CellPress

Feature Review

Structures, Not Strings: Linguistics as Part of the Cognitive Sciences

Martin B.H. Everaert,[1] Marinus A.C. Huybregts,[1] Noam Chomsky,[2] Robert C. Berwick,[3] and Johan J. Bolhuis[4,5,*]

There are many questions one can ask about human language: its distinctive properties, neural representation, characteristic uses including use in communicative contexts, variation, growth in the individual, and origin. Every such inquiry is guided by some concept of what 'language' is. Sharpening the core question – what is language? – and paying close attention to the basic property of the language faculty and its biological foundations makes it clear how linguistics is firmly positioned within the cognitive sciences. Here we will show how recent developments in generative grammar, taking language as a computational cognitive mechanism seriously, allow us to address issues left unexplained in the increasingly popular surface-oriented approaches to language.

Trends

The computations of the mind rely on the structural organization of phrases but are blind to the linear organization of words that are articulated and perceived by input and output systems at the sensorimotor interface (speech/sign). The computational procedure that is universally adopted is computationally much more complex than an alternative that relies on linear order.

Linear order is not available to the systems of syntax and semantics. It is an ancillary feature of language, probably a reflex of properties of the sensorimotor system that requires it for externalization, and constrained by conditions imposed by sensorimotor modalities.

It follows that language is primarily an instrument for the expression of thought. Language is neither speech/sign (externalized expression) nor communication (one of its many possible uses).

Grammar from a Cognitive Science Perspective: Generative Grammar

Language is a structured and accessible product of the human mind. We choose to study language for this reason, as one possible way to gain understanding about the human mind. This particular choice – language as part of the mind, so cognitive science – arose as the result of the seminal discoveries by the mid-20th century regarding the mathematics of computation, which permitted a shift from the more conventional perspective of language as a cultural/social object of study. This new perspective regarding computation [1–4] enabled for the first time a clear formulation of what we should recognize as the most basic property of language: providing a discretely infinite array of hierarchically structured expressions that receive systematic interpretations at two interfaces, roughly, thought and sound [5–8]. We take **externalization** (see Glossary) at the sensory–motor level (for instance, speech) as an ancillary process, reflecting properties of the sensory modality, sign or speech. Therefore, communication, a particular use of externalized language, cannot be the primary function of language, a defining property of the language faculty, suggesting that a traditional conception of language as an instrument of thought might be more appropriate. At a minimum then, each language incorporates via its **syntax** computational procedures (Box 1) satisfying this basic property. As a result, every theory of a particular language constitutes by definition what is called a **generative grammar**: a description of the tacit knowledge of the speaker–hearer that underlies their actual production and perception (understanding) of speech. We take the property of structure dependence of grammatical rules to be central. We will illustrate the puzzling feature that the computational rules of language rely on the much more complex property of hierarchical structure rather than the much simpler surface property of linear order.

Viewing sentences as just linear word strings has long held a prominent place in areas of natural language processing such as speech recognition and machine translation. Warren Weaver

[1]Utrecht Institute of Linguistics, Utrecht University, 3512 JK Utrecht, The Netherlands
[2]Department of Linguistics and Philosophy, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[3]Department of Electrical Engineering and Computer Science and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[4]Cognitive Neurobiology and Helmholtz Institute, Departments of Psychology and Biology, Utrecht University, 3584 CH Utrecht, The Netherlands

CellPress

Box 1. Merge: The Basic Property of Language

Merge is a (dyadic) operation that takes two syntactic objects, call them X and Y, and constructs from them a single new syntactic object, call it Z. X,Y can be building blocks that are drawn from the lexicon or previously constructed objects. Put simply, Merge (X,Y) just forms the set containing X and Y. Neither X nor Y is modified in the course of the operation Merge.

If X and Y are merged there are only two logical possibilities. Either X and Y are distinct, and neither one is a term of the other, or else one of the two elements X or Y is a term of the other, where Z is a term of W if it is a subset of the other or the subset of a term of the other. We can call the former operation 'External Merge': two distinct objects are combined.

(i) Merge (*read*, *that book*) ⇒ {*read, that book*}

If alternatively X is a term of Y or vice versa and X and Y are merged, we call this 'Internal Merge'. So for example, we can (Internal) Merge *which book* and *John read which book*, yielding the following:

(ii) Merge (*which book*, *John read which book*) ⇒ {*which book, John read which book*}

In this case, the result of merging X and Y contains two copies of Y. Following further operations, this structure will surface as in (iii), under a constraint to externalize ('pronounce') only the structurally most prominent copy of *which book*:

(iii) (Guess) which book John read

This sentence may be understood as (iv):

(iv) (Guess) for which book x, John read the book x

Internal **merge** is a ubiquitous property of language, sometimes called displacement. Phrases are heard in one place but they are interpreted both there and somewhere else.

Human language generates a digitally infinite array of hierarchically structured expressions with systematic interpretations at the interfaces with a sensory–motor (sound/sign) and a conceptual–intentional (meaning) system. Thus, language comprises a system to generate hierarchical syntax along with asymmetric mappings to the interfaces, a basic mapping to the conceptual–intentional interface and an ancillary mapping to the sensory–motor interface. Merge is the basic operation underpinning the human capacity for language, UG, connecting these interface systems. Characterizing UG in terms of recursive merge is just a way of saying that whatever is going on in the brain neurologically can be properly understood in these terms.

[5]Department of Zoology and Sidney Sussex College, University of Cambridge, Cambridge, UK

*Correspondence: j.j.bolhuis@uu.nl (J.J. Bolhuis).

famously made the case for a string-based approach to machine translation as a type of code breaking, using statistical methods [9]. This position seems intuitively plausible because it parallels the familiar way foreign language travel guides are organized, with phrases in one language matched to corresponding phrases in another. The intuition is that simply pairing matching sentence strings that are selected on the basis of statistical likelihood suffices, and that accuracy does not require linguistic analysis, simply the compilation of a database of larger and longer sentence pairs along with more powerful computers for data storage and selection. Boosted by exactly this increased computing power along with innovative statistical work at IBM Research (the late Fred Jelinek and John Lafferty among many others) [10], this approach rapidly gained ascendancy in the late 1980s, gradually pushing out rule-based machine translation approaches. But this surface-oriented 'big data' approach is now all encompassing, not only in computational linguistics.

The focus on the non-hierarchical aspects of language is evident in the work of some typologists [11], and is at the basis of usage-based, constructionist linguistic theories [12,13]. These approaches focus on inductive mechanisms that explain the acquisition and use of 'low-level patterns', 'not predictable from general rules or principles', allowing us to 'create novel utterances based on [constructional] schemas'. [14]. Such approaches focus on words or word-like constructions, usage patterns, do not acknowledge the relevance of structure, and view acquisition essentially as statistical [15]. Introductions to psycholinguistics generally do not mention notions such as hierarchy, structure, or constituent.

A variety of evidence can be brought to bear on the study of language. This can include language use, acquisition, cognitive dissociations, other detailed neuroscience investigations, cross-language comparisons, and much else besides. All this follows from the well-confirmed assumption that the human capacity for language rests on shared biological properties. However, for the development of generative grammar one particular type of evidence has proved most useful: the ability of children to rapidly and effortlessly acquire the intricate principles and properties of the language of their environment. All normally developing children acquire most of the crucial elements of their language long before school age. By contrast, adults exhibit a very different developmental path when they attempt to acquire a second language [16]. Often they do not come close to the level of native speakers, even after a much longer time frame for learning. Most researchers would agree that the distinctive ontogenesis of child language arises from the interplay of several factors, including innate mechanisms, language-independent properties, and external experience. On our view, the ability of children to rapidly and effortlessly acquire the intricate principles and properties of their native language can best be explained by looking for innate, language-dedicated cognitive structures (collectively known as **Universal Grammar**) that guide learning.

Defined in this way, the study of language focuses on three questions:

(i) What constitutes knowledge of language? This amounts to understanding the nature of the computational system behind human language.

(ii) How is knowledge of language acquired? This amounts to unraveling the cognitive processes underlying primary language acquisition, so as to understand how primary language acquisition differs from secondary, subsequent language acquisition.

(iii) How is knowledge of language put to use? This amounts to studying the linguistic processes underlying language production, perception, and interpretation – under varying conditions, such as modality, social environment, and speech context – and the way in which language helps fulfill our communicative needs.

A commitment to some answer to (i) is a logical precondition for addressing (ii) and (iii). Inquiry into language acquisition and language use can proceed most effectively insofar as it is based on careful description and understanding of the system that has evolved. The study of language – we believe – has made sufficient progress answering question (i) to attempt to pursue answers to questions (ii) and (iii).

The Infinite Use of Finite Means

One feature of language that distinguishes it from all non-human communication systems we know of is its ability to yield an unbounded array of hierarchically structured expressions, permitting 'infinite use of finite means' [17]. To see how and why, we need to introduce the notion of **recursion**, which underlies this finite–infinite distinction. Much has been written about recursion from different perspectives. There is no need to repeat this here [18–20]. What is more important to understand is that recursion in its original context – based on the recursive function theory developed by Gödel, Church, and Turing [21–24] – served as the formal grounding for generative grammar and the solution to the finite–infinite puzzle. The picture of Turing machine computation provides a useful explanation for why this is so. In a Turing machine, the output of a function f on some input x is determined via stepwise computation from some previously defined value, by carrying forward or 'recursing' on the Turing machine's tape previously defined information. This enabled for the first time a precise, computational account of the notion of definition by induction (definition by recursion), with $f(x)$ defined by prior computations on some earlier input y, $f(y)$, $y < x$ – crucially so as to strongly generate arbitrarily complex structures [19].

Why is recursion important? As it is formulated above, recursion is important because it supplies part of an answer to the seemingly unbounded creativity of language, so central to linguistic

Glossary

Clitic: a syntactic element that cannot occur freely in syntax but is in need of a 'host'. A typical clitic will attach itself to a host, that is, a (fully inflected) word or phrase, for example, French *te* 'you' in *Je t'aime*.
Compositionality: a principle that constrains the relation between form and meaning by requiring that the meaning of a complex expression is built up from the meanings of its constituent expressions and the way they are combined. This principle plays an important role in formal semantic theories.
C(onstituent)-command: c-command is a binary relation between nodes in a tree structure that is defined as follows: node α c-commands node β iff (i) $\alpha \neq \beta$, (ii) α does not dominate β and β does not dominate α, and (iii) every γ that dominates α also dominates β.
Context-free language: a language (set of sentences) generated by a context-free grammar, namely, a grammar whose rules are all restricted to be in the form $X \rightarrow w$, where X is a single phrase name (such as VP or NP), and w is some string of phrase names or words.
Externalization: the mapping from internal linguistic representations to their ordered output form, either spoken or manually gestured.
Gap: any node in the phrase structure that has semantic content but is without phonological content, for example, 'children should be seen and – not heard'.
Generative grammar: generative grammar is a research program that includes different competing frameworks, and takes linguistics as a science whose goal it is to try to provide a precise (explicit and formal) model of a cognitively embedded computational system of human language, and to explain how it is acquired.
Merge: in human language, the computational operation that constructs new syntactic objects Z (e.g., 'ate the apples') from already constructed syntactic objects X ('ate'), Y ('the apples'), without changing X or Y, or adding to Z, that is, set formation.
Negative concord items: negative polarity items with a more restricted distribution. They can only be licensed by clausemate sentential negation and can sometimes express

theorizing since the mid-20th century. This essential property of language provides a means for expressing indefinitely many thoughts and for reacting appropriately in an indefinite range of new situations [25].

This approach to the unbounded character of language may be contrasted with the conventional empiricist position that assumes inductive generalizations from observable distributional regularities to be sufficient for learning and use of language. For American structuralism, a standard concept was that of Leonard Bloomfield, the leading theoretician, for whom language is 'an array of habits to respond to situations with conventional speech sounds and to respond to these sounds with actions' [26]. Another leading figure, Charles Hockett, attributed language use to 'analogy', and this meant that we construct and understand novel sentences on the basis of those we have constructed and understood before. For Hockett, 'similarity' played the central role in language learning, production, and use [25]. This line of thought is still at the forefront of many modern-day stochastic learning algorithms, generalized learning procedures, and natural language **parsers**. The crucial question, however, is whether a notion of analogy can be properly defined so as to adequately explain how children acquire language (Box 2).

Syntax: What You See is Not What You Get

Given the view set out above, Aristotle's *dictum* that 'language is sound with meaning' could arguably be reformulated as 'language is meaning with sound', since the mappings of expressions to the two interfaces are asymmetric, as noted above. The mapping to the systems of inference, interpretation, and the like we assume to be simple, principled, and close to invariant, following structural principles unexceptionally and possibly in harmony with the methodological principle of **compositionality** [27]. The mapping to the sensory modalities (speech, sign) is more complex, clearly subject to parameterization and is more likely to have exceptions [28]. Linking a cognitive system to one or other of the sensory modalities amounts to the difficult problem of relating two different categories of systems with different properties and different evolutionary histories. But the syntactic operations that map linguistic objects to the semantic interface do not use the simple properties of sequential string order, that is, linear precedence. Instead, they rely exclusively on the hierarchical structural position of phrases, that is, hierarchical structural distance and hierarchical structural relations (Box 3). In the following we illustrate the reliance of language on hierarchical structure rather than linear precedence in all areas of language – by providing examples from semantics, syntax, morphology, and **phonology**.

The Syntax of Semantics

A simple textbook illustration of the reliance of language on hierarchical structure is provided by syntactic properties of **negative polarity items** (NPIs) such as the English word *anybody* or **negative concord items** such as the Japanese word *nani-mo* ('anything'). These items require an overt negative element such as *not* or *nakat*. If we omit the negative items, the sentences become ill-formed ('*'), cf. (1a,b) and (2a,b):

(1) a. The book I bought did *not* appeal to *anybody*.
b. *The book I bought appealed to *anybody*.

(2) a. Taroo-wa *nani-mo* tabe-*nakat*-ta.
Taroo-TOP what-MO eat-NEG-PST
'Taro didn't eat anything'
b. *Taroo-wa *nani-mo* tabe-ta.
Taroo-TOP what-MO eat-PST

From (1a,b) one might also conclude, wrongly, that the English NPI *anybody* must appear in the sentence sequentially after *not*. This conclusion is immediately refuted by the Japanese example

negation on their own as in fragment answers.
Negative polarity items: a word or word group that is restricted to negative contexts – needing the scope of a negation (or more precisely a monotone decreasing word/phrase).
Parasitic gap (PG): is a gap (a null variable) that depends on the existence of another gap RG, sharing with it the same operator that locally binds both variables. PG must conform to a binding condition asserting that PG cannot be c-commanded by RG.
Parsers: a natural language parser is a program for analyzing a string of words (sentence) and assigning it syntactic structure in accordance with the rules of grammar. Ideally, the relation between basic parsing operations and basic operations of grammar approximates the identity function. Probabilistic parsers use statistical information to provide the most likely grammatical analyses of new sentences.
Phonology: the study of the abstract sound patterns of a particular language, usually according to some system of rules.
Phrase structure rules: rewrite rules that generate phrase structure. These have the general form of (i), where X is the name of the phrase and Y Z W defines its structure. Y, Z, and W are either phrases, and therefore must themselves occur to the left of the arrow in other rules of this type, or non-phrasal (terminal) categories (such as noun, verb, or determiner). (i) X → YZW
Prosody: the description of rhythm, loudness, pitch, and tempo. It is often used as a synonym for suprasegmentals, although its meaning is narrower: it only refers to the features mentioned above.
Recursion: a property of a finitely specified generative procedure that allows an operation to reapply to the result of an earlier application of the same operation. Since natural language is unbounded, at least one combinatorial operation must be applicable to its own output (via recursion or some logical equivalent). And given such an operation, any derivational sequence for a generable string will determine a hierarchical structure, thus providing one notion of structure generation ('strong generation') distinct from the weakly generated string.

Box 2. Simple Rules

Consider the following noun phrases (i), and their description in terms of context-free **phrase structure rules** (ii), and accompanying figures (Figure I):

(i) a man (ii) a man on the moon

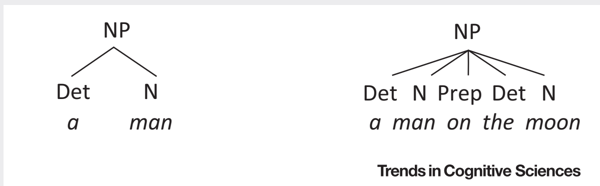


Figure I. Structures for (i) and (ii) on the basis of Grammar G.

(G) a. N(oun) P(hrase) → Det(erminer) N(oun)
b. NP → Det N Prep Det N

Our 'grammar' in (Ga,b) (in which '→' means 'consists of') would allow one to create an enormous variety of noun phrases given a vocabulary of determiners, nouns, and prepositions. However, observing that (iii) is also possible, we would have to add a rule (Gc) to our grammar:

(iii) *a girlfriend of the man from the team*

(G) c. NP → Det N Prep Det N Prep Det N

But now we are missing a linguistically significant generalization: every noun phrase can have a prepositional phrase tacked on the end, which is accounted for by replacing grammar G by the following simpler set of rules:

(G′) a. NP → Det N (PP) (noun phrases consist of a determiner and a noun and may be followed by a prepositional phrase)
b. PP → Prep NP (prepositional phrases consist of a preposition followed by a noun phrase)

(G′) is a simpler grammar. But note that (G′) represents (part of) a grammar yielding a 'discrete infinity' of possible phrases, allowing us to generate ever longer noun phrases taking prepositional phrases. We could only circumvent this unboundedness by returning to a grammar that explicitly lists the configurations we actually observe, such as (G). But such a list would be arbitrarily limited and would fail to characterize the linguistic knowledge we know native speakers have. This recursive generation of potential structures ('linguistic competence') should not be incorrectly equated with real-time production or parsing of actual utterances ('linguistic performance'). Note that this distinction is no different from the rules for addition or multiplication. The rules are finite, but the number of addition or multiplication problems we can solve is unbounded (given enough internal or external resources of time and memory).

Grammar (G′) also reflects the fact that phrases are not simple concatenations of words, but constitute structured objects. (G′), contrary to (G), therefore correctly reflects properties of constituency as illustrated in (v):

(v) He gave me [a book [about [the pope]]]
It is [the pope]$_x$, he gave me [a book [about X]]
It is [about the pope]$_x$, he gave me [a book X].

Selectional properties: the semantic restrictions that a word imposes on the syntactic context in which it occurs: a verb such as *eat* requires that its subject refers to an animate entity and its object to something edible.

Syntax: the rules for arranging items (sounds, words, word parts, phrases) into their possible permissible combinations in a language.

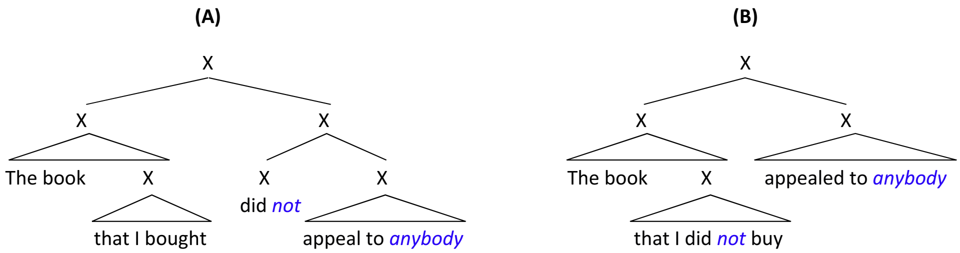
Universal Grammar (UG): is the theory of the genetic component of the faculty of language, the human capacity for language that makes it possible for human infants to acquire and use any internalized language without instruction and on the basis of limited, fragmentary, and often poor linguistic input. UG is the general theory of internalized languages and determines the class of generative procedures that satisfy the basic property, besides the atomic elements that enter into these computations.

in (2a), where *nakat* follows the negative concord item *nani-mo*. Example (3) also shows that the operative constraint cannot be linear order since (3) is ill-formed despite the fact that *not* appears sequentially before *anybody*, just as it does in the well-formed example (1a).

(3) *The book I did *not* buy appealed to *anybody*.

What is the correct constraint governing this pattern? It depends on hierarchical structure and not on sequential or linear structure [29,30].

Consider Figure 1A, which shows the hierarchical structure corresponding to example (1a): the hierarchical structure dominating *not* also immediately dominates the hierarchical structure



Trends in Cognitive Sciences

Figure 1. Negative Polarity. (A) Negative polarity licensed: negative element c-commands negative polarity item. (B) Negative polarity not licensed. Negative element does not c-command negative polarity item.

containing *anybody*. (This structural configuration is called **c(onstituent)-command** in the linguistics literature [31].) When the relationship between *not* and *anybody* adheres to this structural configuration, the sentence is well-formed.

In sentence (3), by contrast, *not* sequentially precedes *anybody*, but the triangle dominating *not* in Figure 1B fails to also dominate the structure containing *anybody*. Consequently, the sentence is not well-formed.

The reader may confirm that the same hierarchical constraint dictates whether the examples in (4–5) are well-formed or not, where we have depicted the hierarchical sentence structure in terms of conventional labeled brackets:

(4) $[_{S1}$ $[_{NP}$ The book $[_{S2}$ I bought$]_{S2}]_{NP}$ did *not* $[_{VP}$ appeal to *anyone*$]_{VP}]_{S1}$

(5) *$[_{S1}$ $[_{NP}$ The book $[_{S2}$ I did *not* buy$]_{S2}]_{NP}$ $[_{VP}$ appealed to *anyone*$]_{VP}]_{S1}$

Only in example (4) does the hierarchical structure containing *not* (corresponding to the sentence *The book I bought did not appeal to anyone*) also immediately dominate the NPI *anybody.* In (5) *not* is embedded in at least one phrase that does not also include the NPI. So (4) is well-formed and (5) is not, exactly the predicted result if the hierarchical constraint is correct.

Even more strikingly, the same constraint appears to hold across languages and in many other syntactic contexts. Note that Japanese-type languages follow this same pattern if we assume that these languages have hierarchically structured expressions similar to English, but linearize these structures somewhat differently – verbs come at the end of sentences, and so forth [32]. Linear order, then, should not enter into the syntactic–semantic computation [33,34]. This is rather independent of possible effects of linearly intervening negation that modulate acceptability in NPI contexts [35].

The Syntax of Syntax

Observe an example as in (6):

(6) Guess which politician your interest in clearly appeals to.

The construction in (6) is remarkable because a single *wh*-phrase is associated both with the prepositional object **gap** of *to* and with the prepositional object gap of *in*, as in (7a). We talk about 'gaps' because a possible response to (6) might be as in (7b):

(7) a. Guess which politician your interest in GAP clearly appeals to GAP.
b. response to (7a): Your interest in Donald Trump clearly appeals to Donald Trump

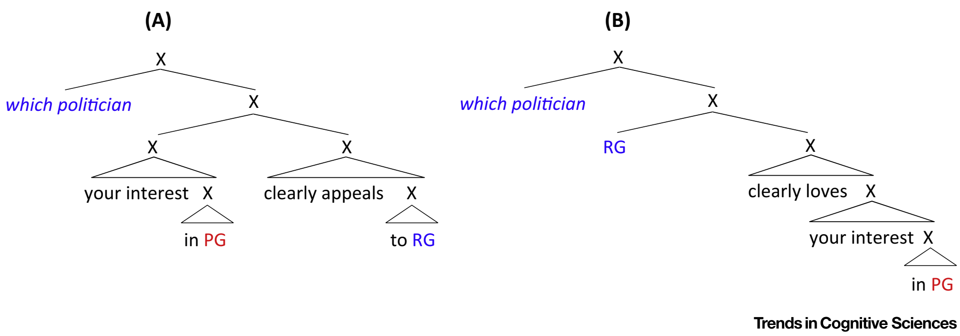


Figure 2. Parasitic Gap. (A) Well-formed parasitic gap construction: *which politician* c-commands both real gap (RG) and parasitic gap (PG). RG does not c-command PG (and PG does not c-command RG either). (B) Ill-formed parasitic gap construction: *which politician* c-commands both real gap (RG) and parasitic gap (PG). RG c-commands PG.

The construction is called '**parasitic gap**' (PG) because the 'first' gap in the nominal expression, the subject, is parasitic on the 'real gap' (RG) in the verbal expression: (8b) is well-formed and occurs independently of (6), while (8a) is ill-formed and does not occur independently of (6).

(8) a. *Guess which politician $[_S$ $[_{NP}$ your interest in PG$]_{NP}$ clearly appeals to Jane$]_S$

b. Guess which politician $[_S$ $[_{NP}$ your interest in Jane$]_{NP}$ clearly appeals to RG$]_S$

In other words, the gap in (8a) cannot exist unless it co-occurs with the independently licensed gap of (8b), resulting in (6/7a). Parasitic gap constructions are rarely attested, virtually absent from the empirical record. Nevertheless, language learners attain robust knowledge of parasitic gap constructions. Although such constructions had been observed to exist long ago (J.R. Ross, PhD thesis, Massachusetts Institute of Technology, 1967; [36]), the properties of parasitic gaps were predicted to exist on theoretical grounds [37], and were (re)discovered as a result of precise generative analysis [38–42]. Applying analytical or statistical tools to huge corpora of data in an effort to elucidate the intriguing properties of parasitic gaps will not work.

However, not every co-occurrence of RG and PG yields a grammatical result:

(9) a. *Guess which politician clearly loves your interest in.

b. Guess which politician $[_S$ RG clearly loves $[_{NP}$ your interest in PG$]_{NP}]_S$

Hierarchical structure and structure dependence of rules are basic factors in explaining parasitic gaps and the asymmetry between (6) and (9), a subject–object asymmetry. The PG is parasitic on an independently occurring RG but may not be linked to a RG that is in a structurally higher position. This is illustrated in Figure 2A and 2B for (6) and (9), respectively.

In Figure 2A *who* is structurally higher than both the RG and the PG, but the PG, being embedded in the noun phrase subject, is not structurally higher than the RG. In Figure 2B, by contrast, the RG in the subject position is in a hierarchically higher position than the PG in lower prepositional object position. The contrasting filler-gap cases of (6) and (9) cannot be characterized by their linear properties. It would be incorrect to state that PGs must precede their licensing RGs, as shown by (10):

(10) Who did you [[talk to RG] without recognizing PG]?

Crucially, the RG licensing the PG is not in a structurally higher position in (10): the verb phrase dominating the RG does not dominate the adverbial phrase containing the PG. Why this restriction precisely holds we leave undiscussed here, but is discussed at length in the literature on parasitic gaps.

The same concepts apply across empirical domains in language. For example, adopting these concepts enables us to explain certain unexpected and surprising phenomena in Dutch. Compare (11a) to its counterpart (11b) with a phonologically weak pronoun (**clitic**).

(11)	a.	Ik ben speciaal voor *het klimaat*	naar *de Provence* toe	gereden.
		I am especially for the climate	to the Provence	driven.
		'I drove to Provence especially for the climate'.		
	b.	Ik ben *er* speciaal voor	naar toe	vertrokken.
		I am it especially for	to	driven.
		'I drove there especially for it'.		

The clitic *er* 'it/there' is linked to two gaps, the NP complements of the preposition *voor* and the complex preposition/postposition *naar*. . .*toe*. A single clitic position *er* simultaneously binds two structural positions that have different **selectional properties** but meet the structural conditions of standard parasitic gaps. This old puzzle of structuralist and generative grammar, sometimes referred to as 'Bech's Problem' [43–45], may now turn out to be explainable as a special case of a parasitic gap construction (if a language-specific property of Dutch morphology is added to the equation). The simple lesson to take home is that a few assumptions about the structure of language suffice to give a unified account of superficially unrelated and disparate phenomena that are left unexplained in models that are restricted to concepts such as linear precedence. In fact, proposals that restrict themselves to just linear order are both too weak (incorrectly permitting ill-formed PGs) and too strong (incorrectly ruling out well-formed PGs). They are therefore neither sufficient nor necessary to deal with natural language and should be dismissed.

The Syntax of Morphology

Sound and meaning in morphology can also be shown to be dependent on hierarchical structure. But there is an asymmetry. As discussed above, computational rules of language invariably keep to the complex property of hierarchical structure and never use the far simpler option of linear order. But, of course, linear order must be available for externalization since the sensory–motor system requires that whatever structure is generated must pass through some type of filter that makes it come out in linear order.

For further evidence of the relevance of hierarchical structure, consider the compounds in (12) and their respective structures in Figure 3A,B.

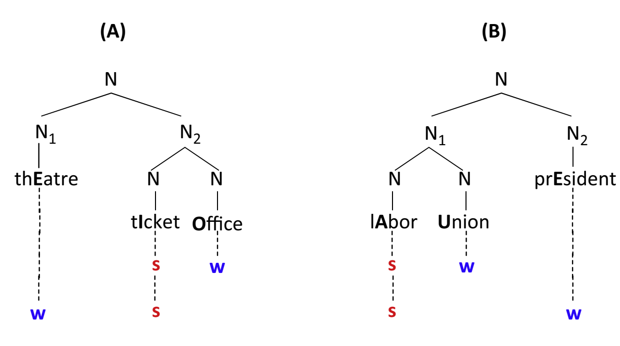


Figure 3. Prosodic Prominence. Right-branching (A) and left-branching (B) nominal compound structures. Bold capital letters in initial syllables of each word denote position of primary word stress. Compound stress rule is applied successively, first to the lower, embedded compound, then to the next higher compound containing it. The syllable consistently assigned strong prosodic prominence ('s') on each application of the rule carries compound stress.

(12) a. lábor union president, kítchen towel rack (rack for kitchen towels)
b. theatre tícket office, kitchen tówel rack (towel rack in the kitchen)

The correct interpretations of these compounds, both at the sensory–motor interface (namely, different **prosodies**) and at the semantic interface (namely, different meanings) follow directly from applying the relevant rules to their radically different hierarchical structures. Here we will limit our illustration to prosodic prominence. The rule describing prosodic prominence is given in (13):

(13) Assign prosodic prominence to the first noun N_1 of a compound $[_N N_1 N_2]$ if and only if the second noun N_2 does not branch.
(More precisely: In a compound N, $[_N N_1 N_2]$, assign prosodic prominence ('s') to the primary stressed syllable of N_1 if N_2 does not branch.)

The recursive application of this structure-dependent rule, based on [46–48], to the different hierarchically structured expressions in Figure 3A and 3B yields the correct prosodic prominence patterns in each case. If none of the parts of a compound branches, as in *ticket office* or *labor union*, prosodic prominence ('s') is assigned by (13) to the left-hand noun N_1 (*tícket*, *lábor*) because its right-hand noun N_2 (*óffice*, *únion*) does not branch. As a corollary effect, the N_2 becomes prosodically weak ('w'). The noun *theatre tícket office* (Figure 3A) is a compound N consisting of a simple noun N_1 (*théatre*) and a noun N_2 (*tícket office*), which is itself a compound noun with prosodic prominence already assigned by previous application of (13), as just discussed. It is a right-branching hierarchical structure. Therefore, the N_1 cannot be prosodically prominent because N_2 branches. Consequently, prominence must be assigned to N_2, the inner compound noun. The repeated application of (13) yields the correct result. Analogously, the compound noun *lábor union president* has a left-branching hierarchical structure (Figure 3B). Prosodic prominence, again, falls on the left-hand noun of the inner compound, which, in this case, is the left-hand member of the full compound structure. The reason is that the right-hand member is non-branching and must therefore be prosodically weak. A derivation working from the bottom up guarantees a correct result.

If prosodic prominence would have been constrained by conditions on linear structure we would have expected stress to fall uniformly and rigidly on a fixed linear position in the string. But language does not work that way. Patterns of prosodical prominence are neither random nor rigid but determinate and they universally depend on a more complex hierarchical structure of compounds such as *lábor union president election*, *evening compúter class teacher*, *community centre búilding council*, which have each a different stress pattern that is sensitive to structure and is assigned in accordance with (13). Depending on specific hierarchical structure, stress falls on a word-stressed vowel of the first, second, or penultimate noun but never on the final noun. These results would be totally unexpected if we just assume conditions on linear properties of language.

The Syntax of Phonology

In spoken English certain sequences of words can be contracted, for example, *don't* vs *do not*. Similarly, *want to* can be contracted to *wanna*:

(14) a. I want to persuade the biologist vs c. I wanna persuade the biologist.
b. Who do you want to persuade? vs d. Who do you wanna persuade?

But this contraction is not always possible. There are some cases where one cannot substitute *wanna* for *want to*, as in (15):

(15) a. I want my colleague to persuade the biologist.
b. *I wanna my colleague persuade the biologist.

Here the constraint seems clear: one can only contract to *wanna* if no words intervene between them. Apparently, the phonological process of contraction is sensitive to an adjacency condition. However, some examples such as in (16a) and (17a) below seem to meet this adjacency constraint, yet the contraction is still blocked, as in (16b) and (17b):

(16) a. Who do you want to persuade the biologist?
b. *Who do you wanna persuade the biologist?

(17) a. We expect parents who want to long for luxury
(that is, *want* meaning 'to be needy')
b. *We expect parents who wanna long for luxury

Why is this so? (16a) asks '*Who* should persuade the biologist?' – in other words, *who* is the subject of *persuade*. In (14b) *who* is the object of *persuade*. The hierarchical syntactic structure for these two sentences is therefore different, and it is this difference that allows contraction in (14d) while blocking it in (16b). The syntactic structure of the two examples is representable as (14b′) and (16b′), where we have struck through the original position of *who*, its place of interpretation, before the basic operation of generative grammar has applied that put *who* at the front of the sentence. The crossed-out *who* is not pronounced, which is why the externalized output appears only as *who do you want to persuade.*

(14b′) [Who [do you want [to persuade ~~*who*~~]]]?

(16b′) [Who [do you want [~~*who*~~ to persuade the biologist]]]

Note that in (16b′) the crossed-out *who* (i.e. not pronounced) intervenes between *want* and *to*, just as *my colleague* does in (15a). But as we have seen, the contraction rule that yields *wanna* does not tolerate any elements intervening between *want* and *to*. The complex case of (16b) thus reduces to the simple case of (15b), and contraction is blocked [49,50].

The examples in (17), from [51], show that for contraction c-command between the verb *want* and *to* is also a necessary condition. Contraction is not allowed in (17) because *want* (in the meaning 'to be needy') is part of the subject and, therefore, structurally not higher than *to* (cf. 17b′). Absence of c-command is the relevant factor blocking contraction despite the availability of linear adjacency.

(17b′) We expect [[$_{NP}$ parents who want] to long for luxury]

Once again then, it is ultimately the structural properties of a sentence that run the show. For speakers, the 'hidden' properties, non-pronounced words (like in 16b′) are just as substantial as pronounced words. The linguistic computations of the mind 'hear' what the ear does not. Just as color and edges do not exist out 'in the world' but rather are internal constructions of the mind, language is not a property of external sound sequences and does not exist apart from mind-internal computations (Box 1). In this sense, language behaves just like every other cognitive ability that scientists have so far uncovered.

Summarizing the discussion above, we have shown that for

(i) the mapping to the conceptual–intentional interface, our discussion on negative polarity items and parasitic gaps:
→ hierarchical structure is necessary and sufficient
→ linear structure is irrelevant, that is, order is inaccessible

(ii) the mapping to the sensory–motor interface, our discussion of stress assignment and contraction:

Box 3. Constituents: Weak versus Strong Generative Capacity

We experience language, written or spoken, linearly, and therefore it seems straightforward to take order as a central feature of language. But take the example *a blue striped suit*. We are instantaneously capable of assessing that this phrase is ambiguous between a reading in which the suit is both blue and striped (Figure I) and a reading where the suit is blue-striped (Figure I).

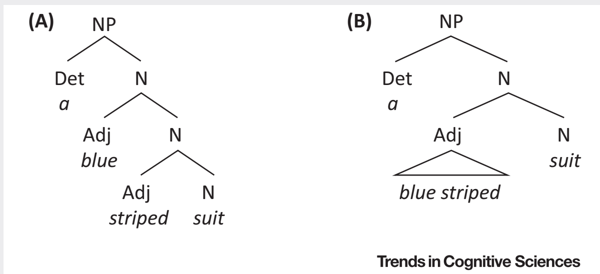


Figure I. Constituency Natural Language. Two structures for the ambiguous *a blue striped suit*, reflecting its syntax and semantics: (A) a reading in which the suit is both blue and striped, and (B) a reading where the suit is blue-striped.

In the trees above this meaning difference is reflected in a different structuring of the same words with the same linear order. In generative grammar these aspects (structure and order) are distinguished by the notions of weak and strong generative capacity. In weak generative capacity, what counts is whether a grammar will generate correct strings of words; strong generative capacity adds the requirement that the right hierarchical structure is accounted for. And this latter point is of the essence for the study of natural language as we just illustrated.

Let us explain the difference more precisely. For example, the **context-free language** characterized as $a^n b^n$ can be correctly generated by the grammars G_A and G_B in (i).

(i) a. G_A $S \Rightarrow a\ B$
$B \Rightarrow S\ b$
$S \Rightarrow a\ b$

b. G_B $S \Rightarrow A\ b$
$A \Rightarrow a\ S$
$S \Rightarrow a\ b$

These two grammars are weakly equivalent in that they both generate exactly the same string set, accepting the string *aabb*, but not *aabbb*. However, these two grammars differ in their strong generative capacity. For example, the substring *aab* is a constituent in G_B, but it is not in G_A (Figure II).

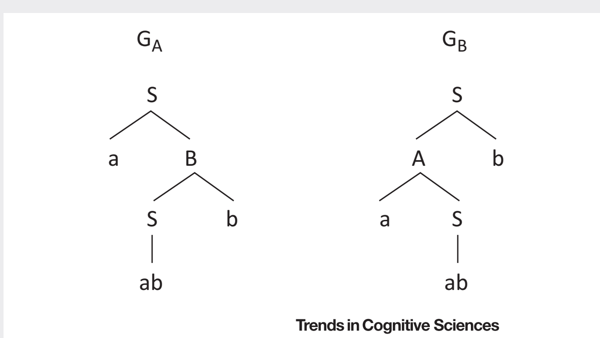


Figure II. Constituency Formal Language. The string *aabb* on the basis of grammar G_A and grammar G_B.

Weak generative capacity may play a significant role in formal language theory, where it is stipulated, as in formal arithmetic. But for natural language the concept of weak generative capacity is unnatural, unformulable, and inapplicable. It is important to realize that many possible phrase structure grammars that weakly generate some set of words or linear pattern fail as soon as strong generative capacity is taken into account. The main text illustrates serious challenges for any system based solely on weak generative capacity, as was forcibly argued from the very beginning of the modern generative enterprise [1,73]. In this respect, natural languages behave very differently from formal languages.

→ hierarchical structure is necessary, but not sufficient
→ linear structure is relevant, that is, order is needed for externalization.

What reaches the mind is unordered, what reaches the ear is ordered.

Language and Communication

The generative research tradition has never assumed that the communicative function of language underpins the essential properties of language. Note that generative grammar does not claim that language cannot be used for communicative purposes, rather that its design features are not to be understood in communicative terms [52]. For many, both linguists and non-linguists, it is difficult to imagine that some of the core properties of human language are not derived from its communicative functions. This seems to follow from the observation that language is so deeply embedded in human social interaction, facilitating the communicative and social needs of a community of speakers to share information. Communication provides a vehicle for sharing information with others. Viewed this way, language is closely intertwined with non-verbal modes of communication, such as gestures, eye contact, pointing, facial expressions, music, and the like, any of which may have communicative significance. For this approach to be well-founded, one must be precise about what 'communication' means. One can, for instance, somewhat naturally talk about flowers communicating with bees. The (often tacit) assumption is that one can pursue non-human comparisons by comparing human communication to animal communication, and more precisely the natural communication systems that use auditory, visual, or audiovisual signals [53]. And it is this notion of communication that one has in mind when one defines language as 'The systematic, conventional use of sounds, signs, or written symbols in a human society for communication and self-expression.' [54].

What then makes such verbal behavior, 'language', different from non-verbal systems of communication? Communicating how to assemble an Ikea bookcase proceeds without (much) language, via a manual consisting of just pictures, or by a video manual combining picture and accompanying speech. But explaining what compositionality or impeachment mean is not done via music, or facial expressions. So could it be that language as we know it might be particularly useful in 'hard' communicative situations, and is, therefore, 'far more complex than any animal communication system'? [55]. On such a view, animal communication systems would not be so far removed from what humans do: less complex, but not qualitatively different. By contrast, we believe that animal communication systems differ qualitatively from human language [56–58]: animal communication systems lack the rich expressive and open-ended power of human language, the creative aspect of normal language use in the Cartesian sense. Moreover, even the 'atoms' of natural language and animal communication systems are crucially different. For animal systems, 'symbols' (e.g., vervet calls) are linked directly to detectable physical events, associated with some mind-independent entity. For natural language it is radically different [59]. The evolutionary puzzle, therefore, lies in working out how this apparent discontinuity arose [60,61], demonstrating how the basic property fits this discontinuity both to the known evolutionary facts and evolutionary theory [62].

As illustrated above, structure dependency is a paramount feature of natural language, which only makes sense if solutions that rely on linear order are not available to the system that computes the mapping to the conceptual–intentional system. But if this is the case, using language for communicative purposes can only be a secondary property, making externalization (e.g., as speech or sign) an ancillary process, a reflection of properties of the sensory–motor system that might have nothing special to do with language in the restricted sense we take it to be: uniquely human (species-specific) and uniquely linguistic (domain-specific). The fact that we share a wide variety of cognitive and perceptual mechanisms with other species, for instance, vocal learning in songbirds, would then come as no surprise [63]. It would also follow that what is externally produced might yield difficulties for perception, hence communication. For example, consider the sentence *They asked if the mechanics fixed the cars*. In response to this statement, one can ask *how many cars?* yielding *How many cars did they ask if the mechanics fixed?* However, one cannot ask *how many mechanics*, yielding *How many mechanics did they ask if fixed the cars*, even though it is a perfectly fine thought. To ask about the number of mechanics, one has to use some circumlocution, one that impedes communication. In this case, communicative efficiency is sacrificed for the sake of internal computational efficiency, and there are many instances of this sort. Examples running in the other direction, where communicative function is favored over internal computational function (Box 1), seem impossible to find. Thus, the functional relationship between efficient language-as-internal-computation versus language-as-communication is asymmetric – in every case that can be carefully posed. The asymmetry is: the mapping to meaning is primary and is blind to order (language as a system for thought), the mapping to sound/sign is secondary and needs order (imposed by externalization of language). The empirical claim is, therefore, that linear order is available for the mapping to sound/sign, but not for the mapping to meaning.

Structures, Not Strings

The examples we have just given illustrate what is perhaps the most significant aspect of language: utterances are not simple linear concatenations of simpler building blocks (words, morphemes, phonemes). Rather, utterances are hierarchically structured objects built out of these simpler elements. We have to take this property into account if we want to correctly describe linguistic phenomena, whether semantic, syntactic, morphological, or phonological in nature. Structure dependence of rules is a general property of language that has been extensively discussed from the 1950s onwards and is not just restricted to the examples we have

Box 4. String Linguistics

To illustrate the type of problems an approach to human language that adopts a purely sequential structure is confronted with, we use Google Translate, a powerful string-based machine translation service that supports the non-hierarchical, linear view on language. Google Translate [used through Firefox on June 8, 2015] maps the French *La pomme mange le garçon*, lit. *the apple eats the boy*, into *the boy eats the apple*, precisely because the 'most likely' output sentence is the product of the probabilities of linear word strings or pairs, and the probability of the latter string vastly dominates the probability of the former. This problem pervades the entire approach. For example, observe Dutch (i) and its Google translation:

(i) De man van mijn tante kust de vrouw.
(ii) The husband of my aunt kissing the woman.

While not perfect – it should be *The husband of my aunt is kissing the woman* – this certainly approximates what one would like. But the system fails dismally when translating the question equivalent: Dutch (iii) becomes (iv), rather than (v).

(iii) Kust de man van mijn tante de vrouw?
(iv) Shore man of my aunt's wife?
(v) Is the husband of my aunt kissing the woman?

Here, *kust* ('kisses'), derived from *kussen* ('to kiss'), is translated as *shore*, having been misinterpreted as the Dutch noun *kust* for *shore/coast*. Moreover, the subject *de man van mijn tante* is analyzed as the possessive of the object *de vrouw*. What has gone wrong? Omitting much detail along with trade secrets, what such systems do is roughly this: given a particular Dutch sentence, notated (Diii), iterate over all English strings of words to find that 'best' English string, E′, which maximizes the probability of E′ × the probability (Diii | E′), that is, the probability of the Dutch (iii) given E′. Note that this statistical decomposition is linear. It will tend to select commonly occurring word pairs, for instance, *kust/coast*, if no longer pairing is readily available or inferred. For example, no English pairing for the Dutch *kust de man* because the 'phrase book' is still not dense enough in the space of pairings.

Adopting the view that hierarchy is only relevant 'when the language user is particularly attentive, when it is important for the task at hand' [71] comes at a price. For a practical business solution the price is right, for a scientific approach to the study of language the price is wrong.

presented so far. These are phenomena that, in our view, must be explained in terms of intrinsic and domain-specific properties of a biolinguistic system.

Native speakers have robust knowledge of the constraints that we discussed above, and often that knowledge is tacit – again analogous to the reconstruction of 'color' and 'edges'. Sometimes relevant examples are rarely attested in adult language, but children acquire them nonetheless. Furthermore, it has been shown repeatedly that infants acquiring language do not solely engage in statistical learning by approximating the target language [64–70]. For these and other reasons, usage-based approaches that reject generative procedures, and apply statistical methods of analysis to unanalyzed data (Box 4), probing into huge but finite lists of data that are not extendable, fail to distinguish these cases properly. By contrast, generative procedures succeed in amalgamating a large, diverse set of individual examples into just a few constraints such as the hierarchical dominance example.

Linear statistical analysis fails to account for how semantic readings are specifically linked to syntactic structures or to explain why ambiguity is constrained in some cases but not in others. A major problem is not just the failure to succeed, but more importantly the apparent unwillingness to come to terms with simple core puzzles of language structure such as those we have noted [71]. There have been a handful of other efforts to provide alternative accounts for structure dependence [72,74], but these have been shown to fail [69]. However, if we are really interested in the actual mechanisms of the internal system we should ask about the properties that determine how and why the syntax–semantics mappings are established in the way they are and not otherwise (see Outstanding Questions).

Concluding Remarks

Approximating observational phenomena is very different from formulating an explanatory account of a significant body of empirical data. Equating likelihood probabilities of language use with grammaticality properties of internal systems does not succeed because structural properties of phrases and the generative capacity of internal systems to build structure cannot be reduced to linear properties of strings. These somewhat elementary but important insights have been recognized since the very origins of generative grammar [1,18], but seem to have been forgotten, ignored, or even denied without serious argument in recent times.

Acknowledgements

J.J.B. is part of the Consortium on Individual Development (CID), which is funded through the Gravitation program of the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research (NWO; grant number 024.001.003).

References

1. Chomsky, N. (1956) Three models for the description of language. *IRE Trans. Inform. Theor.* IT–2, 113–124
2. Miller, G.A. (1956) The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.* 63, 81–97
3. Newell, A. and Simon, H.A. (1956) Logic Theory Machine: a complex information processing system. *IRE Trans. Inform. Theor.* IT–2, 61–79
4. Shannon, C.E. (1956) The zero error capacity of a noisy channel. *IRE Trans. Inform. Theor.* IT–2, 8–19
5. Chomsky, N. (1995) *The Minimalist Program,* MIT Press
6. Reinhart, T. (2006) *Interface Strategies: Optimal and Costly Computations,* MIT Press
7. Rizzi, L. (2012) Core linguistic computations: how are they expressed in the mind/brain? *J. Neuroling.* 25, 489–499
8. Selkirk, E. (2011) The syntax–phonology interface. In *The Handbook of Phonological Theory* (2nd edn) (Goldsmith, J. *et al.*, eds), pp. 435–484, Blackwell
9. Weaver, W. (1947) Translation. In *Machine Translation of Languages* (Locke, W.N. and Booth, D.A., eds), pp. 15–23, MIT Press
10. Brown, P. *et al.* (1988) A statistical approach to language translation. In *COLING '88 Proceedings of the 12th Conference on Computational Linguistics (Vol. 1)*, pp. 71–76, Association for Computational Linguistics
11. Evans, N. and Levinson, S. (2009) The myth of language universals. *Behav. Brain Sci.* 32, 429–492
12. Tomasello, M. (2003) *Constructing A language: A Usage-Based Theory of Language Acquisition,* Harvard University Press
13. Langacker, W. (2008) *Cognitive Grammar: A Basic Introduction,* Oxford University Press
14. Dąbrowska, E. (2015) What exactly is Universal Grammar, and has anyone seen it? *Front. Psychol.* 6, 852
15. Elman, J.L. *et al.* (1996) *Rethinking Innateness: A Connectionist Perspective on Development,* MIT Press
16. Meisel, J. (2011) *First and Second Language Acquisition,* Cambridge University Press

Outstanding Questions

What operating principles are there besides SIMPLEST MERGE (yielding hierarchical, structure-preserving structure without linear order) and MINIMAL SEARCH (a domain-general condition of minimal computation that restricts application of rules of agreement and displacement to strictly local domains and minimal structural distance)?

What can we find out about the neural organization underlying higher-order computation of merge-based hierarchical structure of language and what are its evolutionary roots? Concentrating on the basic property, how does the discontinuity fit the known evolutionary facts and evolutionary theory?

What is the precise division of labor between domain-general and domain-specific learning systems that enter into the explanation of learnability and evolvability of natural language? How does the *Strong Minimalist Thesis* – the conjecture that, optimally, UG reduces to the simplest computational principles that operate in accordance with conditions of computational efficiency – enhance the prospects of explaining the emergence and learning of human language, permitting acquisition of rich languages from poor inputs (*poverty of stimulus*)?

How can we attain a better understanding of the mind-dependent nature, development, and evolutionary origins of the word-like elements ('atoms') of human language that enter into core computational operations of language, yielding its basic property?

What is the role of morphosyntactic features in identifying phrases of exocentric constructions, that is, phrases not containing a head capable of uniquely identifying them, and demarcating minimal domains of computation? How do these features function in the language architecture?

If an improved understanding of the sources of complexity, diversity, and malleability of languages helps us explain their significance for the externalization process, which linearization principles and strategies govern the externalization of the syntactic products generated by the basic combinatorial operation of language?

17. Moro, A. (2014) On the similarity between syntax and actions. *Trends Cogn. Sci.* 18, 109–110
18. Chomsky, N. (1959) On certain formal properties of grammars. *Inform. Control* 2, 137–167
19. Watumull, J. *et al.* (2014) On recursion. *Front. Psychol.* 4, 1–7
20. Lobina, D.J. (2011) 'A running back'; and forth: a review of *Recursion and Human Language*. *Biolinguistics* 5, 151–169
21. Church, A. (1936) An unsolvable problem of elementary number theory. *Am. J. Math.* 58, 345–363
22. Gödel, K. (1986) On undecidable propositions of formal mathematical systems. In *Kurt Gödel: Collected Works Vol. I: Publications 1929–1936* (Feferman, S. *et al.*, eds), pp. 346–371, Oxford University Press
23. Turing, A.M. (1936) On computable numbers, with an application to the Entscheidungsproblem. *Proc. Lond. Math. Soc.* 42, 230–265
24. Kleene, S.C. (1936) General recursive functions of natural numbers. *Math. Ann.* 112, 727–742
25. Chomsky, N. (1966) *Cartesian Linguistics,* Harper & Row
26. Bloomfield, L. (1933) *Language,* Holt
27. Hauser, M.D. *et al.* (2014) The mystery of language evolution. *Front. Psychol.* 5, 401
28. Arregui, K. and Nevins, A. (2012) *Morphotactics: Basque Auxiliaries and the Structure of Spellout,* Springer
29. Giannakidou, A. (2011) Negative polarity and positive polarity: licensing, variation, and compositionality. In *The Handbook of Natural Language Meaning* (2nd edn) (von Heisinger, K. *et al.*, eds), pp. 1660–1712, Mouton de Gruyter
30. Kuno, M. (2008) Negation, focus, and negative concord in Japanese. *Toronto Work. Pap. Ling.* 28, 195–211
31. Reinhart, T. (1981) Definite NP-anaphora and c-command domains. *Ling. Inq.* 12, 605–635
32. Baker, M. (2003) Language differences and language design. *Trends Cogn. Sci.* 7, 349–353
33. Musso, M. *et al.* (2003) Broca's area and the language instinct. *Nat. Neurosci.* 6, 774–781
34. Smith, N. and Tsimpli, I. (1995) *The Mind of a Savant: Language Learning and Modularity,* Oxford University Press
35. Vasishth, S. *et al.* (2008) Processing polarity: how the ungrammatical intrudes on the grammatical. *Cogn. Sci.* 32, 685–712
36. Ross, J.R. (1986) *Infinite Syntax!* Ablex
37. Chomsky, N. (1981) *Lectures on Government and Binding,* Foris Publications
38. Taraldsen, K.T. (1980) The theoretical interpretation of a class of marked extractions. In *The Theory of Markedness in Generative Grammar* (Belletti, A. *et al.*, eds), pp. 475–516, Scuola Normale Superiore di Pisa
39. Engdahl, E. (1983) Parasitic gaps. *Ling. Philos.* 6, 5–34
40. Chomsky, N. (1982) *Some Concepts and Consequences of the Theory of Government and Binding (LI Monograph 6),* MIT Press
41. Huybregts, M.A.C. and van Riemsdijk, H.C. (1985) Parasitic gaps and ATB. In *Proceedings of the NELS XV Conference*, pp. 168–187, GSLA, University of Massachusetts
42. Hoekstra, T. and Bennis, H. (1984) Gaps and parasitic gaps. *Ling. Rev.* 4, 29–87
43. Bech, G. (1952) Über das Niederländische Adverbialpronomen *er*. *Travaux du Cercle Linguistique de Copenhague* 8, 5–32
44. Bennis, H. (1986) *Gaps and Dummies,* Foris Publications
45. Huybregts, M.A.C. (1991) Clitics. In *Grammatische Analyse* (Model, J., ed.), pp. 279–330, Foris Publications
46. Chomsky, N. *et al.* (1956) On accent and juncture in English. In *For Roman Jakobson: Essays on the Occasion of his Sixtieth Birthday* (Halle, M. *et al.*, eds), pp. 65–80, Mouton
47. Chomsky, N. and Halle, M. (1968) *The Sound Pattern of English,* Harper and Row
48. Liberman, M. and Prince, A. (1977) On stress and linguistic rhythm. *Ling. Inq.* 8, 249–336
49. Lakoff, G. (1970) Global rules. *Language* 46, 627–639
50. Chomsky, N. and Lasnik, H. (1978) A remark on contraction. *Ling. Inq.* 9, 268–274
51. Aoun, J. and Lightfoot, D. (1984) Government and contraction. *Ling. Inq.* 15, 465–473
52. Chomsky, N. (2013) What kind of creatures are we? The Dewey Lectures. Lecture I: What is language? Lecture II: What can we understand? *J. Philos.* 12, 645–700
53. Hauser, M.D. (1997) *The Evolution of Communication,* MIT Press
54. Crystal, D. (1992) *An Encyclopedic Dictionary of Language and Languages,* Blackwell
55. Hurford, J. (2008) The evolution of human communication and language. In *Sociobiology of Communication: An Interdisciplinary Perspective* (D'Ettorre, P. and Hughes, D., eds), pp. 249–264, Oxford University Press
56. Hauser, M. *et al.* (2002) The faculty of language: What is it, who has it, and how did it evolve? *Science* 298, 1569–1579
57. Berwick, R.C. *et al.* (2013) Evolution, brain, and the nature of language. *Trends Cogn. Sci.* 17, 89–98
58. Bolhuis, J.J. and Everaert, M.B.H. (2013) *Birdsong, Speech and Language. Exploring the Evolution of Mind and Brain,* MIT Press
59. Chomsky, N. (2013) Notes on denotation and denoting. In *From Grammar to Meaning: The Spontaneous Logicality of Language* (Caponigro, I. and Cecchetto, C., eds), pp. 38–46, Cambridge University Press
60. Berwick, R.C. (2010) All you need is merge: a biolinguistic opera in two acts. In *Biolinguistic Approaches to Language Evolution* (Di Sciullo, A.M. and Boeckx, C., eds), pp. 461–491, Oxford University Press
61. Bolhuis, J.J. *et al.* (2014) How could language have evolved? *PLoS Biol.* 12, e1001934
62. Berwick, R.C. and Chomsky, N. (2016) *Why Only Us: Language and Evolution,* MIT Press
63. Chomsky, N. (2005) Three factors in language design. *Ling. Inq.* 36, 1–22
64. Crain, S. (2012) *The Emergence of Meaning,* Cambridge University Press
65. Lidz, J. and Gagliardi, A. (2015) How nature meets nurture: Universal Grammar and statistical learning. *Annu. Rev. Ling.* 1, 333–353
66. Medina, T.N. *et al.* (2011) How words can and cannot be learned by observation. *Proc. Natl. Acad. Sci. U.S.A.* 108, 9014–9019
67. Gleitman, L. and Landau, B. (2012) Every child an isolate: Nature's experiments in language learning. In *Rich Languages from Poor Inputs* (Piattelli-Palmarini, M. and Berwick, R.C., eds), pp. 91–104, Oxford University Press
68. Yang, C. (2016) Negative knowledge from positive evidence. *Language* 92, in press
69. Berwick, R.C. *et al.* (2011) Poverty of the stimulus revisited. *Cogn. Sci.* 35, 1207–1242
70. Chomsky, N. (2011) Language and other cognitive systems. What is special about language? *Lang. Learn. Dev.* 7, 263–278
71. Frank, S. *et al.* (2012) How hierarchical is language use? *Proc. R. Soc. B* 297, 4522–4531
72. Reali, F. and Christiansen, M.H. (2005) Uncovering the richness of the stimulus: structure dependence and indirect statistical evidence. *Cogn. Sci.* 29, 1007–1028
73. Chomsky, N. (1965) *Aspects of the Theory of Syntax,* MIT Press
74. Perfors, A. *et al.* (2011) Poverty of the stimulus: a rational approach. *Cognition* 118, 306–338