A Compositional Neural Architecture for Language

Andrea E. Martin[1,2]

Abstract

■ Hierarchical structure and compositionality imbue human language with unparalleled expressive power and set it apart from other perception–action systems. However, neither formal nor neurobiological models account for how these defining computational properties might arise in a physiological system. I attempt to reconcile hierarchy and compositionality with principles from cell assembly computation in neuroscience; the result is an emerging theory of how the brain could convert distributed perceptual representations into hierarchical structures across multiple timescales while representing interpretable incremental stages of (de) compositional meaning. The model's architecture—a multidimensional coordinate system based on neurophysiological models of sensory processing—proposes that a manifold of neural trajectories encodes sensory, motor, and abstract linguistic states. Gain modulation, including inhibition, tunes the path in the manifold in accordance with behavior and is how latent structure is inferred. As a consequence, predictive information about upcoming sensory input during production and comprehension is available without a separate operation. The proposed processing mechanism is synthesized from current models of neural entrainment to speech, concepts from systems neuroscience and category theory, and a symbolic-connectionist computational model that uses time and rhythm to structure information. I build on evidence from cognitive neuroscience and computational modeling that suggests a formal and mechanistic alignment between structure building and neural oscillations, and moves toward unifying basic insights from linguistics and psycholinguistics with the currency of neural computation. ■

INTRODUCTION

Natural language enables us to produce and understand words and sentences that we have never encountered before, as long as we (and the words and sentences) play by the rules. This fact is particularly startling if you consider that human language is processed and generated by a biological organ whose general remit is to be driven by statistical regularities in its environment. The human brain manifests a paradox[1] when it comes to language: Despite the clear importance of statistical knowledge and distributional information during language use and language acquisition, our everyday language behaviors exemplify an ability to break free from the very (statistical) vice that bootstrapped us up into the realm of natural language users in the first place.

For example, although we have specific expectations about what a given word should sound or look like, we do not require an exact physical copy, as a machine might, nor do we fail to recognize a word if a person previously unknown to us produced it. Furthermore, although we might learn a word in a phrase or sentence context, or might tend to experience that word more often in one context than in another, we are by no means limited to recognizing or using that word "only" in that context, or only in related contexts, or only in the contexts that we have ever experienced it in. As such, a marvelous expressive capacity is extended to us—the ability to understand and produce sentences we have never encountered before, to generate and express formal structures that lead to contextually specific compositional meanings. Although this capacity may seem pedestrian to us, it sets language apart from other perception–action systems and makes language behavior vexingly difficult to account for from a neuroscientist's and computationalist's point of view. One of the system properties that underlies this capacity in language is compositionality, whereby units or structures compose (and decompose) into meanings that are determined by the constituent parts and the rules used to combine them (Partee, 1975).

The study of language spans a 3000-year tradition in philosophy (e.g., the Rigveda [O'Flaherty, 1981]; Aristotle, Plato, De Saussure [Robins, 2013]) up to the recent formalizations of the last 60 years in linguistics (e.g., Kratzer & Heim, 1998; Hornstein, 1984; Partee, 1975; Chomsky & Halle, 1968; Lenneberg, 1967; Halle, 1962; Chomsky, 1957) and has revealed the pantheon of linguistic forms that the systematicity of mind can take (Phillips, 2020; Phillips & Wilson, 2010; Fodor & Pylyshyn, 1988). The last century has also seen astonishing progress in neuroscience (e.g., Buzsáki, 2006, 2019; Ballard, 2015; Marder, 2012; Gallistel, 1990; Hebb, 1949; Ramon y Cajal, 1928) and in artificial intelligence (e.g., Hinton, Osindero, & Teh, 2006; Rumelhart, McClelland, & PDP Research Group, 1987),

[1]Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands, [2]Donders Centre for Cognitive Neuroimaging, Radboud University, Nijmegen, The Netherlands

yielding powerful, complex models (e.g., Tenenbaum, Kemp, Griffiths, & Goodman, 2011; Doumas, Hummel, & Sandhofer, 2008). But all this remarkable progress has yet to offer a satisfying explanation as to how the defining features of human language arise within the constraints of a neurophysiological system (for discussion, see Brennan & Martin, 2020; Martin & Doumas, 2017, 2019a, 2019b; Baggio, 2018; Martin, 2016; Embick & Poeppel, 2015; Hagoort, 2003, 2013; Friederici, 2002, 2011). Without an explanatory neurophysiological and computational account (Kaplan, 2011; Kaplan & Craver, 2011; Piccinini, 2007) of the quintessential properties of human language—of hierarchical structure and domain, of function application and scope, and, most definitely, of compositionality—our theories of language and the human mind and brain seem startlingly incomplete.

In this paper, I attempt to simultaneously consider the basic constraints of network computation in a neurophysiological system, the core formal properties of language, and the psycholinguistics of language processing. These topics are traditionally treated as individual subjects for theories and models, which, as a result, leads to necessarily independent theories and models. However, the capacity that these theories wish to explain and the problems that these theories and models face are often tacitly common to all domains. Thus, in my view, the topic is best served by an integrated solution, however difficult it may be to achieve. A comprehensive view of language in the mind and brain requires consideration of (and obedience to) the hard constraints on each domain, because in the limit it is these constraints that shape any viable solution. Unless we as a field are interested in psychological models that cannot be implemented in neural systems or in neurophysiological models that have no meaning in linguistics or psychology, our only choice is to develop theories under the constraints of multiple levels of analysis. I advocate the view that we must build models that pay heed to the constraints on computation (see van Rooij, Blokpoel, Kwisthout, & Wareham, 2019; Blokpoel, 2018; van Rooij, 2008), and in this particular case, we must obey the constraints that physiological systems impose while also capturing the formal properties of language we set out to account for. In an attempt to determine how linguistic representations could be expressed in the brain, I apply concepts from neurophysiology and dynamical systems neuroscience, broadly construed (e.g., neural oscillations, cell assemblies, gain modulation [including inhibition], sensory recoding or coordinate transformation, neural trajectories, and manifolds[2]), to psycholinguistics. Then, I try to face the formal facts by considering how basic compositionality could be achieved within existing neurophysiological models of sensory coding for systems that also guide action (e.g., vision → pointing, grasping). I propose that linguistic structure building is a form of perceptual inference—the ability to infer the presence of a stimulus, often from incomplete or partial sensory information. Perceptual inference is based on information stored in neural representations that were acquired through experience (Aggelopoulos, 2015) and that are generated internally in response to stimulation (Martin, 2016). I posit a processing mechanism for perceptual inference via the neural transformation of sensory codes to structured representations. The mechanism operates over a manifold of neural trajectories or the activity of a neural population projected into a space whose dimensions represent unit activation in time. Increasingly abstract structures during language comprehension are inferred via gain modulation, the way in which neurons combine information from two or more sources (Salinas & Sejnowski, 2001). Inhibition, or the interruption, blockade, or restraint of neural activity in time and space (Jonas & Buzsáki, 2007) is a form of gain modulation, and plays a key role in combining and separating information during language processing.

In the first section, I argue that (de)compositionality implies that the neural state space of linguistic representation is inherently multidimensional and thus is best described as dimensions in a manifold of neural trajectories. The neuroscientific and linguistic ways in which these dimensions relate can be described mathematically as transformations that stand in particular relations or morphisms to one another across multiple coordinate systems in cortical time. Coordinates of each dimension range from sensory-registered values (e.g., topographic, retinal, or head-centered values, outside language) to abstractions that correspond to the units of linguistic analysis at hand (e.g., phonetic features, semantic features, possible syntactic relations in a grammar). Abstract structures are built from sensory codes via coordinate transform; a given dimension in the manifold can be weighted according to the demands of behavior, with the resulting activation being a form of neural gain modulation on relevant dimensions, which in turn controls state transitions and further coordinate transforms. In psycholinguistic terms this can be referred to as structure building. States are generated, in line with contemporary and emerging thought in neuroscience (e.g., Buzsáki, 2019; Ballard, 2015; i.e., perceived as higher level structures during comprehension via inductive inference or during speaking or signing, deduced from knowledge of language and its functor with both conceptual and sensory knowledge).

The second section describes a possible implementation of the architectural principles from the first section. It also focuses on a neurophysiological mechanism for how linguistic structures could be generated from sensory input via a gain modulation-based mechanism, which (a) accounts for the unbounded combinatorial nature of language, (b) can encode hierarchy in a sequence and vice versa, and (c) makes predictions about energy expenditure in cortical networks that can be tested empirically. The proposals in the first and second sections synthesize psycholinguistics with the cognitive neuroscience of language via computational principles that have relevance across the cognitive and brain sciences.

I. A NEURAL ARCHITECTURE FOR LINGUISTIC REPRESENTATION AS PERCEPTUAL INFERENCE

Language comprehension can be characterized as a perceptual detection task wherein the percept to be detected is abstract structure, meaning, and intention of the speaker. Percepts or latent structures beyond sensation (see Table 1 for cartoon illustrations of various formal accounts of the representations at stake) must be inferred from noisy and often incomplete sensory representations of physical signals, using existing implicit grammatical, semantic, contextual, and procedural knowledge to make an inference about what the latent structure of the stimulus is likely to be given sensory evidence (Martin, 2016; Marslen-Wilson & Tyler, 1980). Helmholtz (1867) famously characterized perception as an inferential process[3]—one based on sensory input but exceeding that input by using the products of past experience (see also Olshausen, 2014; Yuille & Kersten, 2006; Ernst & Bülthoff, 2004). Thus, the language comprehension system, in contrast with the production system, is inferential and probabilistic, a characterization that perceptual systems in modern neuroscience receive despite internal tensions regarding precise mathematical expression (e.g., Gershman & Niv, 2010; Beck et al., 2008; Ma, Beck, Latham, & Pouget, 2006). To comprehend is to take an exogenous signal or set of sensory cues and combine them with linguistic knowledge—endogenous signals—the representations that sensory cues elicited from memory (Martin, 2016). On this view, language comprehension is a form of "analysis-by-synthesis" (Poeppel & Monahan 2011; Bever & Poeppel, 2010; Marslen-Wilson & Welsh, 1978; Halle & Stevens, 1962), whereby cues in the speech signal activate or trigger inference about higher level representations as projected by grammatical knowledge in the comprehender (for a process model, see Martin, 2016, and for theoretical frameworks of a similar spirit, see Marslen-Wilson & Tyler, 1980; Marslen-Wilson & Welsh, 1978). Comprehension cast this way has a strong probabilistic component, which is in line with dominant theories of word recognition and sentence comprehension over the last several decades (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994; Dell, 1986). But the characterization that I advocate here contrasts strongly with purely statistical, frequentist, or associationist accounts because it embraces the symbolic nature of language and, indeed, capitalizes upon it to perform inference over noisy and variable input. Note that embedding probabilistic activation functions within an analysis-by-synthesis model does not mean that abstract symbolic representations of language are no longer necessary—in fact, such an account claims that symbols are the perceptual targets to be inferred during comprehension and are what is "counted" or induced during statistical learning and during language acquisition (cf. Martin & Doumas, 2017, 2019a, 2019b; Doumas & Martin, 2018; Doumas, Puebla, & Martin, 2017; Martin, 2016; Holland, Holyoak, Nisbett, & Thagard, 1986). Perceptual inference asserts that sensory cues activate latent representations in the neural system that

Table 1. Cartoon Examples of Some of the Representational Systems in Linguistics. Neural Systems Must Implement Functionally Adequate Expressions of These Representations if They are to Remain Faithful to Formal Principles that Shape Language and Behavior

Time flies like an arrow.

/taɪm flaɪz laɪk ən ˈærəʊ/

Dependency parse	Syntactic tree	Role-filler binding calculus	The lambda calculus

Dependency parse:

(ROOT

  (S

    (NP (NNP Times))

    (VP (VBZ flies)

    (PP (IN like)

    (NP (DT an) (NN arrow))))

  (. .)))

http://nlp.stanford.edu:8080/parser/index.jsp
http://mshang.ca/syntree/
Syntax code: [IP [DP[D'[D]][NP [N Time]]] [I'[I][VP [V" [V flies] [AdvP[Adv'[Adv like][^DP an arrow]]]

Role-filler binding calculus:

flies like(time, an arrow)

flies (time)   like(an arrow)

The lambda calculus:

Time flies like an arrow (S)
FLIES LIKE(t, a)

Time (NP)   Flies like an arrow (VP)
t   λx. FLIES LIKE(x, a)

Flies like (V)   An arrow (NP)
λx. λy. FLIES LIKE (x, y)   a

FLIES LIKE (a two-place predicate)
t: TIME (simplifying, a constant)
a: AN ARROW (simplifying, another constant)

have been learned through experience.[4] In line with this idea, there is ever-accumulating evidence that "lower level" cues like speech rate and phoneme perception (e.g., Kaufeld, Ravenschlag, Meyer, Martin, & Bosker, 2020; Kaufeld, Naumann, Meyer, Bosker, & Martin, 2019; Heffner, Dilley, McAuley, & Pitt, 2013; Dilley & Pitt, 2010), morphology (e.g., Gwilliams, Linzen, Poeppel, & Marantz, 2018; Martin, Monahan, & Samuel, 2017), foveal and parafoveally processed orthography (e.g., Cutter, Martin, & Sturt, 2020; Veldre & Andrews, 2018; Schotter, Angele, & Rayner, 2012), as well as "higher level" sentential (e.g., Martin, 2018; Martin & McElree, 2008, 2009, 2011, 2018; Kutas & Federmeier, 2011; Ferreira & Clifton, 1986; van Alphen & McQueen, 2006) and discourse representations (e.g., Nieuwland & Martin, 2012; Nieuwland, Otten, & Van Berkum, 2007; Nieuwland & van Berkum, 2006; Sturt, 2003) can interact to bias perception in constraining ways.

Linguistic Representation in Neural Terms: *N*-dimensional Manifolds of Neural Trajectories

The notion that linguistic structure is a product of perceptual inference implies that there are multiple representations at stake—for our purposes, neural states that are associated with a given sensory input or given conceptual unit to be expressed. Minimally then, we must enter a space where sensory representations can be transformed into nonsensory and increasingly abstract representations, and vice versa. This neural state space, described by an *n*-dimensional manifold, has dimensions that have coordinate systems (see Glossary). The map or relation between coordinate systems and dimensions can be described by a functor. Some dimensions might have group homomorphisms or relationships that preserve algebraic structure between dimensions, for example, between syntactic structure and semantic domain or scope, whereas others do not, for example, between minimal pairs in phonemes and lexical semantic features. Thus, the degree of homomorphism between two given dimensions will shape how activation is propagated between them and the path through the manifold that reflects the transformation of a sensory cue into an abstract structure. The mathematical object *n*-dimensional manifold[5] is a useful description for our purposes, because for each point on a surface (or dimension) of a manifold, there is a homeomorphic relationship with points in a neighboring dimension, meaning that there is a continuous inverse function between dimensions we can apply to describe the transition between trajectories in the manifold. Manifolds can be used to describe neural population activity in time (e.g., Gámez, Mendoza, Prado, Betancourt, & Merchant, 2019; Bressler & Kelso, 2001, 2016; Sporns & Kötter, 2004; Amari, 1991). A neural manifold is composed of neural trajectories, in our case, of a multiple cell assemblies', activation in time. A neural trajectory typically describes the time course of population activity in a high-dimensional space, where each axis represents the firing rate of one neuron in the population; as

activity unfolds over time, a trajectory is traced out in the space (e.g., Gámez et al., 2019). A path through the coordinate systems in the manifold reflects the evolution of a linguistic representation from sensation to abstraction and back again. In this sense, language production and comprehension both are forms of nonlinear dimensionality reduction—when we perceive a word or phrase, we have reduced their acoustic instantiations into an abstract neural coding space by applying our linguistic knowledge to the neural projection of physical stimulation; when we produce a word or phrase, we are reducing the dimensions of conceptual content to a particular sequence of articulatory gestures. In summary, spatiotemporal patterns of brain activity during language processing can be described by a manifold of neural trajectories, and dimensions of that manifold must relate in particular ways to each other than can be described by the mathematical concepts of morphism and the functors between them (viz., structure-preserving functions between coordinate systems and a map between them).

Neural Gain Modulation for Coordinate Transform

If neural representations for language are definitionally multidimensional, then they require coordinate transform to move from one dimension to another in neural spacetime. I propose that this transform can occur via an existing mechanism that is repeatedly used throughout perception–action: gain modulation. As it is unlikely that any specialized brain mechanism could have emerged given the timescale that language appeared (Boeckx & Benítez-Burraco, 2014), the empirical question becomes whether existing neural coordinate transform schemes could apply to linguistic representations and faithfully account for their formal properties.

Gain modulation is the neurophysiological way to relate activity from one modality or representational coordinate system to another (Buzsáki, 2019; Salinas & Abbott, 2001; Salinas & Thier, 2000; Zipser & Andersen, 1988). It is the way neurons combine information from two or more sources, relating disparate information sources in space and time, underlying the integration of information over time. It perhaps most often described in the non-neuroscientific context of volume control; Buzsáki (2019) gives an accessible description of dialing up the volume on your radio. This control of output or volume requires two things, an amplifier and a modulator. The amplification aspect of gain modulation is a change in the response amplitude of a neuron or group of neurons (i.e., a cell assembly) as a function of selectivity, which is assumed to be dependent upon the sensory context and the behavior that is being performed by the organism (Buzsáki, 2019; Haegens, Händel, & Jensen, 2011; Jazayeri & Movshon, 2007; Salinas & Abbott, 2001; Salinas & Thier, 2000; Andersen & Mountcastle, 1983). These changes in activity are interpreted as reflecting the recruitment of the representational dimension of the neural assemblies

that has been selected by the sensory context and behavioral target. Gain modulation (both amplification of a signal and inhibition are subsumed by this term, see Buzsáki, 2019) is hypothesized to underlie coordinate transform between sensory modalities and between sensory and motor systems; it is formalized as the product of a neuron or cell assembly's response function, $f(x)$, and another's, $g(y)$, yielding a new gain field from the value given by the function, $f(x)g(y)$ (Salinas & Abbott, 2001; Salinas & Thier, 2000). The resulting product of this computation over receptive fields is a "gain field," which no longer codes for representation in a purely afferent-driven way. Gain fields are invoked to account for the transformation of neural representations from afferent retinal coordinates to efferent limb-centric coordinates, and vice versa, but also for translation invariance of an object across different locations in the visual field (e.g., Zipser & Andersen, 1988). In a trajectory manifold where dimensions relate to each other through gain modulation, gain-modulated coordinate transform is also referred to as "sensory recoding" (Jazayeri, 2008). For example, in vision, low-level visual information is processed into shape and, ultimately, into object recognition (Olshausen, 2014; Ernst & Bülthoff, 2004). In speech perception, acoustic information must be transduced from entry-level variables like pitch, intensity, and duration, coded in the cochlea and auditory cortex (Smith & Lewicki, 2006; Kim, Rhode, & Greenberg, 1986) to the first abstractions of pitch accent and linguistic stress. Coordinate transform may be a computational requirement of any system with multiple data types or formats from multiple perceptors, effectors, and behavioral goals. In models of sensory recoding, sensory representations can be separated from areas that control the responses to those sensations, allowing the system to "contemplate" or transform information and use it in other modalities and situations (Buzsáki, 2019; Jazayeri, 2008). It is worth noting that, although gain modulation and attention are strongly associated with one another, they are not synonymous; gain modulation is a system-wide factor shaping neural information processing, assembly formation, and communication (Buzsáki, 2019; Salinas & Thier, 2000). Capture of covert and overt attention certainly lead to an increase in gain modulation (e.g., Ling, Liu, & Carrasco, 2009), but the role of gain modulation in the brain is likely to be much more broad than simply as a neuronal instantiation of attention. For our purposes, the representational claims I make in the first section rely on gain fields as filtered amplifiers, which in this case do not so much amplify the input from an afferent or output to an efferent, but rather propagate aspects of the representation, which I refer to as dimensions in a neural trajectory manifold. The modulator will come into play in the second section when I describe how inhibition that then tunes the propagation of features to contact or activate representations in other coordinate systems/dimensions and to enable compositional representations to emerge without violating independence (Martin & Doumas, 2019b; Hummel, 2011). Inhibition will operate

laterally and downward in a feedback manner through the hierarchy of state transitions. It is important to note that there are multiple ways that gain modulation (both amplification and inhibition) can be implemented in the brain (see Buzsáki, 2019; Kaplan, 2011). As such, my proposal is not tied to particular neurophysiological realization of gain modulation.

Concrete Examples of Gain Modulation for Coordinate Transform from Sensation to Abstraction

If you unlock your office door while looking at the lock, the visual signal available to the brain is different than if you do the same behavior without looking. That is, the response amplitude of a given neural population can depend on the direction of gaze, as does the contribution that activation to executing the same door-opening behavior. Nonetheless, you are able to unlock your office door whether looking or not because you can transform the visual information (either currently taken up or from memory) into a coordinate space that motor action can occur in. However, in contrast, locked doors will never be opened without placing a key in the lock, so the activity contributed by the motor system during execution should be comparable whether you are looking at the lock or not, perhaps with more enhancement of internal tactile proprioceptive signals when gaze is away from the target. Gain modulation allows the system to enhance the contribution of motoric codes when visual input is not informative (e.g., if you are looking away from the target).

We can gradually extend the example toward language behavior; imagine you are engaged in conversation with a friend, and you say the same word at different moments. When you produce the word, the values along a given dimension of your friend's neural response to the acoustic energy of your word utterance will necessarily differ from those incurred when your friend produces the word herself and her brain reacts to that production. This difference can also be described as a difference in gain modulation. In the case of language production and comprehension, this separation is particularly useful—we do not want a system that must involuntarily repeat what is understood, nor one where we reprocess our own speech as if it were another's to be comprehended as we talk.

Now, we can take a step further and apply this conceptual analysis to an example that derives linguistic structure from sensation during comprehension. Here, gain modulation takes the form of selective amplification and inhibition, which shape the sensory projection of the envelope and spectral properties of the signal (syllables, phonetic features) into words, phrases, and sentence structures. In the following example, gain acts to combine aspects of representation in one coordinate system and pass that information forward into the dimension of another coordinate system. For instance, the neural response to a sharp edge detected in the speech envelope propagates activity to the stored syllabic or phonemic codes that are consistent with

that edge in context; once that syllable or phoneme is active, the edge is no longer available as an edge alone. The higher level structure of the syllable or phoneme has inhibited it. The propagation of activation through coordinate systems that are interconnected to each other is aided by the inhibition of recently processed representations as they are subsumed by structure. This is the mechanism that lies at the core of the model.

What I describe here requires some suspension of disbelief as the precise nature of the computations are obscured by the unavoidable cartoonification of an example. Pseudocode is available in Table 2.

Table 2. Pseudocode for the Gain-Modulation-Based Formation of Linguistic Representations From Sensory Signals during Language Comprehension

---

High-level Pseudocode for "Analysis-by-synthesis" Language Comprehension

0. Project physical sensation of speech or sign into state space of neural trajectories

1. Apply gain to generate coordinate transform

      1.1   Pass activation through gain-field trajectories*;

      1.2   Inhibit $t − 1$ trajectory and laterally connected trajectories

2. Current manifold state impinges on unfolding bias of sensory signal ($t + 1$), mutual constraint;

Return to 0.

*computations must be based on summation and divisive normalization, and result in nonlinear additive gain modulation.

Specified Pseudocode to Generate a Phrase from Syllables and Words

For each [sensory input segment]++ at t_0

0   Project physical sensation of speech or sign into manifold of neural trajectories

      0.1   syllable envelope, spectral contents enter Dimension 0 of manifold

      0.2   Apply gain from stored linguistic representations (priors in the form of distributional and transitional probabilities) onto coordinates in Dimension 0, creating Dimension 1

      0.3   Inhibit $t − 1$ trajectory and laterally connected trajectories

      0.4   Bias upcoming sensory signal ($t + 1$) through mutual constraint of Dimension 0 onto upcoming sensory input

1   Return Dimension 1 [phonetic, phonological, prosodic coordinates]

      1.1   Pass activation through gain-field trajectories* and apply gain as in 0.2; creating Dimension 2

      1.1   Inhibit $t − 1$ trajectory and laterally connected trajectories

      1.2   Bias upcoming sensory signal ($t + 1$) through mutual constraint of Dimensions 0 and 1 onto upcoming sensory input

2   Return Dimension 2 [lexical and morphological coordinates]

      2.0 Pass activation through gain-field trajectories* and apply gain as in 0.2; creating Dimension 3

      2.1   Inhibit $t − 1$ trajectory and laterally connected trajectories

      2.2   Bias upcoming sensory signal ($t + 1$) through mutual constraint of Dimensions 0, 1, and 2 onto upcoming sensory input

3   Return Dimension 3 [lexico-syntactic and lexico-semantic relations]

      3.1   Pass activation through gain-field trajectories* and apply gain as in 0.2; creating Dimension 4

      3.2   Inhibit $t − 1$ trajectory and laterally connected trajectories

      3.3   Bias upcoming sensory signal ($t + 1$) through mutual constraint of Dimensions 0, 1, 2, and 3 onto upcoming sensory input

4   Return Dimension 4 [phrase-level syntactic and semantic relations]

      4.1   Pass activation through gain-field trajectories* and apply gain as in 0.2; creating Dimension 5

      4.2   Inhibit $t − 1$ trajectory and laterally connected trajectories

      4.3   Bias upcoming sensory signal ($t + 1$) through mutual constraint of Dimensions 0, 1, 2, 3, and 4 onto upcoming sensory input

5   Return Dimension 5 [clause- and sentence-level relations]

To comprehend the sentence from speech or sign:

Time flies like an arrow.

The first dimension in manifold trajectory space is the neural projection of the modulation spectrum and envelope of the sensory stimulus; whether this is best described as neural representations of syllables, phonemes, or minimally articulatory–phonetic features (see Anumanchipalli, Chartier, & Chang, 2019; Cheung, Hamilton, Johnson, & Chang, 2016) is an empirical question (see Figure 1). In any case, this first dimension cues the invocation of abstracted functionally phonemic representations in cortical time, which we can represent, for our cartoon purposes, in a sequence of the international phonetic alphabet:

/taɪm flaɪz laɪk ən ˈærou/;

where in an incremental manner, this process is happening iteratively as each burst of signal-related activity occurs. To pass to the next dimension of the manifold, syllabically segmented representations receive gain from internal lexical representations; this gain signal synthesizes the second dimension with activation from assemblies that store lexical knowledge, selecting a lexical representation directly from memory to become active in the manifold (see Figure 2A for a static representation and Figure 2B for a visualization of the process iterated in time). This process essentially serves to transform the activation pattern in the coordinate system of phonetics and phonology into lexical coordinates. Sequences of segmented syllables can be organized by thresholded lexical uniqueness point in the stream; this characterization will serve as our simplification of lexical access. Once the lexical dimension of neural trajectories has been reached, inhibition is passed down to the constituent codes on "lower" dimensions. We can denote an unfolding lexical representation as it emerges from syllable segmentation as
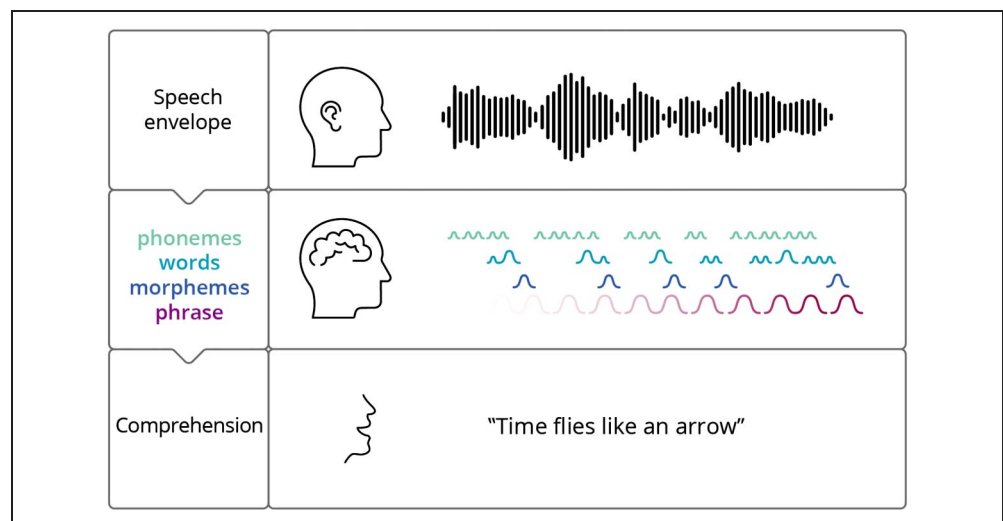
/time/ /flies/…

with each lexical and morphological dimension, in turn delimiting the abstraction process in time. Once the lexical dimension has been achieved locally (i.e., patently not assuming that a phrase or sentential structure is atemporal, they must be computed incrementally), synthesis with morphemic dimensions yields supralexical syntactic structure or the phrase dimension (Figure 2A and 2B):

/time flies/…

Inhibition is likely to play a significant role in the selection of locations in each dimension that make up the trajectory. For example, for minimal pairs of phonetic features or phonemes, activation of a given unit inhibits its paired contrast. From the lexical and morphemic dimensions upward, inhibition is needed not only to select targets, but to suppress those targets' individuation as they are synthesized into upper dimensions, such as from morphemes to words. In the second section, inhibition will play a key role in multiplexing across dimensions during phrasal computation when the system is parsing and producing sentences in time. From the instantiation of the cascaded phrases /time flies/, /flies like/, and /an arrow/, where the activation state in that dimension, expressed in each dimension's coordinates with functors across the dimensions as the neural trajectory of the sentence in the manifold progresses.

Although the path dependence applies in earlier dimensions, on a mesoscopic scale we can see how it shapes the unfolding trajectory. Path dependence is the delimitation of the current trajectory by the past, and in other words, by path choices at earlier stages of processing. In a system with path dependence, information about the relation between a given state and the state space manifold can be recovered via path integration (Gallistel, 1990). Path dependence and integration may turn out to be how perceptual inference evolves in the neural manifold but the role of these concepts in linguistic structure-building trajectories must be empirically established.

Figure 1. A cartoon illustration of the inference problem for the brain during language comprehension. From the speech envelope and spectral contents therein, the brain must generate linguistic structures and meanings that do not have a one-to-one correlate in the acoustic signal. Information on different timescales, putatively encoded in the excitatory and inhibitory cycles of neuronal assemblies, must be synthesized together into meaningful linguistic structures to achieve comprehension.
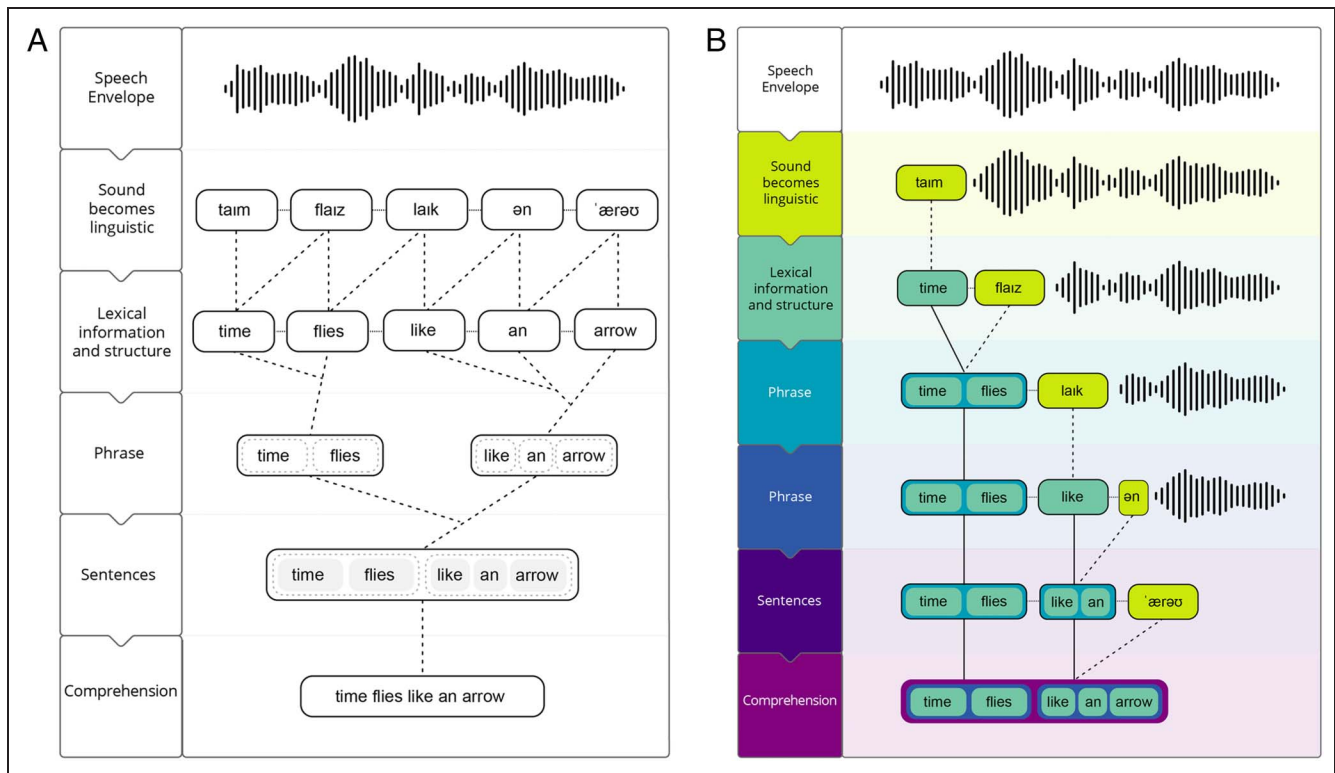
Figure 2. (A) A schematic of the broad strokes representational concepts associated with each dimension in the manifold for the sentence "Time flies like an arrow." Illustrated here are the levels of representation also referred to in the pseudocode. I do not mean to imply that other or more specific and articulated linguistic representations (e.g., phonetic and phonemic representations, constituency grammar representations, formal semantic representations, and flavors of representation that far exceed the illustrations Table 1 in specificity) are not at play in the mind and brain. I believe they are, but gloss over and simplify them for the sake of communicating the arguments I make in this paper, which are about the neuroscientific, cognitive, and linguistic computational levels, and the beginning of an algorithmic account of how levels of representations are transformed into each other. (B) A visualization of the coarse timestep increments for A.

The multidimensional coordinate system for language I sketch here is inspired by theories of neural representation in sensory processing (Ma, 2012; Jazayeri, 2008; Andersen, Essick, & Siegel, 1985), evidence for neural coding schemes in multisensory perception and in perception–action models (Bressler & Kelso, 2001, 2016; Jazayeri & Movshon, 2007; Ghazanfar & Schroeder, 2006; Andersen, Snyder, Bradley, & Xing, 1997), and in auditory and speech processing networks (Cheung et al., 2016; Chang et al., 2011; Lakatos, Chen, O'Connell, Mills, & Schroeder, 2007; Ghazanfar & Schroeder, 2006). Recent studies using electrocorticography have revealed that both speech-gestural, acoustic-phonetic, and speaker identity-related aspects of the speech signal are coded to support multisensory integration (Cheung et al., 2016; Ghazanfar & Schroeder, 2006). Such a coding scheme could be operated on by a gain-modulated coordinate transform that transduces an acoustic signal to at least the level of syllable and the onsets of larger linguistic structures. Namely, populations that selectively respond to acoustic features also appear to infer phonemes based on those features, even when the phoneme itself is not present in the input (Fox, Sjerps, & Chang, 2017; Leonard, Baud, Sjerps, & Chang, 2016; Chang et al.,

2011). From evidence like this, we can assert that cortical networks encode information in a multidimensional coordinate system such that, contingent upon the route of activation (i.e., behavior), activation weights in one dimension have more gain relative to another dimension but remain coregistered through homomorphism with one another.[6] Through neural gain modulation, representations that are more relevant (i.e., have higher likelihoods) in a given context can dominate and guide behavior, giving the system the flexibility needed to dynamically amplify aspects of representations in relation to the sensory context and behavioral goal (see Engel & Steinmetz, 2019, for a review). In such a coding scheme for language, knowledge of the lexicon, and grammatical, semantic, and contextual knowledge can be "shared" across modalities and recruited during the assembly of representations for articulation, as well as during the perceptual inference and generation of structures during comprehension.

Furthermore, gain modulation also offers a built-in system for predictive coding (Friston, 2005); activation can be passed to assemblies that represent likely upcoming representations through multiplication of present response functions, divisive normalization of less

expected or less relevant dimensions (Carandini & Heeger, 2012), and inhibition of recently perceived dimensions. Predictive coding would be a form of neural gain application to future representations or representational dimensions as a function of the present stimulus. In summary, neural systems can achieve a form of representational efficiency by coding perceptual targets as intersections in a multidimensional space.

Principles from neurophysiological models of sensory coding must produce patterns of activation that abide by the requirements of language: The representations called upon during production and comprehension are coordinate transforms across dimensions the neural trajectory manifold of cell assemblies. One set of coordinates is based on the sensory information of a given linguistic processing unit and the concomitant motor program to produce that unit within a context specified by the highest unit being planned. Another set of coordinates relates the morphism[7] of the sensory space with the abstract structural and conceptual knowledge it is related to (likely an algebraic, not an exclusively geometric space, see Phillips, 2020). In such a space, the unit being produced or comprehended is represented in relation to the other representations in memory that they cue (as in Martin, 2016, a function of grammatical knowledge modulated by referential and other aspects of the perceptual context). Production or comprehension then becomes a behavioral target; coordinate systems that play a role in producing (i.e., sensorimotor, motor) are more active during that behavior than during the opposed behavior (i.e., comprehension), but the mapping between systems persists. As in models of attention and perception, gain modulation allows for behavior to be guided by one coordinate dimension over another as a function of task demands (Jazayeri & Movshon, 2007; Carrasco, Ling, & Read, 2004).

Necessary Computational Principles for Higher-Level Linguistic Structures

To represent linguistic structures in the system described above, it is likely that individual neurons and even neural networks or assemblies must participate in the coding of multiple dimensions in the manifold—coordinate transformation via gain modulation is the brain's way of reading out or translating information represented in one assembly in the context of, or combined with, information represented by another (Buzsáki, 2010, 2019). To pull this off in a unit-limited system, individual units will have to play double duty. Fortunately, there is ample evidence that neurons can participate in multiple larger networks, even "at the same time" by firing at different frequencies as part of different networks (Bucher, Taylor, & Marder, 2006; Weimann & Marder, 1994; Hooper & Moulins, 1989). There are likely many cellular mechanisms that underlie overlapping neural circuits and the "switching" of neurons on and off within an assembly; it appears that the system can employ a number of these mechanisms concurrently to achieve rhythmic homeostasis (Marder,

2012). In the second section, I return to how gain modulating through coordinate dimensions in the manifold links up with a mechanism for learning and representing structures in a computational model (itself a theory of representation) and with principles from neurophysiology and linguistic theory. Essentially, to make the architecture proposed in the first section sufficient to support compositionality, the representations at each level of linguistic description must be functionally orthogonalized. From a computational point of view, there are two ways to achieve this, one is to hardcode vector orthogonalization, which in my opinion yields data structures that are not flexible enough to account for the productivity and generalization seen in natural language, and the other more plausible way is to use a time-based neural processing mechanism (viz., an algorithm or series of algorithms that control the neural transform path over time) in a way that maintains independence between representational layers as per formal linguistic needs.

II. A MECHANISM FOR COMPOSITION: BRAIN RHYTHMS STRUCTURE SPEECH INPUT INTO LANGUAGE VIA GAIN-MODULATED MULTIPLEXING

Neurobiological models have focused on identifying the functional and anatomical circuits that underlie speech and language processing in the brain (Friederici & Singer, 2015; Skipper, 2015; Hagoort, 2013; Hickok & Poeppel, 2007). However, within the last decade, a wealth of results has emerged that points toward a process model based on neural oscillations and their role in speech and language processing has emerged (e.g., Meyer, Sun, & Martin, 2019; Obleser & Kayser, 2019; Meyer, 2018; Murphy, 2018; Ding, Melloni, Zhang, Tian, & Poeppel, 2016; Keitel & Gross, 2016; Friederici & Singer, 2015; Gross et al., 2013; Hagoort, 2013; Arnal & Giraud, 2012; Giraud & Poeppel, 2012; Peelle & Davis, 2012; Ghitza, 2011; Obleser, Meyer, & Friederici, 2011; Morillon, Kell, & Giraud, 2009; Bastiaansen, Oostenveld, Jensen, & Hagoort, 2008; Luo & Poeppel, 2007; Bastiaansen & Hagoort, 2006; Hald, Bastiaansen, & Hagoort, 2006; Bastiaansen, van der Linden, Ter Keurs, Dijkstra, & Hagoort, 2005). These accounts have taken the first steps in attempting to link real-time signals for cortical networks during speech and language processing to the neural mechanisms that render language comprehension from the acoustic signal of speech (e.g., Ding et al., 2016; Giraud & Poeppel, 2012). Inquiry into the classes of neural architectures and computations that the brain could carry out to achieve perception of linguistic structure from sequential sensory input is ongoing (see Martin & Doumas, 2017, 2019a, 2019b; Meyer et al., 2019; Martin, 2016); here, I offer an account of (de)compositionality in a computational framework that uses oscillatory activation to combine and separate information in a system bounded by cycles of activation and inhibition.

Phase Synchronization and Temporal Multiplexing of Information as Structure Building

A prominent feature of neural oscillations is the potential correspondence with multiple timescales of information processing, expressed either in aspects of time (latency, onset, duration), in the periodicity of processing, in power, or in phase information. From animal models of basic neurophysiological mechanisms, temporal multiplexing, often empirically operationalized at cross-frequency coupling or phase (phase–phase, phase–amplitude) coherence, is implicated as a stalwart processing mechanism, carrying information that either occurs on different timescales or is relevant on different timescales for perception, action, and behavior (Fries, 2009; Schroeder & Lakatos, 2009; Lakatos et al., 2007). Evidence suggests that synchronization between cell assemblies as reflected in neural oscillations and phase coherence generalizes widely to other areas of perception and memory in humans (Hanslmayr & Staudigl, 2014; Fries, 2009; van Rullen & Koch, 2003) as well as to speech processing (Obleser & Kayser, 2019; Assaneo & Poeppel, 2018; Rimmele, Morillon, Poeppel, & Arnal, 2018; Keitel & Gross, 2016; Giraud & Poeppel, 2012). Questions that the emerging field of neural oscillations during speech and language processing grapples with include (a) whether neural oscillations are indeed the computations at work, or just "read-out" of those computations, (b) whether endogenous neural oscillations encode abstract stimuli beyond the Fourier transform (Cole & Voytek, 2017), and (c) if there is a functional interpretation for a given frequency band, and if so, what it is and is it a type or a token.[8] Regardless of the answers to these difficult questions, one thing is clear: Brains make use of information that occurs on different timescales in the environment and within the individual (cp. Buzsáki, 2019). I will take for granted, then, a link between the syllable envelope or speech rhythm and the theta oscillation (~4–7 Hz), and between the fine acoustic featural structure of speech and the gamma oscillation (~30–90 Hz), and assume that this link reflects the perceptual mechanism that renders speech into language (cp. Giraud & Poeppel, 2012). A strong version of such a hypothesis is that slower rhythms (i.e., delta and theta oscillations) give structure that is regularly phase reset by informationally dense (relatively infrequent) linguistic units, such as stressed syllables demarcating lexical and phrase codas (Alday & Martin, 2017; Halgren et al., 2018; Ghitza, 2013), and higher frequency bursts of activity reflect the application of grammatical rules or stored lexical knowledge to infer a larger structure coded by a new assembly that has come online. In this characterization, gamma activity is associated with the retrieval of memory-based linguistic representations by minimal or thresholded acoustic cues (Meyer et al., 2019; Martin, 2016), which may require increased interregional communication. Gamma has been associated with interregional coherence in cognition (Buzsáki & Schomburg, 2015; Lisman & Jensen, 2013), and seems to be tied to perisomatic inhibition (Buzsáki & Wang, 2012). Gamma magnitude is also modulated by slower rhythms and occurs with the irregular firing of single neurons, and is implicated in the transient organization of cell assemblies (Buzsáki & Wang, 2012). These characteristics align with the inference of higher level linguistic representations from sensory input being a punctate perceptual event that has ongoing consequences for whole brain dynamics. Once higher-level linguistic structure has been inferred, further coordination of assemblies must occur via inhibition, passing inhibition to recently processed constituent representations and to related competitor representations. This process would result in gamma modulations, which in turn shapes the processing of upcoming sensory input in the context of recent events/activated representations. If higher-level structures have ongoing consequences for future processing, as in, they shape upcoming sensory processing through biases and prediction, then gamma modulations should be observable as a function of the generation of higher-level linguistic structure and the degree to which upcoming input is constrained by it (Nelson et al., 2017). In a model that generates linguistic structure internally, knowledge about what goes with what or what is likely to come next is encoded in the structures themselves. The system has access to predictive information by virtue of the way that it represents structures and infers them from incomplete sensory input. The predictive aspect of the model's architecture would crucially rely on it not being feed forward, on passing inhibition laterally and downward, and on the ability to learn from internal dynamics. An instantiation of this latter ability can be seen in a settling network architecture that uses the systematic perturbations of internal states to learn representations (Martin & Doumas, 2019a; Doumas & Martin, 2018; Doumas et al., 2008).

On this view, ongoing slow rhythms are coupled with high-frequency activity that reflects inference, the activation of abstract grammatical knowledge in memory (likely both procedural and semantic memory; Buzsáki, 2019; Martin, 2016; Ballard, 2015).[9] "Entrainment" to higher level structure is actually driven by internal evoked responses to sensory input—the cascade of perceptual inference via gain modulation and inhibition (Meyer et al., 2019; Martin, 2016) resulting in path dependence as discussed in the first section. Linguistic structures, merely by virtue of their neural coding structure, then can constrain sensory processing forward in time in what could be described as predictive coding (Haegens & Golumbic, 2018; Spitzer & Haegens, 2017; Arnal & Giraud, 2012; Morillon et al., 2009; Friston, 2005). But how linguistics units combine together via gain modulation over time is hypothetical and must be tested; in Table 2, I offer pseudocode that presents a hypothesis about how such an algorithm might work.

The articulatory gestures that produce speech have a necessarily sequential nature, as our articulators cannot work in parallel and produce more than one gesture at a time. However, coarticulation and other phenomena allow information about both what is upcoming and what

has recently occurred to be spread across the signal. Similarly, in comprehension, acoustic and temporal envelope information enter the system and is segmented into discrete units across multiple timescales for further processing (Ghitza, 2011, 2013; Giraud & Poeppel, 2012; van Rullen & Koch, 2003). In speaking and listening, the representation and processing of information must differ across multiple timescales (roughly: articulatory unit, morpheme, word, phrase). In production a composed message must be sequenced into articulatory gestures, and in comprehension, the acoustic and rhythmic output of those gestures must be composed into a hierarchical structure from a sequence.[10] This gives rise to the need to branch or spread information across linguistic levels of analysis across time—syllables, words, phrases, and sentences tend to occur on disparate timescales, but often, timescale and linguistic content cannot be fully orthogonalized—a syllable can be a morpheme, a word, a phrase, or even denote a sentence. To solve the problem of interpretation and production of structured meaning through sequential channels of speech, sign, or text, the brain needs a mechanism that can spread information about representational content across time. The theta oscillation may be a likely carrier signal for linguistic sensory input, but more carrier signals and coherence between them must exist for the perceptual inference of linguistic structure, which itself is not recoverable from the sensory codes alone.

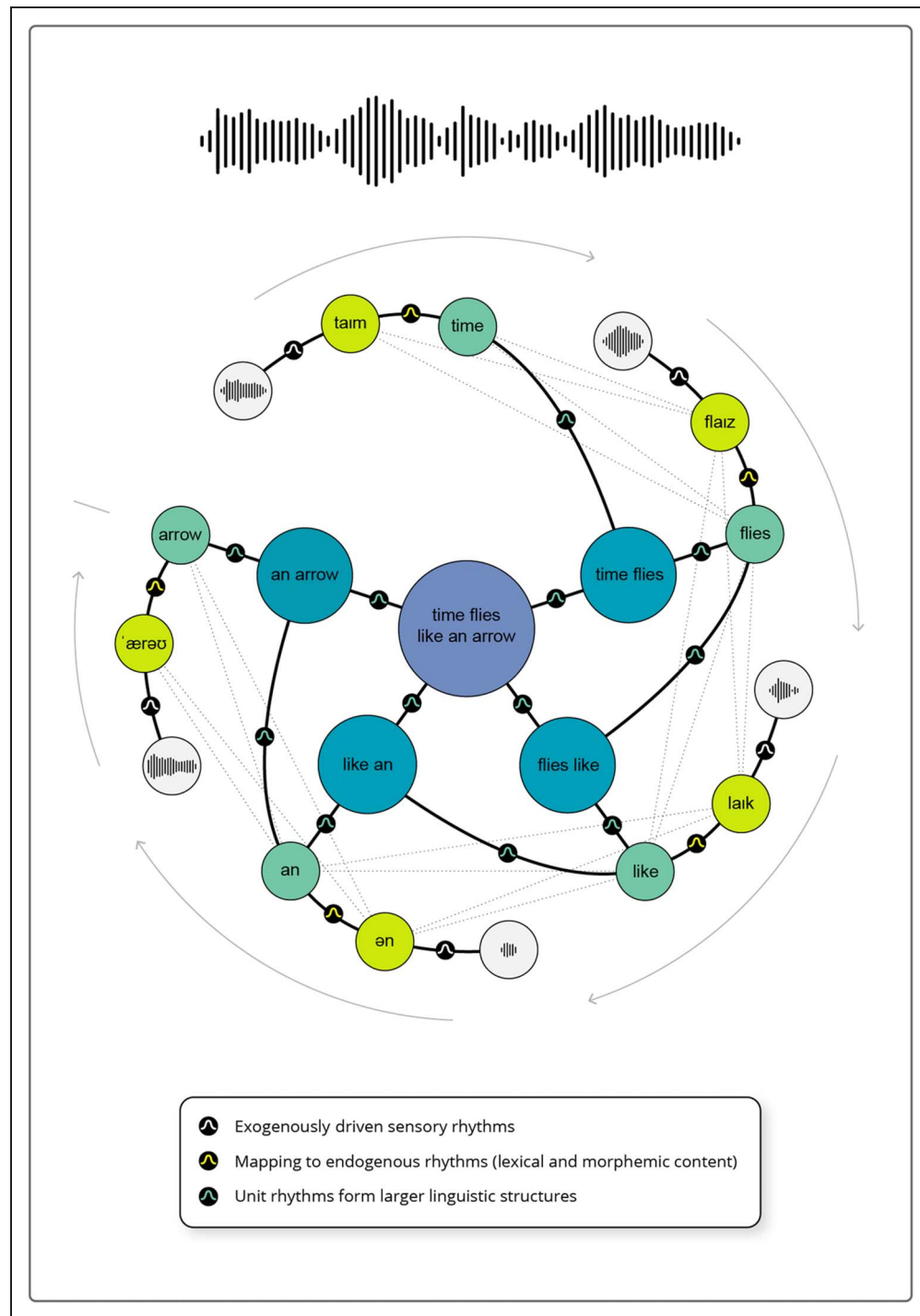## How Time and Rhythm Could Generate Compositional Linguistic Structures

The problem of (de)composing representations in language processing can be conceptually analyzed as, at minimum, two states of the network must be linked together for processing by a third, separable representational state (Doumas & Hummel, 2012). The instantiation of a third state is what allows stored representations to not be defined by this particular instance of composition (see Figure 3 for an illustration of a trajectory in the manifold for a sentence). Such a mechanism allows the system to maintain independent representations of inputs that are composed together during processing as needed during multiplexing and, in principle, to produce a theoretically limitless set of combinations of states. As such, in sequences, implicit ordinal and time-sensitive relationships matter and carry information and, in fact, can be used to signal the hierarchical relationships that have been compressed into that sequence and which can be reconstructed from that sequence (Martin & Doumas, 2017, 2019a, 2019b; Doumas & Martin, 2018; Doumas et al.,2008, 2017). Information represented in the lower layers of the cortical network is directly read in from the neural projections in sensory dimension of the manifold of the sequential input. In such an architecture, higher level representations are dimensions in the manifold that integrate or bind lower level representations over time, which

gives rise to more protracted activity. Hierarchical structures thus mandate an asynchrony of activation in time between layers of the network or across dimensions of the manifold, which correspond to levels of linguistic representation and the products of composing them into meaningful structures. This asynchrony/desynchronization can only be achieved with a modulator, in this case, inhibition carried out by yoked inhibitors.

One way to implement multiplexing computationally is to distribute levels of representation across layers of a neural network (Martin & Doumas, 2017, 2019a; Doumas et al., 2008). Representations that must be sequenced from a hierarchy, and vice versa, can be composed and decomposed only if activation across layers is desynchronized at some point in time. This fact gives rise to "rhythmic computation" in the network, where time is used to carry information about the relationships between representations that are present in the input (see Martin & Doumas, 2017, 2019a; Doumas & Martin, 2018; Doumas et al., 2008, 2017). The mechanism of rhythmic computation is based on a principle of cortical organization that "neurons that fire together, wire together" (Hebb, 1949). Neurons that do not fire in synchrony can stay independent—the proximity in time between firings can be exploited to carry information about the relation between recognized inputs in the sequence. Though all neural networks can be said to contain an implicit notion of time (i.e., in that they have activation functions and learn as a function of iteration and weight updating), few models explicitly use time to carry information. Those that do tend to use the synchrony of firing to bind information (Singer, 1999; von der Malsburg, 1999) and do not use the information carried by asynchrony (Shastri, 1999; von der Malsburg, 1995).

In contrast, Discovery of Relations by Analogy (DORA; a symbolic-connectionist model of relational reasoning; the full computational specifics can be found in Doumas et al., 2008, 2017) exploits the synchrony principle, but it does so by using "both sides" of the distinction (see Figure 2 for a cartoon illustration). To keep representations separable while binding them together for processing, the model is sensitive to asynchrony of firing. The use of time on multiple scales to carry information about the relations between inputs in a sequence is implemented as systematic asynchrony of unit firing across layers of the network. This manner of computation is rhythmic because synchrony and asynchrony of activation is what computes the bindings. Rhythmic activation of separable populations of units computes different levels of representation as they occur in time, which both binds the representations together while keeping them as separable neural codes. The ability to maintain (de)compositionality is a computational feature that is crucial for the kinds of relations that are necessary to represent human language. DORA achieves compositionality by representing information across layers of the network and using a combination of distributed codes (e.g., for features, objects, words, concepts) and localist codes (e.g., for role-filler bindings). The particular

Figure 3. A cartoon of the neural trajectory for the sentence "Time flies like an arrow" as it progresses through the manifold. Time progresses in a clockwise manner. Mutual constraint between dimensions is represented by the dotted lines; the solid lines are the expression of path dependence into larger linguistic structures. The small Gaussian symbols represents the application of gain and inhibition as coordinate transform occurs. Temporal multiplexing is represented in the cascaded and twisting nature of the solid line arms, such that there is also desynchronization between levels of linguistic representation. Dimension or level of linguistic representation is represented by the different color circles.

implementation of the conjunctive localist nodes in DORA make the tacit assumption that words and phrases are composed with one another via vector addition and not a multiplicative operator (e.g., a tensor product). It is not known whether vector addition is a sufficient operator for compositionality in natural language, but addition has clear advantages over tensors for formal reasons relating to variable value independence (Martin & Doumas, 2019b; Doumas & Hummel, 2005; Hummel, 2011; Hummel & Holyoak, 1997, 2003; Holyoak & Hummel, 2000).

In the conceptual terminology of neural oscillations, temporal asynchrony corresponds to neural desynchronization. In our case, this would be expressed as desynchronization between dimensions in the manifold. These desynchronizations tune the path of the evolving trajectory of the linguistic structure in question and create, over time, phase sets that denote or group units that are interpreted together. Importantly, synchrony and asynchrony of unit firing in time are not orthogonal mechanisms; they are the same function or variable with different input values (e.g., $\sin(x)$ and $\sin(2x)$) that can carry

different information. Binding or forming representations through synchrony alone would effectively superimpose a variable and its value onto a single, undecomposable representation (Singer, 1999, von der Malsburg, 1999). Martin and Doumas (2017) showed that DORA and, in principle, any model that represents and processes information in a similar way, better predicts cortical rhythms to spoken sentence comprehension (Ding et al., 2016) than models that do not represent structures or exploit time explicitly (e.g., traditional recurrent neural networks). Energy expenditure in cortical and artificial networks was consistent with formal linguistic descriptions of the structure of language and offers evidence that the human brain represents and processes information in a way that is more similar to a hierarchical compositional system than to an unstructured one. As in the DORA architecture, inhibition plays a key role in how information that has recently been processed is suppressed or controlled and how information is combined and separated.

Applying Rhythmic Computation to Producing and Comprehending a Phrase

In the DORA instantiation, a phrase can be formed from vector representations of the input words via a conjunctive code on a different layer of the network. This conjunctive code represents the phrase; the individual input words have distributed representations in DORA. Under such a coding scheme, a phrase is separable from word-level representations whose distributed representations are functionally independent from the conjunctive code of the phrase. In comprehension, the activation of the phrase can only occur after the onset of the first word and persists throughout the duration of the second word. In production, the activation of the conceptual proposition and thus phrasal relations precedes the activation of individual words. This difference in the time course of activation makes the prediction that compositional representations should be detectable earlier in production than in comprehension. It is also consistent with the idea that, during comprehension, representations serve as cues to each other in a form of perceptual inference, and during production, the path from meaning to its ultimate expression can be incrementally and dynamically composed as long as local domains like words and phrase are internally coherent. An incremental, cascaded, treelet-like grammar could capture these processing dynamics (Kempen, 2014; Pauls & Klein, 2012; Hagoort, 2003; Marcus, 2001; Vosse & Kempen, 2000). Temporal multiplexing is expressed in this instantiation by firing the distributed codes for the words in the phase set of the phrase, essentially the phrase node stays active for the duration of all of the words that make up that phrase, but the yoked inhibitor of the word-level representation turns off the activity of the individual words after their sensory time has elapsed. As a result, the model oscillates, with pulses of activation related to activating distributed word codes and a slower pulse of activation that codes the phrase (see Martin & Doumas, 2017). Inhibition and yoked

integrative inhibitors are used to turn the word units off as they pass activation to phrase nodes (for a detailed description of the DORA model, including pseudocode, see Martin & Doumas, 2017, 2019a; Doumas & Martin, 2018; Doumas et al., 2008, 2017).

In the terminology from the first section, forming a phrase from words (see phrase-specified pseudocode examples in Table 2) draws on an iterative process whereby a path is formed through the neural trajectory manifold; each dimension is claimed to correspond with levels of linguistic representation. Gain modulation controls the progression of transforms through the path, and each time step brings sensory representations toward latent structure in comprehension, whereas in production, each time step moves progressively toward articulatory gestures. Temporal multiplexing works here to combine dimensions—projection of activation through gain fields and concomitant inhibitory signals on the "path not chosen" or "path not taken" shape by the coding of upcoming sensory input in comprehension and upcoming articulatory gestural movements in production. As in the DORA implementation, the concept of phase set is useful in conceptualizing how trajectories in the manifold form word and phrase patterns. Desynchronization, fueled by inhibition, is what allows the phase sets to form in the manifold.

Predictions

There are a measure of coarse-grained predictions that arise from the claims I make here. I have summarized four general predictions for oscillatory activity related in language processing in Table 3. These are expected patterns

Table 3. Predictons

1. If linguistic structure is represented as claimed in the model, then low-frequency power and phase synchronization should increase as structure accrues.

2. Lower level linguistic representations (i.e., those closer to sensory representations during comprehension) should be treated differently by the brain as a function of the unfolding trajectory context.

3. Linguistic content and the encoding of the timescale of its occurrence should be separable in the brain, if not orthogonalizable.

4. If coordinate systems exist for levels of linguistic representation and there is path dependence between levels, then perturbations or experimental manipulations on a lower level should have bounded effects on the next level's representational coding.

5. The relationship between the neural signals that index the coordinate systems for linguistic representation should be better fit by models that use a modified gain function than ones that use another method for combining sources of neural information.

in neural oscillations if the core claims of the architecture are attested. Here, I outline a second set more closely related to psycholinguistics, which concerns how behavior (production vs. comprehension) should modulate gain-controlled neural responses.

The chief prediction regarding structure and meaning from the architecture is that low-frequency power and phase synchronization should increase as structure and meaning build up in time. This has been attested in the literature (Brennan & Martin, 2020; Kaufeld, Ravenschlag, et al., 2020; Kaufeld, Naumann, et al., 2019; Meyer, 2018; Ding et al., 2016; Meyer, Henry, Gaston, Schmuck, & Friederici, 2016; Bastiaansen et al., 2005, 2008; Bastiaansen & Hagoort, 2006) but needs more careful investigation. It is likely that low-frequency phase organization reflects the increasingly distributed nature of the neural assemblies being (de)synchronized as structure and meaning are inferred, rather than reflecting a phrasal or sentential oscillator. If perceptual inference is a product of neural trajectory, then lower level linguistic representations should be treated differently by the brain as a function of the context they occur in. This model also predicts that there should be more phase synchronization between assemblies involved in coordinate transform between dimensions than between assemblies that are not participating in coordinate transform.

In terms of behavioral tuning, the first prediction is that different dimensions should compete or interfere as a function of behavior—when preparing to speak, semantic competitors, both at the combinatorial and word level, should be more detrimental to processing than perceptually overlapping stimuli—which should interfere only later. For example, when preparing to say "coffee," "tea" should be more problematic to process than "coffin."

Table 4. A Summary of the Theses and Axioms Put Forth in this Paper

Computational-level Thesis

1. Linguistic representations in the brain are the product of cue-based perceptual inference, an internally driven generative model for externalizing formatted thought.

2. The perceptual inference of linguistic representations is a series of transformations of sensory input into other coordinate systems; although elicited by sensation, linguistic representations are an internally driven cascade of transformations that become distinct in neural spacetime from sensation.

3. Grammatical knowledge of different levels of granularity is encoded in the possible trajectories of the manifold.

4. Morphisms exist between coordinate transforms, as morphisms between categories are described in mathematical linguistics (composition preserves the morphism from syntactic algebra to the semantic one; Partee et al., 2012).

5. There is a mapping between each dimension and coordinate system in the manifold. A functor describes this mapping. Most dimensions are not isomorphic and thus not injective in relation to each other (viz., there is no 1:1 mapping between dimensions, nor their coordinates).

6. Grammatical "rules" can be generalized to new inputs and outputs via low dimensional projections (core) into higher dimensional spaces (periphery). *candidate algorithms: mapping, relational generalization from (see Doumas et al., 2008; Hummel & Holyoak, 1997, 2003).

Algorithmic-level Thesis

1. Perceptual inference of linguistic structure is a coordinate transform that is achieved through gain modulation, of which inhibition is an important form that is used to separate dimensions and build structure through desynchronization of relevant population activity.

2. Perceptual inference for language is achieved through the synthesis of sensory information with stored knowledge. Priors about the relationship between sensory objects and abstract structures exist on short and long timescales and bias the inference process.

3. "Temporal multiplexing" describes the propagation of activation through the manifold of trajectories. It refers to information on one timescale cueing information at other timescales. It depends on inhibition and performs inference in an iterative fashion such that mutual constraint is achieved between stimulus and increasingly abstracted internal states.

Implementational Axioms

1. Low-frequency oscillations (viz., delta) are more likely indicative of the increasingly distributed nature of cell assemblies than any timescale-related activation of linguistic structure.

2. Trajectories are specified by priors about both specific sensory objects and abstract structures; these trajectories can be interpolated and extrapolated to support novel composition and productivity.

The reverse should be true in comprehension. Similarly, when processing adjective–noun phrases like "green tea," "tree" should be more intrusive during comprehension than during production. Such predictions also imply that inhibitory control will be needed for lemma selection in production, but for segmentation in comprehension. The main prediction that is unique to this theory of language and that is derived necessarily from the symbolic connectionist systems that it is inspired by (Doumas et al., 2008; Hummel & Holyoak, 1997, 2003) is that activation in the system, as it corresponds to levels of linguistic representation, is additive. Because the model relies on vector addition and asynchrony of firing through yoked inhibitor nodes to dynamically bind variables and values, it predicts that activation patterns of words becoming a phrase should be additive, not interactive or multiplicative. This claim is similar to Sternberg's additive factors logic (Sternberg, 1969) and distinguishes the model from others, especially those based on tensor products. By comparing representations along dimensions and exploiting their intersections to find latent structure and other orders of representation (Doumas et al., 2017, 2008), we may be able to explain how generative unbounded combinatoriality can exist in the human mind and brain.

Table 4 summarizes the claims I have made in this paper. There is accumulating evidence that is consistent with the theses put forth here, from evidence for cross-linguistic-level cue integration (listed under "Linguistic Representation in Neural Terms") to the modulation of oscillatory signatures during sentence processing (listed under "Cortical Rhythms Structure Speech Input into Language through Gain-Modulated Multiplexing"). A next step or, more likely, a longer term goal is to see if this model can offer a satisfying explanation for a wider range of behavioral effects in psycholinguistics.

Conclusion

In this paper, I have argued that the core properties of human language—the formation of compositional, hierarchical structures whether spoken, signed, or heard—can be accounted for, in principle, by a theory of the space-time trajectories of neural assemblies controlled by gain modulation. In the first section, I described how the representations that underlie language processing could be expressed as dimensions in a neural trajectory manifold, where a particular trajectory is a function of grammatical knowledge impinging upon sensation in a path-dependent way and is determined by behavior (viz., production or comprehension). The multiplexing mechanism presented in the second section operates over the spatiotemporal patterns in the manifold and cascades the inference of latent structures via gain modulation of sensory input into the coordinates of abstract representation. Inhibition of lower level structures by higher level ones gives rise to oscillatory patterns of activation during language processing and is what allows the system to preserve independence

between lower level input units and the higher level structures they form. The mechanism described in the pseudocode in Table 2 is synthesized from a computational model of relational cognition (DORA; Doumas et al., 2008) and basic principles of neurophysiology; it uses oscillatory activation to combine and separate information in a neural network and is able to predict human cortical rhythms to the same stimuli. Through this synthesis, I have tried to turn the core computational properties of human language, which have traditionally made language difficult to account for within existing neurobiological and cognitive theories into the lynch pins by which language's physical expression, that is, its extension across multiple timescales, becomes the currency of neural computation.

GLOSSARY

argument – input to a function or variable.

cell assembly – network of neurons whose excitatory connections have been strengthened in time and this strength is the basis of their unit (Buzsáki, 2006, 2019; Hebb, 1949).

compositionality – the property of a system whereby the meaning of complex expression is determined by its structure and the meanings of its constituents (Partee, 1984).

coordinates – values that neural representations or population codes can be expressed in that are derived from the mode of processing or computation that a given neuron or cell assembly is participating in. For example, coordinates range from being topographic in nature, derived from external visual space, or become sensory in nature as in retinal- or head-centered, toward latent coordinate systems that describe the systems of abstract structure that are generated to guide behavior. As sensory coordinates are gain-modulated by representations of stored linguistic knowledge, the neural coordinates describing linguistic representation necessarily become abstractions in a high-dimensional space; however, it is likely that the neural coordinates of a given dimension of the manifold for language processing correspond to units of linguistic analysis (e.g., phonetic features, lexical semantic features, possible syntactic relations in a grammar).

coordinate transform – modifying a set of coordinates by performing an operation on the coordinate axes; changes the reference frame of a representation from and between afferent/efferent spaces in sensation and action and moves toward latent coordinate systems to guide complex behavior.

domain – the set of possible values of the independent variable or variables of a function; in linguistics, the influence spheres of elements in a structure (Kracht, 1992).

function – a relation or expression over one or more variables; a relation that takes an element of a set and associates it with another set.

functor – a map or function between categories (Phillips & Wilson, 2010); encodes an invariant link between categories (Bradley, 2018).

gain modulation – nonlinear way in which neurons combine information from two or more sources (Salinas & Sejnowski, 2001).

inhibition – interruption, blockade, or restraint of neural activity in both space and time (Jonas & Buzsáki, 2007).

latent variable or structure – a variable or structure that "lies hidden" and is not directly perceived but rather inferred from other observed variables.

manifold – a collection of points forming a set; a topological space that resembles a Euclidean space at each point (Lee, 2010).

morphism – structure-preserving map from one object to another of the same type; relations between algebras may be described by functions mapping one algebra in another, a "morphism" is a mapping conceived of as a dynamic transformation process (Partee, ter Meulen, & Wall, 2012).

neural oscillations – brainwaves, brain rhythms, repetitive patterns of activity in neural space and time caused by excitatory and inhibitory cycles in cell assemblies (Buzsáki, 2006, 2019).

neural trajectory – activity of a neural population overtime; plotted in a space where each dimension is the activity of a unit or a sub-population. Dimensions can be summaries of a given assembly's activation in time when participating in larger assembly computation.

path dependence – when the set of possible trajectories is delimited by past trajectories and choices about them.

path integration – the estimation of the path to the starting point from the current position in the state space (Gallistel, 1990).

perceptual inference – the ability to infer sensory stimuli from information stored in internal neural representations acquired through experience (Aggelopoulos, 2015).

phase synchronization – state or process where two or more cyclic signals oscillate in such that their phase angles stand in a systematic relation to one another (Pikovsky & Rosenblum, 2007).

phonology – the sound system of a language (Larson, 2009).

predicate – expression of one or more variables defined in a domain; quantifying a variable; something that is affirmed or denied about an object or proposition.

scope – the domain over which an operator affects interpretation of other phrases.

semantics – the meanings of a language's words and how those meanings combine in phrases and sentences (Larson, 2009).

syntactic structure – basic structural elements of a language and their possible combinations in phrases and sentences (Larson, 2009).

Notes

1. This paradox is particularly striking from a neuroscientific perspective; in a particularly charming turn of phrase, the brain has been described as a log transform of its environment (see Buzsáki, 2019, Chap. 12). Perhaps more generously, it could be described as a log transform of the perception–action demands of the environment and the latent states those entail. I add to this description that some of those latent states seem to be learned via statistics, but nonetheless come to represent symbolic structures and rules that operate over them.

2. I use these concepts and jargon in an attempt to synthesize knowledge from putatively disparate disciplines in the hope of showing how ideas in one discipline fit or line up with notions from another. I include a glossary of working definitions for all the terms I will use (see Glossary).

3. Helmholtz's actual term is "psychic energy."

4. I do not offer a satisfying account of learning here, but I will note that I see a promising account entailed in the Discovery of Relations by Analogy (DORA) model of Doumas et al. (2008). In DORA, learning of structured representations from experience occurs because of a few key principles. First, DORA is not a feed-forward architecture but rather a settling network; it compares internal states and gleans information from the settling rates to equilibrium after perturbation by a stimulus. To achieve comparison of internal states, inhibition is passed within a neural processing bank, but not between banks such that two spatiotemporal patterns can be compared. This architecture allows the comparison (and also orthogonalization) of sensory representations for two stimuli; the distributed and relational features in common can be symbolized into a structure that is latent in both stimuli but now able to be activated orthogonally from the stimuli (although it will pass activation to related stimuli when the structure is active if the model is interacting with long-term memory). The orthogonalizing features can also be learned from. Comparison, combined with time-based binding, as well the Mapping and Relational Generalization algorithms from Hummel and Holyoak (1997, 2003) in my view represent important insights about boundary conditions on learning mechanisms for structured representations in neural systems.

5. I note that it is not a core claim of my approach that the system is Euclidean or non-Euclidean in nature; the most common descriptions in neuroscience tend to be Euclidean at the moment while the dynamics are assumed to be nonlinear.

6. Gain fields speak to a classic representational conundrum for speech processing: the degree of involvement of articulatory motor programs in speech perception (Assaneo & Poeppel, 2018; Cheung et al., 2016; Skipper, 2015; Hickok, 2012; Hickok & Poeppel, 2007; Skipper, Nusbaum, & Small, 2005). Upon perception of a given unit, the articulatory motor program may receive activation because it is highly related to the sensory unit, but as the network or assembly is not in production mode, but rather has the behavioral goal of detecting a linguistic signal, gain modulation amplifies nonmotor aspects of representation related to the perceived unit in motor cortices, which then dominate processing. Such a scheme may account for why motor areas have been observed to be activated during speech perception

(Cheung et al., 2016; Skipper et al., 2005) even when clinical data suggest that motor representations are not required for speech perception to occur (Hickok, 2012).

7. Internally generated representations should not be injective (see Partee, ter Meulen, & Wall, 2012) with the stimulus properties—that is, onsets of internal representations or the rhythms that represent higher level linguistic structures should not be evoked by stimulus rhythms in a one-to-one way, in other words, they do not have to stand in a one-to-one relationship to spectral and envelope response. In fact, to be divorced from stimulus properties and thus generalizable, rhythms reflecting internal generation of structure must not be injective with sensory rhythms to avoid the superposition catastrophe.

8. This is an important question that needs to be explored carefully and is beyond the scope of the current thesis. I think there are reasons to see frequency bands as tokens of processes with physiological bounds that render them into functional types. Without the space to reason this conjecture out based on existing literature via conceptual analysis, I can only say that I do not think they are strict types with fixed functional interpretations that map in an injective way onto cognition.

9. Evidence that speech production might also be structured by time and rhythm comes from magnetoencephalographic studies of overt and covert speech production. Tian and Poeppel (2013, 2014) showed that the production of syllables where the lag between production and self-comprehension was artificially delayed by more than 100 msec were judged as being produced by someone else and that auditory cortex responded to these syllables as if they were no longer self-produced. Although these findings suggest that timing and rhythm might structure production and contribute to suppressing neural responses to one's own speech, the functional role of cortical entrainment in naturalistic language production is largely unknown (but see Giraud et al., 2007).

10. For example, if your friend says a phrase or a sentence, when she produces the corresponding bursts of energy, the intended composed meaning will be activated earlier in her neural system than yours, and consequently, the values along a given dimension of your friend's neural response during preparation of the utterance will differ from those incurred when you perceive the phrase. Conversely, the compositional structure will be available later in time in your cortical networks because it must be inferred from sensory input as latent structures.

REFERENCES

Aggelopoulos, N. C. (2015). Perceptual inference. *Neuroscience & Biobehavioral Reviews*, 55, 375–392.

Alday, P. M., & Martin, A. E. (2017). Decoding linguistic structure building in the time-frequency domain. In *The 24th Annual Meeting of the Cognitive Neuroscience Society (CNS 2017)*.

Amari, S.-I. (1991). Dualistic geometry of the manifold of higher-order neurons. *Neural Networks*, 4, 443–451.

Andersen, R. A., Essick, G. K., & Siegel, R. M. (1985). Encoding of spatial location by posterior parietal neurons. *Science*, 230, 456–458.

Andersen, R. A., & Mountcastle, V. B. (1983). The influence of the angle of gaze upon the excitability of the light-sensitive neurons of the posterior parietal cortex. *Journal of Neuroscience*, 3, 532–548.

Andersen, R. A., Snyder, L. H., Bradley, D. C., & Xing, J. (1997). Multimodal representation of space in the posterior parietal cortex and its use in planning movements. *Annual Review of Neuroscience*, 20, 303–330.

Anumanchipalli, G. K., Chartier, J., & Chang, E. F. (2019). Speech synthesis from neural decoding of spoken sentences. *Nature*, 568, 493–498.

Arnal, L. H., & Giraud, A. L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences*, 16, 390–398.

Assaneo, M. F., & Poeppel, D. (2018). The coupling between auditory and motor cortices is rate-restricted: Evidence for an intrinsic speech-motor rhythm. *Science Advances*, 4, eaao3842.

Baggio, G. (2018). *Meaning in the brain*. Cambridge, MA: MIT Press.

Ballard, D. H. (2015). *Brain computation as hierarchical abstraction*. Cambridge, MA: MIT Press.

Bastiaansen, M., & Hagoort, P. (2006). Oscillatory neuronal dynamics during language comprehension. *Progress in Brain Research*, 159, 179–196.

Bastiaansen, M. C., van der Linden, M., Ter Keurs, M., Dijkstra, T., & Hagoort, P. (2005). Theta responses are involved in lexical—Semantic retrieval during language processing. *Journal of Cognitive Neuroscience*, 17, 530–541.

Bastiaansen, M. C., Oostenveld, R., Jensen, O., & Hagoort, P. (2008). I see what you mean: Theta power increases are involved in the retrieval of lexical semantic information. *Brain and Language*, 106, 15–28.

Beck, J. M., Ma, W. J., Kiani, R., Hanks, T., Churchland, A. K., Roitman, J., et al. (2008). Probabilistic population codes for Bayesian decision making. *Neuron*, 60, 1142–1152.

Bever, T. G., & Poeppel, D. (2010). Analysis by synthesis: A (re-) emerging program of research for language and vision. *Biolinguistics*, 4, 174–200.

Blokpoel, M. (2018). Sculpting computational-level models. *Topics in Cognitive Science*, 10, 641–648.

Boeckx, C. A., & Benítez-Burraco, A. (2014). The shape of the human language-ready brain. *Frontiers in Psychology*, 5, 282.

Bradley, T. D. (2018). What is applied category theory? *arXiv preprint arXiv:1809.05923*.

Brennan, J. R., & Martin, A. E. (2020). Phase synchronization varies systematically with linguistic structure composition. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, 375, 20190305.

Bressler, S. L., & Kelso, J. S. (2001). Cortical coordination dynamics and cognition. *Trends in Cognitive Sciences*, 5, 26–36.

Bressler, S. L., & Kelso, J. A. (2016). Coordination dynamics in cognitive neuroscience. *Frontiers in Neuroscience*, 10, 397.

Bucher, D., Taylor, A. L., & Marder, E. (2006). Central pattern generating neurons simultaneously express fast and slow rhythmic activities in the stomatogastric ganglion. *Journal of Neurophysiology*, 95, 3617–3632.

Buzsáki, G. (2006). *Rhythms of the brain*. Oxford, United Kingdom: Oxford University Press.

Buzsáki, G. (2010). Neural syntax: Cell assemblies, synapsembles, and readers. *Neuron*, 68, 362–385.

Buzsáki, G. (2019). *The brain from inside out*. Oxford, United Kingdom: Oxford University Press.

Buzsáki, G., & Schomburg, E. W. (2015). What does gamma coherence tell us about inter-regional neural communication? *Nature Neuroscience*, 18, 484–489.

Buzsáki, G., & Wang, X. J. (2012). Mechanisms of gamma oscillations. *Annual Review of Neuroscience*, 35, 203–225.

Carandini, M., & Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, 13, 51–62.

Carrasco, M., Ling, S., & Read, S. (2004). Attention alters appearance. *Nature Neuroscience*, 7, 308–313.

Cole, S. R., & Voytek, B. (2017). Brain oscillations and the importance of waveform shape. *Trends in Cognitive Sciences*, 21, 137–149.

Chang, E. F., Edwards, E., Nagarajan, S. S., Fogelson, N., Dalal, S. S., Canolty, R. T., et al. (2011). Cortical spatio-temporal dynamics underlying phonological target detection in humans. *Journal of Cognitive Neuroscience*, 23, 1437–1446.

Cheung, C., Hamilton, L. S., Johnson, K., & Chang, E. F. (2016). The auditory representation of speech sounds in human motor cortex. *eLife*, *5*, e12577.

Chomsky, N. (1957). Syntactic structures (The Hague: Mouton, 1957). *Review of Verbal Behavior by BF Skinner, Language*, *35*, 26–58.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

Cutter, M. G., Martin, A. E., & Sturt, P. (2020). Capitalization interacts with syntactic complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*, 1146–1164.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*, 283–321.

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

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

Doumas, L. A. A., & Hummel, J. E. (2005). Approaches to modeling human mental representations: What works, what doesn't and why. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 73–94). Cambridge: Cambridge University Press.

Doumas, L. A. A., & Hummel, J. E. (2012). Computational models of higher cognition. In K. J. Holyoak & R. G. Morrison (Eds.), *The Oxford handbook of thinking and reasoning* (Vol. 19). New York: Oxford University Press.

Doumas, L. A. A., Hummel, J. E., & Sandhofer, C. M. (2008). A theory of the discovery and predication of relational concepts. *Psychological Review*, *115*, 1–43.

Doumas, L. A. A., & Martin, A. E. (2018). Learning structured representations from experience. *Psychology of Learning and Motivation*, *69*, 165–203.

Doumas, L. A. A., Puebla, G., & Martin, A. E. (2017). How we learn things we didn't know already: A theory of learning structured representations from experience. *BioRxiv: 198804*.

Embick, D., & Poeppel, D. (2015). Towards a computational(ist) neurobiology of language: Correlational, integrated and explanatory neurolinguistics. *Language, Cognition and Neuroscience*, *30*, 357–366.

Engel, T. A., & Steinmetz, N. A. (2019). New perspectives on dimensionality and variability from large-scale cortical dynamics. *Current Opinion in Neurobiology*, *58*, 181–190.

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, *8*, 162–169.

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

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*, 3–71.

Fox, N. P., Sjerps, M. J., & Chang, E. F. (2017). Dynamic emergence of categorical perception of voice-onset time in human speech cortex. *Journal of the Acoustical Society of America*, *141*, 3571–3571.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, *6*, 78–84.

Friederici, A. D. (2011). The brain basis of language processing: From structure to function. *Physiological Reviews*, *91*, 1357–1392.

Friederici, A. D., & Singer, W. (2015). Grounding language processing on basic neurophysiological principles. *Trends in Cognitive Sciences*, *19*, 329–338.

Fries, P. (2009). Neuronal gamma-band synchronization as a fundamental process in cortical computation. *Annual Review of Neuroscience*, *32*, 209–224.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *360*, 815–836.

Gallistel, C. R. (1990). *Organization of learning (learning, development, and conceptual change)*. Cambridge, MA: MIT Press.

Gámez, J., Mendoza, G., Prado, L., Betancourt, A., & Merchant, H. (2019). The amplitude in periodic neural state trajectories underlies the tempo of rhythmic tapping. *PLoS Biology*, *17*, e3000054.

Gershman, S. J., & Niv, Y. (2010). Learning latent structure: Carving nature at its joints. *Current Opinion in Neurobiology*, *20*, 251–256.

Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Sciences*, *10*, 278–285.

Ghitza, O. (2011). Linking speech perception and neurophysiology: Speech decoding guided by cascaded oscillators locked to the input rhythm. *Frontiers in Psychology*, *2*, 130.

Ghitza, O. (2013). The theta-syllable: A unit of speech information defined by cortical function. *Frontiers in Psychology*, *4*, 138.

Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, *15*, 511–517.

Giraud, A. L., Kleinschmidt, A., Poeppel, D., Lund, T. E., Frackowiak, R. S., & Laufs, H. (2007). Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron*, *56*, 1127–1134.

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., et al. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biology*, *11*, e1001752.

Gwilliams, L., Linzen, T., Poeppel, D., & Marantz, A. (2018). In spoken word recognition, the future predicts the past. *Journal of Neuroscience*, *38*, 7585–7599.

Haegens, S., & Golumbic, E. Z. (2018). Rhythmic facilitation of sensory processing: A critical review. *Neuroscience & Biobehavioral Reviews*, *86*, 150–165.

Haegens, S., Händel, B. F., & Jensen, O. (2011). Top-down controlled alpha band activity in somatosensory areas determines behavioral performance in a discrimination task. *Journal of Neuroscience*, *31*, 5197–5204.

Hagoort, P. (2003). How the brain solves the binding problem for language: A neurocomputational model of syntactic processing. *Neuroimage*, *20(Suppl. 1)*, S18–S29.

Hagoort, P. (2013). MUC (memory, unification, control) and beyond. *Frontiers in Psychology*, *4*, 416.

Halgren, M., Fabó, D., Ulbert, I., Madsen, J. R., Erőss, L., Doyle, W. K., et al. (2018). Superficial slow rhythms integrate cortical processing in humans. *Scientific Reports*, *8*, 2055.

Hald, L. A., Bastiaansen, M. C., & Hagoort, P. (2006). EEG theta and gamma responses to semantic violations in online sentence processing. *Brain and Language*, *96*, 90–105.

Halle, M. (1962). Phonology in generative grammar. *Word*, *18*, 54–72.

Halle, M., & Stevens, K. (1962). Speech recognition: A model and a program for research. *IRE Transactions on Information Theory*, *8*, 155–159.

Hanslmayr, S., & Staudigl, T. (2014). How brain oscillations form memories—A processing based perspective on oscillatory subsequent memory effects. *Neuroimage*, *85*, 648–655.

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. New York: Wiley.

Heffner, C. C., Dilley, L. C., McAuley, J. D., & Pitt, M. A. (2013). When cues combine: How distal and proximal acoustic cues are integrated in word segmentation. *Language and Cognitive Processes*, *28*, 1275–1302.

Helmholtz, H. V. (1867). *Handbuch der physiologischen Optik* (Vol. 9). Leipzig: Voss.

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, *13*, 135–145.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*, 393–402.

Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, *18*, 1527–1554.

Holyoak, K. J., & Hummel, J. E. (2000). The proper treatment of symbols in a connectionist architecture. In E. Dietrich & A. B. Markman (Eds.), *Cognitive dynamics: Conceptual change in humans and machines* (pp. 229–263). Mahwah, NJ: Erlbaum.

Holland, J. H., Holyoak, K. J., Nisbett, R. E., & Thagard, P. (1986). *Induction: Processes of inference, learning, and discovery*. Cambridge, MA: MIT Press.

Hooper, S. L., & Moulins, M. (1989). Switching of a neuron from one network to another by sensory-induced changes in membrane properties. *Science*, *244*, 1587–1589.

Hornstein, N. (1984). *Logic as grammar*. Cambridge, MA: MIT Press.

Hummel, J. E. (2011). Getting symbols out of a neural architecture. *Connection Science*, *23*, 109–118.

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*, *104*, 427–466.

Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review*, *110*, 220–264.

Jazayeri, M. (2008). Probabilistic sensory recoding. *Current Opinion in Neurobiology*, *18*, 431–437.

Jazayeri, M., & Movshon, J. A. (2007). Integration of sensory evidence in motion discrimination. *Journal of Vision*, *7*, 7.1–7.7.

Jonas, P., & Buzsáki, G. (2007). Neural inhibition. *Scholarpedia*, *2*, 3286.

Kaplan, D. M. (2011). Explanation and description in computational neuroscience. *Synthese*, *183*, 339.

Kaplan, D. M., & Craver, C. F. (2011). The explanatory force of dynamical and mathematical models in neuroscience: A mechanistic perspective. *Philosophy of Science*, *78*, 601–627.

Kaufeld, G., Naumann, W., Meyer, A. S., Bosker, H. R., & Martin, A. E. (2019). Contextual speech rate influences morphosyntactic prediction and integration. *Language, Cognition, and Neuroscience*, 1–16.

Kaufeld, G., Ravenschlag, A., Meyer, A. S., Martin, A. E., & Bosker, H. R. (2020). Knowledge-based and signal-based cues are weighted flexibly during spoken language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*, 549–562.

Keitel, A., & Gross, J. (2016). Individual human brain areas can be identified from their characteristic spectral activation fingerprints. *PLoS Biology*, *14*, e1002498.

Kempen, G. (2014). Prolegomena to a neurocomputational architecture for human grammatical encoding and decoding. *Neuroinformatics*, *12*, 111–142.

Kim, D. O., Rhode, W. S., & Greenberg, S. R. (1986). Responses of cochlear nucleus neurons to speech signals: Neural encoding of pitch, intensity and other parameters. In B. C. J. Moore & R. D. Patterson (Eds.), *Auditory frequency selectivity* (pp. 281–288). Boston: Springer.

Kracht, M. (1992). The theory of syntactic domains. *Logic Group Preprint Series*, *75*.

Kratzer, A., & Heim, I. (1998). *Semantics in generative grammar* (Vol. 1185). Oxford: Blackwell.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647.

Lakatos, P., Chen, C. M., O'Connell, M. N., Mills, A., & Schroeder, C. E. (2007). Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron*, *53*, 279–292.

Larson, R. K. (2009). *Grammar as science*. Cambridge, MA: MIT Press.

Lee, J. (2010). *Introduction to topological manifolds* (Vol. 202). New York: Springer Science & Business Media.

Lenneberg, E. H. (1967). *The biological foundations of language*. New York: Wiley.

Leonard, M. K., Baud, M. O., Sjerps, M. J., & Chang, E. F. (2016). Perceptual restoration of masked speech in human cortex. *Nature Communications*, *7*, 13619.

Ling, S., Liu, T., & Carrasco, M. (2009). How spatial and feature-based attention affect the gain and tuning of population responses. *Vision Research*, *49*, 1194–1204.

Lisman, J. E., & Jensen, O. (2013). The theta-gamma neural code. *Neuron*, *77*, 1002–1016.

Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, *54*, 1001–1010.

Ma, W. J. (2012). Organizing probabilistic models of perception. *Trends in Cognitive Sciences*, *16*, 511–518.

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, *9*, 1432–1438.

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

Marcus, G. (2001). *The algebraic mind*. Cambridge, MA: MIT Press.

Marder, E. (2012). Neuromodulation of neuronal circuits: Back to the future. *Neuron*, *76*, 1–11.

Marslen-Wilson, W. D., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, *8*, 1–71.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, *10*, 29–63.

Martin, A. E. (2016). Language processing as cue integration: Grounding the psychology of language in perception and neurophysiology. *Frontiers in Psychology*, *7*, 120.

Martin, A. E. (2018). Cue integration during sentence comprehension: Electrophysiological evidence from ellipsis. *PLoS One*, *13*, e0206616.

Martin, A. E., & Doumas, L. A. A. (2017). A mechanism for the cortical computation of hierarchical linguistic structure. *PLoS Biology*, *15*, e2000663.

Martin, A. E., & Doumas, L. A. A. (2019a). Predicate learning in neural systems: Using oscillations to discover latent structure. *Current Opinion in Behavioral Sciences*, *29*, 77–83.

Martin, A. E., & Doumas, L. A. A. (2019b). Tensors and compositionality in neural systems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*, 20190306.

Martin, A. E., & McElree, B. (2008). A content-addressable pointer mechanism underlies comprehension of verb-phrase ellipsis. *Journal of Memory and Language*, *58*, 879–906.

Martin, A. E., & McElree, B. (2009). Memory operations that support language comprehension: Evidence from verb-phrase ellipsis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 1231–1239.

Martin, A. E., & McElree, B. (2011). Direct-access retrieval during sentence comprehension: Evidence from sluicing. *Journal of Memory and Language*, *64*, 327–343.

Martin, A. E., & McElree, B. (2018). Retrieval cues and syntactic ambiguity resolution: Speed–accuracy tradeoff evidence. *Language, Cognition and Neuroscience*, *33*, 769–783.

Martin, A. E., Monahan, P. J., & Samuel, A. G. (2017). Prediction of agreement and phonetic overlap shape sublexical identification. *Language and Speech*, 60, 356–376.

Meyer, L. (2018). The neural oscillations of speech processing and language comprehension: State of the art and emerging mechanisms. *European Journal of Neuroscience*, 48, 2609–2621.

Meyer, L., Henry, M. J., Gaston, P., Schmuck, N., & Friederici, A. D. (2016). Linguistic bias modulates interpretation of speech via neural delta-band oscillations. *Cerebral Cortex*, 27, 4293–4302.

Meyer, L., Sun, Y., & Martin, A. E. (2019). Synchronous, but not entrained: Exogenous and endogenous cortical rhythms of speech and language processing. *Language, Cognition, and Neuroscience*.

Morillon, B., Kell, C. A., & Giraud, A. L. (2009). Three stages and four neural systems in time estimation. *Journal of Neuroscience*, 29, 14803–14811.

Murphy, E. (2018). Interfaces (travelling oscillations) + recursion (delta-theta code)= language. In E. Luef & M. Manuela (Eds.), *The talking species: Perspectives on the evolutionary, neuronal and cultural foundations of language* (pp. 251–269). Graz, Austria: Unipress Graz Verlag.

Nelson, M. J., El Karoui, I., Giber, K., Yang, X., Cohen, L., Koopman, H., et al. (2017). Neurophysiological dynamics of phrase-structure building during sentence processing. *Proceedings of the National Academy of Sciences, U.S.A.*, 114, E3669–E3678.

Nieuwland, M. S., & Martin, A. E. (2012). If the real world were irrelevant, so to speak: The role of propositional truth-value in counterfactual sentence comprehension. *Cognition*, 122, 102–109.

Nieuwland, M. S., Otten, M., & Van Berkum, J. J. (2007). Who are you talking about? Tracking discourse-level referential processing with event-related brain potentials. *Journal of Cognitive Neuroscience*, 19, 228–236.

Nieuwland, M. S., & Van Berkum, J. J. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, 18, 1098–1111.

Obleser, J., & Kayser, C. (2019). Neural entrainment and attentional selection in the listening brain. *Trends in Cognitive Sciences*, 23, 913–926.

Obleser, J., Meyer, L., & Friederici, A. D. (2011). Dynamic assignment of neural resources in auditory comprehension of complex sentences. *Neuroimage*, 56, 2310–2320.

O'Flaherty, W. D. (1981). *The Rig Veda: An anthology: One hundred and eight hymns*. New York: Penguin Books.

Olshausen, B. A. (2014). 27 Perception as an inference problem. In G. R. Mangun & M. S. Gazzaniga (Eds.), *The cognitive neurosciences* (pp. 295–304). Cambridge, MA: MIT Press.

Partee, B. (1975). Montague grammar and transformational grammar. *Linguistic Inquiry*, 6, 203–300.

Partee, B. (1984). Compositionality. *Varieties of Formal Semantics*, 3, 281–311.

Partee, B. H., ter Meulen, A. G., & Wall, R. E. (2012). *Mathematical methods in linguistics* (Vol. 30). Berlin: Springer Science & Business Media.

Pauls, A., & Klein, D. (2012). Large-scale syntactic language modeling with treelets. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers* (Vol. 1, pp. 959–968). https://www.aclweb.org/anthology/P12-1101/.

Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology*, 3, 320.

Pikovsky, A., & Rosenblum, M. (2007). Synchronization. *Scholarpedia*, 2, 1459.

Phillips, S. (2020). Sheaving—A universal construction for semantic compositionality. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, 375, 20190303.

Phillips, S., & Wilson, W. H. (2010). Categorial compositionality: A category theory explanation for the systematicity of human cognition. *PLoS Computational Biology*, 6, e1000858.

Piccinini, G. (2007). Computing mechanisms. *Philosophy of Science*, 74, 501–526.

Poeppel, D., & Monahan, P. J. (2011). Feedforward and feedback in speech perception: Revisiting analysis by synthesis. *Language and Cognitive Processes*, 26, 935–951.

Ramon y Cajal, S. (1928). *Degeneration and regeneration of the nervous system*. London: Clarendon Press.

Rimmele, J. M., Morillon, B., Poeppel, D., & Arnal, L. H. (2018). Proactive sensing of periodic and aperiodic auditory patterns. *Trends in Cognitive Sciences*, 22, 870–882.

Robins, R. H. (2013). *A short history of linguistics*. New York: Routledge.

Rumelhart, D. E., McClelland, J. L., & PDP Research Group. (1987). Parallel distributed processing (Vol. 1, p. 184). Cambridge, MA: MIT Press.

Salinas, E., & Abbott, L. F. (2001). Coordinate transformations in the visual system: How to generate gain fields and what to compute with them. *Progress in Brain Research*, 130, 175–190.

Salinas, E., & Sejnowski, T. J. (2001). Book review: Gain modulation in the central nervous system: Where behavior, neurophysiology, and computation meet. *Neuroscientist*, 7, 430–440.

Salinas, E., & Thier, P. (2000). Gain modulation: A major computational principle of the central nervous system. *Neuron*, 27, 15–21.

Schotter, E. R., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Attention, Perception, & Psychophysics*, 74, 5–35.

Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences*, 32, 9–18.

Shastri, L. (1999). Advances in SHRUTI—A neurally motivated model of relational knowledge representation and rapid inference using temporal synchrony. *Applied Intelligence*, 11, 79–108.

Singer, W. (1999). Neuronal synchrony: A versatile code for the definition of relations? *Neuron*, 24, 49–65.

Skipper, J. I. (2015). The NOLB model: A model of the natural organization of language and the brain. In R. M. Willems (Ed.), *Cognitive neuroscience of natural language use* (pp. 101–134). Cambridge, United Kingdom: Cambridge University Press.

Skipper, J. I., Nusbaum, H. C., & Small, S. L. (2005). Listening to talking faces: Motor cortical activation during speech perception. *Neuroimage*, 25, 76–89.

Smith, E. C., & Lewicki, M. S. (2006). Efficient auditory coding. *Nature*, 439, 978–982.

Spitzer, B., & Haegens, S. (2017). Beyond the status quo: A role for beta oscillations in endogenous content (re)activation. *eNeuro*, 4, ENEURO.0170-17.2017.

Sporns, O., & Kötter, R. (2004). Motifs in brain networks. *PLoS Biology*, 2, e369.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, 30, 276–315.

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, 48, 542–562.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, 331, 1279–1285.

Tian, X., & Poeppel, D. (2013). The effect of imagination on stimulation: The functional specificity of efference copies in speech processing. *Journal of Cognitive Neuroscience*, *25*, 1020–1036.

Tian, X., & Poeppel, D. (2014). Dynamics of self-monitoring and error detection in speech production: Evidence from mental imagery and MEG. *Journal of Cognitive Neuroscience*, *27*, 352–364.

van Alphen, P. M., & McQueen, J. M. (2006). The effect of voice onset time differences on lexical access in dutch. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 178–196.

van Rooij, I. (2008). The tractable cognition thesis. *Cognitive Science*, *32*, 939–984.

van Rooij, I., Blokpoel, M., Kwisthout, J., & Wareham, T. (2019). *Cognition and intractability: A guide to classical and parameterized complexity analysis*. Cambridge, United Kingdom: Cambridge University Press.

van Rullen, R., & Koch, C. (2003). Is perception discrete or continuous? *Trends in Cognitive Sciences*, *7*, 207–213.

Veldre, A., & Andrews, S. (2018). Beyond cloze probability: Parafoveal processing of semantic and syntactic information during reading. *Journal of Memory and Language*, *100*, 1–17.

von der Malsburg, C. (1995). Binding in models of perception and brain function. *Current Opinion in Neurobiology*, *5*, 520–526.

von der Malsburg, C. (1999). The what and why of binding: The modeler's perspective. *Neuron*, *24*, 95–104.

Vosse, T., & Kempen, G. (2000). Syntactic structure assembly in human parsing: A computational model based on competitive inhibition and a lexicalist grammar. *Cognition*, *75*, 105–143.

Weimann, J. M., & Marder, E. (1994). Switching neurons are integral members of multiple oscillatory networks. *Current Biology*, *4*, 896–902.

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

Zipser, D., & Andersen, R. A. (1988). A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature*, *331*, 679–684.