi-directional, incremental procedures, which map between lower level units and chunks composed of sequences of those units. Similarly, the grammar of the language can be viewed as nothing more than a set of incremental procedures mapping, in both directions, between sequences of linguistic units and semantic representation (see, e.g., O'Grady, 2013; Steedman, 2000). A chunk created in comprehension is immediately available for use in production, and vice versa. There is no need for abstract knowledge of the language, aside from the linguistic structure that is generated through chunking operations. Hence, there is no need to postulate an abstract "knowledge of language," independent of language processing operations. Language is viewed as a skill, and language acquisition is viewed as a case of skill learning, rather than a *sui generis* process of inductive inference, in which an abstract grammar is putatively inferred from observed linguistic data.

Note, too, that this type of integrated, instance-based model predicts that current processing, whether in comprehension or production, will be influenced by the recent use of relevant past processing traces, whether from comprehension or production. Thus, we should expect priming to operate freely between production and comprehension. This is evident in, for example, picture–word interference, where, under certain conditions, a picture of a dog is named more rapidly in the presence of the word *dot* (Schriefers, Meyer, & Levelt, 1990). This facilitation cannot occur through the interpretation of the word *dot*, which would interfere with saying *dog*, but because the word-initial phonemes primes the *production* of those same initial phonemes that are required to generate *dog*.

It is also interesting to note that the very possibility of extremely rapid shadowing, sometimes within a latency of 250 ms (Marslen-Wilson, 1973, 1985), seems to imply a direct relationship between the processes by which speech is recognized and produced—for example, a requirement to translate from a set of processing operations concerned with comprehension, to an abstract

---

[6] We would like to thank an anonymous reviewer on an earlier version of this paper for thoughtful suggestions that have been a starting point for this discussion.

(a)

Perception: **a b c** d e f **g h i** j k l **m n o** p q r…

Production: - - - **a b c** d e f **g h i** j k l **m n o**…

(b)

Perception: **a b c** d e f **g h i** j k l **m n o** p q r…

Production: - - - - - - **a b c d** e f g h **j k l m**…

(c)

Perception: **a b c d** e f g h **i j k l** m n o p **q r**…

Production: - - - - **a b c** - d e f - **g h i** j k l…

Fig. 2. *Aligned vs. misaligned chunks in perception and production in shadowing.* We assume that producing a chunk begins only when it has been perceived—and this requires encountering the entire chunk. (a) If the input is chunked into threes, then production can start directly after the first chunk has been perceived. Thus, there is a delay represented by the three 'dashes'—i.e., the output sequence is delayed by three timesteps. But when the input and output chunks do not align, there is considerable additional delay, e.g., in (b) where the input is split into threes and the output fours, or (c) vice versa. In both cases, a delay represented by six dashes arises. This will both slow shadowing, but also impose a greater memory load as the material has to be held in memory for longer before it can be produced. If, instead, the chunking operations involved in perceiving speech are the very same operations involved in producing speech, then these problems do not arise.

level of linguistic representation, and back to a separate set of processing operations concerned with production would seem to require potentially time-consuming cognitive processing. Indeed, given the different task demands on comprehension and production, it seems likely that if there were separate sets of processes for these two tasks, they would be optimized for the specifics of those tasks. Thus, there would be no reason to use the same linguistic units for comprehension and production but rather to use units that fit the relevant demands of each task. However, this would create problems in the case of shadowing, creating delays because the different linguistic units would not be aligned. The delays that arise in shadowing from such misalignment of perceptual and motor representations are illustrated in Fig. 2.

The tight coupling of comprehension and production processes in speech has also been assessed directly using a response time methodology. In typical response time studies (Luce, 1986), people are slower when they must make a specific response from a set of possible responses (choice response time), rather than making the same response on all occasions (simple response time). The difference between the two times is often viewed as indicative of the time taken to make the choice, in addition to merely generating a response, and is typically between 100 ms and 150 ms (Luce, 1986). In an ingenious study, Fowler, Brown, Sabadini, and Weihing (2003) played people Vowel–Consonant–Vowel (VCV) stimuli. In the simple response task, they had to shadow the first vowel, followed by a fixed CV pattern for all trials (one of /pa/, /ta/ or /ka/). In the choice response time task, they shadowed the entire

VCV pattern that they heard, and so had to choose between one of the three possible CV patterns afresh on each trial, based on the auditory input. The difference between choice and simple response times was just 26 ms, far less than in a typical choice task, suggesting that the choice of motor output was directly linked to perceptual representation of the speech input.

Conclusion

The Now-or-Never bottleneck is a general cognitive constraint, and hence applies not just to language, but also to perception and action more generally. The perceptual-motor system should, more broadly, involve incremental processing and on-line learning. And, to the extent that there is transfer between cognition-perception and motor activities, such transfer cannot be mediated by abstract knowledge, but arises through sharing of procedures involved in both perception and production.

There is, of course, extensive evidence of a very rich integration between perception and production across a wide range of cognitive domains, including a full range of linguistic representations from low-level speech perception to higher level representations (e.g., Pickering & Garrod, 2013). Since Liberman's path-breaking work on the motor theory of speech perception (e.g., Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; for a review, see Galantucci, Fowler, & Turvey, 2006), it has been observed that very different acoustic inputs, corresponding to the same speech gesture, are classified together, sug-

gesting that perception is recovering articulatory processes. Similarly, many lines of evidence suggest that action perception involves the recovery of the underlying action, fairly independently of the sensory evidence used to infer that action. For example, Shiffrar and Freyd (1990) showed that alternating still photographs of body postures results in the perception of apparent motion compatible with human anatomical constraints, rather than representing the shortest path in 'image space' (although the shortest path interpretation was observed for inanimate objects, and for very short latencies with alternating human figures), suggesting that motion perception can involve recovering possible human movements. The same conclusions emerge from studies looking at enhancement or interference between perceptual and motor representations in motor imitation tasks (e.g., Brass, Bekkering, & Prinz, 2001).

We suggest that, in all of these cases, the cognitive system is constrained by the Now-or-Never bottleneck. Decoding the stream of perceptual or linguistic information must occur in real-time; learning the structure of that input cannot involve "reflection" on previous raw data, which is almost immediately obliterated; and learning to respond to input involves learning a skill, rather than developing a theory. This perspective makes it difficult to see how the brain could acquire abstract linguistic knowledge (e.g., Chomsky, 1965); and, in any case, the tight coupling between speech input and output processes implies that there would be insufficient time to consult such knowledge in real time linguistic interactions, even were it available. Thus, we speculate that learning a language should be viewed as acquiring a perceptuo-motor skill, rather than learning an abstract theory of language structure. It is interesting to wonder how far the same considerations may apply more broadly: how far can we understand our interactions with the physical world, or other people, as generalized perceptual-motor skills, most likely with tightly coupled representations between perception and action, rather than applications of "naive physics" (Baillargeon, Spelke, & Wasserman, 1985) or "theory of mind" (Wellman, 1990)?

Acknowledgments

References

Arnon, I., & Clark, E. V. (2011). When 'on your feet' is better than 'feet': Children's word production is facilitated in familiar sentence-frames. *Language Learning and Development, 7*, 107–129.

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language, 62*, 67–82.

Baddeley, A. D. (2007). *Working memory, thought and action*. Oxford: Oxford University Press.

Baillargeon, R., Spelke, E. S., & Wasserman, S. (1985). Object permanence in five-month-old infants. *Cognition, 20*, 191–208.

Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science, 19*, 241–248.

Behrens, H. (2006). The input–output relationship in first language acquisition. *Language and Cognitive Processes, 21*, 2–24.

Bod, R. (2009). From exemplar to grammar: A probabilistic analogy-based model of language learning. *Cognitive Science, 33*, 752–793.

Brass, M., Bekkering, H., & Prinz, W. (2001). Movement observation affects movement execution in a simple response task. *Acta Psychologica, 106*, 3–22.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review, 113*, 234–272.

Chang, F., Lieven, E., & Tomasello, M. (2008). Automatic evaluation of syntactic learners in typologically-different languages. *Cognitive Systems Research, 9*, 198–213. http://dx.doi.org/10.1016/j.cogsys.2007.10.002.

Chater, N. & Christiansen, M. H. (2015). Squeezing through the Now-or-Never bottleneck: Reconnecting language processing, acquisition, change and structure. *Behavioral & Brain Sciences* (in press).

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

Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. Greenwood Publishing Group.

Christiansen, M. H., & Chater, N. (2015). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral & Brain Sciences*. doi: http://dx.doi.org/10.1017/S0140525X1500031X (in press).

Christiansen, M. H., & Chater, N. (2016). *Creating language: Integrating evolution, acquisition, and processing*. Cambridge, MA: MIT Press.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*, 1–39.

Coughlin, D. (2003). Correlating automated and human assessments of machine translation quality. In *Proceedings of MT summit IX* (pp. 63–70).

Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral & Brain Sciences, 24*, 87–114.

Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review, 104*, 123–147.

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

Fowler, C. A., Brown, J. M., Sabadini, L., & Weihing, J. (2003). Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Language, 49*, 396–413.

Frank, S. L., & Bod, R. (2011). Insensitivity of the human sentence-processing system to hierarchical structure. *Psychological Science, 22*, 829–834.

Fraser, C., Bellugi, U., & Brown, R. (1963). Control of grammar in imitation, comprehension, and production. *Journal of Verbal Learning and Verbal Behavior, 2*, 121–135.

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review, 13*, 361–377.

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

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*, 251–279.

Grimm, A., Muller, A., Hamann, C., & Ruigendijk, E. (2011). *Production–comprehension asymmetries in child language*. Berlin: De Gruyter.

Hammerton, J., Osborne, M., Armstrong, S., & Daelemans, W. (2002). Introduction to special issue on machine learning approaches to shallow parsing. *Journal of Machine Learning Research, 2*, 551–558.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93*, 411–428.

Janssen, N., & Barber, H. A. (2012). Phrase frequency effects in language production. *PLoS One, 7*, e33202.

Kempson, R., Meyer-Viol, W., & Gabbay, D. (2001). *Dynamic syntax: The flow of language understanding*. Malden, MA: Blackwell.

Klein, D., & Manning, C. (2004). Corpus-based induction of syntactic structure: models of dependency and constituency. In *Proceedings of the Association for Computational Linguistics*. ACL.

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

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21*, 1–36.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95*, 492–527.

Luce, R. D. (1986). *Response times*. New York: Oxford University Press.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. Mahwah, NJ: Lawrence Erlbaum Associates.

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

Marslen-Wilson, W. D. (1985). Speech shadowing and speech comprehension. *Speech Communication, 4*, 51–55.

Maslen, R. J., Theakston, A. L., Lieven, E. V., & Tomasello, M. (2004). A dense corpus study of past tense and plural over regularization in English. *Journal of Speech, Language, and Hearing Research, 47*, 1319–1333.

McCauley, S. M., & Christiansen, M. H. (2014). Acquiring formulaic language: A computational model. *Mental Lexicon, 9*, 419–436.

McCauley, S. M., & Christiansen, M. H. (2011). Learning simple statistics for language comprehension and production: The CAPPUCCINO model. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd annual conference of the cognitive science society* (pp. 1619–1624). Austin, TX: Cognitive Science Society.

McCauley, S. M., Monaghan, P., & Christiansen, M. H. (2015). Language emergence in development: A computational perspective. In B. MacWhinney & W. O'Grady (Eds.), *The handbook of language emergence* (pp. 415–436). Hoboken, NJ: Wiley-Blackwell.

O'Grady, W. (2013). The illusion of language acquisition. *Approaches to Bilingualism, 3*, 253–285.

Ohno, T., & Mito, S. (1988). *Just-in-time for today and tomorrow*. Cambridge, MA: Productivity Press.

Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002). BLEU: A method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting on association for computational linguistics* (pp. 311–318). Association for Computational Linguistics.

Payne, V. G., & Isaacs, L. D. (2012). *Human motor development: A lifespan approach*. New York: McGraw-Hill.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition, 113*, 244–247.

Pereira, F. & Schabes, Y. (1992). Inside–outside reestimation from partially bracketed corpora. In *Proceedings of the 30th annual meeting of the ACL* (pp. 128–135). Association for Computational Linguistics.

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

Prinz, W. (1990). A common-coding approach to perception and action. In O. Neumann & W. Prinz (Eds.), *Relationships between perception and action: Current approaches* (pp. 167–201). Berlin, New York: Springer.

Punyakanok, V., & Roth, D. (2001). The use of classifiers in sequential inference. In G. Dietterich, S. Becker, & Z. Ghahramani (Eds.), *Proceedings of the conference on advances in neural information processing systems* (pp. 995–1001). Cambridge, MA: MIT Press.

Reali, F., & Christiansen, M. H. (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language, 57*, 1–23.

Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science, 22*, 425–469.

Rizzolatti, G., & Sinigaglia, C. (2007). *Mirrors in the brain: How our minds share actions and emotions*. Oxford: Oxford University Press.

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

Schmid, H. (1995). Improvements in part-of-speech tagging with an application to German. In *Proceedings of the ACL SIGDAT-workshop*.

Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language, 29*, 86–102.

Shiffrar, M., & Freyd, J. (1990). Apparent motion of the human body. *Psychological Science, 1*, 257–264.

Steedman, M. (2000). *The syntactic process*. Cambridge, MA: MIT Press.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., ... Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences, 106*, 10587–10592.

Studdert-Kennedy, M. (1986). Some developments in research on language behavior. In N. J. Smelser & D. R. Gerstein (Eds.), *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*" (pp. 208–248). Washington, DC: National Academy Press.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Van Rijsbergen, C. J. (1979). *Information retrieval*. London: Butterworth.

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*, 586–587.

Wellman, H. M. (1990). *The child's theory of mind*. Cambridge, MA: MIT Press.

Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: Analysis by synthesis? *Trends in Cognitive Sciences, 10*, 301–308.