

Article

Toward a New Scientific Visualization for the Language Sciences

Luca Onnis ^{1,2,*} and Michael J. Spivey ³

¹ Department of Second Language Studies, University of Hawaii, 1890 East-West Rd., Honolulu, HI 96822, USA

² Center for Second Language Research, University of Hawaii, 1890 East-West Rd., Honolulu, HI 96822, USA

³ School of Social Sciences, Humanities and Arts, University of California, Merced, 5200 North Lake Rd. Merced, CA 95343, USA; E-Mail: spivey@ucmerced.edu

* Author to whom correspondence should be addressed; E-Mail: lucao@hawaii.edu; Tel.: +1 808 956 8610.

Received: 31 December 2011 / Accepted: 29 January 2012 / Published: 20 February 2012

Abstract: All scientists use data visualizations to discover patterns in their phenomena that may have otherwise gone unnoticed. Likewise, we also use scientific visualizations to help us describe our verbal theories and predict those data patterns. But scientific visualization may also constitute a hindrance to theory development when new data cannot be accommodated by the current dominant framework. Here we argue that the sciences of language are currently in an interim stage using an increasingly outdated scientific visualization borrowed from the box-and-arrow flow charts of the early days of engineering and computer science. The original (and not yet fully discarded) version of this obsolete model assumes that the language faculty is composed of autonomously organized levels of linguistic representation, which in turn are assumed to be modular, organized in rank order of dominance, and feed unidirectionally into one another in stage-like algorithmic procedures. We review relevant literature in psycholinguistics and language acquisition that cannot be accommodated by the received model. Both learning and processing of language in children and adults, at various putative ‘levels’ of representation, appear to be highly integrated and interdependent, and function simultaneously rather than sequentially. The fact that half of the field sees these findings as trivially true and the other half argues fiercely against them suggests to us that the sciences of language are on the brink of a paradigm shift. We submit a new scientific visualization for language, in which stacked levels of linguistic representation are replaced by trajectories in a multidimensional

space. This is not a mere redescription. Processing language in the brain *equates* to traversing such a space in regions afforded by multiple probabilistic cues that simultaneously activate different linguistic representations. Much still needs to be done to convert this scientific visualization into actual implemented models, but at present it allows language scientists to envision new concepts and venues for research that may assist the field in transitioning to a new conceptualization, and provide a clear direction for the next decade.

Keywords: dynamical systems; language; language acquisition; psycholinguistics; sentence processing; scientific visualization

“Thus, the ‘atomistic’ attitude to words has been dropped and instead our point of view is rather similar to that of field theory in physics, in which ‘particles’ are only convenient abstractions from the whole movement. Similarly, we may say that language is an undivided field of movement, involving sound, meaning, attention-calling, emotional and muscular reflexes, *etc.*”

—David Bohm (Wholeness and the Implicate Order, 1980)

“How dishonest I feel—as ‘expert’ in atomic reality—whenever I draw for schoolchildren the popular planetary picture of the atom; it was known to be a lie even in their grandparents’ day.”

—Nick Herbert (Quantum Reality, 1985)

1. Scientific Visualizations of Theoretical Frameworks

Recently data visualizations techniques have been developed to better understand complex empirical data. When one performs a data visualization technique, one often sees patterns in the data that would have gone unnoticed without that visualization. Likewise, in scientific enquiry more general, formal descriptions are often accompanied by scientific visualizations, *i.e.*, pictorial descriptions that assist scientists in conceptualizing and communicating a theory [1,2]. These visualizations are particularly useful when the scale of the phenomenon under scrutiny is orders of magnitude outside the human visible range (e.g., the atomic and subatomic structure of matter), or when the studied phenomenon comprises complex abstract relations and properties (e.g., how the human mind learns and processes language). As such, model visualizations often allow researchers to draw analogies and are constitutive of scientific endeavor because they can either hinder or promote paradigm shifts. Most would remember from their high school classes of physics how the concept of atom was conceived and visualized as an indivisible unit of matter (the ‘Solid Sphere’ model, Figure 1A). If one looks at the history of how theories of the atom developed in physics, the Solid Sphere model was so deeply entrenched in physicists’ minds that it persisted long after it became apparent that the atom was not so indivisible. In the late 19th century, Thomson’s “Plum Pudding” model (Figure 1B) visualized the atom as a sphere of positive electricity with negative particles embedded throughout. This interim model was an attempt to shoe-horn new empirical data about the divisibility of the atom into the old conceptual framework of the solid sphere. Only with a radical change in visualization to a “Solar System” model by Rutherford (Figure 1C), in which the atom was mostly empty with a compact center,

did the field transition to a truly new mathematical formulation and renewed progress in physics ensued. The new planetary conceptualization further prompted Bohr to abandon classical mechanical theory and make a theoretical leap to quantum theory, a model later refined by Schrödinger (Figure 1D; for a detailed history of the atom, see [3]).

In this article, we track a similar development of psychological theories of *language*, moving from solid encapsulated modules to a somewhat muddled interim model, and then to a complete reorganization of how to visualize and conceptualize language phenomena. We propose a new scientific visualization of the language system, consistent with a growing body of recent psycholinguistic research. We draw an analogy with the shift in scientific visualizations in physics to argue that researchers in the language sciences have been depending on an obsolete scientific visualization borrowed from box-and-arrow flow charts of the early days of engineering and computer science [4]. In its extreme version, this model assumes a language faculty that is composed of autonomously organized levels of linguistic representation (*i.e.*, phonetic, phonological, lexical, syntactic, semantic, and pragmatic). These levels are frequently assumed to be modular, organized in rank order of dominance, and to feed unidirectionally into one another in stage-based algorithmic procedures (Figure 1E). Contemporary versions of this stage-based modularity account of language have at times relaxed some of these constraints, and moved slightly in the direction of more fluid interaction between modules (e.g., [5–10]). However, we contend that these minor revisions to the box-and-arrow framework are reminiscent of Thomson’s “Plum Pudding” model of the atom, in that they are best seen as an interim model that is gradually taking the field toward a radical reconceptualization.

A host of empirical findings in psycholinguistics [11–14] is revealing that the mind does not represent and process language serially, in modular and independent boxes, as suggested by the proverbial computer metaphor of the mind. Therefore, in this article, we outline the emerging new theoretical framework in which processing language in the brain, such as understanding a spoken sentence in real time, equates to traversing a *multidimensional state space* in regions constrained by multiple probabilistic cues (e.g., sublexical, lexical, semantic, syntactic, pragmatic, *etc.*) that simultaneously and continuously imbue partial activation to various linguistic representations. For instance, as we will detail, phonetic variation can influence a syntactic parse from the bottom up, and a pragmatic inference can alter the perception of an ambiguously heard word in a top-down fashion. This new scientific visualization of language comprehension as a trajectory through a single multi-dimensional space, where all information sources and their constraints are conjoined, may help the field let go of its tendency to cling to the original inspiration of boxes-and-arrows and their adjunctive, incremental revisions.

Multiple dimensions simultaneously contribute to the understanding of an utterance, or to bootstrap the child into its first linguistic productions. In fact, the mind may not even separate representations (e.g., symbols) from processing (e.g., rules applied on those symbols); see [15–17]. Instead, the mind generally combines information sources (e.g., phonology, semantics, syntax, *etc.*) such that every substantive change in a neural activation pattern (which corresponds to a transition in state space) is impelled by *emergent dynamical interactions* among multiple types of cues, rather than by an *individual command* that is generated by a single rule-system or encapsulated module. However, before we flesh out a detailed description of this emerging visualization, the next section first discusses what the traditional visualization of language entails for a characterization of language processing. We

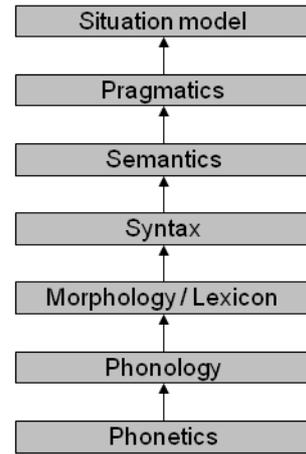
then discuss how an interim model evolved during the 1990s to accommodate new empirical data. Finally, we argue that, like the Plum Pudding Model (Figure 1B) and its mere decade of popularity, this interim interactive box-and-arrow model (Figure 1F) should now be abandoned as largely inadequate.

Figure 1. Visualizations and conceptualizations of the atom have changed dramatically in the history of physics. Physics transitioned from the early Solid Sphere Model (A) to an interim Plum Pudding Model (B) in which the discovery of subatomic particles was ‘shoe-horned’ into the original conceptualization of the solid sphere, to the revolutionary Solar System Model (C) in which most of the atom is empty, to the probabilistic generalization of the solar system model, known as the Electron Cloud Model (D). Changes of visual conceptualizations in the history of linguistic inquiry are also altering our understanding of how the mind represents and processes language. The standard Box-and-Arrow Model of language (E), borrowed from flow chart descriptions in engineering and computer science in the 1950s, posited entirely independent modules with a feed-forward flow of information. The figure represents a model of language comprehension. For language production, the direction of the arrows is simply reversed, yet the model still treats the system as entirely unidirectional in information flow. The interim model (F) includes a tangle of additional arrows to accommodate recent evidence for feedback processes among the putative levels. Blue arrows indicate evidence from adult psycholinguistics; orange arrows indicate evidence from language acquisition. Analogous to the Plum Pudding Model, this interim visualization is as an attempt to ‘shoe-horn’ the flow of multiple sources of information into the old tenet of encapsulated modularity. Dotted frames around the ‘modules’ indicate how these have been progressively understood as less and less encapsulated than in the standard model, due to the highly interactive nature of language processing. The State-Space Model (G), and its statistical generalization (H), is a complete reorganization of how to visualize the way language works. If the various information sources in language are actually interdependent, probabilistic, and continuously integrated, then they should be conceived of as sharing a conjoined state space. For example, different utterances (*i.e.*, trajectories) that use the same verb would all briefly visit that verb’s general region in this conjoined linguistic space. In this figure, only three dimensions are depicted for ease of representation, with the full model understood as a very high-dimensional state space, whose axes are not necessarily orthogonal. In addition, each knowledge dimension may be composed of several sub-dimensions, each bringing a weighted contribution. For instance, prosodic, metric and phonetic information all contribute to phonological knowledge. The solid arrows in Figure 1F indicate feedback and feed-forward influences among linguistic dimensions as suggested in our review of empirical studies.

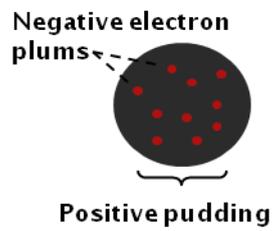
A.



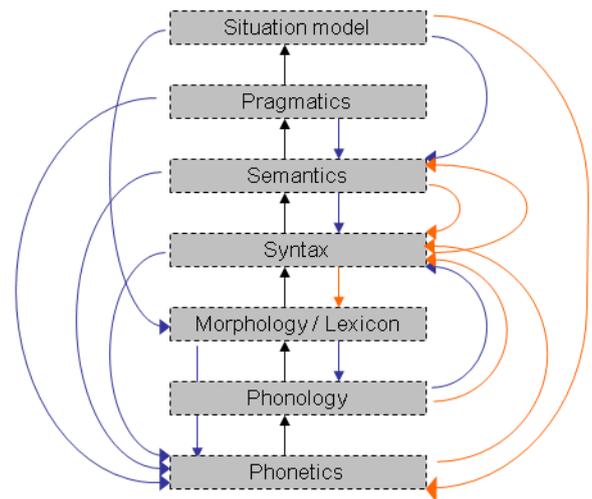
E.



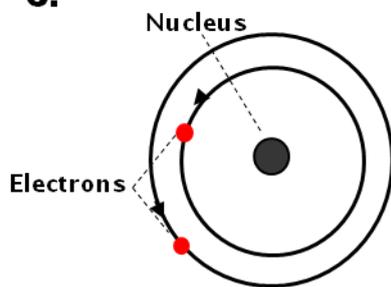
B.



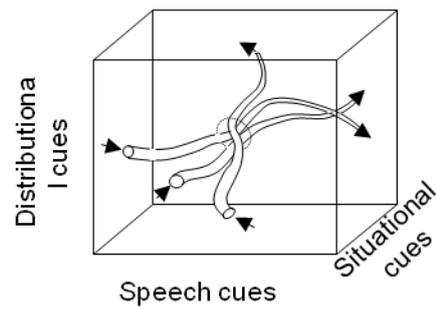
F.



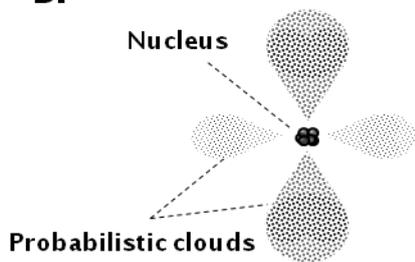
C.



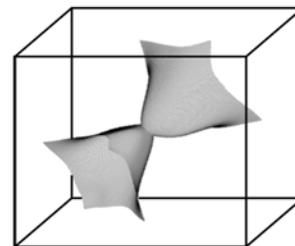
G.



D.



H.



2. The Traditional Visualization of the Language System

In 1951, at a symposium at Cornell University, a group of psychologists and linguists including John Carroll, Charles Osgood and Thomas Sebeok gave birth to a new science and coined it “psycholinguistics” [18]. Although their theoretical perspective stemmed predominantly from behaviorism, the field of psycholinguistics itself quickly came under the influence of the same burgeoning computing theory that was influencing the rest of cognitive psychology, with its engineering descriptions of components and flow-charts [19]. Before long, logical formalisms from linguistics [20] and information-processing models from cognitive psychology [21] ushered into ascendancy the box-and-arrow model on which psycholinguistics textbooks continue to focus (Figure 1F).

Three fundamental assumptions underlie this traditional box-and-arrow model of language:

- (1) linguistic knowledge is encapsulated into discrete modular levels of analysis and representation,
- (2) stage-based, feed-forward processes regulate the flow of information between levels,
- (3) processes are rank-ordered, *i.e.*, earlier processes take priority.

These three assumptions are captured in the box-and-arrow model of Figure 1E. Boxes correspond to the static and encapsulated levels, whereas arrows represent distinct processes operating serially on linguistic knowledge. *Modularity* posits that linguistic levels are informationally encapsulated, *i.e.*, the workings of each level can be explained independently from any other level. The assumption of *stage-based, feed-forward flow of information* posits that the neural subsystems responsible for each level of representation wait until a stable unique representation has been computed before that information is passed on to the next processing stage. The *rank order* assumption further posits that a given level is dominant over others.

Numerous empirical results have been interpreted to support this box-and-arrow scientific visualization. We highlight two sets of findings that form a large part of the evidence brought to bear for feed-forward modularity in its heyday of the 1980’s. The first claim grew out of experiments in syntactic ambiguity resolution. It was thought that the language processor computed a unique syntactic analysis of a sentence by default without any influence from semantic content or contextual plausibility. A sentence like “*The horse raced past the barn fell*” [22] contains a temporary syntactic ambiguity, which produces significantly slowed reading times. One interpretation of this phenomenon is that a reader or listener builds the syntactic structure consistent with the horse doing the racing (rather than being raced by someone) because of a syntactic preference for the structure with the fewest branching nodes [23] and this essentially leads comprehension “down a garden path” that is inconsistent with subsequent input. Therefore, by the end of the sentence, the verb *fell* has nowhere to attach and thus cannot be grammatically integrated into the sentence, producing confusion and long reading times. Thus, the syntactic parse is obligatorily launched down a path because syntax is an independent module (see 1 above) [24,25]; this parse can only be ‘repaired’ at the next clause boundary (if incorrect) because it is a linear feed-forward process (see 2) [26]; and contextual biases from semantics and pragmatics cannot prevent the garden-path effect because syntax has priority over them (see 3) [27,28]. That is, much like syntax would appear to only use what it is given by the lexicon

(it cannot make its own suggestions back to the lexicon), semantics would only initially use what it is given by syntax (it cannot make its own suggestions back to syntax).

The second claim supporting this feed-forward modular box-and-arrow framework for psycholinguistics resulted from experiments in lexical ambiguity resolution. For example, the word “rose” is ambiguous between being a noun and a past tense verb. Immediately after hearing “They all rose”, participants show priming for both versions of the word, responding equally quickly to “stood” and to “flower”, but a couple of hundred milliseconds later, the influence of syntactic context causes priming to be limited to the verb meaning, “stood” [29–31]. It was argued that the lexical module is autonomous, and therefore there is a brief period of time during which only phonological input (no contextual information) can activate lexical entries in the mental lexicon. Somewhat similar limitations on context have been reported in studies of the *reading* of ambiguous words [32,33].

Due to findings like these, the feed-forward modular framework has generally set the psycholinguistic agenda for the past few decades. One could argue that the box-and-arrow scientific visualization (Figure 1E) has guided the search for empirical data to a much greater extent than the empirical data has guided the choice of that model. In fact, in 1987, Arthur Reber lamented this overwhelming *a priori* preference for model over data as spelling doom for the field [34]. This situation began to change in the late 1980’s and throughout the 90’s.

As we briefly recount a few examples of the many dissenting findings from the past 20 years, one can observe that the deconstruction of the box-and-arrow model has proceeded in two steps. First, it has become evident that a much greater number of feed-forward and feed-back processes (arrows) were needed among the putative levels. Over the years, this has led to the gradual development of an interim model of highly-interactive processes, in which a semblance of seriality and modularity is preserved. The second stage of deconstruction chipped away at the very notions of modularity and seriality, arguing that the interactive processes are inherently continuous, not occurring in temporal stages, and that the representations themselves are not *static objects* but temporally *dynamic events*. The result is that both the boxes and the arrows have lost much of their descriptive power and psychological reality. Therefore, now is the time when a new scientific visualization of language is needed.

3. An Interim Scientific Visualization: Projections All the Way Up and Down

Recent psycholinguistic studies have indicated that processing language involves both a bottom-up and a top-down information flow. In the laboratory, this can be shown more eloquently by eliciting some form of temporary linguistic ambiguity, observing how the system settles into one among several potential interpretations of a given input and tracking the fine-grained time course of such interpretation. There is empirical evidence that top-down information flows at several levels of analysis. Lexical effects on speech perception are shown when a target word (e.g., /male/) is spotted earlier and more accurately when preceded by another word (e.g., /calculusmale/) than when preceded by a non-word (e.g., /baltulufmale/). This is the case even when bottom-up information from phonotactics facilitate the non-word condition (/fm/ is less frequent within words than /sm/) [35]. Lexical information facilitates speech perception especially when sounds are ambiguous or degraded [36]. There are top-down effects of syntax on speech perception: when phrases with

alternative possible segmentation (*/take spins/* or */takes pins/*) follow a plural context (*/those women takespins/*), the syntactically congruent target (*/spins/*) is detected faster, even if *pins* is acoustically favored by bottom-up coarticulation [37].

Effects of semantics on speech perception are also documented: monitoring a target word (e.g., */gap/*) is faster and more accurate when the preceding word is semantically related (*/deepeninggap/*) than when it is unrelated (*/pseudonymgap/*). This is the case even if phonotactic regularities favor the semantically unrelated condition (*/mg/*: a low frequency diphone in English words, as contrasted with the more frequent */ng/* [35]. Higher-order information is also effective in speech disambiguation. Pragmatic influences on phonetic segmentation show that information inconsistent with acoustic cues causes listeners to modify their segmentation choice in the direction of the context [38]. When presented with unsegmented near-homophonous phrases (*/plum pie/* or */plump eye/*) the target that was pragmatically congruent with a preceding appended phrase was chosen faster and more accurately. Therefore *The baker looked at the drawing of a...* favored a segmentation into *plum pie*, whereas *The surgeon looked at the drawing of a...* favored a segmentation into *plump eye*. Rohde & Ettliger [39] embedded acoustically ambiguous pronouns (sounding between *he* and *she*) in sentence contexts that were either female-biasing (e.g., *Abigail annoyed Bruce because Xe was in a bad mood*) or male-biasing (*Tyler deceived Naomi because Xe couldn't understand the situation*). Their participants' acoustic judgments of the pronouns were influenced by the biasing contexts, suggesting that interactive processes emerge between the two most separate domains, phonetic perception and discourse-level pragmatics. In addition, time-course analyses suggested that the effect is present at the earliest stage of processing.

Pragmatic effects on semantic interpretation, or “pragmatic normalization”, have been invoked to cover cases where a knowledge-based interpretation is given to sentences expressing unusual propositions [40]. Situational context affects semantics. In the presence of a semantically ambiguous word, e.g., *bulb*, a strongly constraining context such as *The gardener dug a hole. She inserted the bulb carefully into the soil* primed only a contextually supported meaning (e.g., *flower*; [41], see also [42–44]. Conversely, a weakly constraining context such as *The scout patrolled the area. He reported the mine to the commander* primed both senses (e.g., *coal* and *explosive*) of the ambiguous noun *mine* [45]. Word recognition is also facilitated when real world contextual information is provided, suggesting that situational context affects lexical processing [46]. Perhaps most famously, there are clear effects of semantics on syntax. The sentence *The land mine buried in the sand exploded* has exactly the same structure as *The horse raced past the barn fell* but crucially does not induce a syntactic garden-path effect. If syntax were an independent module, it should be equally difficult to process these two sentences. However, the semantic constraints imposed by the lexical items *landmine* and *buried* steer the reader away from the garden path, implicating a more interactive perspective on sentence processing [47]. van Berkum, Brown, & Hagoort [48] obtained pragmatic effects on syntactic processing.

3.1. Boxes and Arrows in Language Acquisition

The box-and-arrow model implicitly underlies many of the assumptions about putative developmental stages of language acquisition. Here the question typically posed is “what source of information

introduces the child to the language acquisition process?” Several proposals have been put forth. Semantics cueing syntax: Young children may discover lexical categories by first noting semantic or referential information. For instance, people and objects are linked to nouns, actions are linked to verbs, and agents of actions are linked to subjects [49,50]. Similarly, it has been proposed that syntax can cue semantics. Verb syntactic frames may help the child narrow down the meaning of verb structures [51,52]. Other proposals have highlighted how higher-order information can be gleaned from low-level cues. There are phonetic cues to the lexical representations: within their first year of life infants become sensitive to language-specific cues such as coarticulation (in /aɪskrɪm/ the diphone /sk/ is more overlapping in “I scream” while the diphone /ɪs/ is more coarticulated in “ice cream”), predominant stress patterns (in English stress-initial words, e.g., *racket*, are much more frequent than stress-final words, e.g., *guitar*; [53], and phonotactics (the segment /-ng/ never starts a word in English; for a review, see [54]). Other studies have shown that phonetics can cue syntax. Prosodic cues that are perceptually available in the first year of life (intonation and stress patterns, phoneme coarticulation) cue the child to discover the structural units of language, word and phrase structure boundaries [55–57]. Sounds at the edge of words can reliably signal noun and verb categories [58]. Noticing transitional probabilities between phonemes and syllables can indicate word boundaries in running speech [59] and distributional relationships among form-based cues can cue syntactic categories. For instance, the child could note co-occurrence relations between certain fragments (*the_cat*, *a_car*, and *has_gone*, *to_play*, but not *the_gone*, *to_cat*) or use distributional word endings regularities (*work-s*, *work-ing*, *work-ed*) to infer lexical categories [60]. Social and situational factors impact perceptual learning: American infants learned a Chinese contrast not present in English only when they were immersed in live social interactions, but crucially not via TV or audio-only [61].

3.2. Limits of the Interim Model

The review above highlights how a strong interconnectivity exists between information sources in language at all putative levels, both in adult language processing and in language acquisition. In principle, a box-and-arrow model augmented with additional processing arrows could still be invoked as consistent with the above data. Under this “interim” paradigm, it may still be assumed that there are identifiably separate levels of representation in which information flows both “bottom-up” and “top-down”. For example, a modular account of single word reading, the dual-route model (e.g., [62]), was updated to share some assumptions about cascaded processing and bi-directional information flow that are embodied in more neurally-inspired connectionist models (e.g., [63]). And the MERGE model of speech perception [64,65] accounts for higher-order word-level influence on phonemic decisions by asserting that while speech perception is feed-forward, phonemic decisions are made by a (later) decision-making mechanism that is sensitive to multiple levels of representation (but cf. [66]). Similarly, the variable-choice reanalysis model of syntactic processing accommodates immediate integration of multiple sources of contextual information, while still attempting to preserve a discrete stage-based selection among syntactic alternatives [67,68]; but cf. [69].

While interactive processing remains an area of intense debate, we identify here some important limitations that derive from thinking about language in terms of separate levels of representation and

processing. We focus on three arguments from the following research areas: (1) language development; (2) continuous linguistic processing in adults; and (3) the neurocognitive architecture of the brain.

Let us consider language acquisition by children first. Thinking in terms of self-contained modular blocks of knowledge operated by stage-based processes implies that the child must first possess mature linguistic competence in at least one putative linguistic level, which would bootstrap the rest—such as lexical segmentation depending on fully-learned phonology. This assumption has led researchers into severe ‘chicken-and-egg’ conundrums. For instance, the very existence of lexical categories (e.g., *dog* and *dinner* are nouns, while *eat* and *run* are verbs) presupposes knowledge of syntactic rules that apply to them, such as how to generate noun and verb phrases. However, to learn those syntactic rules, one must already have those lexical categories in place. In this and other types of chicken-and-egg paradoxes, the question that has been (ill-)posed is “what knowledge comes *absolutely first* in language development?” In fact, both perceptual [70] and statistical learning abilities [71] of infants and toddlers develop gradually and simultaneously over time. Likewise, syntactic knowledge in toddlers such as verb argument structure [72] and pragmatic knowledge [73] seem far from adult competence. Several corpus and behavioral studies now suggest that children’s early language tends to contain mostly phrases and utterance fragments that have been heard and used before. Very little substitution of lexical items or application of abstract syntactic patterns occurs in early language productions [74].

Since no level of representation can be independently mature at any stage to single-handedly kick start the learning process of the others, a more viable solution is that different representations co-develop gradually, *mutually assisting each other* via multi-directional correlations. This characterization of language development is inconsistent with the modular view of language knowledge as encapsulated into independent boxes.

Another related problem with the single bootstrapping ‘level’ is that invariably no single cueing approach provides a perfect correlation with a given linguistic structure. For instance, distributional learning can fall prey to spurious correlations such as *John eats meat*, *John eats slowly*, *The meat is good*, which would erroneously lead the child to infer that *The slowly is good* should be grammatical [50,75]. Likewise, prosodic cueing is only partially successful because phonological phrase boundaries do not map perfectly into syntactic phrase boundaries [76–78]. Similarly, Pinker [79] discussed the limitations of syntactic information for discovering detailed semantic aspects of verbs (the syntactic bootstrapping argument). He pointed to the semantics that could be gleaned from each of the three following examples: *Sew the shirt* (indicated activity over an object), *Sew me a shirt*, (activity of creating an object for a beneficiary), *Sew a shirt out of the rags* (activity of transformation of a material into an object). Pinker’s conclusion was that only a broad semantic inference of *activity* could be gleaned from the set of syntactic constructions in which the verb *sew* occur, but nothing like an enumerated list of semantic features. The semantic bootstrapping argument too seems to suffer from limitations, as many early nouns and verbs do not refer to things and events in the world, many words in English are both verbs and nouns, and all other lexical categories in the best cases display many-to-many mappings with the physical world (colors, as indicated by adjectives, can refer to quite disjointed sets of objects). Thus, all cueing approaches mentioned above would seem to fail if taken in isolation, and indeed this fact has been used as an argument to disqualify a given source of information as useful *at all*. However, these counter-arguments, instead of being used against each other, should be taken as

suggesting that the putative ‘levels’ of linguistic knowledge can only develop if they are conceived as part of a highly interactive system, *i.e.*, if they are allowed to modify and to be modified at the same time, by developing several multi-directional projections and connectivities. We contend that one way to avoid the paradoxes and problems expounded above is to abandon the idea that there exist modular levels of representations. Thus, new research can begin examining how various sources of information contribute to the co-emergence of dimensions of representation [80]. Since each dimension of knowledge emerges gradually, multiple dimensions can provide statistical support *to one another* in highly interactive ways, and perfect correlations are not necessary.

A second serious problem with the box-and-arrow model strikes at the very core of the modularity and seriality assumptions. It has been shown that the uptake of linguistic information in real time is continuous (not stage-like) and is simultaneously mediated by several sources of information. For example, as a spoken word unfolds over the course of a few hundred milliseconds, the continuous uptake of acoustic-phonetic information instigates partial activation of numerous lexical representations that share the same initial phonetic features [46,81–83]. This “cohort” of activated lexical representations quickly winnows as the latter portion of the word is finished. Thus, during the first hundred milliseconds of hearing the word “candy” being spoken, the brain partially activates mental representations for *candy, candle, candid, candelabra, cannelloni, cancer, etc.* During the second hundred milliseconds, a smaller set of just *candy, candle, candid and candelabra* is active. Finally, near the end of the word, only *candy* remains substantially activated. This means that the flow of information from the phonetics “level” to the lexical “level” is uninterrupted. Moreover, as feedback from the lexical “level” to the phonetics “level” appears to be similarly continuous [66,84], the motivation for treating these “levels” as though they are separate systems arranged in a meaningful sequence melts away, encouraging us instead to treat phonetics, the lexicon, and perhaps even semantic properties of words as co-existing in a unified high-dimensional state space [85].

Similar to spoken word recognition, the field of sentence processing has been gradually approaching consensus with regard to the rapid integration of cues from bottom-up and top-down processes. Just as the continuous uptake of the acoustic signal during spoken word recognition involves the integration of multiple cues at the time scale of dozens of milliseconds, so does the resolution of syntactic ambiguity involve the continuous integration of multiple cues at the time scale of hundreds of milliseconds [86–89]. The research points to an account of real-time language comprehension that continuously integrates lexical, syntactic, semantic, discourse, visual, and even situational variables in real time. In light of this work, language comprehension no longer looks like the functioning of a rule-based flowchart, with subprocessors waiting until they finish constructing a symbolic representation before sending it to the next subprocessor. Partial, incomplete information (in the form of probabilistic biases) is shared continuously, and interpretation does not depend on waiting until a ‘syntactic module’ has completed its autonomous parsing process.

A third argument in favor of seeing language as a fully connected multi-dimensional space comes from neuroscience and the architecture of the brain. Electrophysiological recordings show that distributed representations are widely used in the cortex [90]. A distributed representation uses multiple active neuron-like processing units to encode information, and the same unit can participate in multiple representations. Units in distributed representations may represent single features (such as that a sound is plosive) or combinations of features. The implications for language are that it becomes hard to

conceive of words as individual symbolic entities stored in separate mental drawers. For example, a word beginning with /sp/ activates partial neuronal patterns that are simultaneous representations of different words from different semantic categories (e.g., spice, spy, spoon), and syntactic classes (spying, spongiform, spite, spoiler). If phonetic activations are part-and-parcel of a word, it is hard to conceive of all the nouns being grouped in a separate lexical store without at least partially activating those adjectives and verbs that have overlapping phonetic features.

Further evidence from neuroscience in support of interactive processes is that bidirectional connectivity is ubiquitous in cortex [91,92], with communication of activation flowing simultaneously in both bottom-up and top-down directions. Furthermore, there is increasing neuropsychological and electrophysiological evidence that some brain regions often assumed to be language-specific (e.g., Broca's area) are implicated in processing of other non-linguistic cognitive processing such as music [93,94]. Moreover, neural signatures of order violations in the learning of sequenced patterns are similar to those evoked by structural violations in natural language [95]. Therefore, the localization of certain linguistic functions is coextensive with at least some other non-linguistic domains. These findings make informationally-encapsulated modular linguistic subsystems much less plausible.

We have argued that linguistic information involved in language processing and language learning appear to be highly integrated and interconnected, and its uptake is continuous at the time scales of language comprehension and of language acquisition. Given such a characterization, it appears increasingly inadequate to visualize language processing as a box-and-arrow flow-chart with processes (arrows) coming in and out of static encapsulated levels. The notion of a module only makes sense if there is a delay in the information flow such that a given subsystem is forced, during at least some short period of time, to make decisions without assistance from contextual information sources. If the information flow between putative modules is continuous, then there is no period of time during which a given subsystem is making decisions based purely on its own internal algorithms. Under such circumstances, referring to the subsystem as a "module" is no longer coherent. With so many incoming arrows (Figure 1F), and with their influence being continuous in time, any given box is clearly doing quite a bit more than its label implies. For example, if the syntax module's parsing decisions are being immediately influenced by biases from phonetics [96], semantics [97], and the situation model [98], such that those contextual biases are essentially functioning as part of the parsing algorithm itself, then is it really useful to still think of it as a syntax module?

Our impression is that the sciences of language may be at a historical moment similar to when physicists in the 19th century could no longer fit the findings of subatomic entities into the obsolete uncuttable Solid Sphere model (Figure 1A), or even into the modified version of that theory, the Plum Pudding Model (Figure 1B). They ultimately had to develop a completely new scientific visualization of atomic structure. Likewise, in the following section we spell out a new scientific visualization for language that is congruent with the continuous, interactive and overlapping uptake of linguistic information during real-time language processing and acquisition.

4. A New Scientific Visualization: The Multidimensional Space Model

Multidimensional spaces have been used in cognitive science to examine the graded similarity among cognitive categories of various types, including visual objects [99], concepts [100], semantics [101,102],

phonetics [103,104] and syntax [105,106]. Given the numerous findings recounted above of mutual interaction among linguistic information sources, it seems likely that what is needed is a single conjoined state space that combines phonetic cues (including intonational cues), visual cues (including facial expressions and gestures associated with speech [107,108]), distributional/syntactic cues, and other situational cues from the environment. In this high dimensional state space, there would be frequently traveled ruts corresponding to highly common forms of utterances, and there would be graded constraints (forces of attraction) that usually prevent certain types of trajectories corresponding to ungrammatical or nonsensical utterances [109,110]. The wide range of state-space trajectories lying in between those hackneyed ruts and those infelicitous pathways constitute the creative nuances of productive sentence construction that [20] were brought to our attention fifty years ago.

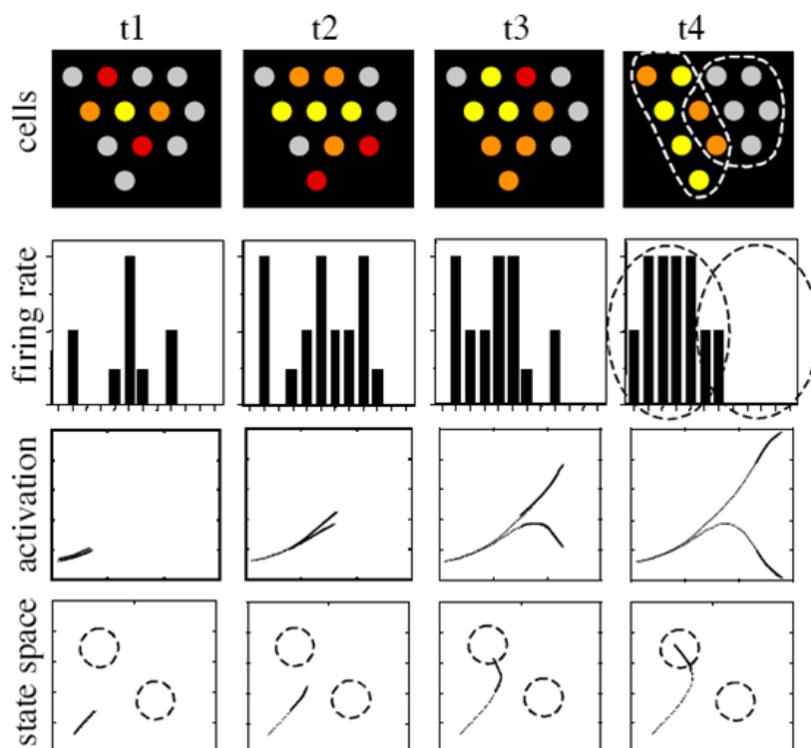
When one switches from treating language as a linear series of complex logic gates to the new approach: a nonlinear dynamical system such as this, there are some important changes in a number of model properties. For example, capacity limitations take on a very different character in a continuous state space than cognitive scientists have grown accustomed to with their computer metaphor of the mind. With a multidimensional state space, temporary ambiguity in a time-dependent signal (such as a word or a sentence) is not implemented as several mutually exclusive symbolic alternatives that are held in working memory, where discretely enumerable capacity limitations could become a serious impediment. Rather, temporary ambiguity becomes nothing more than having the state of the system move into near-equal proximity to multiple attractors in state space. Whether this location in state-space is equidistant from 2, 3, or 19 alternative attractors, the concerns regarding a combinatorial explosion exceeding some capacity limitation do not threaten to overload the system's memory buffer. There is always only one set of coordinates being instantiated in the system, regardless of how multiply-ambiguous that location is with respect to nearby attractors. Thus, the state space framework will encounter resolvability problems when the branching alternatives become too numerous (cf. the "curse of dimensionality", [111]), essentially causing it to lose the ability to distinguish between some of those alternatives. However, in the state space framework, this abundance of ambiguous alternatives will never "crash the program" by violating memory constraints or be forced to arbitrarily disregard some alternatives because the buffer is full—which is a concern with rule-and-symbol systems performing incremental processing on input that contains temporary ambiguities [112,113].

There is another important difference between the theoretical framework of boxes and arrows and that of a trajectory through a multidimensional state space. This is that while the former must be viewed as a metaphor that guides research, the latter can be viewed as a concrete mathematical description of a real physical phenomenon. In Figure 2, the top row of panels depicts an idealized set of neurons that can have a range of spike rates, with grey being the baseline spike rate and red, orange and yellow indicating higher spike rates. The seven neurons on the left constitute a coherent group that responds to a particular stimulus (e.g., the neural ensemble that represents the word "candle"). The partially overlapping group of cells on the right (see dashed circles in top right panel) constitute a coherent group that responds to a similar stimulus (e.g., the neural ensemble that represents the word "candy"). From left to right, these panels show a change in activation pattern that corresponds to both neural ensembles (or "population codes") being slightly active, then more active, then the distribution of activation begins to shift toward the left-hand neural ensemble, and finally only the left-hand neural ensemble is significantly active. Idealized here as only a handful of neurons for simplicity, the second,

third and fourth rows of Figure 2 all depict equivalent data visualizations of the same physical process whereby populations of neurons change their pattern of activation over time [114]; [12,115,116]. That is, the dynamic patterns of activated concepts (third row), and the trajectory through an attractor landscape (fourth row), are not mere visual metaphors in the way that the box-and-arrow framework is. They can be taken as mathematical descriptions that quantitatively approximate concrete physical phenomena, and not as metaphorical qualitative abstractions of the physical phenomena.

Finally, if we step away from the theoretical perspectives that rely so much on static representational objects that are held in a memory buffer, and focus instead on the observation that a sentence's production and its comprehension take place over a period of time, we can better appreciate this alternative, more neurally plausible, account for how a spoken sentence is understood. While a person hears a sentence, it is obviously not the case that the brain is doing nothing until the sentence is finished and then it constructs a static mental representation. As the sentence unfolds over time, the listener's brain is continuously undergoing changes in its patterns of neural activation that are significantly driven by this environmental auditory input. If we describe these average firing rates over time as locations in a high dimensional state space, then the changes over time plot a continuous trajectory through that state space. Thus, the understanding of a sentence is here conceived of as an event in the mind, not an object [109,117,118]. If some of the dimensions that one could identify in this state space are largely phonetic, and others are largely semantic, and still others largely syntactic, one can imagine collections of trajectory traces (from different periods of time) that bundle near a particular region of state space (which corresponds to a cluster of highly similar patterns of neural activation) "belonging" to particular words [105,106]. See Figure 1.G. When a word is understood, it means that the listener's brain has briefly achieved a pattern of neural activation that maps roughly onto that region in state space. Hence, understanding a sentence involves having this moving average of neural firing rates change over time such that it corresponds to a continuous trajectory through the state space, smoothly traveling from one word's region to another and to another, spending at least as much time *in between* those attractors as *inside* them. These word regions then form clusters that act as fuzzy categories of Noun, Verb, Adjective, *etc.* [105], and within each word's region are sub-regions that correspond to situations in which that same word is being used as the Subject of the sentence or the Object or with even more subtle nuances [106]. Thus, grammatical structure, linguistic productivity, and even systematicity can be achieved in this state space with trajectories that have nested loops on themselves and precise entry and exit points in each word's region [119,120]. See also [121]. In this framework, it becomes useful to conceive of words not as the *operands* on which linguistic algorithms operate, but instead as the *operators* themselves, whose directional proclivities in state space are what implement the linguistic algorithms [106].

Figure 2. Two population codes (circled with dashed lines in the rightmost top panel) are mutually competing, and one will win. The figure depicts a theoretical mapping of the equivalence, over four 100ms time periods, among a neural population code's distributed activity pattern (top row, where red-orange-yellow means greater activity), its cell firing rates (second row), activation of the mental representations that correspond to the two population codes (third row) and the trajectory through state space that settles into one attractor (*i.e.*, stable population code) rather than the other. The spike rates of these cells can be treated as coordinates in a 12-dimensional space, of which the bottom row depicts a 2-dimensional compression. During the first time period (column t1), the two partially overlapping groups of cells each have a few of their members somewhat active and spiking (responding to combinations of phonetics, semantics, syntax, *etc.*). In the third row, the two representations are slowly emerging during this initial time period (where the activation curves are starting to rise over time). In the fourth row, the state of the system is just beginning to move toward the two attractors (dashed circles) in state space. During the second time period (column t2), both population codes (*i.e.*, representations) are continuing to increase in their respective activation patterns, and the trajectory is moving even closer to the two attractors. (Solid portions of a curve/trajectory indicate activity change during that time period, and dashed portions indicate the history of that activity change.) During the third time period (column t3), competition causes the left-hand group of cells to continue increasing its activity while the right-hand group decreases, thus the state-space trajectory begins to move toward one attractor and away from the other. During the fourth time period (column t4), one population code is nearly saturated while the other is nearly silent (dashed circles), its representation has risen to a high level of activation while the other is nearly inactive, and the trajectory has fully entered the perimeter of a particular attractor.



One possible objection to the proposed view of language processing as a large state space is that it puts no constraints on what regions of the space are equivalent for what purpose and no constraints on what the dimensionality of the space might be. Will each single utterance, in each different conversational situation, addressed to each different recipient, constitute a different trajectory in the state space? Is it not the scientific goal to understand the different types (not different tokens) of sentences or utterances or syntactic or semantic structures? Thus, this model might reasonably appear unsatisfactory because an arbitrarily large number of different states can be traversed, and the present version of the model fails to place boundaries about which trajectories are allowed in the state space and which are not. The attractiveness of the Rutherford model of the atom was not in its arguing that subatomic particles can inhabit any orbit. Rather, Planck's constant was used to provide constraints on where these orbits could be, how many could co-exist and which types of orbit-configurations had similar properties.

This 'anything goes' objection to a dynamical systems approach to language is a fair criticism of the current state of the art, but it is not sufficient to cast doubt on the future of this research program. It seems clear from comparing Panels E and F of Figure 1 that the highly-constrained modular approach to language has required so many revisions over the years that it may simply be a misleading default framework to have as our standard. Instead of assuming massive restrictions and then having to relax nearly all of them in the model, perhaps better progress in the language sciences will be made by assuming almost no restrictions in the model and then adding the ones we empirically find over time.

Some New Concepts and Venues for the Language Sciences

We have started reviewing a re-conceptualization implied in a dynamical systems view of language processing. This re-conceptualization was anticipated back in the 1990s in the broader context of cognitive science [122], and enough empirical evidence has accumulated in support since then (partly summarized above) to make it a viable framework for the language sciences. This view promises to go far beyond reinterpreting existing psycholinguistic phenomena and issues, by leading to new alleys of research. In particular, it provides a set of tools for re-conceptualizing language learning. Four such concepts are likely to become center stage in the new science of language: (a) The role of multicausality and circular causality; (b) the notion of emergence structure; (c) the concept of nonlinearity; and (d) a greater attention to understanding individual differences in language. These concepts provide a unifying conceptual toolbox. In reviewing these concepts, we also briefly discuss how novel experimental and statistical techniques might be better suited for the tasks at hand.

Circular causality implies that no single element, source of information, or mechanism in the course of learning and processing language has causal priority independently. Although this may sound a truism intuitively, current widespread statistical methods do not incorporate these concepts. Two of the core assumptions behind (multiple) regression analysis, for example, are the *independent* contributions of predicting variables to predict a dependent variable. Notice that the decision of which is the dependent and independent variable rests heavily on the theoretical assumption of the experimenter. The current assumption built into most of our scientific efforts is thus one of imposed unidirectional causality, not of multicausality. Another fundamental assumption in regression is the absence of collinearity among predicting variables. A strong correlation between variables makes it difficult or

impossible to estimate their individual regression coefficients reliably. In everyday data collection however, collinearity is extremely common. Dynamical systems theory turns collinearity into an asset, by acknowledging that multiple variables affect each other over time. In studying language, many processes may be heavily subjected to this principle. For example, speech perception abilities get reorganized as a result of exposure to the onset of reading. [123] showed that perception of native and nonnative speech contrasts can be respectively sharpened and attenuated as a result of (late and novel) experience with phoneme-to-grapheme conversion rules as a product of reading instruction. Children between 6 and 8 years who are good readers for their age show greater reduction of perception of non-native speech contrasts, suggesting that speech perception in this period is affected by a newly (and relatively late) acquired skill.

Causal loops may be effective at different time scales, from the microlevel of combining several cues online for sentence comprehension, to the macrolevel of days, months or years of learning, and may have persistent global effects on the system. Recent studies of young children raised in bilingual contexts have observed bidirectional effects of one language on the other in many varieties of linguistic behavior, from pronunciation to morphosyntax, from grammaticality judgments to patterns of lexical storage, activation, and retrieval [124,125]. The interpenetration of the two language systems makes it impossible for either the L1 or the L2 of a bilingual to be identical in all respects to the language of a monolingual. Grosjean notably made the case that a bilingual is not two monolinguals in one person [126].

Causal loops with interactive feedbacks also get established between interactants. For example, children whose mothers responded immediately to their infants' vocalizations produced more mature and adult-like vocalizations than when the social interaction was not contingent [127]. In order to study two interactants as a unique dyadic system, new quantitative techniques may be required. Dale and Spivey [128] do so by applying recurrence quantification analysis (RQA) to the study of syntactic coordination between child and caregiver in a large corpus of child-parent interactions: "The method is based on analyzing sequences of syntactic elements, time series of grammatical usage, allowing comparison of two such sequences and revealing patterns of recurrence. The ordered sequences of concern here are time series of syntactic class usage by child and caregiver. The approach therefore provides a window on how structures used by the child "recur" in those used by caregiver (and *vice versa*)" ([128] p. 395). These and other studies suggest a strong role of social interaction in shaping the construction of phonemic and syntactic knowledge. Importantly, the same set of analytic and conceptual techniques based on dynamical analyses can now be applied to study both internal (cognitive) phenomena and social (communicative/interactive) dimensions, promising to bind together both internalist and externalist traditions in the social sciences in a new scientific visualization of language and communication (the theme of the Special Issue of this volume).

Emergence via multiple cues. Emergence in science refers to the making of new forms through ongoing processes intrinsic to the system. Emergent properties are often higher-order patterns from lower-level processes and their generation is not inbuilt *a priori* in the system. Think of soap bubbles; despite their differences in size and amount of airflow blown into, soap bubbles invariably 'come out' in systematic shapes (e.g., spherical when in the air, hexagonal when packed next to each other in a single plane). Yet there is no 'program' dictating the shape of bubbles. Rather, the particular shape emerges as an optimal self-organizing solution to keeping all particles on the surface of the bubble in

an equal state of chemical ‘tension’ in relation with neighboring particles. When ported to the study of language, the concept of emergence may have far-reaching consequences. The confluence of several low-level cues (e.g., phonetic) may give rise to higher-order structured representations (e.g., phonemes, syllables, morphemes, *etc.*). This suggests, for example, that infants’ segmentation strategies across languages (syllable, time, or mora-based) may emerge as optimal solutions to probabilistic cues inherent in speech. Rather than assuming that pre-existing inbuilt switches are turned on early in infancy, such segmentation preferences may emerge as the best self-organized solution to the task of identifying words in running speech. The very notion of a phoneme as an abstract entity analogous to a written letter of the alphabet poses several challenges and has been questioned. The perception of speech relies on many probabilistically weighted acoustic cues, and integration of information across several acoustic dimensions is a central characteristic of auditory processing [36]. For example, as many as 16 acoustic dimensions may characterize the perception of voicing contrasts, as in the difference between English /ba/ and /pa/ [129]. Furthermore, the variety of actual pronunciations for any linguistic chunk that speakers may hear is seemingly unlimited and may vary along many continuous parameters, including individual speakers’ voices, regional variations and subtleties of the social and pragmatic context. Since babies learn to recognize such a range of minute phonetic and temporal differences and can control many of them in their own speech as they grow, it is problematic to assume that linguistic memory could store only a canonical, abstract representation based on a minimal number of “distinctive” phonetic features in serial order. Thus, the comprehension and production of speech is a good example of how language does not appear to live in a low-dimensional phonological code. Speech decoding requires a far richer and more concrete representation of acoustic events necessary for storing concrete fragments and for the emergence of invariant linguistic properties (see [130]).

Furthermore, while lexicon and syntax have been traditionally treated as requiring separate mechanisms and underlying neural bases, their emergence in language acquisition appears to be tightly coupled by mutual, progressive loops [131,132]. Here, emergence is closely tied to the concept of circular causality discussed above, because grammar is an emergent property of slow, incremental local abstractions over an increasingly larger repertoire of words and phrases. Another ambitious program for emergentism is to demonstrate that the existence of relatively small variations in the options available to all languages may result from optimal emerging solutions among the constraints on communication and biological constraints on processing sequences of sounds [133]. For example, the widespread appearance of systems of 5 vowels (e.g., Latin, Italian, Spanish, Swahili, *etc.*) is possibly an optimal amount, given constraints on speech production and perception. Hence 5-vowel systems are widespread and stable over long periods of historical time. Similarly, constraints on sequential learning abilities may favor certain word orders over others [134]. For a thorough discussion of where putative language universals may come from, see [135].

Nonlinearity. Small quantitative changes in one or more components of the language system or its input can lead to reorganization and large qualitative differences in behavior [136,137]. Nonlinear behavior will help us explain strong individual differences in normally developing children, unaccounted for by the standard model, as well as departures from normal language abilities. For example, apparently subtle low-level phonological deficits [138,139] and/or processing deficits for sequential

patterns and sounds [140] may have profound and detrimental cascading effects on higher-order morpho-syntactic development, resulting in syndromes such as Specific Language Impairments.

Furthermore, in a dynamical language system, multiple forces (biological, social, and maturational constraints) are likely to shape speakers' linguistic experience in non-linear and unique ways [141]. Therefore, in contrast to more traditional approaches to language, there is likely to be growing interest in individual differences. Implicitly or explicitly, it is widely assumed that all native speakers uniformly possess the same core language competence, with variations in performance relegated to extralinguistic processing factors such as working memory capacity. However, considerable differences have been found in processing core aspects of syntactic knowledge. Dabrowska and Street [142] showed that participants with more schooling performed better than less educated participants in recognizing the agent (the 'doer') of simple passive sentences. Less educated participants seemed to rely on simple processing heuristics rather than syntactic arguments. More surprisingly, the authors found that second language learners performed better than less educated native speakers in the same syntactic task. This superior result was attributed to possible effects of explicit grammatical instruction on the syntactic productivity of the non-native speakers (similar results were obtained for the comprehension of complex syntactic structures [143], and with native and non-native speakers of Japanese [144]). In a dynamical system then, it becomes important to investigate the specific trajectories of learning as they are shaped monthly or even weekly by specific exposure to environmental circumstances [145]. In such a characterization, language learning ceases to be the inevitable cause of a single maturational program pre-wired in the brain.

5. Conclusions

We have presented a new visual and conceptual apparatus for interpreting and explaining language in the mind. The importance of such a change in how we visually depict our models of language should not be underestimated. Similar changes in scientific visualization were a driving force in early 20th century physics. During the emergence of quantum physics in the 1920's, scientific visualizations of the atom reveal that physicists were beginning to "worry less about what atoms *are*" and instead to "think more about what they *do*" [146]. More recently, the growing field of scientific visualization is developing a wide range of techniques for identifying patterns in complex multi-dimensional data sets that lead to the formulation and testing of theories in a variety of disciplines [1,2,147]. The scientific visualization offered by a dynamical-systems approach to language processing and acquisition encourages us to focus not so much on *which* locations in state space are visited, *i.e.*, the representations that are constructed, but instead on *how* language travels through its state space, *i.e.*, the continuous trajectories that are pursued. What does the future hold for the language sciences? We can envision two possible scenarios. In one scenario the two views of language expounded here are seen as competing explanations. As such, the ongoing advances of the newer perspective on language promise to assist the rest of cognitive science in a transition from its classical framework, in which symbolic representations were openly described as the indivisible atoms of the mind [148–150] to the more continuous and distributed framework of dynamical cognition [12,15–17,100]. In a second scenario, the two frameworks will coexist as complementary descriptions at different levels of abstraction; they will ask and answer different questions about the nature of language. Again, an

equivalent precedent in the history of physics may be invoked, that being the time when physicists reached a consensus about the dual nature of light (as both a particle and a wave), because experimental evidence supports both theories. Whether the new language framework proposed here will supplant the standard one or whether the two frameworks will coexist is hard to tell. This paper offers a first step toward conceptualizing language in a novel way that arguably goes beyond several received assumptions of language, naturally accommodates a number of findings that are problematic for the traditional perspective and proposes some possible venues for novel research.

Acknowledgments

We are grateful to Shimon Edelman and Giacomo Cacciapaglia for comments on earlier versions of the manuscript. We also thank two anonymous reviewers for comments that helped us improve the manuscript. This work was supported by NIH grant R03-HD051671 to LO and NSF grant BCS-0721297 to MJS.

References

1. Ochs, E.; Jacoby, S.; Gonzales, P. Interpretive journeys: How physicists talk and travel through graphic space. *Configurations* **1994**, *2*, 151–171.
2. Oestermeier, U.; Hesse, F.W. Verbal and visual causal arguments. *Cognition* **2000**, *75*, 65–104.
3. Pullman, B. *The Atom in the History of Human Thought*; Reisinger, A., Transl.; Oxford University Press: Oxford, UK, 1998; ISBN 0195150406.
4. Forster, K. Levels of processing and the structure of the language processor. In *Sentence Processing: Psycholinguistic Studies Presented to Merrill Garrett*; Cooper, W., Walker, E., Eds.; Erlbaum Press: Hillsdale, NJ, USA, 1979; pp. 27–85.
5. Chomsky, N. *The Minimalist Program*; MIT Press: Cambridge, MA, USA, 1995.
6. Forster, K.; Hector, J. Cascaded versus noncascaded models of lexical and semantic processing: The turtle effect. *Mem. Cogn.* **2002**, *30*, 1106–1117.
7. Frazier, L.; Clifton, C. *Construal*; MIT Press: Cambridge, MA, USA, 1995.
8. Jackendoff, R. *Foundations of Language*; Oxford University Press: New York, NY, USA, 2002.
9. Sorace, A.; Serratrice, L. Internal and external interfaces in bilingual language development: Beyond structural overlap. *Int. J. Biling.* **2009**, *13*, 195–210.
10. Staub, A. Word recognition and syntactic attachment in reading: Evidence for a staged architecture. *J. Exp. Psychol.* **2011** (in press).
11. Elman, J.L. Language as a dynamical system. In *Mind as Motion*; Port, R., van Gelder, T., Eds.; MIT Press: Cambridge, MA, USA, 1995; pp. 195–223.
12. Spivey, M.J. *The Continuity of Mind*; Oxford University Press: New York, NY, USA, 2007.
13. Trueswell, J.; Tanenhaus, M. *Approaches to Studying World-Situated Language Use: Bridging the Language-as-Product and Language-as-Action Traditions*; MIT Press: Cambridge, MA, USA, 2005.
14. Kutas, M.; DeLong, K.A.; Smith, N.J. A look around at what lies ahead: Prediction and predictability in language processing. In *Predictions in the Brain: Using Our Past to Generate a Future*; Bar, M., Ed.; Oxford University Press: New York, NY, USA, 2011; pp. 190–207.

15. Elman, J.L.; Bates, E.A.; Johnson, M.H.; Karmiloff-Smith, A.; Parisi, D.; Plunkett, K. *Rethinking Innateness: A Connectionist Perspective on Development*; MIT Press: Cambridge, MA, USA, 1996.
16. Thelen, E.; Smith, L. *A Dynamic Systems Approach to the Development of Cognition and Action*; MIT Press: Cambridge, MA, USA, 1994.
17. Barsalou, L.W.; Breazeal, C.; Smith, L.B. Cognition as coordinated non-cognition. *Cogn. Process.* **2007**, *8*, 79–91.
18. Osgood, C.; Sebeok, T. Psycholinguistics: A survey of theory and research problems. *J. Abnorm. Soc. Psychol.* **1954**, *49*, 1–203.
19. Newell, A.; Shaw, J.C.; Simon, H.A. Elements of a theory of human problem solving. *Psychol. Rev.* **1958**, *65*, 151–166.
20. Chomsky, N. *Syntactic Structures*; Mouton: The Hague, The Netherlands, 1957.
21. Neisser, U. *Cognitive Psychology*; Appleton-Century-Crofts: New York, NY, USA, 1967.
22. Bever, T.G. The cognitive basis for linguistic structures. In *Cognition and the Development of Language*; Hayes, J.R., Ed.; Wiley: New York, NY, USA, 1970.
23. Frazier, L. On Comprehending Sentences: Syntactic Parsing Strategies. Ph.D. Dissertation, University of Massachusetts, Waltham, MA, USA, 1979.
24. Frazier, L.; Fodor, J. The sausage machine: A new two-stage parsing model. *Cognition* **1979**, *6*, 291–325.
25. Pritchett, B.L. Garden path phenomena and the the grammatical basis of language processing. *Language* **1988**, *64*, 539–576.
26. Fodor, J.; Inoue, A. The diagnosis and cure of garden paths. *J. Psycholinguist. Res.* **1994**, *23*, 407–434
27. Ferreira, F.; Clifton, C. The independence of syntactic processing. *J. Mem. Lang.* **1986**, *25*, 348–368.
28. Rayner, K.; Carlson, M.; Frazier, L. The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *J. Verbal Learn. Verbal Behav.* **1983**, *22*, 358–374
29. Tanenhaus, M.K.; Leiman, J.M.; Seidenberg, M.S. Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *J. Verbal Learn. Verbal Behav.* **1979**, *18*, 427–440.
30. Seidenberg, M.S.; Tanenhaus, M.K.; Leiman, J.; Bienkowski, M. Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cogn. Psychol.* **1982**, *14*, 489–537
31. Swinney, D.A. Lexical access during sentence comprehension: Reconsideration of context effects. *J. Verbal Learn. Verbal Behav.* **1979**, *18*, 645–659.
32. Dopkins, S.; Morris, R.K.; Rayner, K. Lexical ambiguity and eye fixations in reading: A test of competing models. *J. Mem. Lang.* **1992**, *31*, 461–476.
33. Duffy, S.A.; Kambe, G.; Rayner, K. The Effect of Prior Disambiguating Context on the Comprehension of Ambiguous Words: Evidence from Eye Movements. In *On the Consequences of Meaning Selection*; Gorfein, D., Ed.; American Psychological Association Press: Washington, DC, USA, 2001.
34. Reber, A. The rise and (surprisingly rapid) fall of psycholinguistics. *Synthese* **1987**, *72*, 325–339.

35. Mattys, S.L.; White, L.; Melhorn, J.F. Integration of multiple speech segmentation cues: A hierarchical framework. *J. Exp. Psychol.* **2005**, *134*, 477–500.
36. McClelland, J.L.; Mirman, D.; Holt, L.L. Are there interactive processes in speech perception? *Trends Cogn. Sci.* **2006**, *10*, 363–369.
37. Mattys, S.L.; Melhorn, J.F.; White, L. Effects of syntactic expectations on speech segmentation. *J. Exp. Psychol.* **2007**, *33*, 960–977.
38. Mattys, S.L.; Melhorn, J.F. Sentential, lexical, and acoustic effects on the perception of word boundaries. *J. Acoust. Soc. Am.* **2007**, *122*, 554–567.
39. Rohde, H.; Ettliger, M. Effects of pragmatic inference on phonetic category perception. In *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, Ohlsson, S.; Catrambone, R., Eds.; Mahwah, New Jersey: Lawrence Erlbaum Associates: Portland, OR, USA, 11–14 August 2010.
40. Sanford, A.J.; Sturt, P. Depth of processing in language comprehension: Not noticing the evidence. *Trends Cogn. Sci.* **2002**, *6*, 382–386.
41. Vu, H.; Kellas, G.; Paul, S.T. Sources of sentence constraint in lexical ambiguity resolution. *Mem. Cogn.* **1998**, *26*, 979–1001.
42. Fitneva, S.A.; Spivey, M.J. Context and language processing: The effect of authorship. In *World Situated Language Use: Psycholinguistic, Linguistic and Computational Perspectives on Bridging the Product and Action Traditions*; Trueswell, J., Tanenhaus, M., Eds.; MIT Press: Cambridge, MA, USA, 2004.
43. Kawamoto, A.H. Nonlinear dynamics in the resolution of lexical ambiguity: A parallel distributed processing account. *J. Mem. Lang.* **1993**, *32*, 474–516.
44. Tabossi, P.; Colombo, L.; Job, R. Accessing lexical ambiguity: Effects of context and dominance. *Psychol. Res.* **1987**, *49*, 161–167.
45. Martin, C.; Vu, H.; Kellas, G.; Metcalf, K. Strength of discourse context as a determinant of the subordinate bias effect. *Q. J. Exp. psychol.* **1999**, *52*, 813–839.
46. Allopenna, P.D.; Magnuson, J.S.; Tanenhaus, M.K. Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *J. Mem. Lang.* **1998**, *38*, 419–439.
47. Tanenhaus, M.K.; Trueswell, J. Sentence comprehension. In *Handbook of Perception and Cognition*; Miller, J., Eimas, P., Eds.; Academic Press: San Diego, CA, USA, 1995; Volume 11, Speech and Language Communication.
48. Van Berkum, J.J.A.; Brown, C.M.; Hagoort, P. Early referential context effects in sentence processing: Evidence from event-related brain potentials. *J. Mem. Lang.* **1999**, *41*, 147–182.
49. Grimshaw, J. Form, Function, and the language acquisition device. In *The Logical Problem of Language Acquisition*; Baker, C.L., McCarthy, J., Eds.; MIT Press: Cambridge, MA, USA, 1981.
50. Pinker, S. *Language Learnability and Language Development*; Harvard University Press: Cambridge, MA, USA, 1984.
51. Gleitman, L. The structural sources of verb meanings. *Lang. Acquis.* **1990**, *1*, 1–55.
52. Naigles, L. Children use syntax to learn verb meanings. *J. Child Lang.* **1990**, *17*, 357–374.
53. Jusczyk, P.W.; Cutler, A.; Redanz, N. Preference for the predominant stress patterns of English words. *Child Dev.* **1993**, *64*, 675–687.

54. Jusczyk, P.W. *The Discovery of Spoken Language*; MIT Press: Cambridge, MA, USA, 1997.
55. Jusczyk, P.; Hirsh-Pasek, K.; Kemler-Nelson, D.; Kennedy, L.; Woodward, A.; Piwoz, J. Perception of acoustic correlates of major phrasal units by young infants. *Cogn. Psychol.* **1992**, *24*, 252–293.
56. Morgan, J.L.; Demuth, K. *Signal to Syntax*; LEA: Mahwah, NJ, USA, 1996.
57. Soderstrom, M.; Seidl, A.; Kemler Nelson, D.; Jusczyk, P. The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *J. Mem. Lang.* **2003**, *49*, 249–267.
58. Onnis, L.; Christiansen, M.H. Lexical categories at the edge of the word. *Cogn. Sci.* **2008**, *32*, 184–221.
59. Saffran, J.R.; Aslin, R.N.; Newport, E.L. Statistical learning by 8-month old infants. *Science* **1996**, *274*, 1926–1928.
60. Maratsos, M.; Chalkley, M. The internal language of children's syntax. In *Children's Language*; Nelson, K.E., Ed.; Gardner Press: New York, NY, USA, 1980; Volume 2.
61. Kuhl, P.K.; Tsao, F.M.; Liu, H.M. Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 9096–9101.
62. Coltheart, M.; Rastle, K.; Perry, C.; Langdon, R.; Ziegler, J. DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychol. Rev.* **2001**, *108*, 204–256.
63. Plaut, D.C.; McClelland, J.L.; Seidenberg, M.S.; Patterson, K. Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychol. Rev.* **1996**, *103*, 56–115.
64. Norris, D.; McQueen, J.M.; Cutler, A. Merging information in speech recognition: Feedback is never necessary. *Behav. Brain Sci.* **2000**, *23*, 299–370.
65. McQueen, J.M.; Jesse, A.; Norris, D. No lexical-prelexical feedback during speech perception or: Is it time to stop playing those Christmas tapes? *J. Mem. Lang.* **2009**, *61*, 1–18.
66. Magnuson, J.S.; McMurray, B.; Tanenhaus, M.K.; Aslin, R.N. Lexical effects on compensation for coarticulation: The ghost of Christmash past. *Cogn. Sci.* **2003**, *27*, 285–298.
67. Van Gompel, R.P.G.; Pickering, M.J.; Pearson, J.; Liversedge, S.P. Evidence against competition during syntactic ambiguity resolution. *J. Mem. Lang.* **2005**, *52*, 284–307.
68. Van Gompel, R.P.G.; Pickering, M.J.; Traxler, M.J. Reanalysis in sentence processing: Evidence against current constraint-based and two-stage models. *J. Mem. Lang.* **2001**, *45*, 225–258.
69. Farmer, T.; Anderson, S.; Spivey, M.J. Gradiency and visual context in syntactic garden-paths. *J. Mem. Lang.* **2007**, *57*, 570–595.
70. Aslin, R.N.; Smith, L.B. Perceptual development. *Annu. Rev. Psychol.* **1988**, *39*, 631–682.
71. Gómez, R.L.; Maye, J. The developmental trajectory of nonadjacent dependency learning. *Infancy* **2005**, *7*, 183–206.
72. Tomasello, M. Do young children have adult syntactic competence? *Cognition* **2000**, *74*, 209–253.
73. Ninio, A.; Snow, C. The Development of Pragmatics: Learning to Use Language Appropriately. In *Handbook of Language Acquisition*; Bhatia, T.K., Ritchie, W.C., Eds.; Academic Press: New York, NY, USA, 1999; pp. 347–383.
74. Tomasello, M. Construction Grammar for kids. *Constructions* **2006**, *1*, 1–11.

75. Hsu, A.; Chater, N. The logical problem of language acquisition goes probabilistic: No negative evidence as a window on language acquisition. *Cogn. Sci.* **2010**, *34*, 972–1016.
76. Fisher, C.; Tokura, H. Prosody in speech to infants: direct and indirect acoustic cues to syntactic structure. In *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*; Morgan, J.L., Demuth, K., Eds.; Erlbaum: Hillsdale, NJ, USA, 1996; pp. 343–364.
77. Gerken, L.A.; Jusczyk, P.W.; Mandel, D.R. When prosody fails to cue syntactic structure: Nine-month-olds' sensitivity to phonological versus syntactic phrases. *Cognition* **1994**, *51*, 237–265.
78. Jusczyk, P.W.; Kemler Nelson, D.G. Syntactic units, prosody, and psychological reality during infancy. In *Signal to Syntax*; Morgan, J.L., Demuth, K., Eds.; Erlbaum: Hillsdale, NJ, USA, 1996; pp. 389–408.
79. Pinker, S. How could a child use verb syntax to learn verb semantics? *Lingua* **1994**, *92*, 377–410.
80. MacWhinney, B. *The Emergence of Language*; LEA: Mahwah, NJ, USA, 1999.
81. Marslen-Wilson, W. Functional parallelism in spoken word-recognition. *Cognition* **1987**, *25*, 71–102.
82. Spivey, M.J.; Grosjean, M.; Knoblich, G. Continuous attraction toward phonological competitors: Thinking with your hands. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 10393–10398.
83. Vitevitch, M.S.; Luce, P.A. Probabilistic phonotactics and neighborhood activation in spoken word recognition. *J. Mem. Lang.* **1999**, *40*, 374–408.
84. Elman, J.L.; McClelland, J.L. Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *J. Mem. Lang.* **1988**, *27*, 143–165.
85. Rodd, J.M.; Gaskell, M.G.; Marslen-Wilson, W.D. Modelling the effects of semantic ambiguity in word recognition. *Cogn. Sci.* **2004**, *28*, 89–104.
86. Knoeferle, P.; Crocker, M.W. The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. *Cogn. Sci.* **2006**, *30*, 481–529.
87. MacDonald, M.C.; Pearlmutter, N.J.; Seidenberg, M.S. The lexical nature of syntactic ambiguity resolution. *Psychol. Rev.* **1994**, *101*, 676–703.
88. McRae, K.; Spivey-Knowlton, M.J.; Tanenhaus, M.K. Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *J. Mem. Lang.* **1998**, *38*, 283–312.
89. Tabor, W.; Tanenhaus, M. Dynamical models of sentence processing. *Cogn. Sci.* **1999**, *24*, 491–515.
90. Rao, S.C.; Rainer, G.; Miller, E.K. Integration of what and where in the primate prefrontal cortex. *Science* **1997**, *276*, 821–824.
91. Douglas, R.J.; Koch, C.; Mahowald, M.; Martin, K.A.C.; Suarez, H.H. Recurrent excitation in neocortical circuits. *Science* **1995**, *269*, 981–985.
92. Crick, F.; Asanuma, C. Certain aspects of the anatomy and physiology of the cerebral cortex. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*; McClelland, J.L., Rumelhart, D.E., Eds.; MIT Press: Cambridge, MA, USA, 1986; Volume 2, pp. 333–371.
93. Patel, A.D.; Gibson, E.; Ratner, J.; Besson, M.; Holcomb, P.J. Processing syntactic relations in language and music: An event-related potential study. *J. Cogn. Neurosci.* **1998**, *10*, 717–733.

94. Tillmann, B.; Janata, P.; Bharucha, J. Activation of the inferior frontal cortex in musical priming. *Ann. N. Y. Acad. Sci.* **2003**, *999*, 209–211.
95. Christiansen, M.H.; Conway, C.M.; Onnis, L. Neural responses to structural incongruencies in language and statistical learning point to similar underlying mechanisms. In *Proceedings of the 29th Annual Meeting of the Cognitive Science Society*, Austin, TX, USA, 29 July–1 August 2007; McNamara, D.S., Trafton, J.G., Eds.; Cognitive Science Society: Austin, TX, USA, 2007; pp. 173–178.
96. Farmer, T.A.; Christiansen, M.H.; Monaghan, P. Phonological typicality influences on-line sentence comprehension. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 12203–12208.
97. Trueswell, J.C.; Tanenhaus, M.K.; Garnsey, S. Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *J. Mem. Lang.* **1994**, *33*, 285–318.
98. Tanenhaus, M.K.; Spivey-Knowlton, M.J.; Eberhard, K.M.; Sedivy, J.E. Integration of visual and linguistic information in spoken language comprehension. *Science* **1995**, *268*, 1632–1634.
99. Edelman, S. Representation is representation of similarities. *Behav. Brain Sci.* **1998**, *21*, 449–498.
100. McClelland, J.L.; Rogers, T.T. The parallel distributed processing approach to semantic cognition. *Nat. Rev. Neurosci.* **2003**, *4*, 310–322.
101. Regier, T. *The Human Semantic Potential. Spatial Language and Constrained Connectionism*; MIT Press: Cambridge, MA, USA, 1996.
102. Snider, J.; Osgood, C. *Semantic Differential Technique: A Sourcebook*; Aldine Atherton: New York, NY, USA, 1969.
103. Nycz, J. The dynamics of near-merger in accommodation *Proc. ConSOLE XIII* **2005**, 273–285.
104. Shepard, R. Psychological representation of speech sounds. In *Human Communication: A Unified View*; David, E.E., Denes, P.B., Eds.; McGraw-Hill: New York, NY, USA, 1972; pp. 67–113.
105. Elman, J.L. Distributed representations, simple recurrent networks, and grammatical structure. *Mach. Learn.* **1991**, *7*, 195–224.
106. Elman, J.L. An alternative view of the mental lexicon. *Trends Cogn. Sci.* **2004**, *7*, 301–306.
107. Hubbard, A.L.; Wilson, S.; Callan, D.E.; Dapretto, M. Giving speech a hand: Gesture modulates activity in auditory cortex during speech perception. *J. Hum. Brain Mapp.* **2009**, *30*, 1028–1037.
108. Kelly, S.D.; Ozyurek, A.; Maris, E. Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychol. Sci.* **2010**, *21*, 260–267.
109. Culicover, P.W.; Nowak, A. *Dynamical Grammar: Minimalism, Acquisition, and Change*; Oxford University Press: Oxford, UK, 2003.
110. Tabor, W.; Hutchins, S. Mapping the Syntax/Semantics Coastline. In *Proceedings of the 22nd Annual Meeting of the Cognitive Science Society*, Philadelphia, PA, August 13–15 2000; Gleitman, L.; Joshi, A., Eds.; Lawrence Erlbaum Associates: Mahwah, NJ, USA; pp. 511–516.
111. Bellman, R.E. *Adaptive Control Processes*; Princeton University Press: Princeton, NJ, USA, 1961.
112. Marcus, M.P. *Theory of Syntactic Recognition for Natural Languages*; MIT Press: Cambridge, MA, USA, 1980.
113. Jurafsky, D. A probabilistic model of lexical access and disambiguation. *Cogn. Sci.* **1996**, *20*, 137–194.

114. Rolls, E.T.; Tovee, M.J. Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *J. Neurophysiol.* **1995**, *73*, 713–726.
115. Port, R.F. Language is a social institution: Why phonemes and words do not have explicit psychological form. *Ecol. Psychol.* **2010**, *22*, 304–326.
116. Wright, J.J.; Liley, D.T.J. Dynamics of the brain at global and microscopic scales: Neural networks and the EEG. *Behav. Brain Sci.* **1996**, *19*, 285–309.
117. Slobin, D.I. From “thought and language” to “thinking to speaking”. In *Rethinking Linguistic Relativity*; Gumperz, J.J., Levinson, S.C., Eds.; Cambridge University Press: Cambridge, MA, USA, 1996; pp. 70–96.
118. Kempson, R.M.; Gabbay, D.M.; Meyer-Viol, W. *Dynamic Syntax: The Flow of Language Understanding*; Wiley-Blackwell Publishing: Hoboken, NJ, USA, 2001.
119. Tabor, W. Fractal encoding of context-free grammars in connectionist networks. *Expert Syst.* **2000**, *17*, 41–56.
120. Tabor, W.; Juliano, C.; Tanenhaus, M.K. Parsing in a dynamical system: An attractor-based account of the interaction of lexical and structural constraints in sentence processing. *Lang. Cogn. Process.* **1997**, *12*, 211–271.
121. Botvinick, M.; Plaut, D.C. Doing without schema hierarchies: A recurrent connectionist approach to routine sequential action and its pathologies. *Psychol. Rev.* **2004**, *111*, 395–429.
122. Van Gelder, T.; Port, R. It’s about time: Overview of the dynamical approach to cognition. In *Mind as Motion: Explorations in the Dynamics of cognition*; Port, R., van Gelder, T., Eds.; Bradford Books/MITP: Boston, MA, USA, 1995; pp. 1–43.
123. Burnham, D.; Tyler, M.; Horlyck, S. Periods of speech perception development and their vestiges in adulthood. In *An Integrated View of Language Development: Papers in Honor of Henning Wode*; Burmeister, P., Piske, T., Rohde, A., Eds.; Wissenschaftlicher Verlag Trier: Trier, Germany, 2002; pp. 281–300.
124. Cook, V.J. *Effects of the Second Language on the First*; Multilingual Matters: Clevedon, UK, 2003; p. 268.
125. Davies, A. *The Native Speaker: Myth and Reality*; Multilingual Matters: Clevedon, UK, 2003.
126. Grosjean, F. *Studying Bilinguals*; Oxford University Press: Oxford, UK, 2008.
127. Goldstein, M.H.; King, A.P.; West, M.J. Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 8030–8035.
128. Dale, R.; Spivey, M.J. Unraveling the dyad: Using recurrence analysis to explore patterns of syntactic coordination between children and caregivers in conversation. *Lang. Learn.* **2006**, *56*, 391–430.
129. Lisker, L. “Voicing” in English: A catalogue of acoustic features signaling /b/ vs. /p/ in trochees. *Lang. Speech* **1986**, *29*, 3–11.
130. Port, R. Rich memory and distributed phonology. *Lang. Sci.* **2010**, *32*, 43–55.
131. Bannard, C.; Lieven, E.; Tomasello, M. Modeling children’s early grammatical knowledge. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 17284–17289.
132. Bates, E.; Goodman, J. On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia and real-time processing. *Lang. Cogn. Process.* **1997**, *12*, 507–584.
133. Christiansen, M.H.; Kirby, S. *Language evolution*; Oxford University Press: Oxford, UK, 2003.

134. Christiansen, M.H.; Chater, N. Language as shaped by the brain. *Behav. Brain Sci.* **2008**, *31*, 489–558.
135. Evans, N.; Levinson, S.C. The myth of language universals: Language diversity and its importance for cognitive science. *Behav. Brain Sci.* **2009**, *32*, 429–448.
136. Elman, J.L. Learning and development in neural networks: The importance of starting small. *Cognition* **1993**, *48*, 71–99.
137. McMurray, B. Defusing the childhood vocabulary explosion. *Science* **2007**, *317*, doi:10.1126/science.1144073.
138. Joanisse, M.F.; Seidenberg, M.S. Specific language impairment in children: An impairment in grammar or processing? *Trends Cogn. Sci.* **1998**, *2*, 240–246.
139. Ziegler, J.C.; Pech-Georgel, C.; George, F.; Alario, F.-X.; Lorenzi, C. Deficits in speech perception explain language learning impairment. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 14110–14115.
140. Tallal, P.; Piercy, M. Defects of non-verbal auditory perception in children with developmental aphasia. *Nature* **1973**, *241*, 468–469.
141. Kuhl, P.K. A new view of language acquisition. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 11850–11857.
142. Dąbrowska, E.; Street, J. Individual differences in language attainment: Comprehension of passive sentences by native and non-native English speakers. *Lang. Sci.* **2006**, *28*, 604–615.
143. Chipere, N. Native speaker variations in syntactic competence: Implications for first language teaching. *Lang. Aware.* **2001**, 107–124.
144. Sasaki, Y. Individual variation in a Japanese sentence comprehension task: Form, functions and strategies. *Appl. Linguist.* **1997**, *18*, 508–537.
145. Smith, L.; Thelen, E. Development as a dynamic system. *Trends Cogn. Sci.* **2003**, *7*, 343–347.
146. Lindley, D. *Uncertainty*; Doubleday: New York, NY, USA, 2007.
147. Card, S.; Mackinlay, J.; Shneiderman, B. *Readings in Information Visualization: Using Vision to Think*; Morgan Kaufmann Publishers: San Francisco, CA, USA, 1999.
148. Newell, A. Physical symbol systems. *Cogn. Sci.* **1980**, *4*, 135–183.
149. Baker, M.C. *The Atoms of Language*; Basic Books: New York, NY, USA, 2001.
150. Anderson, J.R.; Lebiere, C. *The Atomic Components of Thought*; Lawrence Erlbaum Associates: Mahwah, NJ, USA, 1998.