How hierarchical is language use?

Stefan L. Frank1,*, Rens Bod2 and Morten H. Christiansen3

1Department of Cognitive, Perceptual and Brain Sciences, University College London, 26 Bedford Way, London WC1H 0AP, UK
2Institute for Logic, Language and Information, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands
3Department of Psychology, Cornell University, 228 Uris Hall, Ithaca, NY 14853-7601, USA

It is generally assumed that hierarchical phrase structure plays a central role in human language. However, considerations of simplicity and evolutionary continuity suggest that hierarchical structure should not be invoked too hastily. Indeed, recent neurophysiological, behavioural and computational studies show that sequential sentence structure has considerable explanatory power and that hierarchical processing is often not involved. In this paper, we review evidence from the recent literature supporting the hypothesis that sequential structure may be fundamental to the comprehension, production and acquisition of human language. Moreover, we provide a preliminary sketch outlining a non-hierarchical model of language use and discuss its implications and testable predictions. If linguistic phenomena can be explained by sequential rather than hierarchical structure, this will have considerable impact in a wide range of fields, such as linguistics, ethology, cognitive neuroscience, psychology and computer science.

Keywords: language structure; language evolution; cognitive neuroscience; psycholinguistics; computational linguistics

1. INTRODUCTION

Sentences can be analysed as hierarchically structured: words are grouped into phrases (or 'constituents'), which are grouped into higher-level phrases, and so on until the entire sentence has been analysed, as shown in (1).

(1) [Sentences [ can [be analysed] ] as [hierarchically structured] ]

The particular analysis that is assigned to a given sentence depends on the details of the assumed grammar, and there can be considerable debate about which grammar correctly captures the language. Nevertheless, it is beyond dispute that hierarchical structure plays a key role in most descriptions of language. The question we pose here is: How relevant is hierarchy for the use of language?

The psychological reality of hierarchical sentence structure is commonly taken for granted in theories of language comprehension [1–3], production [4,5] and acquisition [6,7]. We argue that, in contrast, sequential structure is more fundamental to language use. Rather than considering a hierarchical analysis as in (1), the human cognitive system may treat the sentence more along the lines of (2), in which words are combined into components that have a linear order but no further part/whole structure.

(2) [Sentences] [can be analysed] [as hierarchically structured]

Naturally, there may not be just one correct analysis, nor is it necessary to analyse the sentence as either (1) or (2). Intermediate forms are possible, and a sentence’s interpretation will depend on the current goal, strategy, cognitive abilities, context, etc. However, we propose that something along the lines of (2) is cognitively more fundamental than (1).

To begin with, (2) provides a simpler analysis than (1), which may already be reason enough to take it as more fundamental—other things being equal. Sentences trivially possess sequential structure, whereas hierarchical structure is only revealed through certain kinds of linguistic analysis. Hence, the principle of Occam’s Razor compels us to stay as close as possible to the original sentence and only conclude that any structure was assigned if there is convincing evidence.

So why and how has the hierarchical view of language come to dominate? The analysis of sentences by division into sequential phrases can be traced back to a group of thirteenth century grammarians known as Modists who based their work on Aristotle [8]. While the Modists analysed a sentence into a subject and a predicate, their analyses did not result in deep hierarchical structures. This type of analysis was influential enough to survive until the rise of the linguistic school known as Structuralism in the 1920s [9]. According to the structuralist Leonard Bloomfield, sentences need to be exhaustively analysed, meaning that they are split up into subparts all the way down to their smallest meaningful components, known as morphemes [10]. The ‘depth’ of a structuralist sentence analysis became especially manifest when Noam Chomsky, in his Generative Grammar framework [11], used tree diagrams to represent hierarchical structures. Chomsky urged that a tree diagram should (preferably) be binary (meaning that every phrase consists...
of exactly two parts) which led to even deeper—and thus
more hierarchical—trees. Together with the introduction
of hypothetical ‘empty’ elements that are not phonetically
realized, the generative approach typically led to sentence
tree diagrams that were deeper than the length of the
sentences themselves.

While the notion of binary structure, and especially of
empty elements, has been criticized in various linguistic
frameworks [12,13], the practice of analysing sentences in
terms of deep hierarchical structures is still part and
parcel of linguistic theory. In this paper, we question
this practice, not so much for language analysis but for
the description of language use. We argue that hierarchical
structure is rarely (if ever) needed to explain how
language is used in practice.

In what follows, we review evolutionary arguments
as well as recent studies of human brain activity (i.e.
cognitive neuroscience), behaviour (psycholinguistics)
and the statistics of text corpora (computational linguis-
tics), which all provide converging evidence against the
primacy of hierarchical sentence structure in language
use.1 We then sketch our own non-hierarchical model
that may be able to account for much of the empirical
data, and discuss the implications of our hypothesis for
different scientific disciplines.

2. THE ARGUMENT FROM EVOLUTIONARY
CONTINUITY

Most accounts of language incorporating hierarchical struc-
ture also assume that the ability to use such structures is
unique to humans [16,17]. It has been proposed that the
ability to create unbounded hierarchical expressions may
have emerged in the human lineage either as a consequence
of a single mutation [18] or by way of gradual natural
selection [16]. However, recent computational simulations
[19,20] and theoretical considerations [21,22] suggest that
there may be no viable evolutionary explanation for such
a highly abstract, language-specific ability. Instead, the struc-
ture of language is hypothesized to derive from non-linguistic
constraints amplified through repeated cycles of cultural
transmission across generations of language learners and
users. This is consistent with recent cross-linguistic analyses
of word-order patterns using computational tools from evol-
utionary biology, showing that cultural evolution—rather
than language-specific structural constraints—is the key
determinant of linguistic structure [23].

Similarly to the proposed cultural recycling of cortical
maps in the service of recent human innovations such as
reading and arithmetic [24], the evolution of language is
assumed to involve the reuse of pre-existing neural mech-
anism. Thus, language is shaped by constraints inherited
from neural substrates predating the emergence of
language, including constraints derived from the nature
of our thought processes, pragmatic factors relating to
social interactions, restrictions on our sensorimotor
apparatus and cognitive limitations on learning, memory
and processing [21]. This perspective on language evol-
sion suggests that our ability to process syntactic
structure may largely rely on evolutionarily older,
domain-general systems for accurately representing the
sequential order of events and actions. Indeed, cross-species
comparisons and genetic evidence indicate that humans
have evolved sophisticated sequencing skills that were
subsequently recruited for language [25]. If this evolution-
ary scenario is correct, then the mechanisms employed for
language learning and use are likely to be fundamentally
sequential in nature, rather than hierarchical.

It is informative to consider an analogy to another cul-
turally evolved symbol system: arithmetic. Although
arithmetic can be described in terms of hierarchical struc-
ture, this does not entail that the neural mechanisms
employed for arithmetic use such structures. Rather, the
considerable difficulty that children face in learning arith-
metic suggests that the opposite is the case, probably
because these mathematical skills reuse evolutionary
ly older neural systems [24]. But why, then, can children
master language without much effort and explicit instruc-
tion? Cultural evolution provides the answer to this
question, shaping language to fit our learning and process-
ing mechanisms [26]. Such cultural evolution cannot
of course, alter the basics of arithmetic, such as how
addition and subtraction work.

3. THE IMPORTANCE OF SEQUENTIAL SENTENCE
STRUCTURE: EMPIRICAL EVIDENCE

(a) Evidence from cognitive neuroscience

The evolutionary considerations suggest that associations
should exist between sequence learning and syntactic
processing because both types of behaviour are subserved
by the same underlying neural mechanisms. Several lines
of evidence from cognitive neuroscience support this
hypothesis: the same set of brain regions appears to be
involved in both sequential learning and language, includ-
ing cortical and subcortical areas (see [27] for a review).
For example, brain activity recordings by electroencephalo-
graphy have revealed that neural responses to grammatical
violations in natural language are indistinguishable from
those elicited by incongruencies in a purely sequentially
structured artificial language, including very similar
topographical distributions across the scalp [28].

Among the brain regions implicated in language,
Broca’s area—located in the left inferior frontal gyrus—is
of particular interest as it has been claimed to be
dedicated to the processing of hierarchical structure in
the context of grammar [29,30]. However, several recent
studies argue against this contention, instead underscoring
the primacy of sequential structure over hierarchical
composition. A functional magnetic resonance imaging
(fMRI) study involving the learning of linearly ordered
sequences found similar activations of Broca’s area to
those obtained in previous studies of syntactic violations
in natural language [31], indicating that this part of the
brain may implement a generic on-line sequence processor.
Moreover, the integrity of white matter in Broca’s area cor-
relates with performance on sequence learning, with higher
degrees of integrity associated with better learning [32].

If language is subserved by the same neural mechan-
isms as used for sequence processing, then we would
expect a breakdown of syntactic processing to be associ-
ated with impaired sequencing abilities. This prediction
was tested in a population of agrammatic aphasics, who
have severe problems with natural language syntax in
both comprehension and production. Indeed, there was
evidence of a deficit in sequence learning in agrammatism
[33]. Additionally, a similar impairment in the process-
ing of musical sequences by the same population points
to a functional connection between sequencing skills and language [34]. Further highlighting this functional relationship, studies applying transcranial direct current stimulation during training [35], or repetitive transcranial magnetic stimulation during testing [36], have found that sequencing performance is enhanced by such stimulation of Broca’s area.

Hence, insofar as the same neural substrates appear to be involved in both the processing of linear sequences and language, it would seem plausible that syntactic processing is fundamentally sequential in nature, rather than hierarchical.

(b) Evidence from psycholinguistics

A growing body of behavioural evidence also underlines the importance of sequential structure to language comprehension and production. If a sentence’s sequential structure is more important than its hierarchical structure, the linear distance between words in a sentence should matter more than their relationship within the hierarchy. Indeed, in a speech-production study, it was recently shown that the rate of subject–verb number–agreement errors, as in (3), depends on linear rather than hierarchical distance between words [37,38].

Moreover, when reading sentences in which there is a conflict between local and distal agreement information (as between the plural balls and the singular coat in (3)) the resulting slow-down in reading is positively correlated with people’s sensitivity to bigram information in a sequential learning task: the more sensitive learners are to how often items occur adjacent to one another in a sequence, the more they experience processing difficulty when distracting, local agreement information conflicts with the relevant, distal information [39]. More generally, reading slows down on words that have longer surface distance from a dependent word [40,41].

Local information can take precedence even when this leads to inconsistency with earlier, distal information: the embedded verb tossed in (4) is read more slowly than thrown in (5), indicating that the player tossed is (at least temporarily) taken to be a coherent phrase, which is ungrammatical considering the preceding context [42].

(4) The coach smiled at the player tossed a frisbee.
(5) The coach smiled at the player thrown a frisbee.

Additional evidence for the primacy of sequential processing comes from the difference between crossed and nested dependencies, illustrated by sentences (6) and (7) (adapted from [43]), which are the German and Dutch translations, respectively, of Johanns helps the men teach Hans to feed the horses (the subscripted indices show dependencies between nouns and verbs).

(6) Johanns hat die Männer: Hans3 die Pferde füttern3 lehren3 helfen3
(7) Johanns heeft de mannen: Hans3 de paarden helpen3 leeren3 voeren3

Nested structures, as in (6), result in long-distance dependencies between the outermost words. Consequently, such sentences are harder to understand [43], and possibly harder to learn [44], than sentences with crossed dependencies, as in (7). These effects have been replicated in a study employing a cross-modal serial-reaction-time (SRT) task [45], suggesting that processing differences between crossed and nested dependencies derive from constraints on sequential learning abilities. Additionally, the Dutch/German results have been simulated by recurrent neural network (RNN) models [46, 47] that are fundamentally sequential in nature.

A possibly related phenomenon is the grammaticality illusion demonstrated by the nested dependencies in (8) and (9).

(8) *The spider: that the bullfrog: that the turtle: followed: mercilessly ate: the fly.
(9) The spider: that the bullfrog: that the turtle: followed: chased: ate: the fly.

Sentence (8) is ungrammatical: it has three subject nouns but only two verbs. Perhaps surprisingly, readers rate it as more acceptable [47, 48] and process the final (object) noun more quickly [49], compared with the correct variant in (9). Presumably, this is because of the large linear distance between the early nouns and the late verbs, which makes it hard to keep all nouns in memory [48]. Results from SRT learning [45], providing a sequence-based analogue of this effect, show that the processing problem indeed derives from sequence–memory limitations and not from referential difficulties. Interestingly, the reading-time effect did not occur in comparable German sentences, possibly because German speakers are more often exposed to sentences with clause–final verbs [49]. This grammaticality illusion, including the cross-linguistic difference, was explained using an RNN model [50].

It is well known that sentence comprehension involves the prediction of upcoming input and that more predictable words are read faster [51]. Word predictability can be quantified by probabilistic language models, based on any set of structural assumptions. Comparisons of RNNs with models that rely on hierarchical structure indicate that the non-hierarchical RNNs predict general patterns in reading times more accurately [52–54], suggesting that sequential structure is more important for predictive processing. In support of this view, individuals with higher ability to learn sequential structure are more sensitive to word predictability [55]. Moreover, the ability to learn non-adjacent dependency patterns in an SRT task is positively correlated with performance in on-line comprehension of sentences with long-distance dependencies [56].

(c) Evidence from computational models of language acquisition

An increasing number of computational linguists have shown that complex linguistic phenomena can be learned by employing simple sequential statistics from human-generated text corpora. Such phenomena had, for a long time, been considered parade cases in favour of hierarchical sentence structure. For example, the phenomenon known as auxiliary fronting was assumed to be unlearnable without taking hierarchical dependencies into account [57]. If sentence (10) is turned into a yes–no question, the auxiliary is is fronted, resulting in sentence (11).

(8) *The spider: that the bullfrog: that the turtle: followed: mercilessly ate: the fly.
(9) The spider: that the bullfrog: that the turtle: followed: chased: ate: the fly.

Sentence (8) is ungrammatical: it has three subject nouns but only two verbs. Perhaps surprisingly, readers rate it as more acceptable [47, 48] and process the final (object) noun more quickly [49], compared with the correct variant in (9). Presumably, this is because of the large linear distance between the early nouns and the late verbs, which makes it hard to keep all nouns in memory [48]. Results from SRT learning [45], providing a sequence-based analogue of this effect, show that the processing problem indeed derives from sequence–memory limitations and not from referential difficulties. Interestingly, the reading-time effect did not occur in comparable German sentences, possibly because German speakers are more often exposed to sentences with clause–final verbs [49]. This grammaticality illusion, including the cross-linguistic difference, was explained using an RNN model [50].

It is well known that sentence comprehension involves the prediction of upcoming input and that more predictable words are read faster [51]. Word predictability can be quantified by probabilistic language models, based on any set of structural assumptions. Comparisons of RNNs with models that rely on hierarchical structure indicate that the non-hierarchical RNNs predict general patterns in reading times more accurately [52–54], suggesting that sequential structure is more important for predictive processing. In support of this view, individuals with higher ability to learn sequential structure are more sensitive to word predictability [55]. Moreover, the ability to learn non-adjacent dependency patterns in an SRT task is positively correlated with performance in on-line comprehension of sentences with long-distance dependencies [56].

(c) Evidence from computational models of language acquisition

An increasing number of computational linguists have shown that complex linguistic phenomena can be learned by employing simple sequential statistics from human-generated text corpora. Such phenomena had, for a long time, been considered parade cases in favour of hierarchical sentence structure. For example, the phenomenon known as auxiliary fronting was assumed to be unlearnable without taking hierarchical dependencies into account [57]. If sentence (10) is turned into a yes–no question, the auxiliary is is fronted, resulting in sentence (11).
A language learner might derive from these two sentences that the first occurring auxiliary is fronted. However, when the sentence also contains a relative clause with an auxiliary is (as in The man who is eating is hungry), it should not be the first occurrence of is that is fronted but the one in the main clause. Many researchers have argued that input to children provides no information that would favour the correct auxiliary fronting [58,59]. Yet children do produce the correct sentences of the form (12) and rarely the incorrect form (13) even if they have (almost) never heard the correct form before [60].

(12) *Is the man who eating is hungry?
(13) *Is the man who is eating hungry?

According to [60], hierarchical structure is needed for children to learn this phenomenon. However, there is evidence that it can be learned from sequential sentence structure alone by using a very simple, non-hierarchical model from computational linguistics: a Markov (trigram) model [61]. While it has been argued [62] that some of the results in [61] were owing to incidental facts of English, a richer computational model, using associative rather than hierarchical structure, was shown to learn the full complexity of auxiliary fronting, thus suggesting that sequential structure suffices [63]. Likewise, auxiliary fronting could be learned by simply tracking the relative sequential positions of words in sentences [64].

Linguistic phenomena beyond auxiliary fronting were also shown to be learnable by using statistical information from text corpora: phenomena known in the linguistic literature [65] as subject wh-questions, wh-questions in situ, complex NP-constraints, superiority effects of question words and the blocking of movement from wh-islands, can be learned on the basis of unannotated, child-directed language [66]. Although in [66] hierarchical sentence structure was induced at first, it turned out that such structure was not needed because the phenomena could be learned by simply combining previously encountered sequential phrases. As another example, across languages, children often incorrectly produce uninflected verb forms, as in He go there. Traditional explanations of the error assume hierarchical syntactic structure [67], but a recent computational model explained the phenomenon without relying on any hierarchical processing [68].

Besides learning specific linguistic phenomena, computational approaches have also been used for modelling child language learning in a more general fashion: in a simple computational model that learns to comprehend and produce language when exposed to child-directed speech from text corpora [69], simple word-to-word statistics (backward transitional probabilities) were used to create an inventory of chunks consisting of one or more words. This incremental, online model has broad cross-linguistic coverage, and is able to fit child data from a statistical learning study [70]. It suggests, like the models above, that children’s early linguistic behaviour can be accounted for using distributional statistics on the basis of sequential sentence structure alone.

4. TOWARDS A NON-HIERARCHICAL MODEL OF LANGUAGE USE

In this section, we sketch a model to account for human language behaviour without relying on hierarchical structure. Rather than presenting a detailed proposal that allows for direct implementation and validation, we outline the assumptions that, with further specification, can lead to a fully specified computational model.

(a) Constructions

As a starting point, we take the fundamental assumption from Construction Grammar that the productive units of language are so-called constructions: pieces of linguistic forms paired with meaning [71]. The most basic constructions are single-word/meaning pairs, such as the word fork paired with whatever comprises the mental representation of a fork. Slightly more interesting cases are multi-word constructions: a frequently occurring word sequence can become merged into a single construction. For example, knife and fork might be frequent enough to be stored as a sequence, whereas the less frequent fork and knife is not. There is indeed ample psycholinguistic evidence that the language-processing system is sensitive to the frequency of such multi-word sequences [72–75]. In addition, constructions may contain abstract ‘slots’, as in put X down, where X can be any noun phrase.

Importantly, constructions do not have a causally effective hierarchical structure. Only the sequential structure of a construction is relevant, as language comprehension and production always require a temporal stream as input or output. It is possible to assign hierarchical structure to a construction’s linguistic form, but any such structure would be inert when the construction is used.

Although a discussion of constructions’ semantic representations lies beyond the scope of the current paper, it is noteworthy that hierarchical structure seems to be of little importance to meaning as well. Traditionally, meaning has been assumed to arise from a Language of Thought [76], often expressed by hierarchically structured formulae in predicate logic. However, an increasing amount of psychological evidence suggests that the mental representation of meaning takes the form of a ‘mental model’ [77], ‘image schema’ [78] or ‘sensorimotor simulation’ [79], which have mostly spatial and temporal structure (although, like sentences, they may be analysed hierarchically if so desired).

(b) Combining constructions

Constructions can be combined to form sentences and, conversely, sentence comprehension requires identifying the sentence’s constructions. Although constructions are typically viewed as having no internal hierarchical structure, perhaps their combination might give rise to sentence-level hierarchy? Indeed, it seems intuitive to regard a combination of constructions as a part–whole relation, resulting in hierarchical structure: if the three constructions put X down, your X, and knife and fork are combined to form the sentence put your knife and fork down (or, vice versa, the sentence is understood as consisting of these three constructions) it can be analysed hierarchically as put [your [knife and fork]] down, reflecting hypothesized part–whole relations between
proposed by Ding

similar vein, the very simple neural network model mechanism could account for embedded structure. In a deal with parallel streams or embedded phrases, the found across languages. Although this model does not network that learns when to withhold or pronounce to) is repeated. Here, the control mechanism is a neural [81] assumes that there is a single stream in which a mechanism that takes care of timely switches between the [81], discussed in more detail in §5d.

It is still an open question how to implement a control mechanism that takes care of timely switches between the different streams. A recent model of sentence production [81] assumes that there is a single stream in which a sequence of words (or, rather, the concepts they refer to) is repeated. Here, the control mechanism is a neural network that learns when to withhold or pronounce each word, allowing for the different basic word orders found across languages. Although this model does not deal with parallel streams or embedded phrases, the authors do note that a similar (i.e. non-hierarchical) mechanism could account for embedded structure. In a similar vein, the very simple neural network model proposed by Ding et al. [82] uses continuous activation decay in two parallel sequential processing streams to learn different types of embedding without requiring any control system. A comparable mechanism has been suggested for implementing embedded structure processing in biological neural memory circuits [83].

So far, we have assumed that the parallel sequential streams remain separated and that any interaction is caused by switching between them. However, an actual (artificial or biological) implementation of such a model could take the form of a nonlinear, rich dynamical system, such as a RNN. The different sequential streams would run simultaneously in one and the same piece of hardware (or ‘wetware’), allowing them to interact. Although such interaction could, in principle, replace any external control mechanism, it also creates interference between the streams. This interference grows more severe as the number of parallel streams increases with deeper embedding of multiple constructions. The resulting performance degradation prevents unbounded depth of embedding and thus naturally captures the human performance limitations discussed in §3b.

(c) Language understanding
As explained above, the relationship between a sentence and its constructions can be realized using only sequential mechanisms. Considering the inherent temporal nature of language, this connects naturally to the language-processing system’s sensory input and motor output sequences. But how are sentences understood without resorting to hierarchical processing? Rather than proposing a concrete mechanism, we argue here that hierarchical structure rarely needs to play a significant role in language comprehension.

Several researchers have noted that language can generally be understood by making use of superficial cues. According to Late Assignment of Syntax Theory [84] the initial semantic interpretation of a sentence is based on its superficial form, while a syntactic structure is only assigned at a later stage. Likewise, the ‘good-enough comprehension’ [85] and ‘shallow parsing’ [86] theories claim that sentences are only analysed to the extent that this is required for the task at hand, and that under most circumstances a shallow analysis suffices.

A second reason why deep, hierarchical analysis may not be needed for sentence comprehension is that language is not strictly compositional, which is to say that the meaning of an utterance is not merely a function of the meaning of its constructions and the way they are combined. More specifically, a sentence’s meaning is also derived from extra-sentential and extra-linguistic factors, such as the prior discourse, pragmatic constraints, the current setting, general world knowledge, and knowledge of the speaker’s intention and the listener’s state of mind. All these (and possibly more) sources of information directly affect the comprehension process [51,87,88], thereby reducing the importance of sentence structure. Indeed, by applying knowledge about the structure of events in the world, a recent neural network model displayed systematic sentence comprehension without any compositional semantics [89].

5. IMPLICATIONS FOR LANGUAGE RESEARCH
Our hypothesis that human language processing is fundamentally sequential rather than hierarchical has important implications for the different research fields with a stake in language. In this section, we discuss some of the general implications of our viewpoint, including specific testable predictions, for various kinds of language research.

(a) Linguistics
We have noted that, from a purely linguistic perspective, assumptions about hierarchical structure may be useful for describing and analysing sentences. However, if language use is best explained by sequential structure, then linguistic phenomena that have previously been explained in terms of hierarchical syntactic relationships may be captured by factors relating to sequential constraints, semantic considerations or pragmatic context. For example, binding theory [90] was proposed as a set of syntactic constraints governing the interpretation of referring expressions such as pronouns (e.g. her, them) and reflexives (e.g. herself, themselves). Increasingly, though, the acceptability of such referential resolution is being explained in non-hierarchical terms, such as constraints imposed by linear sequential order [91] in combination with semantic and pragmatic factors [92,93] (see [26] for discussion). We anticipate that many other types of purported syntactic constraints may similarly be amenable to reanalyses that deemphasize hierarchical structure in favour of non-hierarchical explanations.
The astonishing productivity and flexibility of human language has long been ascribed to its hierarchical syntactic underpinnings, assumed to be a defining feature that distinguishes language from other communication systems [16–18]. As such, hierarchical structure in explanations of language use has been a major obstacle for theories of human evolution that view language as being on a continuum with other animal communication systems. If, however, hierarchical syntax is no longer a hallmark of human language, then it may be possible to bridge the gap between human and non-human communication. Thus, instead of searching for what has largely been elusive evidence of hierarchical syntax-like structure in other animal communication systems, ethologists may make more progress in understanding the relationship between human language and non-human communication by investigating similarities and differences in other abilities likely to be crucial to language, such as sequence learning, pragmatic abilities, social intelligence and willingness to cooperate (cf. [94,95]).

**Cognitive neuroscience**

As a general methodological implication, our hypothesis would suggest a re-appraisal of the considerable amount of neuroimaging work which has assumed that language use is best characterized by hierarchical structure [96,97]. For example, one fMRI study indicated that a hierarchically structured artificial grammar elicited activation in Broca's area whereas a non-hierarchical grammar did not [30]. However, if our hypothesis is correct, then the differences in neural activation may be better explained in terms of the differences in the types of dependencies involved: non-adjacent dependencies in the hierarchical grammar and adjacent dependencies in the non-hierarchical grammar (cf. [31]). We expect that it may be possible to reinterpret the results of many other neuroimaging studies—purported to indicate the processing of hierarchical structure—in a similar fashion, in terms of differences in the dependencies involved or other constraints on sequence processing.

As another case in point, a recent fMRI study [98] revealed that activation in Broca's area increases when subjects read word-sequences with longer coherent constituents. Crucially, comprehension of the stimuli was not required, as subjects were tested on word memory and probe-sentence detection. Therefore, the results show that sequentially structured constituents are extracted even when this is not relevant to the task at hand. We predict that, under the same experimental conditions, there will be no effect of the depth of hierarchical structure (which was not manipulated in the original experiment). However, such an effect may appear if subjects are motivated to read for comprehension, if sentence meaning depends on the precise (hierarchical) sentence structure, and if extra-linguistic information (e.g. world knowledge) is not helpful. 

**Psychology**

The presence of long-distance dependencies in language has long been seen as important evidence in favour of hierarchical structure. Consider (14) where there is a long-distance dependency between spider and ate, interspersed in standard accounts by the hierarchically embedded relative clause that the bullfrog chased (which includes an adjacent dependency between bullfrog and chased).

(14) The spider that the bullfrog chased ate the fly.

From our perspective, such long-distance dependencies between elements of a sentence are not indicative of operations on hierarchical syntactic structures. Rather, they follow from predictive processing, that is, the first element’s occurrence (spider) results in anticipation of the second (dependent) element (ate). Thus, the difficulty of processing multiple overlapping non-adjacent dependencies does not depend on hierarchical structure but on the nature of the overlap (nested, as in (6), or crossed, as in (7)) and the number of overlapping dependencies (cf. [99]). Preliminary evidence from an SRT experiment supports this prediction [45]. However, further psychological experimentation is necessary to test the degree to which this prediction holds true of natural language processing in general.

Another key element of our account is that multi-word constructions are hypothesized to have no internal hierarchical structure but only a sequential arrangement of elements. We would therefore predict that the processing of constructions should be unaffected by their possible internal structure. That is, constructions with alleged hierarchical structure, such as [take [a moment]], should be processed in a non-compositional manner similar to linear constructions (e.g. knife and fork) or well-known idioms (e.g. spill the beans), which are generally agreed to be stored as whole chunks. Only the overall familiarity of the specific construction should affect processing. The fact that both children and adults are sensitive to the overall frequency of multi-word combinations [72–75] supports this prediction, but further studies are needed to determine how closely the representations of frequent ‘flat’ word sequences resemble that of possibly hierarchical constructions and idioms.

**Computer science**

Our hypothesis has potential implications for the subareas of computer science dealing with human language. Specifically, it suggests that more human-like speech and language processing may be accomplished by focusing less on hierarchical structure and dealing more with sequential structure. Indeed, in the field of Natural Language Processing, the importance of sequential processing is increasingly recognized: tasks such as machine translation and speech recognition are successfully performed by algorithms based on sequential structure. No significant performance increase is gained when these algorithms are based on or extended with hierarchical structure [101,102].

We also expect that the statistical patterns of language, as apparent from large text corpora, should be detectable to at least the same extent by sequential and hierarchical statistical models of language. Comparisons between particular RNNs and probabilistic phrase-structure grammars revealed that the RNNs' ability to model statistical patterns of English was only slightly below that of the hierarchical grammars [52–54]. However, these studies were not designed for that particular comparison so they are by no means conclusive.
6. CONCLUSION

Although it is possible to view sentences as hierarchically structured, this structure appears mainly as a side effect of exhaustively analysing the sentence by dividing it up into its parts, subparts, sub-subparts, etc. Psychologically, such hierarchical (de)composition is not a fundamental operation. Rather, considerations of simplicity and evolutionary continuity force us to take sequential structure as fundamental to human language processing. Indeed, this position gains support from a growing body of empirical evidence from cognitive neuroscience, psycholinguistics and computational linguistics.

This is not to say that hierarchical operations are non-existent, and we do not want to exclude their possible role in language comprehension or production. However, we expect that evidence for hierarchical operations will only be found when the language user is particularly attentive, when it is important for the task at hand (e.g. in meta-linguistic tasks) and when there is little relevant information from extra-sentential/linguistic context. Moreover, we stress that any individual demonstration of hierarchical processing does not prove its primacy in language use. In particular, even if some hierarchical grouping occurs in particular cases or circumstances, this does not imply that the operation can be applied recursively, yielding hierarchies of theoretically unbounded depth, as is traditionally assumed in theoretical linguistics. It is very likely that hierarchical combination is cognitively too demanding to be applied recursively. Moreover, it may rarely be required in normal language use.

To conclude, the role of the sequential structure of language has thus far been neglected in the cognitive sciences. However, trends are converging across different fields to acknowledge its importance, and we expect that it will be possible to explain much of human language behaviour using just sequential structure. Thus, linguists and psychologists should take care to only invoke hierarchical structure in cases where simpler explanations, based on sequential structure, do not suffice.

We would like to thank Inbal Arnon, Harald Baayen, Adele Goldberg, Stewart McCauley and two anonymous referees for their valuable comments. S.L.F. was funded by the EU 7th Framework Programme under grant no. 253803, R.B. by NWO grant no. 277-70-006 and M.H.C. by BWF grant no. 2011107.

ENDNOTES

1 For linguistic arguments and evidence for the primacy of sequential structure, see [14,15].

2 See [100] for an alternative non-hierarchical account of multi-word familiarity effects.

REFERENCES


10 Bloomfield, L. 1933 Language. Chicago, IL: University of Chicago Press.


12 Bresnan, J. 1982 Language and meaning in the analysis of structurally ambiguous sentences. Cogn. Psychol. 13, 90008-1)


78 Johnson, M. 1987 The body in the mind: the bodily basis of meaning, imagination, and reason. Chicago, IL: University of Chicago Press.
91 Culicover, P. W. Submitted. The role of linear order in the computation of referential dependencies.


100 Baayen, R. H. & Hendrix, P. 2011 Sidestepping the combinatorial explosion: towards a processing model based on discriminative learning. In *LSA workshop 'Empirically examining parsimony and redundancy in usage-based models'*.  
