Five Decades After Chomsky An Experienced-Based Awakening
Review by Dom Massaro

Creating Language: Integrating Evolution, Acquisition, and Processing (MIT Press) Mar 18, 2016
by Morten H. Christiansen and Nick Chater

Over 50 years ago, Noam Chomsky (1959, 1965) threw down the nativist gauntlet on language, and has withstood potshots from many quarters. His debut involved a critique of B.F Skinner's (1957) account of language acquisition. Central to Skinner's account was the important influence of the linguistic environment. Fundamental reinforcement principles would effect the association of spoken words with meaning. Quine (1960, 1990/1992) had not yet popularized the Gavagai challenge in which a child has to determine which of many possible meanings correspond to an uttered word. Similarly, Rescorla and Wagner had not yet broadened Rescorla–Wagner model of association learning with the concept of reinforcement as information or surprise value (Rizley & Rescorla, 1972). Skinner viewed the child as being reinforced when she realizes that a particular word has a particular meaning. The child learns to distinguish the words *milk* and *spinach* because she is reinforced to reply appropriately to her father's request because she likes one and not the other.

Chomsky's Shadow Sets the Stage
According to Chomsky's critique, children have at their beck and call an unlimited set of sentences and a child could never acquire this skill based on a paucity of language input along with reinforcement principles. What was needed was an innate Universal Grammar to allow language development in children and language processing in adults. Even though the assumed properties of this facility have changed over the decades, the bottom line is that language could never be learned without being bootstrapped by this innate ability.

With hindsight bias and the intervening five decades of research, we might counter Chomsky's critique with the apparent impact granted by experience and the natural ability to generalize from one event to another. But somehow it has been difficult for the study of language use to distance itself from Chomsky's shadow. Only recently with increasingly sophisticated empirical and theoretical research and now with the advent of Morten Christiansen and Nick Chater's (hereafter C&C) compelling volume is there the impression that the field has sunny days ahead. This book, grounded in research on brain, evolution, culture, and language acquisition and use, evolved over a two-decade collaboration. Chapters 2-5 and Chapter 7 of the book rework and consolidate several of their previous publications. Two of these publications were BBS articles with peer commentary which we can expect to have already broadened an existing large scope of study.

C&C maintain a strong case for the negative impact that the so-called Chomsky's hidden agenda of generative grammar had on the study of language. From its origins Chomsky's influence has tended to isolate language studies from processing, acquisition, and evolution. As is well-known and rehearsed, advocates of generative grammar believe that language performance cannot inform the idealized linguistic competence of a language user. In retrospect, this seems especially odd because generative grammar stimulated many early psycholinguistic experiments looking for its psychological reality. The plethora of click studies alone (in which participants are to locate where a click occurred in a sentence) generated considerable research attempting to uncover the constituent structure of sentences (Freund, 1975). Chomsky saw little value in studying language acquisition because he claimed that language arrived almost full-blown for a one- or two-year-old growing up in a typical language environment. Similarly, for Chomsky, there was little of interest in the evolution of language because it came about whole without an important precedent.

The major contribution of C&C's book is the articulation of the interplay of evolution, processing, and acquisition in a coherent account of language. C&C succeed at their ambitious goal of integrating these three areas of inquiry by describing their interplay and showing how tightly they interact. Figure 1 (Figure 1.6 from their book) illustrates this interaction among these important dimensions for understanding language and how it is acquired and used. Acquisition constrains what can evolve and fits what is learned to the processing mechanism; and processing limits what can be acquired, and constrains what can evolve. Evolution fits language to the processing mechanisms and to the learner. I would like to believe that this book and the momentum of its approach should have a large impact, and eventually supersede Chomsky's and improve the science of language.

Insert Figure 1 about here

Beyond Universal Grammar
To their credit, C&C take universal grammar (UG) and its falsification very seriously, C&C certainly have not yet converted and probably will not convert many additional followers of the "language is special" camp, as witnessed by some of the commentaries on their two BBS target articles (C&C, 2008, 2016). However, their systematic dissection of how UG might have evolved is very instructive and worth summarizing here.

They consider two explanations of how the arbitrary principles of UG could be genetically specified: adaptationist and non-adaptationist. The first assumes that UG evolved gradually through natural selection. The second gives a minor role to natural selection and allows the possibility of other events such as a "lucky" mutation or two during its evolutionary history. C&C take seriously Pinker and Bloom's adaptationist view that the evolution of the language faculty consisted of

many arbitrary constraints to maintain a standardized communication code. C&C describe various limitations of this explanation, one involving the likely occurrence of language change during its evolution. We see language change everyday with the arrival of new vocabulary and the disappearance of previously frequent words, such as their examples of the disappearance of *FAX* and the recent arrival of *selfie*. We could imagine a similar change in the early appearance of language as people moved to different niches and found new points of discourse. C&C bolster their argument with computer simulations, showing that a fixed language will lead to fixed genes that will optimize processing of that language but with a changing language, neutral genes are favored. Neutral genes prepare the language user with multi-purpose strategies to handle a language that is continuously changing. (I like the idea of neutral genes because it would be compatible with the possibility of naturally acquired written literacy, which might be conceptualized as just another language change that children from birth onward might master when embedded in the imminent technology and artificial intelligence, Massaro, 2012a).

C&C's primary critique of the non-adaptationist account is simply that the likelihood of stumbling on a UG fix for language is extremely small. From a somewhat biased perspective, the non-adaptationist account is much too much like "then a miracle occurs". If early language users were succeeding in language use why would they need some additional boost from UG to make it possible? Thomas Kuhn's (1962) insights about scientific progress might be helpful here. UG has been consistently modified (patched up?) since almost the time of its conception but will not likely be abandoned until a new paradigm replaces it. Now armed with sophisticated experimental techniques across the life span and the increasing availability of large language corpora, we are witnessing a revolution in the science of what it takes to participate in a linguistic community. Almost daily we are told again that babies are expert pattern recognizers, association engines, and statistical learning machines, important processes for acquiring language (e.g., Wang & Saffran, 2014). C&C add to this dialog by bringing to bear a strong case for experience-based processing and learning.

C&C request that we replace the question of the evolution of language users with the question of the evolution of languages. Employing the metaphor of biological evolution, the evolution of languages could follow an analogous path. Given the obvious advantage language would ensure, chatty individuals would be selected over those reticent to participate in the language game. In addition, C&C offer a huge counterpoint to modularity of the brain. Learning and using language might simply involve exercising existing brain mechanisms in this new domain of gossiping. They discuss important research by Anderson (2010) who analyzed a plethora of fMRI experiments to determine which regions of the brain participate in various behaviors. Supporting the claim of non-specialized processes, those regions involved in language processing are also active in a variety of other non-linguistic task domains, including attention, memory, reasoning, and action execution.

C&C have argued that our biological adaptation for language has been negligible. The wide diversity of the thousands of languages support the idea of language adapted to the user rather than biological adaptation accommodating a specific type of language. If this were the case, then it would again make the possibility of naturally acquired literacy using technology more of a possibility because there were not selective adaptations of humans for speech or gesture but rather speech and gesture had to be accommodated for the language user. Similarly, using technology, written language could be adapted to the infant, toddler, and preschooler (Massaro, 2012a; 2015).

C&C advocate that language learning consists of learning a systematic body of linguistic entities rather than learning specific items in a piecemeal fashion. The interconnections among linguistic items is most apparent in word learning, such as the past tense of verbs. We more easily learn repeated patterns like *leap* and *leapt* or *sleep* and *slept* than other one of a kind verbs like *go* and *went* or *lie* and *lay*.

Recursion and Beyond
Before closing their magnum opus, C&C delve into recursion, which appears to be the last bastion of the nativist claim that language is special. All languages putatively have it—this is what is unique about language. Admittedly, there continues to be an unending controversy over what recursion actually is and whether it is truly universal across languages (e.g., Everett, 2005). The authors point out that much of recursion like right branching sentences can be accounted for by a simpler iterative processing. A right branching sentence, *this is the rat that ate the malt that lay in the house that Jack built*, is not necessarily recursive. Repeating the construction in this sentence could more simply be generated by iterative processing in terms of a loop that repeats a given structure. A recursive structure must contain self-reference or call itself.

Doubly-embedded recursive sentences like, *The cat the dog the mouse bit chased ran away*, do not reduce to simple iteration. And many psycholinguistic experiments have found that this type of sentence is extremely difficult to process and understand. C&C's analysis highlighted for me the irony that the uniqueness claim about recursion in language rests on exactly those recursive sentences that stumble the typical language user. This is similar to the paleontologist telling us that teeth are a special adaptation for food because they crack when they are used to crush frozen food. C&C devote this chapter to interpreting experimental differences in processing different types of sentences as best accounted for by the language users' experience in processing similar structures. The success of this analysis is a major victory for task specific learning, which is a general principle of learning and not one limited to language processing.

Twenty-First Century Psycholinguistics

C&C thoroughly review a wide berth of psycholinguistic results, involving corpora analyses and experimental manipulation of linguistic tokens. They offer reasonable explanations of a variety of findings, primarily based on language experience. The authors propose that frequency of exposure is an important force in learning language and ease of language processing. Consider the differences in the difficulty of processing the two sentences

The reporter that attacked the senator admitted the error.
The reporter that the senator attacked admitted the error.

Experimental research has shown that the first sentence is easier to process than the second and, of course, there are different explanations of this finding from both the generative grammar and psycholinguistic camps. C&C's experienced-based approach claims that the processing differences are simply due to "their relative distribution in the experience of individual language comprehenders (p. 173). Supporting this conclusion, their analysis of a large corpus of over 11 million words in both spoken and written English found that examples of the first subject relative sentence occurred over twice as often as examples of the second object relative sentence. Thus, the argument is that simple experienced frequency of prior exposure is at least partially responsible for this difference. A nativist might simply reply that frequency of prior exposure is simply a performance influence, and has little to do with understanding what is being said (which is the responsibility of generative grammar). But this reply seems much less important now because there is more agreement on what is responsible for actual performance.

More generally, language corpora are revolutionizing language inquiry, and their increasingly ease of deployment (e.g., Language Goldmine, 2016) is winning many converts, myself included. One persistent source of "evidence" that has long been central to the nativist claim of language is special is the poverty of the stimulus (Chomsky, 1980). The growing child simply does not hear enough language to account for her creative language use. Yet who would have thought that a typical child most likely has heard millions of words during at least a thousand hours of speech before she reaches her terrible twos (Roy, 2015). Strengthing the empiricist account is the recent finding that the vocabulary used during parental speech to children is highly correlated with the child's vocabulary (Massaro, 2016).

The authors also provide convincing evidence for general processes rather than language-unique processes in language understanding. This is also true, they claim, for individuals with speech language impairment (SLI). They propose that there is a plethora of perceptual/cognitive mechanisms at play in language processing, and some subset of these might account for SLI. They show, for example, that poor sequence learning appears to account for some of the language processing deficits in SLI. More generally, it is possible that a deficit in

general purpose procedural learning might account for SLI and thus would also be transparent in non-linguistic tasks.

A Personal Critique

Notwithstanding my admiration for the C&C book, the following critique is best understood by a caveat and full-disclosure that I dedicated much of my research career to an outsider's study of language processing. I spent most of my career as one of the few lone voices against the dominant claim speech is special. Early on in my postdoc, I mentioned our speech research to David Green, the noted auditory psychophysicist, and encountered the reply "Oh, that's very different." My dissident role had very little impact on the field but this book and the plethora of research it reviews appears to substantiate the value of my early research trajectory. This value is only somewhat diminished by my neglect of evolutionary principles and neurological underpinnings of mind and behavior. With respect to the former, I argued that psychological explanation required proximal influences on behavior regardless of the history of distal influences (Massaro, 1979). For the latter, I took a stance against the adequacy of a completely reductionist account of behavior and promoted the value of a more global functional account (Massaro, 1986).

Four decades ago as a junior professor, I convinced a small cohort of graduate students mentored by my new senior faculty colleagues to apply an information processing analysis to understanding language. Given the graduate students' areas of visual perception, verbal learning, and eyelid conditioning, and residing in the Midwest, Chomsky's shadow did not exert much influence and we could happily proceed with our quest. Our concern was to "view the understanding of language as a sequence of psychological (mental) processes that occur between the initial presentation of the language stimulus and the meaning in the mind of the language processor." (Massaro, 1975b, pp. 4-5). Central to this information processing framework, our primary concern was with real-time processing. The theoretical framework was grounded in structure and process. Memory structures constrain the processing that was possible. As an example, research indicated that the initial speech signal is stored in a preperceptual auditory store that lasts only about a quarter second. Some transformation is therefore necessary to create a more stable encoding (in this case, recognizing a so-called perceptual unit) that can be used by a succeeding stage of processing. This framework anticipated C&C's *Now or Never Bottleneck* and *Chunk-and-Pass* perspective, which assume that "the rich perceptual input is recoded as it arrives to capture the key elements of the sensory information as economically and distinctively as possible" (p. 97).

It is instructive to view our so-called information processing model juxtaposed with C&C's scheme (see Figure 2). Constrained by the information-processing framework, it seemed necessary to include both structural (memory) and functional (process) components to model how the language comprehender advances from the language input to understanding. C&C set themselves a

somewhat more general charge of simply listing four increasing abstract levels of processing.

Insert Figure 2 about here

Quantifying Language Processing
With their broad coverage, C&C tended to neglect important research findings on the speech side of language. This neglect is unwarranted: They do not have to fall into Hockett's trap of basically equating language with speech, but certainly speech is the primary materialization of the world's languages. They correctly claim that the sounds of speech are transient but give the duration of its initial sensory representation as less than 100 ms when it's best estimated at around a quarter of a second (Massaro, 1972). This larger estimate was based on auditory backward recognition masking (ABRM). Previously, research had focused on detection masking in which a neighboring intense sound blocks hearing or detecting another soft sound. Whether the neighboring sound comes before, during, or after the target sound is not critical. This type of detection experiment is the one that gives an answer of somewhere between 50 and 100 ms, which C&C use as their estimate. ABRM, on the other hand, does not interfere with detection because the neighboring sound is the same intensity as the target sound to be identified. Usually, the target sound varies on some auditory characteristic such as the pitch or timbre of a sound or is a speech category difference. Participants must identify the target, and the interference only occurs when the neighboring sound comes after, not before, the target sound. Interference can occur up to about 250 ms after the onset of the target, thus giving the larger estimate of the initial sensory representation in speech.

Why agonize about a difference of 150 ms or so? Given the shorter estimate, there would have to be as many as 10 transformations or recodings per second of speech. This number is excessively large and might easily overwhelm the processing system. Given the lack of invariance between the acoustic input and various phonemes, I used the larger duration to propose that a larger perceptual unit existed. This would not only bring the number of transformations per unit time into a reasonable range, various research findings indicated the syllable unit could restore a reasonable degree of constancy between it and auditory representation. I claimed V, CV, VC syllables, where V is a vowel and C is a consonant or consonant cluster) as perceptual units for speech rather than the phoneme (Massaro, 1972, 2011). The phoneme has a clear lack of invariance between signal and percept whereas these syllables restore most of the invariance needed for reliable pattern recognition (Massaro & Oden, 1980).

Meeting the Gavagai Challenge
The plethora of research literature the C&C review goes a long way to foster understanding how the child is easily capable of solving the Gavagai problem.

There are many, many different sources of information that serve as potential constraints to allow the child to associate the appropriate meaning with the appropriate linguistic utterance. Given C&C's operating thesis of domain general underpinnings of language use, I am troubled when they argue a case for distinguishing between the processes involved in understanding language and those used in pattern recognition in other domains. This distinction is between navigating the natural world (N-induction) and coordinating with our fellow interlocutors (C-induction). C&C state "these two types of problems are very different" (p. 69). In N-induction we are measuring up to an immutable standard; In C-induction the standard is socially converted. If language has evolved to be learnable, the authors claim it may not present the same challenge of induction that typical pattern recognition involves. For objects, the perceiver has to induce what pattern is most likely (given the myriad of possible cues). For language, the perceiver simply has to determine the answer that the community of language learners has already agreed on. The authors make this argument because language is learned and learners have similar learning biases. These biases somehow give the learner an advantage because "the first wild guesses that the learner makes about how some linguistic structure works are likely to be the right guesses." (Christiansen & Chater, 2008, p. 507).

It is not apparent to me how the language task is any easier because in one situation the community of language users enforce the correct answer and in the other the physical world does. Even if some first wild guesses were correct, the language learner would still have to learn and remember which were correct and which were not. In signal detection terminology, a bias does not improve sensitivity. The perceiver of both objects and language has to infer what has been encountered based on the available information. I do not see how language evolved to facilitate pattern recognition more than the pattern recognizer evolving to make sense of fairly constant aspects of the physical world. Given that successive generations experience a fairly stable ecological niche but encounter a language that is continually changing, we might, in fact, actually expect more specialized innate constraints for N-induction than C-induction. Surely, the information will differ in the two domains but I believe the information processing will follow the same format—inferring a meaningful event whether it be some object input or some linguistic input. Our research, as well as that from many others, has demonstrated exactly analogous processes for understanding language and for recognizing objects and events (Massaro, 1998; Movellan & McClelland, 2001). Ironically, C&C's call for the apparent ease of language processing could almost be interpreted as a nativist argument, which the authors would abhor. Perhaps the authors might be convinced of this critique since they so strongly advocate fundamental perceptual, cognitive, and learning processes for language understanding as opposed to language-specific ones.

The Need for Formal Models of Language Understanding
If there is a downside to C&C's contribution, it is a plethora of discursive narrative without any formal models. Their Chapter 4 on the *Now or Never Bottleneck*

(NNB) and *Chunk-and-Pass* processing might be read to allow many (sometimes) contradictory propositions. The main idea behind the NNB is that language input is highly transient and it must be chunked and passed to a higher more abstract level of representation. This is all very reasonable and reflects the progress speech and psycholinguistic science has contributed over the past decades. But the devil is in the details. Without a formal model as a guide, C&C might be interpreted to imply discrete categorization from level to level when they say, "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 are recoded…" (p 107). Just a few pages later, however, they say their proposal "fits with proposals … where local ambiguity resolution is temporally delayed until later disambiguating information arrives" (p. 113). Given these statements, the authors do not take a strong stance on what type of processing and recoding occur at each increasing abstract level. It is well-known, and recognized by the authors, that later information can resolve ambiguity in earlier arriving information. But to take advantage of the later information, the previous recoded information cannot be discrete or categorical. If language perceivers are to benefit from two sources of information, the sources must be graded and not discrete (Massaro, 1987).

Like the Gavagai challenge, language understanding might appear to be an insurmountable problem which is perhaps why Chomsky was so successful in convincing the field of a nativist solution. According to C&C, the high quality of language processing follows from the utilization of multiple constraints. They propose that the perceiver performs parallel integration of multiple cues at multiple levels of language processing. Now that Bayesian reasoning has been featured in cognitive science, their solution is palatable to most of the current players. There is now much convincing evidence of a Bayesian type integration in speech perception and reading (Massaro, 1998; Massaro & Jesse, 2005). What is ironic, however, is that the authors do not review a single study illustrating Bayesian-like integration in language processing. Their elegant corpus analyses do undoubtedly reveal the ecological validity of multiple constraints or cues in typical spoken and written language. However, they do not succeed in describing experiments demonstrating that multiple cues are actually used together to facilitate language processing.

C&C repeatedly postulate that multiple cues are integrated—Chapter 5 actually includes "Multiple-Cue Integration" in its title. I have defined several possibilities of how multiple cues could be used and formally operationalized integration as the simultaneous use of two or more cues to categorize a single presentation of a language event (Massaro, 1987). One of the most popular illustrations of this type of integration is when the sound of speech and the facial movements of the talker are used together to identify a speech syllable (the so-called McGurk effect). Our research has taken great pains to distinguish among the almost plethora of ways multiple sources of information can be used. So, for example, using the most informative cue on each trial would not be equivalent to an

optimal Bayesian-type integration. Within our Fuzzy Logical Model of Perception (FLMP), each cue is assigned a truth value indicating the degree to which it supports each potential categorization. These truth values supporting a given categorization are then multiplied and evaluated against all other possible categorizations. Truth values are a good metric because they more naturally represent graded information compared to probabilities which can easily be interpreted as discrete (Massaro, 1998). As formulated, the FLMP is mathematically equivalent to Bayes theorem, which can be interpreted as an optimal use of multiple cues.

When we began our language studies, the predominant experimental strategy (except for a few notable exceptions) was to manipulate only a single source. I advocated the approach of manipulating several sources of information independently of one another in language processing tasks (Massaro, 1975a). This paradigm was necessary because manipulating just a single source would not illuminate how that source interacted with other sources. In addition, by neutralizing other sources in a single-source experiment, participants might easily zero in on that source even though they don't normally use it in a productive manner. Manipulating multiple sources, for example, we evaluated how both the letter quality and the frequency of orthographic patterns influenced letter and word recognition. In both speech and reading domains. Based on our research and that of others, we have also claimed that the robustness of language processing is due to the efficient use of multiple top down and bottom up sources of information (Massaro, 1979).

C&C discuss top-down and bottom-up sources of information when they describe how both contextual and acoustic information are used to identify word recognition. To rationalize the evolution and development of an arbitrary relationship between the form of a word and its meaning, they cite mathematical evidence that two sources of information provide maximal constraint when they are independent of one another. Thus, they propose that multiple cues do not help if they are redundant. In language processing as in other forms of pattern recognition, however, their redundancy in fact is a necessary condition. In auditory/visual perception, for example, the two sources of information are necessarily redundant because they come from the same speech utterance. Similarly, constraining context will facilitate word processing even when the form of the word also predicts its meaning. The reason is that perceiver treats these two sources as mostly independent of one another (Massaro & Stork, 1998), and benefits from having two evaluations relative to just one. The perceiver get partial information from each cue and a Bayesian type integration provides more information given both cues than just one. Using this framework, we have provided a good quantitative description of how these two sources of information are integrated in both speech perception and reading (Massaro, 2012b).

Our research in auditory-visual bimodal speech perception not only demonstrates the value of two sources of information relative to just one, it also adds

convincing support for experience in language processing. Adults and children of various ages identified speech syllables with either consistent or inconsistent auditory and visual properties. All participants utilized both modalities but the younger the children the less they benefitted from the visible speech. Speech perception and language more generally are slowly acquired skills and consistent with C&C's claim that language acquisition can be interpreted as learning to process language. We make a further distinction between information and information processing operationalized as the informativeness of a source of information and how the sources are used together. Clearly, informativeness increases with experience but there is evidence for an optimal integration of sources in auditory/visual speech perception across ages three to eighty-three (Massaro, 1998).

C&C proselytize the benefit from having multiple sources of information. Having two discrete sources of information, however, would not benefit a correct resolution of the linguistic input. If the two sources agreed, there could not be any advantage having only one source. If the two sources disagree, they would be no information to guide which source should be followed. Thus continuous or graded encodings are necessary to Assuming graded information from each source was one of the central assumptions of the FLMP to predict the integration of top-down and bottom-up sources of information in both speech perception and reading (Massaro, 1979; Massaro & Oden, 1980). The framework of the FLMP provides not only a coherent description but a testable quantitative one.

The FLMP is grounded in fuzzy logic in which a proposition has a degree of truth rather than just true or false. The NNB and the Chunk-and-Pass operation could therefore be formalized within the FLMP framework to pass continuous rather than categorical information from one level to the next more abstract level. To illustrate the FLMP and how it is tested in experiments, consider how Massaro and Oden (1995) analyzed Pitt's (1995) study of the joint influence of phonological information and lexical context in an experimental paradigm developed by Ganong (1980). A speech continuum is made between two alternative CVC syllables *gift* and *kiss*, and the contextual information was varied to support one alternative or the other. The initial consonant of the CVC syllable was varied in six steps between /g/ and /k/. The following context was either /Ift/ or /Is/. The context /Ift/ favors or supports initial /g/ because *gift* is a word whereas *kift* is not. Similarly, the context /Is/ favors or supports initial /k/ because *kiss* is a word whereas *giss* is not. Pitt improved on earlier studies by collecting enough observations to allow a subject-by-subject evaluation of the ability of specific models of language processing to account for the results. Previous tests of models using this task have been primarily dependent on group averages, which may not be representative of the individuals that make the averages up. The results showed a strong effect of both the initial consonant and the following context and a significant interaction between these two variables.

According to the perceptual unit assumption, the CVC is composed of two successive units. Applying the FLMP, it is assumed that the initial CV is not categorically perceived but rather is represented as the degree to which it corresponds to a /gI/ or /kI/. This information is passed on to the next level of word identification given the additional source of information /Ift/ or /Is/. These two sources of information are combined to give an overall degree of support for the words *gift* or *kiss*. The FLMP provided a good quantitative fit for each of the 12 subjects in the task. Assuming independence of bottom-up and top-down sources of information is critical for a coherent description of their joint influence (Massaro & Cohen, 1991). Later arriving context does not change the representation of the earlier arriving bottom-up information. The updated information is simply represented at a more abstract level. Thus, in the *gift-kiss* example, the two sources are combined at the word level, not at the initial CV syllable level. Evidence for this independence was found in a signal detection analysis revealing that the information about the bottom-up source remained independent of the top-down source (Massaro & Oden, 1995).

In summary, several assumptions are necessary to account for research findings within the NNB and Chunk and Pass framework. First, categorizations at each level must be graded rather than discrete to take advantage of multiple sources of information at multiple levels. Second, new information does not feed down and modify the representation or categorization of an earlier level. The new information as in the word ending in the *gift-kiss* experiment simply functions as an additional independent source of information. Of course, the word recognition can also be graded as substantiated by continuous rating judgments. This form of processing also describes how the perceiver traverses through more cognitively complex language understanding.

Form Meaning Correspondences
C&C provide a thorough treatment of the issue of whether the correspondence between the form of a spoken word and its meaning is arbitrary or systematic in some way. They define absolute and relative iconicity. Absolute iconicity occurs when a linguistic feature directly imitates some semantic characteristic. Thus, choochoo represents the sound that a train makes (or at least used to make). These words are usually described as onomatopoeic words. However, Perry, Perlman, and Lupyan (2015) have established a role for absolute iconicity beyond obvious imitation, which they define as a cross-modal correspondence in the analog properties of a word's meaning and its spoken symbol. Measures of iconicity can be derived from adult's judgments of how much English words sound like what they mean. Onomatopoeic words like *slurp* would be rated as highly iconic. Words like *teeny* and *huge* are iconic because they sound small and big, respectively. Words like *cat* and *dog* do not appear to have any crossmodal correspondence between their sound and meaning.

C&C define relative iconicity as a statistical regularity between sounds and meanings in the absence of imitation (p. 139; see also Monaghan, Shillcock,

Christiansen & Kirby, 2014). They carried out a series of corpus analyses to explore the role of relative iconicity in vocabulary and syntax. Although statistically significant relative to chance, the amount of variance accounted for by sound–meaning mappings in this case was very small with less than .2% of the variance accounted for. To find evidence for a role of iconicity, Monaghan et al. (2014) continued to find significant correlations even when all monosyllabic words that shared morphophonic and etymological roots were eliminated from the analysis. Moreover, even if these shared morphophonic and etymological roots among words are completely accounted for, there may be other constraints besides iconicity that are contributing to this correlation. Although the authors appear to use the terms systematicity and relative iconicity interchangeably, I suggest that the term iconicity be reserved for a cross-modal correspondence between form and meaning that can be rationalized as having some imitative property. As observed by Winter et al. (submitted), the English cluster gl– bears no obvious resemblance to the meaning of shiny visual phenomena, but it repeatedly occurs in words such as *glitter, glimmer* and *glisten* (Winter et al., submitted).

It is reasonable to expect that iconicity would play a larger role early in language acquisition than later when the vocabulary would necessarily become more arbitrary as it increases in size. Like other investigators, C&C use adult's estimated age of acquisition (AoA) as a measure of the age at which words are acquired. Although one might think that judging the age when specific words were learned would be unreliable, research has shown that age of acquisition ratings and overall frequency occurrence have very large and roughly equivalent correlations with reaction time and accuracy in a lexical decision task (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012; Kuperman & Van Dyke, 2013). C&C attempt to make the case, however, for a larger influence of their measure of relative iconicity on AoA estimates at ages 2 and 3 than at later ages, but their results shown in their Figure 5.2 are not very convincing. Other results, on the other hand, have found evidence for absolute iconicity influencing early vocabulary learning (Perry, Perlman, & Lupyan, 2015).

Continuing to explore systematicity in English, C&C's corpus analyses also evaluated the phonological properties of noun and verb categories, and found that these properties could predict the two categories above chance. They then determined distributional properties of the words by quantifying the likelihood of a word following one of 20 most frequent words in the corpus. This cue also predicted noun-verb category membership well above chance. They then claim to show that the simple combination of the phonological and distribution cues "good classification can be found" (p. 153). However, the claimed synergy that the authors show in their Figure 5.3 is not really there. Performance given both sources is never better than performance given just the most informative source. Their method of simply representing the sources in a multidimensional space might be responsible if they do use an appropriate integration algorithm (Massaro & Friedman, 1990).

This is a productive research program documenting the possible multiple cues supporting language acquisition and processing. As recognized by the authors, uncovering predictive properties in the language does not address how these properties are actually processed by the language perceiver. It is important for researchers to keep in mind a distinction between ecological and functional validity in terms of whether an ecological cue is actually functional in language processing (Massaro, 1987).

Retrospective
C&C's last chapter builds a strong foundation for integration rather than fragmentation in language studies. Integration corresponds to language disciplines as well as behavioral principles that are not unique to language. The present volume certainly reinforces Skinner's original faith in the environment's influence on language processing. It remains to be seen whether their general framework will supplant an alternative nativist account. Like the last decade of the question of climate change, we might have to accept a prolonged process mimicking evolution itself.

References

Childes (2016). (http://childes.psy.cmu.edu/

Chomsky, N. (1959). "A Review of B. F. Skinner's Verbal Behavior" in Language, 35, No. 1 (1959), 26-58.

Chomsky, N. (1980). Rules and representations. Oxford: Basil Blackwell.

Everett, D. L. (2005) Cultural Constraints on Grammar and Cognition in Pirahã. Current Anthropology, vol. 46 issue 4. 621-646.

Freund, A. (1975). Word and phrase recognition in speech processing. In D. W. Massaro (Ed.) Understanding Language: An Information Processing Analysis of Speech Perception, Reading and Psycholinguistics (pp. 357-389). New York: Academic Press

Ganong, W. F., III. (1980). Phonetic categorization in auditory word recognition. Journal of Experimental Psychology: Human Perception and Performance, 6, 110-125.

Kuhn, T. S. (1962). The structure of scientific revolutions. ChicagoL University of Chicago Press.

Kuperman, V., Drieghe, D., Keuleers, E., & Brysbaert, M. (2013). How strongly do word reading times and lexical decision times correlate? Combining data from eye movement corpora and megastudies. Quarterly Journal of Experimental Psychology, 66, 563-580.

Kuperman, V., & Van Dyke, J. A. (2013). Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. Journal of Experimental Psychology: Human Perception and Performance.

Language Goldmine (2016). http://languagegoldmine.com/

Massaro, D.W. (1972). "Preperceptual images, processing time, and perceptual units in auditory perception," Psychological Review, 79, 124-145.

Massaro, D.W. (1975a). Experimental Psychology and Information Processing, Chicago: Rand McNally.

Massaro, D.W. (1975b). Understanding Language: An Information Processing Analysis of Speech Perception, Reading and Psycholinguistics. New York: Academic Press.

Massaro, D. W. (1979). Reading and Listening (tutorial paper) In P.A. Kolers, M.E. Wrolstad,& H. Bouma (Eds.), Processing of Visible Language, (Vol. 1, pp. 330-354). New York, NY: Plenum Press.

Massaro, D. W. (1986). The computer as a metaphor for psychological inquiry: Considerations and recommendations. Behavior Research Methods, Instruments and Computers, 18, 73-92.

Massaro, D. W. (1987). Speech perception by ear and eye: A Paradigm for psychological inquiry. Hillsdale, NJ: Erlbaum.

Massaro, D. W. (2011). Reflections on the syllable as the perceptual unit in speech perception by Dom Massaro
http://www.talkingbrains.org/2011/02/reflections-on-syllable-as-perceptual.html

Massaro, D. W. (2012a). Acquiring Literacy Naturally: Behavioral science and technology could empower preschool children to learn to read naturally without instruction, American Scientist, 100, 324-333.

Massaro, D. W. (2012b). Speech Perception and Reading: Two Parallel Modes of Understanding Language and Implications for Acquiring Literacy Naturally. American Journal Psychology, 125, 307-320.

Dominic William Massaro. Method And System For Acquisition Of Literacy. United States Patent Number U.S. 9,028,255 B2. Date May 12, 2015.

Massaro, D. W. (2016). Multiple Influences on Vocabulary Acquisition: Parental Input Dominates. Paper to be presented at INTERSPEECH 2016: Understanding speech processing humans and machines. San Francisco, CA. September 8-12, 2016.

Massaro, D. W., & Cohen, M. M. (1991). Integration versus Interactive Activation: the joint influence of stimulus and context in perception. Cognitive Psychology, 23, 558-614.

Massaro, D. W., & Friedman, D. (1990). Models of integration given multiple sources of information. Psychological Review, 97, 225-252.

Massaro, D. W., & Oden, G. C. (1980). Speech Perception: A Framework for Research and Theory. In N.J. Lass (Ed.). Speech and Language: Advances in Basic Research and Practice (Vol. 3, pp. 129-165). New York: Academic Press.

Massaro, D. W., & Oden, G. C. (1995). Independence of Lexical Context and Phonological Information in Speech Perception. Journal of Experimental Psychology: Learning Memory and Cognition, 21, 1053-1064.

Massaro, D. W., & Perlman, M. (submitted). Quantifying Iconicity's Role in Language Acquisition: Implications for Vocabulary Learning. Journal of Child Language.

MRC Psycholinguistic Database (http://www.psych.rl.ac.uk/).

Perry, L. K., Perlman, M., & Lupyan, G. (2015). Iconicity in English and Spanish and Its Relation to Lexical Category and Age of Acquisition. PLoS ONE, 10(9), e0137147.

Perry, L. K., Perlman, M., Winter, B., Lupyan, G., & Massaro, D. W. (2016). The role of iconicity in child-directed speech. Poster presentation at the 41st Boston University Conference on Language Development, to be held on November 4-6, 2016.

Pitt, M. A. (1995a). The locus of the lexical shift in phoneme identification. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21, 1037-1052.

Quine, W. V. O. (1960). Word and object: An inquiry into the linguistic mechanisms of objective reference. Cambridge, MA: MIT Press.

Quine W. V. O. (1990/1992) Pursuit of Truth (Harvard Univ Press, Cambridge, MA).

Rizley, R. C., & Rescorla, R. A. Associations in second-order conditioning and sensory preconditioning. Journal of Comparative and Physiological Psychology, 1972, 81, 1-11.

Roy, B. C., Michael C. Frank, Philip DeCamp, Matthew Miller and Deb Roy. (2015). Predicting the Birth of a Spoken Word. Proceedings of the National Academy of Sciences of the United States of America (PNAS). http://www.pnas.org/content/early/2015/09/15/1419773112

Skinner, B. F. (1957). Verbal Behavior. Acton, MA: Copley Publishing Group. p. 11. ISBN 1-58390-021-7

Wang, T., & Saffran, J. R. (2014). Statistical learning of a tonal language: The influence of bilingualism and previous linguistic experience. Front. Psychol. 5:953. doi: 10.3389/fpsyg.2014.00953.

Winter, B., Perlman, M., & Perry, L. K., Lupyan, G. (submitted). Which words are most iconic? Iconicity in English sensory words.
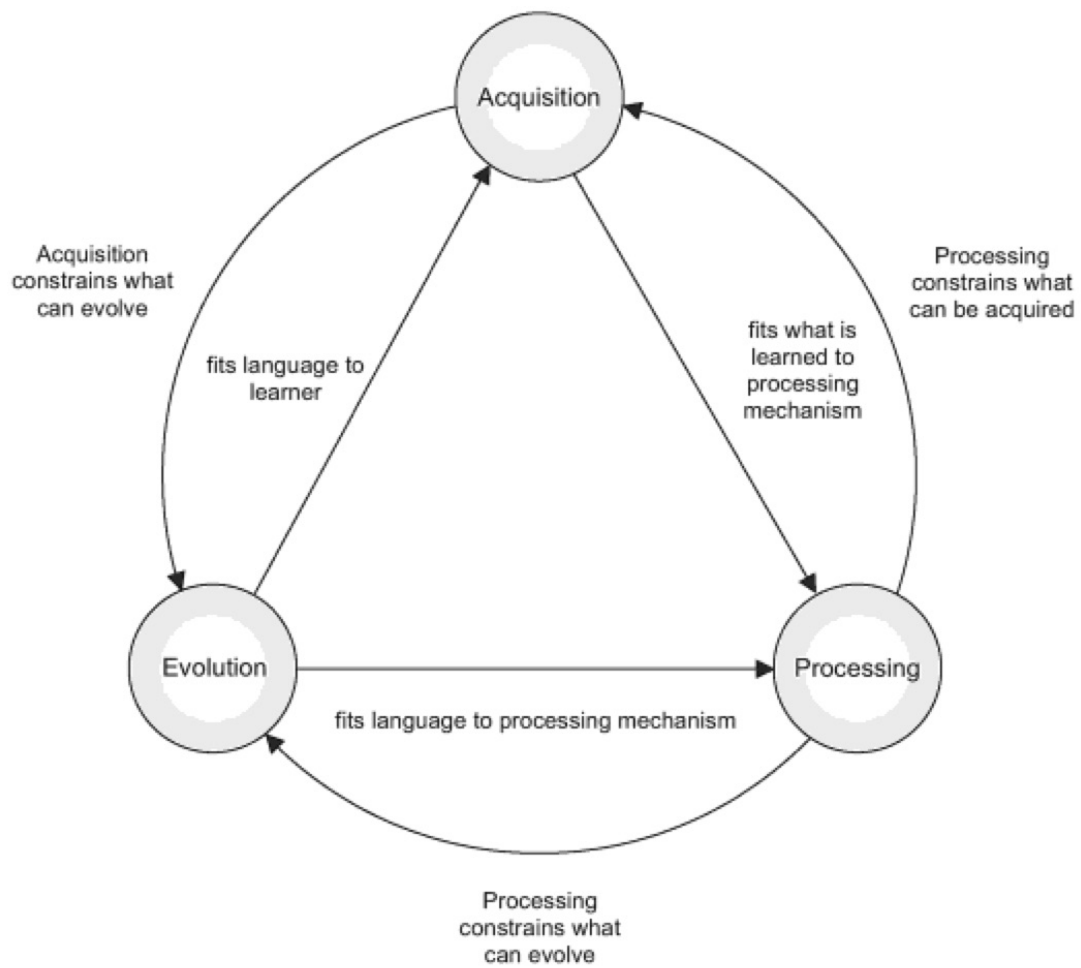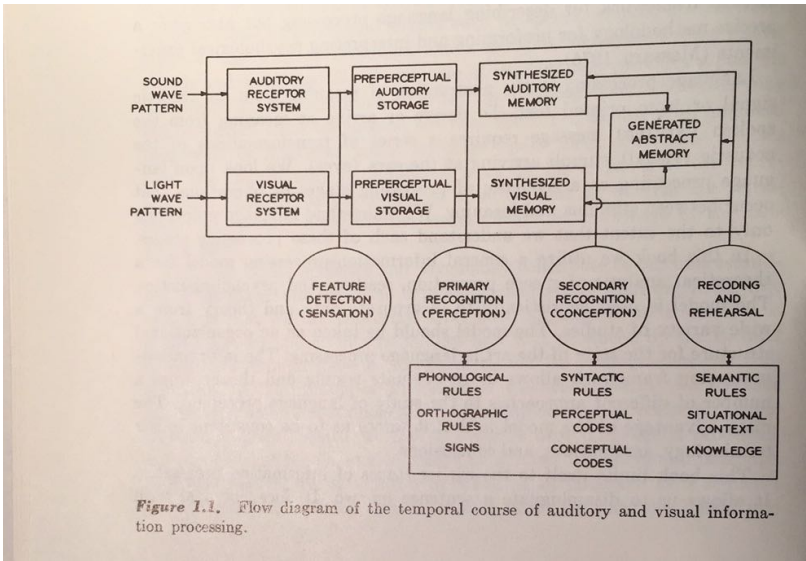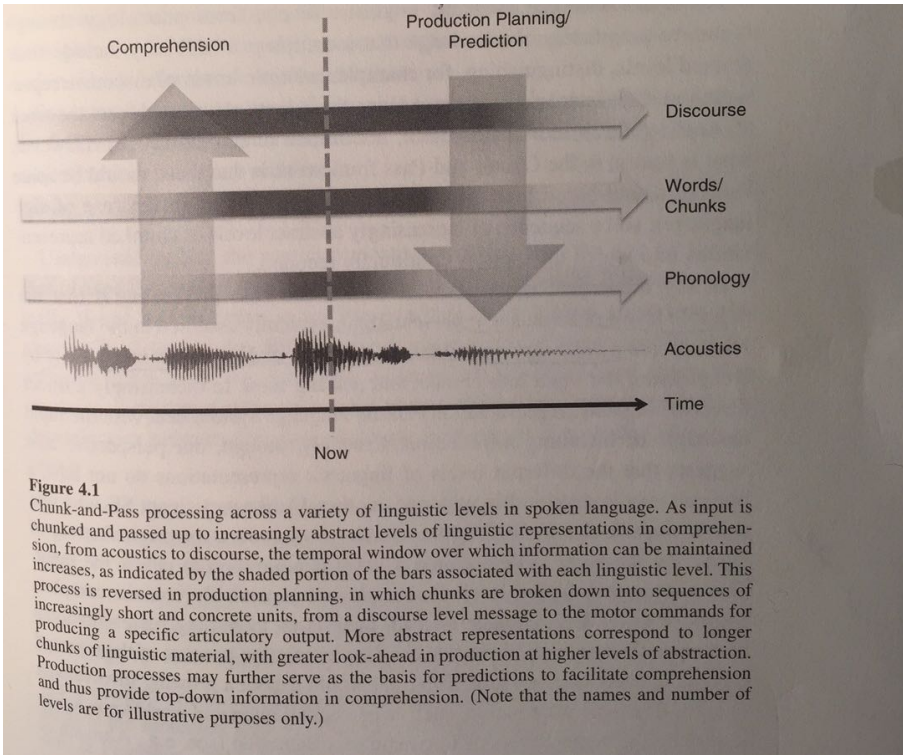
Figure 1. The interrelations between the evolution acquisition, and processing of language. From Christiansen and Chater (2016).

Figure 1.1. Flow diagram of the temporal course of auditory and visual information processing.

From Massaro (1975)

Figure 4.1
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.)

From Christensen and Chater (2016).

Figure 2. Comparison of Christiansen and Chater's (2016) processing scheme to Massaro's (1975).