Syntax and the cognitive neuroscience of syntactic structure building

Jon Sprouse (corresponding author)
Department of Linguistics
University of Connecticut
jon.sprouse@uconn.edu

Norbert Hornstein
Department of Linguistics
University of Maryland
nhornste@umd.edu

Abstract

In this chapter we argue that modern syntactic theories are well-suited to provide a cognitive theory of the structure-building computations that neural systems must perform in order to process language. Therefore a plausible research program for cognitive neuroscience would be to search for a theory of (i) how neural systems could perform these computations, and (ii) which neural systems are performing these computations during any given language processing event. As syntacticians this strikes us as the natural evolution of the goals of the cognitive revolution of the 1950s in general, and the goals of generative syntax in particular. However, we are also aware that this is not how many cognitive neuroscientists would describe current syntactic theory. As such, we provide two concrete examples of the cognitive structure-building computations proposed (from two distinct syntactic theories), and discuss the prospects and challenges of using these theories as a roadmap for a large-scale collaboration between syntacticians and neuroscientists in the search for neuronal instantiations of these computations.

Keywords: syntax, minimalism, cognitive neuroscience, neuronal computation

1. Introduction

One goal of cognitive neuroscience, if not *the* goal of cognitive neuroscience, is to uncover how neural systems can give rise to the computations that underlie human cognition. Assuming, as most do, that the relevant biological description can be found at the level of neurons, then another way of stating this is that cognitive neuroscience is (at least) the search for the neuronal computations that underlie human cognition (e.g., Carandini, 2012; Carandini & Heeger, 2012). To the extent that this is an accurate formulation of the goal(s) of the field, any research program in cognitive neuroscience will have three components: (i) a cognitive theory that specifies the potential computations that underlie cognition, (ii) a neuroscientific theory that specifies how neurons (or populations of neurons) carry out different types of computations, and (iii) a linking theory that maps between the cognitive theory and the neuroscientific theory (Marantz, 2005; Poeppel & Embick, 2005; Poeppel, 2012). We take all of this to be relatively uncontroversial; however, we mention it explicitly because we believe that modern syntactic theories, under a certain conception, are well positioned to provide the first component (a theory of computations) for a cognitive neuroscientific theory of syntactic structure building. Our goal in this chapter is to

make a case for this belief. We hope to demonstrate that the potential for a productive cross-fertilization exists between theoretical syntacticians and neuroscientists, and suggest that developments in syntactic theory over the past two decades make this an optimal time to engage seriously in this collaboration.

For ease of exposition, we will call our view *the computational view of syntax.* This is simply the idea that the operations that have been proposed in syntactic theory (e.g., `merge` in Minimalism, `substitution` in Tree Adjoining Grammar) are a plausible cognitive theory of the structure-building computations that neural systems must perform in order to process language. Therefore a plausible research program for cognitive neuroscience would be to search for a theory of (i) how neural systems could perform these computations, and (ii) which neural systems are performing these computations during any given language processing event. As syntacticians this strikes us as the natural evolution of the goals set forth in the 1950s when the field of cognitive science first coalesced, and a natural evolution of the goals of generative linguistics in particular. However, we are also aware that this is not how many would describe current syntactic theory. Therefore we will attempt to make our case in a series of steps. In section 2 we provide a brief history of the field of syntax. The goal of this section is to contextualize modern syntactic theories such that it becomes clear that modern theories are not simply lists of grammatical rules (although older theories were), but instead theories of cognitive computations. In section 3 we present two concrete examples of potential structure-building computations (from two distinct contemporary syntactic theories) to illustrate the computational view of syntax. In section 4 we lay out several of the properties of modern syntactic theories that we believe make them well suited for the computational view of syntax. We believe that these properties will be easily recognizable to all cognitive neuroscientists as the properties of a theory of cognitive computations. In section 5 we discuss the large-scale collaboration between syntacticians, psycholinguists, and neuroscientists that will be necessary to construct a cognitive neuroscience of syntactic structure building. In section 6 we discuss some of the challenges that this collaboration might face. Section 7 concludes.

Before making our case for the computational view of syntax, a small clarification about the scope of this chapter may be in order. We have explicitly chosen to focus on the issue of *why* syntactic theories will be useful for a cognitive neuroscience of language, and not *how* syntactic theorizing is conducted today. In other words, this chapter is intended to layout arguments in favor of a large-scale collaboration between syntacticians and neuroscientists, and is not intended to be a review chapter on syntax. We assume that if our arguments are successful, syntacticians within these collaborations can carry the burden of doing the syntax. That being said, for readers interested in reviews of topics in contemporary syntax, we can recommend the review chapters in the recently published *Cambridge Handbook of Generative Syntax* (2013, edited by Marcel den Dikken), which contains 26 excellent review chapters covering everything from the history and goals of syntactic theory, to overviews of several major contemporary theories, to reviews of specific phenomena in syntax.

2. A brief history of syntactic theory

Syntactic theory starts from two critical observations. The first is that there is no upper bound on the number of possible phrases/sentences within any given language (i.e. languages are for all practical purposes "infinite"). This implies that successful language learning is not just the

memorization of a set of expressions (otherwise infinity would be impossible), but rather the acquisition of a grammar, which is just a finite specification of a recursive set of combinatory rules. The second observation is that any child can acquire any language (e.g., a child born to US citizens living in Kenya will successfully learn Swahili if exposed to Swahili speakers during childhood). Given that the first observation suggests that languages should be viewed as grammars, the second observation translates as any child can acquire any grammar. These two observations lead to the two driving questions for the field of syntax:

1. What are the properties of the grammars of all of the world's languages?

2. What are the mental mechanisms that allow humans to learn human languages?

The goal of Generative Syntax (GS) over the last 60 years has been to explore the properties of human grammars (question 1) in such a way as to make it possible to explore the mental mechanisms that are required for successful language acquisition (question 2). As with any specialized science, the pursuit of these dual driving questions has led to the development of specific research programs and technical terminology, both of which have at times been opaque to other cognitive scientists working outside of syntax. Our goal in this section is to provide a brief history of the way the field has pursued these driving questions (to contextualize the modern syntactic theories discussed in section 2.2), and to clarify some of the major points of miscommunication that have historically arisen between syntacticians and other cognitive scientists.

Generative Syntax began by describing specific rules found in particular languages (and so contained in the grammars of these languages). This is hardly surprising; for if one is interested in the kinds of rules natural language grammars contain, a good way to begin is by looking for particular examples of such rules. Thus, in the earliest period of GS, syntacticians built mini-grammars describing how various constructions in particular languages were built (e.g. relative clauses in Chamorro, questions in English, topic constructions in German, reflexivization in French, etc.) and how they interacted with one another to generate a reasonably robust "fragment" of the language.

With models of such grammars in hand, the next step was to factor out the common properties of these language particular grammars and organize them into rule types (e.g. movement rules, phrase structure rules, construal rules). This more abstract categorization allowed for the radical simplification of the language particular rules investigated in the prior period, with constructions reducing to congeries of simpler operations (although analogies are dangerous, this seems similar to the way other sciences often discover that seemingly distinct phenomena are in fact related, such as the unification of planetary motion, projectile motion, and tidal motion as instances of gravitational attraction in physics). By the mid 1980s there were several reasonably well-articulated candidate theories of syntax (e.g. Government and Binding Theory, Lexical-Functional Grammar, Tree Adjoining Grammar), each specifying various rule types and their properties, and each illuminating commonalities across constructions and across languages.

The simplification of grammatical rule types also led to progress on the second driving question. By reducing syntactic theories to only a few rule types, syntacticians could reduce the number of learning mechanisms required to learn human grammars (here we use the term "learning mechanisms" as a cover term for all of the components of learning theories: biases to

attend to certain input, specifications of hypothesis spaces, algorithms for searching hypothesis spaces, etc). With fewer learning mechanisms in the theory, syntacticians were able to investigate (and debate) the nature of the learning mechanisms themselves. Although there are a number of dimensions along which learning mechanisms might vary, syntactic theory has often focused on two in particular. The first is specificity: the learning mechanisms can either be domain-general, meaning that they are shared by several (or all) cognitive domains, or they can be domain-specific, meaning that they are specific to language learning. The second dimension is nativity: the learning mechanisms can either be innate, meaning that they arise due to the genetic make-up of the organism, or they can be derived, meaning that they are constructed from the combination of experience and other innate mechanisms. This leads to a 2x2 grid that can be used to classify any postulated learning mechanism (see also Pearl and Sprouse 2013):

Specificity

	Domain-specific	Domain-general
Innate	Universal Grammar	e.g., statistical learning
Derived	e.g., learning to read	e.g., n-grams

(row label: Nativity)

What is particularly interesting about this grid is that it helps to clarify some of the miscommunications that have often arisen between syntacticians and other cognitive scientists surrounding terms like "innate", "domain-specific", and worst of all, "Universal Grammar". This grid highlights the fact that the classification of any given learning mechanism is an empirical one. In other words, given a rule type X, and a learning mechanism Y that could give rise to X, which cell does Y occupy in the grid? It may be the case that one or more of the cells are never used. Second, this grid highlights the fact that a complete specification of all of the rule types underlying human grammars, and all of the learning mechanisms deployed to learn human grammars, could involve any combination of the 4 types of learning mechanisms. As cognitive scientists, syntacticians are interested in all of the mechanisms underlying human syntax, not just the ones that get all of the attention in debates. Finally, this grid makes it clear what exactly syntacticians mean when they use the term "Universal Grammar." Universal Grammar (UG) is just a special term for potential learning mechanisms that are simultaneously domain-specific and innate. Despite this rhetorical flourish, we hope it is clear that syntacticians view UG mechanisms (if they exist at all) as only a subset of the learning mechanisms that give rise to human language.[1]

---

[1] As a quick side note on Universal Grammar, the reason that UG receives so much attention, both within the syntax literature and across cognitive science, is that the other three types of learning mechanisms are generally uncontentious. For example, it is widely assumed that learning cannot occur in a blank slate (i.e. every learning system needs some built in biases if there is to be any generalization beyond the input), therefore at least one learning mechanism must be innate. Nearly every postulated neural architecture (both symbolic and sub-symbolic) assumes some form of statistical learning, which is presumably a learning mechanism (or set of

The progress made in the 1980s on simplifying the rule types in human grammars also laid the foundation for the current research program within modern Generative Syntax: to distill the computational commonalities found among the various *kinds* of rules (i.e., the computational features common to movement rules, phrase building rules, and construal rules). Here again the dimension of domain-generality and domain-specificity plays a role in theoretical discussions, but this time at the level of cognitive computation rather than at the level of learning mechanisms. As syntacticians have made progress distilling the computational properties of grammatical rules, they have found that some of the suggested computations appear similar to computations in other domains of cognition (e.g., the binding, or concatenation, of two mental representations), while others still retain some amount of domain-specificity (see section 3 for a concrete example). Current GS work is pursuing this program in full force: attempting to identify the basic computations, and determine which are specific to the syntax, and which are shared with other cognitive domains.

Note the odyssey described above: the field of syntax moved from the study of very specific descriptions of particular rules in particular languages, to very general descriptions of the properties of linguistic computations and their relationship, and finally to cognitive computation more generally. This shift in the "grain" of linguistic analysis (in the sense of Embick & Poeppel, 2005) has had two important effects. First, it has reduced the special "linguistic" character of syntactic computations, making them more similar to the cognitive computations we find in other domains. Second, it has encouraged investigation of how syntactic computations might be used in real time tasks such as parsing, production, and learning. Both these effects have had the consequence of bringing syntactic theory much closer to the empirical interests of others working in cognitive neuroscience.

Unfortunately, this shift in syntactic theory and it implications for cognitive neuroscience has not always been widely appreciated. Although the field of syntax was a central player in cognitive science when the field coalesced in the 1950s, in the intervening decades, syntax and the other domains of cognitive science have drifted apart. Some of this drift is the inevitable consequence of scientific specialization, and some of it reflects the internal logic of the different research programs (i.e. that the rule-based theories of the past were a necessary step in the evolution of syntactic theories). However, some of the drift reflects the view that syntactic theory has little to contribute to other domains of language research (including cognitive neuroscience). We worry that part of this problem may be that syntacticians have done a less-than-adequate job of conveying the general computational character of modern syntactic theories. In the absence of

---

mechanisms) that is domain-general and innate. The domain-general/derived cell is likely filled with the more complex statistical learning mechanisms required by different domains of cognition, such as the ability to track the probabilities of different sized sequences (n-grams). Similarly, the domain-specific/derived cell could potentially contain the learning mechanisms tailored to specific areas of higher-order cognition, such as reading (or maybe even language itself), but built from cognitive mechanisms available more broadly. It is the final cell, domain-specific/innate, that is the most contentious (and therefore, to some, the most interesting). In syntax, we call learning mechanisms that potentially fall into this cell Universal Grammar to highlight their significance. Currently, as we note below, a very hot area of syntactic investigation aims to reduce these domain specific innate mechanisms to a minimum without losing explanations for the linguistic phenomena and generalizations that syntacticians have discovered over the last 60 years of syntactic research.

such discussions, it would not be surprising to learn that some cognitive neuroscientists still view syntax in terms of the phrase structure rules and transformations that typified syntactic theory in the 1950s and 1960s (and in varying forms up through the 1980s), rather than the more cognitively general computations common in current practice.[2] In the next two subsections, we provide two examples of how contemporary syntax might fruitfully make contact with cognitive neuroscience.

3. Two concrete examples of syntactic structure-building computations

While early formulations of syntactic theories postulated complex rules that applied to entire constructions (often permuting, adding, or deleting multiple words in different positions in the constructions), as noted in section 2.1, there has been a steady evolution toward theories that postulate a small number of structure-building operations that can be applied mechanistically (or derivationally) to construct more elaborate syntactic structures in a piece-wise fashion. With very few exceptions, the primitives of contemporary syntactic theories are (i) units, and (ii) the computations that apply to those units. Here are two concrete examples:[3]

The syntactic theory known as *Minimalism* (or the Minimalist Program) postulates a single structure-building computation called `merge`, which takes two units and combines them to form a third. The units in Minimalism are lexical and sub-lexical items (something akin to the notion of word or morpheme, although the details can vary by analysis). `Merge` applies to these units directly, and also applies recursively to the output of previous instances of `merge`. In this way, `merge` can be used to iteratively construct complex syntactic structures from a basic inventory of lexical atoms. Of course, `merge` cannot freely concatenate any two units together.

---

[2] The rule & transformation view of syntax has other problems as well. This conception of syntax is considered problematic for the computational view of syntax, because there are well known empirical results from the 1950s and 1960s that appear to demonstrate that rule-based syntactic theories of that sort are poor models for real-time sentence processing (or, more specifically, poor predictors of complexity effects in language processing, as captured by the Derivational Theory of Complexity; for reviews see Fodor, Bever, & Garrett, 1974; but see Phillips, 1996 for a useful re-evaluation of these claims). This problem is compounded by the fact that syntactic theories are at best only theories of syntactic structure building, with little to nothing to say about other components that are necessary for a complete theory of sentence processing, such as ambiguity resolution, memory/resource allocation, semantic structure building, and discourse structure building. Therefore if one views syntactic theory as a rule-based theory, then it might appear to be a poor theory of only one small corner of language processing. Even as syntacticians, we understand why other cognitive scientists might find this version of syntactic theory difficult to engage with.

[3] There are, of course, a number of other syntactic theories that propose different types of computations (and different types of units). For example, Head-Driven Phrase Structure Grammar (HPSG) proposes a computation similar to merge, but without the possibility of internal merge (non-local dependencies involve a special slash unit instead). Construction grammar proposes a tree-unification computation similar to substitution in TAG, but operating over much larger units (entire constructions) and with the possibility of multiple unification points in a single construction. We assume that a full-fledged research program on the computational view of syntax would investigate all of these possible theories.

This means that restrictions on `merge` must be built into the lexical items themselves (only certain lexical items are compatible with each other), and in the case of merging units with the output of previous merges, this means that the outputs of `merge` must also contain restrictive properties. This is accomplished through a labeling computation, let's call it `label`, that applies a label to the output of `merge`, which can then be used to determine what that output can be merged with in the future.

The goal of syntactic theory is to capture the major properties of human syntactic structures with the proposed units and computations. For concreteness, we will illustrate how `merge` and `label` succeed in capturing two such properties. The first is the distinction between local dependencies and non-local dependencies. A local dependency is simply the relationship between two adjacent items in a sentence. Local dependencies are captured by `merge` by concatenating two distinct elements together. A non-local dependency is a relationship between two elements that are not adjacent in a sentence, such as the word *what* and *buy* in the question *What did John buy?*. Non-local dependencies can be modeled by `merge` by concatenating a phrase with an element that is already properly contained within that phrase. Syntacticians call the former instantiation `external merge`, because the two elements are external to each other, and the latter instantiation `internal merge`, because one element is properly contained within the other (Chomsky, 2004). The second property is the distinction between structures that contain verbs and their arguments (e.g. *eat bananas*), and structures that contain modifiers (e.g. *eat quickly*). The former, which we can call non-adjunction structures, are built from a combination of `merge` and `label`; the latter, which we can call adjunction structures, are built from `merge` alone (no `label`) (Hornstein, 2009). In this way, the two primitive computations `merge` and `label` can be used to construct syntactic structures capable of modeling the variety of structures one finds within natural language.

The syntactic theory known as *Tree Adjoining Grammar* (TAG) postulates two structure-building computations called `substitution` and `adjunction`. The units in TAG are small chunks of syntactic structure, or trees (hence the name of the theory). The `substitution` computation allows two elementary trees to be concatenated into locally-dependent, *non-adjunction* structures. The `adjunction` computation, as the name implies, allows two trees to be concatenated into locally-dependent, *adjunction* structures. TAG captures non-local dependencies that are only a single clause in length with a single elementary tree (so, *What did John buy?* is a single tree without any application of `substitution` or `adjunction`). For dependencies that are more than one clause in length, the `adjunction` computation is applied to a special type of tree called an auxiliary tree to extend the dependency length. In this way, the two primitive computations `substitution` and `adjunction` can be used to construct syntactic structures from elementary and auxiliary trees, and give rise to the important distinctions of human syntax (for accessible introductions to TAG, see Frank, 2002; and Frank, 2013).

Box 1: Structure-building computations in Minimalism and Tree Adjoining Grammar

The structure-building computation in Minimalism is called `merge`. It takes two syntactic objects, and concatenates them into a third object. When the two syntactic objects are distinct, it

is called `external merge`. When one of the objects are contained within the other, it is called `internal merge`:

External merge: [eat] + [bananas] = [[eat] [bananas]]
Internal merge: [did John buy what] + [what] = [[what] [did John buy what]]

The `label` computation determines the properties of the new syntactic object constructed by merge by applying a label based on the properties of one of the merged objects (the head). `Label` is mandatory for the merge of argument relationships (e.g. verbs and their arguments), but appears to be optional for the merge of adjuncts (e.g., verbs and modifiers):

Merge with Label: $[_V$ eat$]$ + $[_{NP}$ bananas$]$ = $[_{VP} [_V$ eat$] [_{NP}$ bananas$]]$
Merge without Label: $[_{VP} [_V$ run$]]$ + $[_{AdvP}$ quickly$]$ = $[\ [_{VP} [_V$ run$]] [_{AdvP}$ quickly$]]$

Tree Adjoining Grammar proposes two structure-building operations. `Substitution` combines two elementary trees to form argument relationships, while `adjunction` combines elementary trees and adjunct trees to form adjunction structures:

Substitution: $[_{DP}$ John$]$ + $[_{TP} [_{DP} \ ] [_{VP}$ eats bananas$]]$ = $[_{TP} [_{DP}$ John$] [_{VP}$ eats bananas$]]$
Adjunction: $[_{TP} [_{DP}$ John$] [_{VP} [_V$ runs$]]]$ + $[_{VP} [_{VP} \ ]$ quickly$]$
                = $[_{TP} [_{DP}$ John$] [_{VP} [_{VP} [_V$ runs$]]$ quickly$]]$

Although both theories capture the same set of phenomena in human syntax, and both theories postulate structure-building computations, they do so using (i) different computations, (ii) different units, and (iii) different combinations of computations for each phenomenon. For non-adjunction structures that involve only local dependencies, Minimalism uses `external merge` and `label`, while TAG uses `substitution` with two elementary trees. For adjunction structures, Minimalism uses `external merge` alone, while TAG uses `adjunction` with two elementary trees. For non-local dependencies, Minimalism uses `internal merge` and `label`, while TAG uses `adjunction` with one elementary tree and one auxiliary tree. On the one hand, the similarities between these two syntactic theories (i.e., both use two basic computations to capture a wide range of characteristics of human syntax) suggest that both are tapping into deeper truths about human structure-building computations. On the other hand, the subtle differences in the character of the proposed computations suggest that one might be able to derive competing predictions from each theory about the presence or absence of computations in different constructions. This combination of abstract similarities and subtle differences strike us as a potentially fruitful starting point for a search for neuronal structure-building computations.

4. Additional properties of syntactic theories that one would expect from a theory of cognitive computations

In addition to focusing on structure-building computations, there are a number of additional properties of contemporary syntactic theories that make them ideal candidates for the computational view of syntax. Here we review three.

First, contemporary syntactic theories attempt to minimize the number of computations while maximizing the number of phenomena captured by the theory. This is a general desideratum of scientific theories in general (it is sometimes called unification, or reductionism, or just Occam's razor), and syntax, as a science, has adopted it as well. In fact, the name Minimalism was chosen to reflect the fact that years of investigations using earlier theories had yielded enough information about the properties of language as a cognitive system that it was finally possible to fruitfully incorporate unification/reduction/Occam's razor as a core principle of the research program. Other syntactic theories have been less blunt about this in their naming conventions, but the principles are obvious in the shape of the theories. Commitment to Occam has led to syntactic theories based on simple computations with wide applicability across the thousands of syntactic constructions in human languages. One nice side benefit of the ratio of computations to constructions is that it may make the search for neurophysiological correlates of these computations more fruitful, especially given concerns about spurious correlations in high-dimensional neurophysiological data.

Second, syntactic theories attempt to minimize the number of domain-specific computations, and maximize the number of domain-general computations (to the extent possible given the overall minimization of the number of computations). This is an important, and often overlooked, point within syntax. The `merge` computation in Minimalism and the `substitution` computation in TAG are both plausibly domain-general computations similar to the binding computations that occur in multiple cognitive domains (vision, hearing, etc), albeit operating over language-specific representations. The formulation of these plausibly domain-general computations stems directly from the premium that syntactic theories now place on unification/reductionism. In contrast, the `label` computation and the `adjunction` computation are potentially domain-specific, as there are no obvious correlates in other cognitive domains, though that could just be a consequence of our current state of knowledge. The question of whether plausibly domain-specific computations like `label` and `adjunction` can be learned or must be innate is an open area of research in language acquisition.

Finally, syntactic theories have mapped a sizable portion of the potential hypothesis space of syntactic structure building computations. As we have mentioned above, with few exceptions, every contemporary syntactic theory has the potential to serve as a theory of cognitive structure building computations. While the sheer number of competing theories may seem daunting from outside of syntax, from inside of syntax we believe this is a necessary step in the research. We need to explore every possible combination of unit-size and computation type that captures the empirical facts of human languages (and to be clear, not every combination does) in order to provide neuroscientists with a list of possible cognitive computations. To be sure, there is more work to be done on this front. And it goes without saying that syntacticians actively debate the empirical coverage of the different theories, and also how well each theory can achieve empirical coverage without inelegant stipulations. But from the perspective of cognitive neuroscience, the value is in the hypothesis space – each theory represents a different hypothesis about the types of fundamental structure building computations (and the distribution of those functions across different sentences in any given language).[4]

---

[4] Inside of the field of syntax there is a recurring debate about whether different syntactic theories (e.g. Minimalism and TAG) are in some sense notational variants of one another. There are various mathematical proofs demonstrating that many theories are identical in terms of weak generative capacity (i.e., the ability to create certain strings of symbols and not others; e.g. Joshi

5. The collaboration necessary to engage in this program

The research program that the computational view of syntax suggests will require close collaboration between different types of researchers. The first step is for syntacticians to identify the structure-building computations that are deployed at each point in constructions from human syntax. From these analyses, syntacticians could identify two types of interesting cases. The first interesting case would be constructions that predict the same type of structure building computation at the same location in all theories (e.g., Minimalism predicts merge at the same location in the construction as TAG predicts substitution). These areas of convergence may be fruitful places to begin the search for neuronal computations. A second interesting case would be constructions that require diverging computations across theories (e.g., Minimalism predicts merge but TAG predicts adjunction). If these analyses could be identified across a large number of constructions, it should be possible to construct a type of comparison/subtractive logic that could uncover neuronal correlates of these computations. It seems to us that phenomena that vary along the major dimensions of human syntax, such as non-adjunction versus adjunction structures, or local versus non-local dependencies, will be most likely to lead to these types of convergences and divergences. But over the long term, every phenomenon of syntax should be investigated (to the extent possible given some of the challenges discussed in section 4 below).

   The second step is for syntacticians and theoretical neuroscientists to figure out how neural systems deploy the structure-building computations that underlie the phenomenon in each theory. In practice, this step might require several substeps. For example, the typical form of syntactic theories is "bottom-up": the most deeply embedded constituents are constructed first, followed by the next most deeply embedded, and so on. This is largely the reverse order from sentence comprehension and production. Given that the empirical studies required by later steps will be based on comprehension (and perhaps production), it may be necessary to convert the "bottom-up" computations of syntactic theories into the "left-to-right" or "top-down" computations of parsing theories. There exist several computational models for how to relate bottom-up grammars with left-right parsers. This step will most likely involve collaboration among mathematical linguists (to rigorously formalize the syntactic computations (e.g., Stabler, 1997; Collins & Stabler, 2011)), mathematical psycholinguists to convert those computations into parsing computations (e.g., Marcus, 1980; Berwick & Weinberg, 1984; Stabler, 2011; Stabler, 2013; and for issues beyond structure-building: Hale, 2003; Kobele et al., 2013) and neuroscientists to identify candidate neurocomputational systems. Although this sounds straightforward, it is likely that the space of possible computations will expand at each step, from syntactic computations to mathematically formalized computations, from formalized computations to parsing computations, and from parsing computations to neuronal computations.

---

et al., 1991; Stabler, 1997; Michaelis, 1998). However, it is an open question whether these theories are equivalent in other terms, such as strong generative capacity (the types of structures that they can generate) or empirical adequacy for human languages. It is interesting to note that inside of syntax this debate is often couched in terms of theoretical "elegance", i.e., how elegantly one theory captures a specific phenomenon relative to another theory. However, the research program suggested here would make such debates purely empirical: the "correct" syntactic theory would be the one that specifies the correct distribution of syntactic computations (and therefore their neuronal instantiations) across all of the constructions of a given language.

It is quite possible that this step will result in hypothesis spaces for the possible neuronal computations for each syntactic theory relevant to each phenomenon.

Once the structure-building computations have been translated into potential neuronal computations (or hypothesis spaces of potential neuronal computations), the final step is to look for evidence of those computations in neural systems. Again, although we state this as a single step in principle, we assume that it will be a multi-faceted process in practice, drawing on neuroscientists of all stripes: electrophysiologists (EEG/MEG), neuroimagers (fMRI), and even neurosurgeons (ECoG). As syntacticians, this step is the furthest beyond our area of expertise, but we could imagine a process like the following. First, (extracranial) electrophysiological work (either EEG or MEG) could be used to identify the gross neuronal signatures in either the amplitude domain (ERP/ERF) or frequency domain (oscillations) that occur at the critical regions in the sentences of interest. Depending on the similarities and differences predicted by the different syntactic theories, and the different classes of neuronal populations that follow from the formalization of those theories in the previous step, the neuronal signatures (ERP/ERFs or oscillations) may be useful in eliminating competing computations from consideration. Recently, there have been exciting examples of work of this type in both syntax and semantics research. For example, Pylkkanen and colleagues have been searching for neurological correlates of fundamental semantic combinatory processes in the time-amplitude domain using MEG, with results pointing to increased activity in left anterior temporal lobe (LATL) and ventromedial prefrontal cortex (vmPFC) (e.g., Brennan & Pylkkanen, 2008; and Bemis & Pylkkanen, 2011; and many others). As another example, Bastiaansen and colleagues have been searching for neurological correlates of both syntactic and semantic combinatory processes in the time-frequency domain using EEG, with results pointing to the gamma frequency band (>30Hz) for semantic processes and the lower beta frequency band (13-18Hz) for syntactic processes (e.g., Bastiaansen et al., 2002; Bastiaansen et al., 2010; for a review see Bastiaansen et al., 2012).

Once electrophysiological correlates have been identified, localization studies, either with MEG (if the orientation of the generators is appropriate) or concurrent EEG and fMRI, could be used to identify cortical areas associated with the neuronal activity of interest. There is a large and ever-growing literature on localization in language processing, and other chapters in this volume provide enlightening reviews of that literature. However, we would like to point to Pallier et al., 2011 as an example of localization work that shares the same spirit as the program advocated here. Pallier et al. searched for brain areas that respond to the size of the syntactic constituent being processed (from 1 to 12 words), in essence using the number of syntactic computations deployed as a measure of complexity, and found activity in a number of regions, including left inferior frontal gyrus (LIFG), left anterior superior temporal sulcus (LaPSTS), and left posterior superior temporal sulcus (LpSTS). Finally, when suitable location and electrophysiological hypotheses are established, intracranial recordings (ECoG) could be used to identify the single unit information necessary to begin to identify the specific neuronal computation and observe its implementation.

We admit that the brief sketch above of the collaboration suggested by the computational view is based on our incomplete understanding of the various fields that would be part of the collaboration. We also admit that the space of possible neuronal computations is likely much larger than the space of extant structure-building operations, making the search for the actual neuronal computations that much more difficult. But it seems to us that the size of the hypothesis space is irrelevant to the question of how to move the fields of syntax and neuroscience forward (and together). This is either the right hypothesis space to be searching, or it isn't. It seems to us

that multiple domains of cognition are converging on both the need for identifying neuronal computations, and the plausibility of conducting such a search in the 21st century (e.g., Carandini, 2012; Poeppel et al., 2012). We believe that the wider field of syntax is ready to join the search that researchers such as Bastiaansen, Dehaene, Pallier, Pylkkanen, and colleauges have begun.

6. Challenges to this research program

Beyond the obvious challenge of engaging in the interdisciplinary work laid out in section 3 above, there are numerous smaller challenges that will need to be addressed for the collaboration to be successful. In this section we will discuss five, in some cases in an attempt to dispel the challenge, and in others simply to raise the issue for future work.

One obvious challenge is the concern from some cognitive scientists that syntactic theories are not built upon solid empirical foundations. This concern has been expressed since the earliest days of syntactic theorizing (e.g. Hill, 1961), and with several high profile publications recently (e.g., Gibson & Fedorenko, 2010; Gibson & Fedorenko, 2013). This concern is driven by the idea that the typical data collection methods are too informal to provide reliable data, therefore the theories built on that data are themselves unreliable. The persistence of this concern speaks to a fundamental failure on the part of syntacticians to make the argument either that the data type that they are collecting (acceptability judgments) are robust enough that the informality of the collection methods have no impact, or that there are unreported safeguards in the informal methods to prevent the kind of unreliability that they are concerned about (see Marantz, 2005; Phillips, 2009 for discussions of these issues). Sprouse & Almeida, 2012 and Sprouse et al., 2013 have begun to address this concern directly by exhaustively re-testing all of the phenomena in a popular Minimalist textbook, and re-testing a large random sample of phenomena from a popular syntax journal, using traditional experimental psychology methods. These re-tests have replicated 98% and 95% of the phenomena respectively, suggesting that the informal methods used in syntax indeed have the reliability that syntacticians claim. Given recent concerns about replicability inside of some areas of psychology, it is heartening to see that large-scale replications inside of syntax yield potential error rates at or below the conventional Type I error rate of 5%.

Despite the substantial evidence that the acceptability judgments that form the basis of syntactic theory are reliable, one could imagine potential collaborators being concerned that a theory built on offline data (like acceptability judgments) would be irrelevant for a theory built on real-time language processing data (like the electrophysiological data required by the research program proposed here). We agree that this could be a reasonable concern *a priori*. However, there is also a growing body of research in the sentence processing literature demonstrating that real-time sentence processing behavior respects grammatical conditions on well-formedness. For example, several studies have shown that complex constraints on the formation of non-local dependencies (called island constraints in the syntax literature) are respected by the parsing mechanism that form these dependencies in real-time (e.g., Stowe, 1986; Traxler & Pickering, 1996). In addition, several studies have demonstrated that these same processing mechanisms respect the sophisticated *exceptions* to these constraints postulated by syntactic theories (e.g., Phillips, 2006; Wagers & Phillips, 2009). Similarly, several studies have demonstrated that complex constraints on the dependencies that give pronouns their referents (called binding constraints in the syntax literature) are also respected by real-time referential processing

mechanisms (e.g., Sturt, 2003; Van Gompel & Liversedge, 2003; Kazanina et al., 2007). Several recent studies also show these effects to be the result of grammatical constraints and not the consequences of non-grammatical processing mechanisms (e.g., Sprouse et al., 2012; Dillon & Hornstein, 2013; Kush et al., 2013; Yoshida et al., 2013). In sum, there is a growing body of convincing evidence that syntactic theories capture structure-building properties that are relevant for real-time sentence processing, despite having initially been empirically based on offline data.

A third potential challenge for the computational view of syntax is that not every syntactician agrees that syntactic theories should serve as a theory of cognitive structure building computations. The potential for a logical distinction between theories of syntax and theories of cognitive structure-building is clearest in examples of non-mentalistic, or Platonic, linguistic theories, which seek to study the mathematical properties of language without making any claims about how those properties are instantiated in a brain. Even within GS, which is mentalistic, it is not uncommon to hear theories of syntax described as theories of knowledge (or competence) and not theories of use (or performance). The computational view of syntax goes beyond simple knowledge description. The computational view sees syntactic theories as making substantive claims about how syntactic structure building is instantiated in the human brain. It may be the case that there is a one-to-many relationship between syntactic theories and neuronal structure-building computations, but the relationship is there (see e.g., Lewis & Phillips, 2013 for a deeper discussion of this challenge).

A final challenge to the computational view of syntax is the problem of isolating structure-building computations from other sentence processing computations in real-time processing data. Real-time language processing data is going to contain signals from both structure building computations and all of the non-structure-building computations that syntactic theory abstracts away from (parsing strategies, resource allocation, task specific strategies in the sense of Rogalsky & Hickok, 2011, etc). This means that the actual construction of neurophysiological experiments discussed in section 3 will require quite a bit ingenuity to isolate the structure-building computations, especially given the high-dimensionality of neural data, and the likelihood of spurious correlations. And even assuming that *logically* isolating a computation of interest is possible in the experimental stimuli, *physically* isolating a neuronal computation in human neural systems is probably orders of magnitude more difficult. To our knowledge there are no existing neuronal computations that can be used as guide (a Rosetta stone of sorts) to mark the beginning or end of a computation being physically performed. We assume that as more and more computations are investigated, combining them in novel ways will eventually allow the physical boundaries of computations to be mapped, but this is currently a promissory note. In short, the narrow focus of syntactic theories on structure building computations is in some ways a blessing, as it provides a hypothesis space for a problem that is potentially tractable, but it is also a curse, because the computations left out of that hypothesis space may either be confounds, or necessary additions to solve the physical localization problem.

7. Conclusion

We believe that modern syntactic theory is well-suited to serve as a cognitive theory of syntactic structure-building computations, and that the time is right for a large-scale collaboration between syntacticians, mathematical linguists and psycholinguists, and theoretical and experimental neuroscientists to identify the neuronal instantiations of those computations. Such a research program will be a collaborative project of unprecedented scope, and will face numerous

theoretical and technological challenges, but there has never been a better time in the histories of cognitive science, linguistics, and neuroscience to try.

References

Bastiaansen, M. C. M., Van Berkum, J. J. A., & Hagoort, P. (2002). Syntactic processing modulates the θ rhythm of the human EEG. *NeuroImage, 17*, 1479-1492.

Bastiaansen, M. C. M., Magyari, L., & Hagoort, P. (2010). Syntactic unification operations are reflected in oscillatory dynamics during on-line sentence comprehension. *Journal of Cognitive Neuroscience, 22*, 1333-1347.

Bastiaansen, M. C. M., Mazaheri, A., & Jensen, O. (2012). Beyond ERPs: Oscillatory neuronal dynamics. In S. J. Luck, & E. S. Kappenman (Eds.), *The Oxford handbook of event-related potential components* (pp. 31-50). New York, NY: Oxford University Press

Bemis, D. K., & Pylkkänen, L. (2011). Simple Composition: An MEG investigation into the comprehension of minimal linguistic phrases. *Journal of Neuroscience, 31(8),* 2801-2814.

Berwick, R. C., & Weinberg, A. S. (1984). *The grammatical basis of linguistic performance*. Cambridge, MA: MIT Press.

Brennan, J. & Pylkkänen, L. (2008). Processing Events: Behavioral and Neuromagnetic Correlates of Aspectual Coercion. *Brain and Language, 106*, 132-143.

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

Carandini, M. (2012). From circuits to behavior: a bridge too far? *Nature Neuroscience, 15(4),* 507-509.

Chesi, C. (2012). *Competence and Computation: Toward a processing friendly minimalist grammar.* Padova: Unipress.

Chomsky, N. (2004). Beyond Explanatory Adequacy. In A. Belletti (Ed). *The Cartography of Syntactic Structure Vol 3: Structures and Beyond* (pp.104-131). Oxford: Oxford University.

Collins, C., & Stabler E. (2012). A formalization of minimalist syntax. http://ling.auf.net/lingbuzz/001691

den Dikken, Marcel (Ed). (2013). *The Cambridge Handbook of Generative Syntax*. Cambridge University Press.

Dillon, B., & Hornstein, N. (2013). On the structural nature of island constraints. In J. Sprouse.

& N. Hornstein (Eds.), *Experimental Syntax and Island Effects* (pp.208-222). Cambridge University Press.

Fodor, J., Bever, T., & Garrett. M. (1974). *The Psychology of Language*. New York: McGraw Hill.

Ferreira, F. (2005). Psycholinguistics, formal grammars, and cognitive science. *The Linguistic Review, 22,* 365-380.

Frank, R. (2002). *Phrase Structure Composition and Syntactic Dependencies*. Cambridge, MA: MIT Press.

Frank, R. (2013). Tree Adjoining Grammar. In M. den Dikken (Ed), *The Cambridge Handbook of Generative Syntax* (pp.226-261). Cambridge University Press.

Gibson, E. & Fedorenko, E. (2010). Weak quantitative standards in linguistics research. *Trends in Cognitive Sciences, 14,* 233-234.

Gibson, E. & Fedorenko, E. (2013). The need for quantitative methods in syntax and semantics research. *Language and Cognitive Processes, 28,* 88-124.

Hale, John. (2003). *Grammar, Uncertainty, and Sentence Processing* (Unpublished doctoral dissertation). Johns Hopkins University, Baltimore, MD.

Hill, A. A. (1961). Grammaticality. *Word, 17*, 1-10.

Hornstein, N. (2009). *A theory of syntax: Minimal operations and universal grammar*. Cambridge University Press.

Joshi, A. K., Vijay-Shanker, K., & Weir, D. J. (1991). The convergence of mildly context-sensitive grammar formalisms. In P. Sells, S. Sheiber, & T. Wasow (Eds.), *Foundational Issues in Natural Language Processing* (pp.31-81). Cambridge, MA: MIT Press.

Kazanina, N., Lau, E. F., Lieberman, M., Yoshida, M., & Phillips, C. (2007). The effect of syntactic constraints on the processing of backwards anaphora. *Journal of Memory and Language, 56*, 384-409.

Kobele, G. M., Gerth, S., & Hale, J.T. (2013). Memory Resource Allocation in Top-Down Minimalist Parsing. In G. Morrill, & M.-J. Nederhof (Eds.), *FG 2012/2013*, *volume 8036 of Lecture Notes in Computer Science* (pp.32-51). Heidelberg: Springer.

Kush, D., Omaki, A. & Hornstein, N. (2013). Microvariation in Islands? In J. Sprouse, & N. Hornstein (Eds.), *Experimental Syntax and Island Effects* (pp.239-264). Cambridge University Press.

Lewis, S. & Phillips, C. (2013). Aligning grammatical theories and language processing models.

Journal of Psycholinguistic Research.

Marantz, A. (2005). Generative linguistics within the cognitive neuroscience of language. *The linguistic Review, 22,* 429-445.

Marcus, M. (1980). *A Theory of Syntactic Recognition for Natural Language*. Cambridge. MA: MIT Press.

Marr, D. 1982. *Vision*. Cambridge, MA: The MIT Press.

Michaelis, J. (1998). Derivational Minimalism is mildly context sensitive. In M. Moortgat (Ed.), *Logical Aspects of Computational Linguistics, Lecture Notes in Artificial Intelligence Volume 1* (pp.179-198). Heidelberg: Springer.

Pallier, C., Devauchelle, A.-D., & Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. *Proceedings of the National Academy of Sciences*, 108(6), 2522—2527.

Pearl, L., & Sprouse, J. (2013). Syntactic islands and learning biases: Combining experimental syntax and computational modeling to investigate the language acquisition problem. *Language Acquisition, 20*, 23-68.

Phillips, C. (1996). *Order and Structure* (Unpublished doctoral dissertation). Massachusetts Institute of Technology, Cambridge, MA.

Phillips, C. (2003). Linear order and constituency. *Linguistic Inquiry, 34,* 37-90.

Phillips, C. (2006). The real-time status of island phenomena. *Language, 82*, 795-823.

Phillips, C. (2009). Should we impeach armchair linguists? In S. Iwasaki, H Hoji, P. Clancy, & S.-O. Sohn (Eds.), *Proceedings from Japanese/Korean Linguistics* 17. Stanford, CA: CSLI Publications.

Poeppel, D. (2012). The maps problem and the mapping problem: Two challenges for a cognitive neuroscience of speech and language. *Cognitive Neuropsychology, 29,* 34-55.

Poeppel, D. & Embick, D. (2005). The relation between linguistics and neuroscience. In A. Cutler (Ed.), *Twenty-First Century Psycholinguistics: Four Cornerstones* (pp.103-118). Mahwah, NJ: Lawrence Erlbaum Associates.

Poeppel, D., Emmorey, K., Hickok, G., & Pylkkänen, L. (2012). Towards a new neurobiology of language. *Journal of Neuroscience, 32(41),* 14125-14131.

Rogalksy, C., & Hickok, G. (2011). The role of Broca's area in sentence comprehension. *Journal of Cognitive Neuroscience, 23*, 1664-1680.

Sprouse, J., & Almeida. D. (2012). Assessing the reliability of textbook data in syntax: Adger's Core Syntax. *Journal of Linguistics, 48,* 609-652.

Sprouse, J., Wagers, M., & Phillips, C. (2012). A test of the relation between working memory capacity and syntactic island effects. *Language, 88(1),* 82-123.

Sprouse, J., Schütze, C. T., & Almeida, D. (2013). A comparison of informal and formal acceptability judgments using a random sample from Linguistic Inquiry 2001-2010. *Lingua, 134,* 219-248.

Stabler, E. (1997). Derivational minimalism. In C. Retoré (Ed.), *Logical Aspects of Computational Linguistics* (pp.68-95)*.* New York: Springer.

Stabler, E. (2011). Top-down recognizers for MCFGs and MGs. In *Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics* (CMCL '11) (pp.39-48). Stroudsburg, PA: Association for Computational Linguistics.

Stabler, E. (2013). Two models of minimalist, incremental syntactic analysis. *Topics in Cognitive Science, 5(3),* 611-633.

Stowe, L. A. (1986). Evidence for on-line gap-location. *Language and Cognitive Processes, 1,* 227–45.

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

Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language, 35,* 454-475.

Van Gompel, R. P. G., & Liversedge, S. P. (2003). The influence of morphological information on cataphoric pronoun assignment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 128-139.

Wasow, T. & J. Arnold. (2005). Intuitions in linguistic argumentation. *Lingua, 115,* 1481-1496.

Yoshida, M., Kazanina, N., Pablos, L., & Sturt, P. (2013). On the origin of islands. *Language and Cognitive Processes.*