PROCEEDINGS OF THE ROYAL SOCIETY B | BIOLOGICAL SCIENCES

How hierarchical is language use?

Stefan L. Frank, Rens Bod and Morten H. Christiansen

References	**This article cites 66 articles, 7 of which can be accessed free** http://rspb.royalsocietypublishing.org/content/early/2012/09/05/rspb.2012.1741.full.ht ml#ref-list-1
P<P	Published online 12 September 2012 in advance of the print journal.
Subject collections	Articles on similar topics can be found in the following collections cognition (178 articles)
Email alerting service	Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click **here**

Advance online articles have been peer reviewed and accepted for publication but have not yet appeared in
the paper journal (edited, typeset versions may be posted when available prior to final publication). Advance
online articles are citable and establish publication priority; they are indexed by PubMed from initial publication.
Citations to Advance online articles must include the digital object identifier (DOIs) and date of initial
publication.

To subscribe to *Proc. R. Soc. B* go to: **http://rspb.royalsocietypublishing.org/subscriptions**

Review

How hierarchical is language use?

Stefan L. Frank[1],*, Rens Bod[2] and Morten H. Christiansen[3]

[1]*Department of Cognitive, Perceptual and Brain Sciences, University College London,*
26 Bedford Way, London WC1H 0AP, UK
[2]*Institute for Logic, Language and Information, University of Amsterdam, Science Park 904,*
1098 XH Amsterdam, The Netherlands
[3]*Department 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

* Author for correspondence (s.frank@ucl.ac.uk).

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 linguistics), 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 structure 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 structure 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 evolutionary 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 mechanisms. Thus, language is shaped by constraints inherited from neural substrates predating the emergence of language, including constraints deriving 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 evolution 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 evolutionary 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 culturally evolved symbol system: arithmetic. Although arithmetic can be described in terms of hierarchical structure, this does not entail that the neural mechanisms employed for arithmetic use such structures. Rather, the considerable difficulty that children face in learning arithmetic suggests that the opposite is the case, probably because these mathematical skills reuse evolutionarily older neural systems [24]. But why, then, can children master language without much effort and explicit instruction? Cultural evolution provides the answer to this question, shaping language to fit our learning and processing 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, including cortical and subcortical areas (see [27] for a review). For example, brain activity recordings by electroencephalography 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 correlates with performance on sequence learning, with higher degrees of integrity associated with better learning [32].

If language is subserved by the same neural mechanisms as used for sequence processing, then we would expect a breakdown of syntactic processing to be associated 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 processing 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].

(3) *The coat with the ripped cuffs by the orange balls ***were*** ...

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 *Johanna helped the men teach Hans to feed the horses* (the subscripted indices show dependencies between nouns and verbs).

(6) Johanna$_1$ hat die Männer$_2$ Hans$_3$ die Pferde füttern$_3$ lehren$_2$ helfen$_1$
(7) Johanna$_1$ heeft de mannen$_2$ Hans$_3$ de paarden helpen$_1$ leren$_2$ voeren$_3$

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$_1$ that the bullfrog$_2$ that the turtle$_3$ followed$_3$ mercilessly ate$_2$ the fly.
(9) The spider$_1$ that the bullfrog$_2$ that the turtle$_3$ followed$_3$ chased$_2$ ate$_1$ 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).

(10) The man is hungry.
(11) Is the man hungry?

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 is eating hungry?
(13) *Is the man who is eating is 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
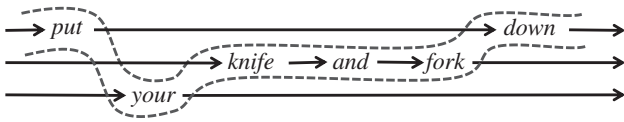
Figure 1. Combining constructions into a sentence by switching between parallel sequential streams. Note that the displayed vertical order of constructions is arbitrary.

constructions. However, such hierarchical combination of constructions is not a *necessary* component of sentence processing. For example, if each construction is taken to correspond to a sequential process, we can view sentence formation as arising from a number of sequential streams that run in parallel. As illustrated in figure 1, by switching between the streams, constructions are combined without any (de)compositional processing or creation of a part–whole relation—as a first approximation this might be somewhat analogous to time-division multiplexing in digital communication [80]. The figure also indicates how we view the processing of distal dependencies (such as between *put* and *down*), discussed in more detail in §5*d*.

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 §3*b*.

(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.

(b) *Ethology*

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 gulf 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]).

(c) *Cognitive neuroscience*

As a general methodological implication, our hypothesis would suggest a reappraisal 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.

(d) *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 psycholinguistic 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[2], but further studies are needed to determine how closely the representations of frequent 'flat' word sequences resemble that of possibly hierarchical constructions and idioms.

(e) *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 nonexistent, 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.

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

1 Frazier, L. & Rayner, K. 1982 Making and correcting errors during sentence comprehension: eye-movements in the analysis of structurally ambiguous sentences. *Cogn. Psychol.* **14**, 178–210. (doi:10.1016/0010-0285 (82)90008-1)

2 Jurafsky, D. 1996 A probabilistic model of lexical and syntactic access and disambiguation. *Cogn. Sci.* **20**, 137–194. (doi:10.1207/s15516709cog2002_1)

3 Vosse, T. & Kempen, G. 2000 Syntactic structure assembly in human parsing: a computational model based on competitive inhibition and a lexicalist grammar. *Cognition* **75**, 105–143. (doi:10.1016/S0010-0277(00)00063-9)

4 Kempen, G. & Hoenkamp, E. 1987 An incremental procedural grammar for sentence formulation. *Cogn. Sci.* **11**, 201–258. (doi:10.1207/s15516709cog1102_5)

5 Hartsuiker, R. J., Antón-Méndez, I. & Van Zee, M. 2001 Object attraction in subject–verb agreement construction. *J. Mem. Lang.* **45**, 546–572. (doi:10.1006/jmla.2000.2787)

6 Borensztajn, G., Zuidema, W. & Bod, R. 2009 Children's grammars grow more abstract with age: evidence from an automatic procedure for identifying the productive units of language. *Top. Cogn. Sci.* **1**, 175–188. (doi:10.1111/j.1756-8765.2008.01009.x)

7 Pinker, S. 1984 *Language learnability and language development*. Cambridge, MA: Harvard University Press.

8 Law, V. 2003 *The history of linguistics in Europe*. Cambridge, UK: Cambridge University Press.

9 Seuren, P. 1998 *Western linguistics: an historical introduction*. Oxford, UK: Oxford University Press.

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

11 Chomsky, N. 1957 *Syntactic structures*. The Hague, The Netherlands: Mouton.

12 Bresnan, J. 1982 *The mental representation of grammatical relations*. Cambridge, MA: The MIT Press.

13 Sag, I. & Wasow, T. 1999 *Syntactic theory: a formal introduction*. Stanford, CA: CSLI Publications.

14 Bybee, J. 2002 Sequentiality as the basis of constituent structure. In *The evolution of language out of pre-language* (eds T. Givón & B. F. Malle), pp. 109–134. Amsterdam, The Netherlands: John Benjamins.

15 Langacker, R. W. 2010 Day after day after day. In *Meaning, form, and body* (eds F. Parrill, V. Tobin & M. Turner), pp. 149–164. Stanford, CA: CSLI Publications.

16 Pinker, S. 2003 Language as an adaptation to the cognitive niche. In *Language evolution: states of the art* (eds M. H. Christiansen & S. Kirby), pp. 16–37. New York, NY: Oxford University Press.

17 Hauser, M. D., Chomsky, N. & Fitch, W. T. 2002 The faculty of language: what is it, who has it, and how did it evolve? *Science* **298**, 1569–1579. (doi:10.1126/science.298.5598.1569)

18 Chomsky, N. 2010 Some simple evo-devo theses: how true might they be for language? In *The evolution of human language* (eds R. K. Larson, V. Déprez & H. Yamakido), pp. 45–62. Cambridge, UK: Cambridge University Press.

19 Kirby, S., Dowman, M. & Griffiths, T. L. 2007 Innateness and culture in the evolution of language. *Proc. Natl Acad. Sci. USA* **104**, 5241–5245. (doi:10.1073/pnas.0608222104)

20 Chater, N., Reali, F. & Christiansen, M. H. 2009 Restrictions on biological adaptation in language evolution. *Proc. Natl Acad. Sci. USA* **106**, 1015–1020. (doi:10.1073/pnas.0807191106)

21 Christiansen, M. H. & Chater, N. 2008 Language as shaped by the brain. *Behav. Brain Sci.* **31**, 489–558. (doi:10.1017/S0140525×08004998)

22 Evans, N. & Levinson, S. C. 2009 The myth of language universals: language diversity and its importance for cognitive science. *Behav. Brain Sci.* **32**, 429–492. (doi:10.1017/S0140525×0999094x)

23 Dunn, M., Greenhill, S. J., Levinson, S. C. & Gray, R. D. 2011 Evolved structure of language shows lineage-specific trends in word-order universals. *Nature* **473**, 79–82. (doi:10.1038/Nature09923)

24 Dehaene, S. & Cohen, L. 2007 Cultural recycling of cortical maps. *Neuron* **56**, 384–398. (doi:10.1016/j.neuron.2007.10.004)

25 Fisher, S. E. & Scharff, C. 2009 FOXP2 as a molecular window into speech and language. *Trends Genet.* **25**, 166–177. (doi:10.1016/j.tig.2009.03.002)

26 Chater, N. & Christiansen, M. H. 2010 Language acquisition meets language evolution. *Cogn. Sci.* **34**, 1131–1157. (doi:10.1111/j.1551-6709.2009.01049.x)

27 Conway, C. M. & Pisoni, D. B. 2008 Neurocognitive basis of implicit learning of sequential structure and its relation to language processing. *Learn. Skill Acquisit. Read. Dyslexia* **1145**, 113–131. (doi:10.1196/annals.1416.009)

28 Christiansen, M. H., Conway, C. M. & Onnis, L. 2012 Similar neural correlates for language and sequential learning: evidence from event-related brain potentials. *Lang. Cogn. Process.* **27**, 231–256. (doi:10.1080/01690965.2011.606666)

29 Grodzinsky, Y. 2000 The neurology of syntax: language use without Broca's area. *Behav. Brain Sci.* **23**, 1–21. (doi:10.1017/S0140525X00002399)

30 Bahlmann, J., Schubotz, R. I. & Friederici, A. D. 2008 Hierarchical artificial grammar processing engages Broca's area. *Neuroimage* **42**, 525–534. (doi:10.1016/j.neuroimage.2008.04.249)

31 Petersson, K. M., Folia, V. & Hagoort, P. 2012 What artificial grammar learning reveals about the neurobiology of syntax. *Brain Lang.* **120**, 83–95. (doi:10.1016/j.bandl.2010.08.003)

32 Flöel, A., de Vries, M. H., Scholz, J., Breitenstein, C. & Johansen-Berg, H. 2009 White matter integrity in the vicinity of Broca's area predicts grammar learning success. *Neuroimage* **47**, 1974–1981. (doi:10.1016/j.neuroimage.2009.05.046)

33 Christiansen, M. H., Kelly, M. L., Shillcock, R. C. & Greenfield, K. 2010 Impaired artificial grammar learning in agrammatism. *Cognition* **116**, 382–393. (doi:10.1016/j.cognition.2010.05.015)

34 Patel, A. D., Iversen, J. R., Wassenaar, M. & Hagoort, P. 2008 Musical syntactic processing in agrammatic Broca's aphasia. *Aphasiology* **22**, 776–789. (doi:10.1080/02687030701803804)

35 De Vries, M. H., Barth, A. C. R., Maiworm, S., Knecht, S., Zwitserlood, P. & Flöel, A. 2010 Electrical stimulation of Broca's area enhances implicit learning of an artificial grammar. *J. Cogn. Neurosci.* **22**, 2427–2436. (doi:10.1162/jocn.2009.21385)

36 Uddén, J., Folia, V., Forkstam, C., Ingvar, M., Fernandez, G., Overeem, S., van Elswijk, G., Hagoort, P. & Petersson, K. M. 2008 The inferior frontal cortex in artificial syntax processing: an rTMS study. *Brain Res.* **1224**, 69–78. (doi:10.1016/j.brainres.2008.05.070)

37 Gillespie, M. & Pearlmutter, N. J. 2011 Hierarchy and scope of planning in subject–verb agreement production. *Cognition* **118**, 377–397. (doi:10.1016/j.cognition.2010.10.008)

38 Gillespie, M. & Pearlmutter, N. J. 2012 Against structural constraints in subject–verb agreement production. *J. Exp. Psychol. Learn. Mem. Cogn.* (doi:10.1037/a0029005)

39 Misyak, J. B. & Christiansen, M. H. 2010 When 'more' in statistical learning means 'less' in language: individual differences in predictive processing of adjacent dependencies. In *Proc. 32nd Annu. Conf. Cognitive Science Society* (eds S. Ohlsson & R. Catrambone), pp. 2686–2691. Austin, TX: Cognitive Science Society.

40 Vasishth, S. & Drenhaus, H. 2011 Locality in German. *Dialogue Discourse* **1**, 59–82. (doi:10.5087/dad.2011.104)

41 Bartek, B., Lewis, R. L., Vasishth, S. & Smith, M. R. 2011 In search of on-line locality effects in sentence comprehension. *J. Exp. Psychol. Learn. Mem. Cogn.* **37**, 1178–1198. (doi:10.1037/A0024194)

42 Tabor, W., Galantucci, B. & Richardson, D. 2004 Effects of merely local syntactic coherence on sentence processing. *J. Mem. Lang.* **50**, 355–370. (doi:10.1016/j.jml.2004.01.001)

43 Bach, E., Brown, C. & Marslen-Wilson, W. 1986 Crossed and nested dependencies in German and Dutch. A psycholinguistic study. *Lang. Cogn. Process.* **1**, 249–262. (doi:10.1080/01690968608404677)

44 Uddén, J., Ingvar, M., Hagoort, P. & Petersson, K. M. 2012 Implicit acquisition of grammars with crossed and nested non-adjacent dependencies: investigating the push-down stack model. *Cogn. Sci.* (doi:10.1111/j.1551-6709.2012.01235.x)

45 De Vries, M. H., Geukes, S., Zwitserlood, P., Petersson, K. M. & Christiansen, M. H. 2012 Processing multiple non-adjacent dependencies: evidence from sequence learning. *Phil. Trans. R. Soc. B* **367**, 2065–2076. (doi:10.1098/rstb.2011.041)

46 Christiansen, M. H. & Chater, N. 1999 Toward a connectionist model of recursion in human linguistic performance. *Cogn. Sci.* **23**, 157–205. (doi:10.1207/s15516709cog2302_2)

47 Christiansen, M. H. & MacDonald, M. C. 2009 A usage-based approach to recursion in sentence processing. *Lang. Learn.* **59**, 126–161. (doi:10.1111/j.1467-9922.2009.00538.x)

48 Gibson, E. & Thomas, J. 1999 Memory limitations and structural forgetting: the perception of complex ungrammatical sentences as grammatical. *Lang. Cogn. Process.* **14**, 225–248. (doi:10.1080/016909699386293)

49 Vasishth, S., Suckow, K., Lewis, R. L. & Kern, S. 2010 Short-term forgetting in sentence comprehension: crosslinguistic evidence from verb-final structures. *Lang. Cogn. Process.* **25**, 533–567. (doi:10.1080/01690960903310587)

50 Engelmann, F. & Vasishth, S. 2009 Processing grammatical and ungrammatical center embeddings in English and German: a computational model. In *Proc. 9th Int. Conf. Cognitive Modeling* (eds A. Howes, D. Peebles & R. Cooper), pp. 240–245. Manchester, UK.

51 Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V. & Hagoort, P. 2005 Anticipating upcoming words in discourse: evidence from ERPs and reading times. *J. Exp. Psychol. Learn. Mem. Cogn.* **31**, 443–467. (doi:10.1037/0278-7393.31.3.443)

52 Frank, S. L. & Bod, R. 2011 Insensitivity of the human sentence-processing system to hierarchical structure. *Psychol. Sci.* **22**, 829–834. (doi:10.1177/0956797611409589)

53 Fernandez Monsalve, I., Frank, S. L. & Vigliocco, G. 2012 Lexical surprisal as a general predictor of reading time. In *Proc. 13th Conf. European chapter of the Association for Computational Linguistics* (ed. W. Daelemans), pp. 398–408. Avignon, France: Association for Computational Linguistics.

54 Frank, S. L. & Thompson, R. L. 2012 Early effects of word surprisal on pupil size during reading. In *Proc. 34th Annu. Conf. Cognitive Science Society* (eds N. Miyake, D. Peebles & R. P. Cooper), pp. 1554–1559. Austin, TX: Cognitive Science Society.

55 Conway, C. M., Bauernschmidt, A., Huang, S. S. & Pisoni, D. B. 2010 Implicit statistical learning in language processing: word predictability is the key. *Cognition* **114**, 356–371. (doi:10.1016/j.cognition.2009.10.009)

56 Misyak, J. B., Christiansen, M. H. & Tomblin, J. B. 2010 Sequential expectations: the role of prediction-based learning in language. *Top. Cogn. Sci.* **2**, 138–153. (doi:10.1111/j.1756-8765.2009.01072.x)

57 Crain, S. 1991 Language acquisition in the absence of experience. *Behav. Brain Sci.* **14**, 597–612. (doi:10.1017/S0140525X00071491)

58 Chomsky, N. 1980 The linguistic approach. In *Language and learning: the debate between Jean Piaget and Noam Chomsky* (ed. M. Piatelli-Palmarini), pp. 109–130. Cambridge, MA: Harvard University Press.

59 Crain, S. & Pietroski, P. 2001 Nature, nurture and universal grammar. *Linguist. Phil.* **24**, 139–186. (doi:10.1023/A:1005694100138)

60 Crain, S. & Nakayama, M. 1987 Structure dependence in grammar formation. *Language* **63**, 522–543. (doi:10.2307/415004)

61 Reali, F. & Christiansen, M. H. 2005 Uncovering the richness of the stimulus: structural dependence and indirect statistical evidence. *Cogn. Sci.* **29**, 1007–1028. (doi:10.1207/s15516709cog0000_28)

62 Kam, X., Stoyneshka, L., Tornyova, L., Fodor, J. & Sakas, W. 2008 Bigrams and the richness of the stimulus. *Cogn. Sci.* **32**, 771–787. (doi:10.1080/03640210802067053)

63 Clark, A. & Eyraud, R. 2006 Learning auxiliary fronting with grammatical inference. In *Proc. 10th Conf. Computational Natural Language Learning*, pp. 125–132. New York, NY: Association for Computational Linguistics.

64 Fitz, H. 2010 Statistical learning of complex questions. In *Proc. 32nd Annu. Conf. Cognitive Science Society* (eds S. Ohlsson & R. Catrambone), pp. 2692–2698. Austin, TX: Cognitive Science Society.

65 Adger, D. 2003 *Core syntax: a minimalist approach.* Oxford, UK: Oxford University Press.

66 Bod, R. & Smets, M. 2012 Empiricist solutions to nativist puzzles by means of unsupervised TSG. In *Proc. Workshop on Computational Models of Language Acquisition and Loss*, pp. 10–18. Avignon, France: Association for Computational Linguistics.

67 Wexler, K. 1998 Very early parameter setting and the unique checking constraint: a new explanation of the optional infinitive stage. *Lingua* **106**, 23–79. (doi:10.1016/S0024-3841(98)00029-1)

68 Freudenthal, D., Pine, J. M. & Gobet, F. 2009 Simulating the referential properties of Dutch, German, and English root infinitives in MOSAIC. *Lang. Learn. Develop.* **5**, 1–29. (doi:10.1080/15475440802502437)

69 McCauley, S. & Christiansen, M. H. 2011 Learning simple statistics for language comprehension and production: the CAPPUCCINO model. In *Proc. 33rd Annu. Conf. Cognitive Science Society* (eds L. Carlson, C. Hölscher & T. Shipley), pp. 1619–1624. Austin, TX: Cognitive Science Society.

70 Saffran, J. R. 2002 Constraints on statistical language learning. *J. Mem. Lang.* **47**, 172–196. (doi:10.1006/jmla.2001.2839)

71 Goldberg, A. E. 2006 *Constructions at work: the nature of generalization in language.* Oxford, UK: Oxford University Press.

72 Arnon, I. & Snider, N. 2010 More than words: frequency effects for multi-word phrases. *J. Mem. Lang.* **62**, 67–82. (doi:10.1016/j.jml.2009.09.005)

73 Bannard, C. & Matthews, D. 2008 Stored word sequences in language learning: the effect of familiarity on children's repetition of four-word combinations. *Psychol. Sci.* **19**, 241–248. (doi:10.1111/j.1467-9280.2008.02075.x)

74 Siyanova-Chantuira, A., Conklin, K. & Van Heuven, W. J. B. 2011 Seeing a phrase 'time and again' matters: the role of phrasal frequency in the processing of multi-word sequences. *J. Exp. Psychol. Learn. Mem. Cogn.* **37**, 776–784. (doi:10.1037/a0022531)

75 Tremblay, A., Derwing, B., Libben, G. & Westbury, C. 2011 Processing advantages of lexical bundles: evidence from self-paced reading and sentence recall tasks. *Lang.*

Learn. **61**, 569–613. (doi:10.1111/j.1467-9922.2010.00622.x)

76 Fodor, J. A. 1975 *The language of thought.* Cambridge, MA: Harvard University Press.

77 Johnson-Laird, P. N. 1983 *Mental models.* Cambridge, UK: Cambridge University Press.

78 Johnson, M. 1987 *The body in the mind: the bodily basis of meaning, imagination, and reason.* Chicago, IL: University of Chicago Press.

79 Barsalou, L. W. 1999 Perceptual symbol systems. *Behav. Brain Sci.* **22**, 577–660.

80 Stern, H. P. E. & Mahmoud, S. A. 2004 *Communication systems: analysis and design.* Upper Saddle River, NJ: Prentice-Hall.

81 Takac, M., Benuskova, L. & Knott, A. In press. Mapping sensorimotor sequences to word sequences: a connectionist model of language acquisition and sentence generation. *Cognition* (doi:10.1016/j.cognition.2012.06.006)

82 Ding, L., Dennis, S. & Mehay, D. N. 2009 A single layer network model of sentential recursive patterns. In *Proc. 31st Annu. Conf. Cognitive Science Society* (eds N. A. Taatgen & H. Van Rijn), pp. 461–466. Austin, TX: Cognitive Science Society.

83 Pulvermüller, F. 2010 Brain embodiment of syntax and grammar: discrete combinatorial mechanisms spelt out in neuronal circuits. *Brain Lang.* **112**, 167–179. (doi:10.1016/j.bandl.2009.08.002)

84 Townsend, D. J. & Bever, T. G. 2001 *Sentence comprehension: the integration of habits and rules.* Cambridge, MA: MIT Press.

85 Ferreira, F. & Patson, N. 2007 The good enough approach to language comprehension. *Lang. Ling. Compass* **1**, 71–83. (doi:10.1111/j.1749-818x.2007.00007.x)

86 Sanford, A. J. & Sturt, P. 2002 Depth of processing in language comprehension: not noticing the evidence. *Trends Cogn. Sci.* **6**, 382–386. (doi:10.1016/S1364-6613(02)01958-7)

87 Kamide, Y., Altmann, G. T. M. & Haywood, S. L. 2003 The time-course of prediction in incremental sentence processing: evidence from anticipatory eye movements. *J. Mem. Lang.* **49**, 133–156. (doi:10.1016/S0749-596X(03)00023-8)

88 Otten, M. & Van Berkum, J. J. A. 2008 Discourse-based word anticipation during language processing: prediction or priming? *Discourse Processes* **45**, 464–496. (doi:10.1080/01638530802356463)

89 Frank, S. L., Haselager, W. F. G. & van Rooij, I. 2009 Connectionist semantic systematicity. *Cognition* **110**, 358–379. (doi:10.1016/j.cognition.2008.11.013)

90 Chomsky, N. 1981 *Lectures on Government and Binding. The Pisa Lectures.* Dordecht, The Netherlands: Foris.

91 Culicover, P. W. Submitted. The role of linear order in the computation of referential dependencies.

92 Culicover, P. W. & Jackendoff, R. 2005 *Simpler syntax.* New York, NY: Oxford University Press.

93 Levinson, S. C. 1987 Pragmatics and the grammar of anaphora: a partial pragmatic reduction of binding and control phenomena. *J. Linguist.* **23**, 379–434. (doi:10.1017/S0022226700011324)

94 Wheeler, B. *et al.* 2011 Communication. In *Animal thinking: contemporary issues in comparative cognition* (eds R. Menzel & J. Fischer), pp. 187–205. Cambridge, MA: MIT Press.

95 Tomasello, M. & Herrmann, E. 2010 Ape and human cognition: what's the difference? *Curr. Direct. Psychol. Res.* **19**, 3–8. (doi:10.1177/0963721409359300)

96 Grodzinsky, Y. & Santi, S. 2008 The battle for Broca's region. *Trends Cogn. Sci.* **12**, 474–480. (doi:10.1016/j.tics.2008.09.001)

97 Friederici, A., Bahlmann, J., Friederich, R. & Makuuchi, M. 2011 The neural basis of recursion and complex syntactic hierarchy. *Biolinguistics* **5**, 87–104.

98 Pallier, C., Devauchelle, A.-D. & Dehaene, S. 2011 Cortical representation of the constituent structure of sentences. *Proc. Natl Acad. Sci.* **108**, 2522–2527. (doi:10.1073/pnas.1018711108)

99 De Vries, M. H., Christiansen, M. H. & Petersson, K. M. 2011 Learning recursion: multiple nested and crossed dependencies. *Biolinguistics* **5**, 10–35.

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'*.

101 Baker, J., Deng, L., Glass, J., Lee, C., Morgan, N. & O'Shaughnessy, D. 2009 Research developments and directions in speech recognition and understanding, part 1. *IEEE Signal Process. Mag.* **26**, 75–80. (doi:10.1109/MSP.2009.932166)

102 Koehn, P. 2009 *Statistical machine translation*. Cambridge, UK: Cambridge University Press.