nature reviews neuroscience

https://doi.org/10.1038/s41583-022-00659-5

Perspective

Check for updates

What neural oscillations can and cannot do for syntactic structure building

Nina Kazanina 1,2 & Alessandro Tavano 3

Abstract

Understanding what someone says requires relating words in a sentence to one another as instructed by the grammatical rules of a language. In recent years, the neurophysiological basis for this process has become a prominent topic of discussion in cognitive neuroscience. Current proposals about the neural mechanisms of syntactic structure building converge on a key role for neural oscillations in this process, but they differ in terms of the exact function that is assigned to them. In this Perspective, we discuss two proposed functions for neural oscillations – chunking and multiscale information integration – and evaluate their merits and limitations taking into account a fundamentally hierarchical nature of syntactic representations in natural languages. We highlight insights that provide a tangible starting point for a neurocognitive model of syntactic structure building.

Sections

---

[1]University of Bristol, Bristol, UK. [2]Higher School of Economics, Moscow, Russia. [3]Max Planck Institute for Empirical Aesthetics, Frankfurt, Germany. e-mail: nina.kazanina@bristol.ac.uk

Introduction

How does the brain understand linguistic utterances? Two points seem certain. First, interpretations cannot be retrieved holistically from long-term memory because utterances can be completely novel and unpredictable[1–3]. Thus, constructing an interpretation of utterances requires the compositional combination of word meanings. Second, a sentence is not merely a linear sequence of words. Rather, from each sentence, we can infer a latent syntactic structure that defines hierarchical relationships between words and word groups[1–4] (Box 1). Interpreting a sentence requires assigning each word into a position in a latent syntactic structure, constructed on the basis of the linguistic input and the comprehender's grammatical knowledge. Psycholinguistics and computational linguists have uncovered some psychological and computational principles behind parsing a string of words into a syntactic representation[5–7]; however, how syntactic structure is inferred in neural terms remains a mystery.

Pioneering work by Ding and colleagues[8] generated an active discussion regarding the role of neural oscillations in syntactic structure building. In their study, people were exposed to a highly regular mini-language, in which words, noun and verb phrases, and sentences all occurred isochronously at a rate of 4 Hz, 2 Hz and 1 Hz, respectively. The participants' magnetoencephalography responses during language exposure were analysed in the spectral domain and contained peaks at the frequencies corresponding to the rates of different units in the syntactic hierarchy; that is, a 4-Hz peak corresponding to words, a 2-Hz peak corresponding to phrases and a 1-Hz peak relating to sentences. On the basis of these findings, the study authors argued that "concurrent neural tracking of hierarchical linguistic structures provides a plausible functional mechanism for temporally integrating smaller linguistic units into larger structures" and proposed that neural oscillatory dynamics – or 'neural oscillations' for short – were essential to implement such tracking[8]. The idea quickly gained popularity and sprouted into a range of ideas about the role of neural oscillations in syntactic processing[9–20].

In terms of the operations carried out by neural oscillations in the domain of syntactic structure building, the proposals above can be classified into two groups. The first group comprises proposals in which neural oscillations linearly chunk (or segment) the linguistic input into syntactically relevant units through alignment of oscillations with the units. The alignment, usually brought about via phase resetting, can be driven by acoustic landmarks and/or endogenously generated word grouping. The second group is made up of proposals in which neural oscillations perform the task of multiscale integration of information across timescales and/or brain regions to establish hierarchical syntactic relationships between words or groups of words. Most broadly, this function is brought about via cross-frequency interaction; that is, a coordinated interplay of rhythmic activity across different frequencies.

The distinction between the two groups of proposals – which we refer to as the 'oscillations for chunking' and 'oscillations for integration' views for short – is not always perspicuous in the cognitive neuroscience of language literature, and these categories are often not clearly separated, leading to serious confusion (although see ref.[21] for a notable exception). Much neurocognitive research that uses spoken speech stimuli (that is, that focuses on the auditory modality) in discussing the role of oscillations in syntactic structure building refers to the process of chunking. Such a process is often assumed to yield the most basic syntactic phrases, which subsequently are combined into larger phrases or sentences. In the first part of this Perspective, we argue that, contrary to this line of work, the chunking function of oscillations cannot be considered essential for syntactic structure building. However, a role for oscillations in integration may be pivotal and, in the second part of this Perspective, we review several approaches that have used an integratory function of neural oscillations for structure building, highlighting their strengths and limitations.

Before we continue, it is worth clarifying what is meant by the term 'neural oscillations' (Box 2). As noted by Giraud[21], this term has been used loosely in cognitive neuroscience research, including research on speech and language. Rather than strictly denoting self-sustained activity from a single oscillatory component, the term 'oscillation' has (also) been used to refer to a process originating in population rhythmic activity of any origin. In Giraud's words[21], 'oscillations' refers to the "discrete and collective activity of neurons (spiking, bursting, post-synaptic potentials etc.) that tend to occur in synchrony when they are commonly engaged in a task, i.e. when they operate as a cohesive functional network" (for a similar point, see ref.[22]), and that is how we use the term here.

Oscillations for chunking

A prevalent view of the role of neural oscillations in syntactic structure building is that they enable the segmenting of strings of words into higher-level syntactic units. Essentially, a designated frequency band – usually delta – is assumed to phase-align with syntactic phrases in the linguistic input, thus chunking the input into such units (for example, see ref.[12]). Physiologically, the idea is grounded in demonstrations of the effects of neural entrainment on sensory and behavioural performance, for example, in research on attention. Theories of rhythmic attention define entrainment as the alignment of neural phase with stimulus phase[23,24]. Phase alignment is assumed to operate through sequential phase resetting of ongoing neural oscillations in response to periodic sensory landmarks in the input (Box 2), thereby adjusting both the period (frequency) of internal oscillations and their phase to track the external stimulus[24,25]. As neural oscillations are assumed to reflect moments of maximal versus minimal neural excitability, at any given point in time, the information conveyed by the external rhythm will be more likely to be gated in if it is aligned with the maximal excitability portion of the neural cycle, and gated out or minimized otherwise[26]. In its application to linguistic chunking, the alignment is achieved via tracking a (semi-)rhythmically occurring acoustic property (usually carried by the speech envelope) that is assumed to correlate with syntactic phrases[10,11]. Alternatively, if acoustic counterparts in speech are considered insufficient, neural entrainment is achieved by using prior grammatical knowledge; that is, oscillations align with endogenously constructed syntactic units[9,12,13]. Meyer and colleagues[13] refer to this distinction as "entrainment proper" versus "intrinsic synchronicity". In either case, the key aspect is that the oscillation frequency is interpreted literally, so that the oscillation period is commensurate with the duration of the chunk to be formed. Building a unit at some level of the syntactic hierarchy that has a physical duration of *n* seconds necessitates an oscillation with period 1/*n* (Fig. 1).

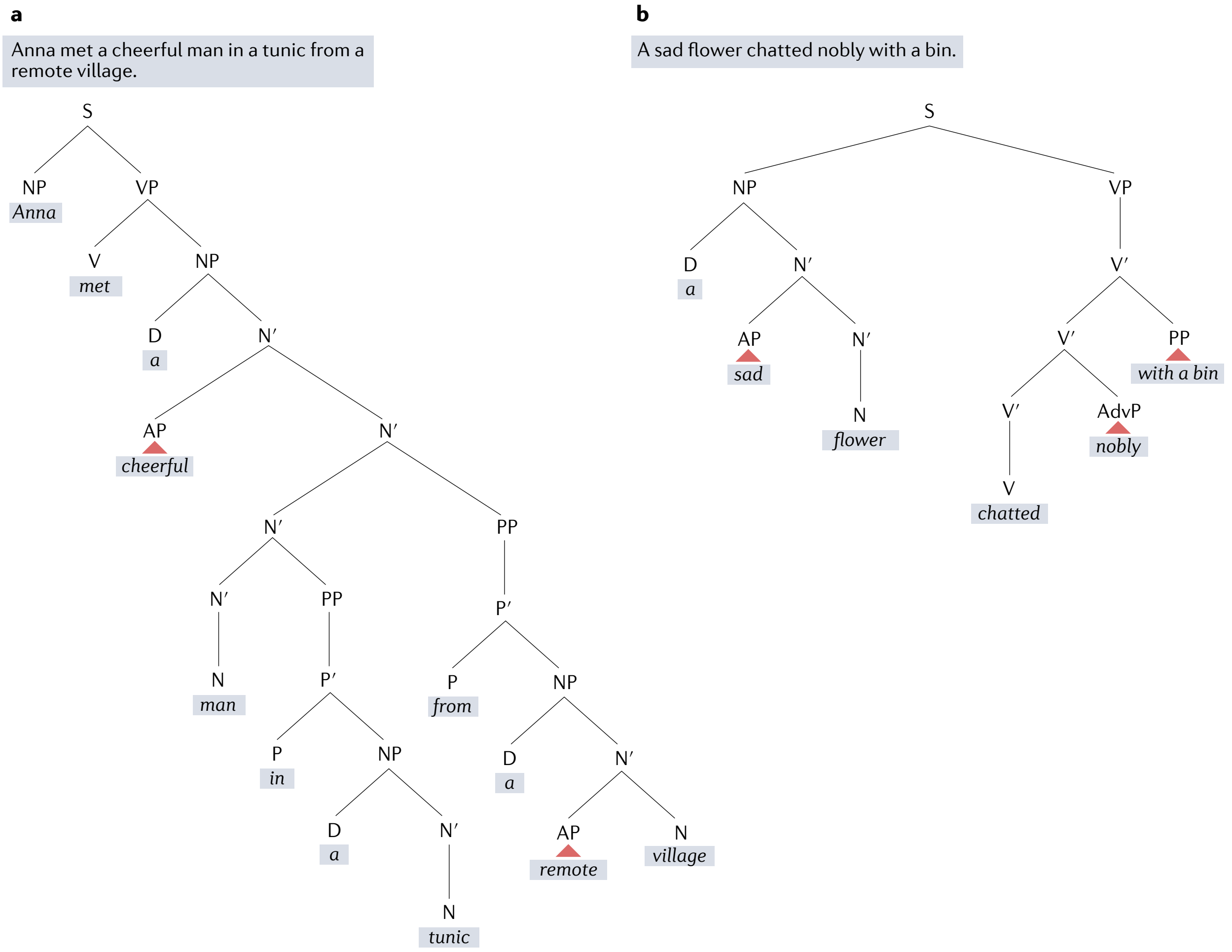
Historically, this 'chunking' view derives from two sources, one theoretical and one empirical. On the theoretical side, an analogy is attempted with the neurocognitive model of speech processing of Giraud and Poeppel[27], in which oscillations segment continuous speech into syllables (see also ref.[28]). In this model, the phase of theta oscillations in the 4–7-Hz range aligns with the amplitude modulation of the speech envelope[29], which in turn strongly correlates with syllabic unit boundaries[27]. Entrainment to the speech envelope and chunking into syllable sequences are possible because the syllable durations robustly fall within the theta band[30] and the syllable sequences are linear; that is, syllables – be they within or across words – strictly follow one another.

Box 1

Hierarchy in syntax

How do we understand sentences, including novel sentences (see the figure, part **a** (top)) or even sentences that are both novel and make no common sense (see the figure, part **b** (top))? We understand them because we possess the knowledge of grammar that guides how words must be related to one another and how the sentence should be interpreted.

Although words in a sentence appear in the input linearly, the grammatical relations between them may not respect the linear order: in the figure, part **b**, 'nobly' forms a relationship with the verb 'chatted' immediately to the left ('chatted nobly') but not with the preposition 'with' immediately to the right ('nobly with' does not make sense on its own). More significantly, (groups of) words that are syntactically related may be far apart in the linear input; for example, in 'The boy who fed the dog ran away', 'the boy' and 'ran away' are grammatically related to each other despite being at the opposite ends of the sentence (as witnessed both by the sentence meaning and by the elements commonly being described as the main subject and verb). This demonstrates that sentence-internal relationships are not determined by the linear order of the words; rather, sentences are organized using a multilayer hierarchical representation (see ref.[2] for an excellent comprehensive discussion or ref.[106] for a more succinct demonstration of this point with links to



Anna met a cheerful man in a tunic from a remote village.

A sad flower chatted nobly with a bin.

c

Anna met a cheerful man in a tunic from a remote village ...	
and Maya met a gloomy one in a kaftan from a city.	One = man
and Maya met a gloomy one from a city.	One = man in a tunic
and Maya met a gloomy one.	One = man in a tunic from a remote village

(continued)

graph theory). The hierarchical organization is often represented using syntactic tree diagrams in which linear adjacency loses its defining role (see the figure, parts **a**,**b** (bottom)). Instead, the trees highlight the fact that some words combine more closely to form so-called syntactic phrases or constituents. Any sequence of words united by a single node is a constituent. Using the text in part **a** of the figure as an example, we see that 'remote village', 'a remote village', 'from a remote village', 'a tunic', 'in a tunic', 'man in a tunic', 'man in a tunic from a remote village', 'cheerful man in a tunic from a remote village', 'a cheerful man in a tunic from a remote village', 'met a cheerful man in a tunic from a remote village' and, finally, 'Anna met a cheerful man in a tunic from a remote village' are each a syntactic phrase or constituent. By contrast, 'a remote', 'from a remote' and 'a tunic from a' are not. Syntactic constituents behave as a unit in that they can further combine with other words or with each other to produce even larger constituents until the whole sentence is encompassed. Syntactic constituency represents an important aspect of sentence structure. The trees in parts **a**,**b** of the figure also illustrate an important point of syntactic constituents being nested within one another: the phrase 'a tunic' is a subpart of the phrase 'in a tunic', which is a subpart of the larger phrase 'man in a tunic', which is itself nested within the even larger phrase 'man in a tunic from a remote village', and so on.

For some sentences, the syntactic structure is rather deep, with many intermediate nodes and constituents. Hence, one can ask a reasonable question of whether all these intermediate levels are indeed necessary and 'psychologically real'. Looking at how the sentence in part **a** of the figure can be legitimately continued proves useful (see the figure, part **c**). In the continuations, the word 'one' replaces one or more words from the original sentence, and the interpretation that native speakers of English assign to 'one' is shown on the right. In particular, in the middle and bottom lines, 'one' corresponds to a group of words. That a group of words can be replaced by a single word demonstrates that the brain indeed treats each of these word sequences as a unit; that is, the fact that only selected subsets of words can be replaced by a single word (while many others cannot) attests to their unit-like status in the hierarchical syntactic structure as shown in the tree in part **b** of the figure (for the curious, the generalization is that 'one' can replace any N-bar (N′) constituent in the tree[107]).

In the trees in parts **a**,**b** of the figure, the triangles indicate that some of the structures are not shown in full (in the interest of space). The grammatical knowledge here comes in the form of rules such as noun phrase (NP)+verb phrase (VP)→sentence (S), V+NP→VP and so on that combine smaller elements within the sentence into larger ones. A, adjective; Adv, adverb; Adv P, adverb phrase; AP, adjective phrase; D, determiner; N, noun; P, preposition; P′, P-bar; PP, prepositional phrase; V, verb; V′, V-bar.

In work such as refs.[10,11], both assumptions are extended to syntactic phrases. The relevant oscillations for syntax are those with a period corresponding to the physical (usually acoustic) duration of syntactic phrases. With regard to linearity, the sentence is viewed as a sequence of separate, linearly ordered syntactic phrases of roughly comparable length that follow each other without overlap (Fig. 1); the process of syntactic structure building is reduced to the linear grouping of words into syntactic phrases rather than building a hierarchy in which phrases can be nested within and dominate each other (Box 1).

The empirical source for the chunking view is rooted in the finding of spectral responses to latent syntactic properties of sentences in the frequency-tagging study of Ding and colleagues[8]. At the heart of the frequency-tagging method lies the observation that when a specific stimulus or feature occurs regularly in the input, the neural population or populations coding for it rhythmically responds or respond at the same frequency, thus 'tagging' the stimulus or feature[31]. This response presents itself in the electroencephalography or magnetoencephalography power spectrum as a peak at the stimulus frequency that can be teased apart from background neural activity. Ding et al. used highly regular sentences consisting of two two-word phrases, namely a noun phrase followed by a verb phrase as in 'dry fur rubs skin'; all words were monosyllabic and isochronously presented at a rate of 250 ms per syllable[8]. This 'toy' language was designed to have strict isomorphism between levels of syntactic representation and their corresponding acoustic durations; that is, each word, phrase and sentence was always 250 ms, 500 ms and 1,000 ms long, respectively. Ding et al. found spectral peaks ('frequency tags') at rates of 4 Hz, 2 Hz and 1 Hz, corresponding to three levels of syntactic hierarchy, and concluded that "cortical activity of different timescales concurrently tracked the time course of abstract linguistic structures at different hierarchical levels, such as words, phrases and sentences"[8]. (A similar design has since been used in other articles, including refs.[32–36], and slower peaks corresponding to the rate of larger syntactic groupings have been found, although the studies differed in whether a specifically syntactic origin for such peaks could be endorsed.) Subsequently, this conclusion was generalized beyond a toy, regular language and adapted into the claim that structure building in natural language is achieved via a set of neural oscillations each corresponding to a given level of the syntactic hierarchy that chunk linguistic input into units of relevant size (Fig. 1). A concurrent assumption is that the oscillations work in parallel and combine temporally smaller linguistic units into larger ones.

However, the approach in which syntactic structure building is based on the 'oscillation for chunking' idea is inadequate when one considers the syntactic properties of natural language. Indeed, here we present five arguments challenging the view that the syntactic structure is built by neural oscillations chunking the linguistic input. We argue that the assumption of semi-regular duration of syntactic phrases and the assumption of linearity in syntax are untenable on linguistic grounds. We also demonstrate that the assumption of an isomorphic match between levels of syntactic representation and temporal integration window sizes (neural oscillation rhythms) misses the essence of syntactic structure and presents an unrealistic basis for a neural mechanism of syntactic structure building. Finally, we argue that the chunking view misconstrues the process of syntactic structure building.

A narrow rhythm for phrases cannot be set

At the heart of syntactic structure building in all chunking approaches sits the notion of a phrase-level oscillation that yields word strings corresponding to syntactic phrases (Fig. 1). How tenable is this assumption? Indeed, is there a 'phrase rhythm'?

A prerequisite for using oscillations to yield chunks of any given level (syllable, word, phrase and so on) is that the chunks have a property of being regular or quasi-regular. For example, Giraud and Poeppel's claim[27] of neural tracking of syllables via a theta oscillation critically relies on the fact that across languages the mean syllabic rate of speech is 5–8 Hz (ref.[30]). In other words, syllable length is quasi-regular, and any variation in length falls in the theta range. But unlike syllables, syntactic phrases, including phrases of the same type such

Box 2

Neural oscillations

In 1924, Hans Berger recorded for the first time neural activity directly from the surface of a human brain and immediately recognized rhythmicity as an important feature, isolating two spontaneous oscillatory components: a slower one — at approximately 10 Hz — termed 'alpha waves' and a faster one — at approximately 20 Hz — termed 'beta waves'[108]. The interest in neural oscillatory activity has slowly increased in neuroscience. In addition to the alpha band (8–13 Hz) and the beta band (13–30 Hz), other frequency bands of interest were established: delta (1–4 Hz), theta (4–8 Hz), gamma (greater than 30 Hz)[109] and, more recently, sub-delta (less than 1 Hz)[110] (note that the precise band values differ somewhat in the literature; for example, the delta band is defined as 1.5–4 Hz in ref.[111], as 2–4 Hz in ref.[112] and as 0.5–3 Hz in ref.[113]).

Most straightforwardly, neural oscillatory activity is generated by an oscillator, a dynamical physical system that self-generates a periodic displacement between states about a central position (equilibrium point), giving rise to a sustained rhythmic behaviour[114,115]. The sustained attribute is important: the cycle generated by the oscillator must be recurrent, as opposed to being simply a phasic, transient response or a damped oscillatory response (a single resonating response to a stimulus). Oscillation is an intrinsic property of synchronously activated neurons constituting an ensemble[116]. When an ensemble is activated, neural oscillations emerge as fluctuations of local field potentials (LFPs)[117]. The LFP is a collective, mesoscopic-scale signal recorded from the extracellular space around neurons in the vicinity of a depth recording electrode; it corresponds to a summation of excitatory and inhibitory dendritic potentials of these neurons (reviewed in ref.[118]). LFPs can be directly recorded using intracortical electrodes. Electroencephalography signals recorded on the scalp relate to the same biophysical process that generates LFPs, just at a macroscopic scale[119].

An endogenous neural oscillator synchronizes its member neurons to fire with a given natural frequency (also known as the eigenfrequency), regulating the probability with which each member neuron fires relative to the phase of the ensemble's oscillation and yielding alternating high-excitability and low-excitability phases. This leads to the possibility of neural entrainment (that is, phase alignment between an internal rhythm and an external rhythm so that the most informative portion of the external stimulus can be optimally processed within high-excitability phases of the internal rhythm). Typically, successful neural entrainment requires the neural oscillator's eigenfrequency to be sufficiently close to that of the input to match that of the rhythmic input[25].

It is important to emphasize that entrainment does not require the external stimulus to be strictly periodic. A certain amount of jitter in its period is tolerated by the neural oscillator[24], and the ensuing flexibility is clearly a benefit when the signal to be tracked is only quasi-regular, as with speech or music. At the same time, it remains unclear whether neural oscillations recorded in response to quasi-regular stimuli are indeed generated by an oscillator-like mechanism or represent a chain of transient responses[120–122]. Serial phase resetting in a true neural oscillator — for example, caused by a succession of salient acoustic landmarks such as rapid and large changes in a speech amplitude envelope[123] — can be mimicked by the superposition of responses to sensory stimuli in a resonant neural system or by the concatenation of evoked responses (see the figure). Finally, an important question remains as to whether neural oscillations causally shape perception, cognition and action[124] or are simply a by-product of how the brain captures salient sensory events[125].

The figure shows that acoustic landmarks reset the phase of an oscillation to specific, periodic reference sensory points, such as rapid changes in acoustic energy, which end up effectively controlling the frequency and phase of a neural oscillator (that is, neural oscillations align with the stimulus acoustic envelope; left panel). The phase-reset effects of a neural oscillator can be closely mimicked by the superposition of the resonant responses to a succession of salient transitions (for example, silence-to-sound transitions), as each response arrives at the same phase of the neural cycle (middle panel). A similar effect could be obtained via the concatenation of evoked responses occurring in close proximity (right panel).

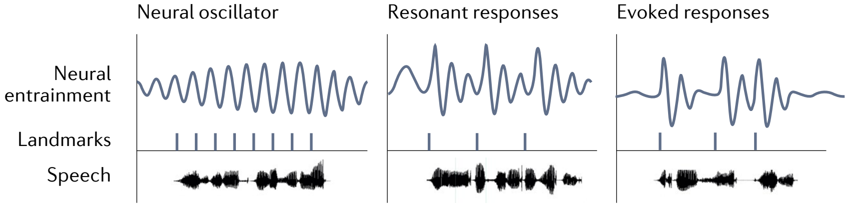


Figure adapted with permission from ref.[121], Elsevier.

Fig. 1 | Neural oscillations chunk linguistic input into syntactic phrases. Figures (abridged) from three studies exemplifying the oscillations for chunking approach. **a**, An example sentence from Kaufeld et al.[10] segmented into a linear sequence of syllables, words or phrases. Each of the levels of the linguistic hierarchy is tracked via a dedicated band of oscillations: 3.5–5 Hz for syllables, 1.9–2.8 Hz for words and 0.8–1.1 Hz for phrases. An acoustic waveform for the sentence is shown at the top. **b**, An equivalent example from Keitel et al.[11]. The frequency bands for syllables, words and phrases are set as 2.8–4.8 Hz, 1.8–3 Hz and 0.6–1.3 Hz, respectively. **c**, An example sentence from Meyer[12] illustrating a sequence of words being chunked into syntactic phrases via delta-band cycles (0.5–4 Hz). In each part, the frequency band boundaries are shown as defined in each original study. Part **a** adapted with permission from ref.[10], Society for Neuroscience. Part **b** adapted with permission from ref.[11], PLOS. Part **c** adapted with permission from ref.[12], Wiley.

as noun phrases, may differ markedly in length (even within a single discourse episode). First, syntactic phrases consist of words or morphemes and thus inherit word or morpheme length differences. For example, the noun phrase 'worms' has fourfold fewer syllables than the noun phrase 'caterpillars'. Second, the number of words within a phrase may differ: 'worms' versus 'young worms' versus 'long young worms'. It takes only seconds to find real-life examples of very long phrases containing many multisyllabic words: for example, 'the California public employees' retirement system' and 'the nine-day Albuquerque International Balloon Fiesta' (which contain 17 and 18 syllables, respectively); both examples are from the Penn Treebank[37]. Still, despite massive variation in length, these examples are all noun phrases that are completely interchangeable in terms of their syntactic function; that is, one can be used in the same syntactic positions as the others. According to the assumption of a neural oscillation per level of syntactic hierarchy, these noun phrases must all be the outcome of the same neural process; that is, chunking using the same phrase-level oscillation. However, marked length differences render this scenario unfeasible.

The variation in the length of syntactic phrases becomes even more apparent if syntactic constituents of any type (not restricted to noun phrases) are considered. Our analysis of 88,378 multiword syntactic constituents from the Natural Language Toolkit parsed treebank confirms that the number of syllables per phrase ranges widely, from 2 to at least 35 syllables[38] (Supplementary information). Long phrases are not infrequent; for example, ~20% of all phrases contain at least 12 syllables. Even more revealingly, 5% of all phrases contain at least 25 syllables. Moreover, no preferred phrase length is observable beyond the fact that shorter phrases are more frequent; for example, two-, three-, four- and five-syllable-long phrases account for 16%, 14%, 11% and 9% of all syntactic constituents, respectively. A roughly monotonously decreasing frequency for longer syntactic phrases is expected due to nesting of syntactic constituents (see later). A wide range of possible phrase lengths with no preferred length undermines the claim that syntactic phrases can be neurally tracked using a dedicated frequency band.

How could phrases be tracked in Ding et al.'s study[8] then? The key point is that Ding et al. chose stimuli that had a highly restricted syntactic and phonological make-up. Because of this, the split into non-overlapping syllabic or word units, phrase units and sentence units was conceivable. In effect, these levels were defined in physical terms as corresponding to one-, two- or four-syllable units, respectively (recall that syllables were presented isochronously at 250 ms per syllable, yielding rates of 4 Hz, 2 Hz and 1 Hz for each level). The term 'sentence rhythm' could be reserved for exactly four-syllable units and the term 'phrase rhythm' could be used for exactly two-syllable units because of the circumscribed nature of the linguistic materials chosen in Ding et al.'s isochronous mini-language. That is, the 'sentence rhythm' and 'phrase rhythm' could be precisely defined only because all sentences followed the same syntactic pattern; for example, a two-word noun phrase followed by a two-word verb phrase. As an important aside, this aspect of the design is critical in that it provides a viable explanation for Ding et al.'s findings that does not rely on the chunking function of oscillations. Owing to the nature of the linguistic stimuli, the spectral peaks at 2 Hz and 1 Hz may simply reflect an evoked response corresponding to the parser's regular building of phrases and sentences and clearing them out of working memory (figure 4 in Ding et al.[8] illustrates the response to phrases in the time domain and supports this interpretation). Hence, the rhythmic activity found in Ding et al.'s study[8] and similar studies may be "a by-product of the existence of hierarchical units in language, not the mechanism by which those units come to exist"[39].

Returning to the main thread, we note that the fact that an artificial language can be constructed with constant word or phrase durations as in Ding et al.'s study tells us little about whether natural language has sufficiently regular word or phrase durations needed for the oscillatory chunking account[8]. Indeed, an attempt to extend Ding et al.'s terminology and logic to natural language is ill-conceived and leads to arbitrary decisions. Focusing on the phrase level, we find that the construct of a 'phrase rhythm' simply does not extend to natural language. For example, in one study[10], phrases were defined as adjective–noun or verb–noun combinations, as shown by square brackets in '[Timid heroes] [pluck flowers] and the [brown birds] [gather branches]'. In the stimulus set used for the study (naturally spoken Dutch sentences), such two-word combinations corresponded to the 0.8–1.1-Hz frequency range, which was designated as the phrase band. However, it is unclear why the three-word phrase 'the brown bird' was not included in the phrase level and whether it would fit into the range. One could equally plausibly argue that the entire clauses 'timid heroes pluck flowers' and 'the brown birds gather branches' should be considered as phrases as well. In other words, to get a regular phrase rate out of natural language (even though the sentence structure chosen in the study was still rather templatic), the definition of the phrase-level oscillation had to be reverse engineered from an arbitrary and restrictive definition of phrases specifically as adjective–noun or noun–verb combinations, making 0.8–1.1 Hz a questionable range.

Finally, the notion of a phrase rhythm remains problematic not only if each oscillation period is fixed as in the case of a classical oscillator (an extreme scenario for a biological system[21]) but also under a weaker assumption of quasi-periodicity, when the period is perturbed by a small amount of random jitter. Quasi-periodic oscillators have often been assumed to have a role in language processing; in particular internal neural oscillators for syllable tracking have been modelled as a voltage-controlled oscillator component of classic phase-locked loop circuits to cover a range of frequencies between 3 Hz and 9 Hz, roughly corresponding to the theta range[40,41]. However, the oscillation that tracks syntactic phrases would need to span several traditionally distinct frequency bands. If we use earlier examples of 19- versus 1-syllable-long noun phrases and assume a mean speech rate of four syllables per second, the oscillation would need to range from 0.21 Hz to 4 Hz, spanning the sub-delta, delta and theta frequency bands. For even longer phrases or sentences of at least 40 syllables, the lower bound falls below 0.1 Hz. Whereas there is evidence of linguistically relevant oscillatory activity in the infraslow sub-delta frequencies below 0.1 Hz (refs.[42,43]), it is not clear whether population activity in the theta, delta and sub-delta bands can be functionally equivalent. Hence, the feasibility of an approach to tracking syntactic constituents that relies on functional equivalence of neural oscillations across these bands cannot be taken for granted and needs careful reassessment.

Word rhythm and phrase rhythm are not easily dissociable

The chunking approaches assume a neat hierarchy of linguistic levels, from syllable to word to phrase to sentence, with each level represented by a progressively slower rhythm or quasi-rhythm[39] (Fig. 1). In particular, a standard assumption is that because phrases generally consist of several words, the phrase rhythm will be slower than the word rhythm (or equal to it in the case of single-word phrases). In reality, phrases – including multiword phrases – can contain fewer syllables than single words, as in 'a long bug' versus 'caterpillar', leading to an overlap in the rate of the corresponding rhythms. The absence of an isomorphism between the levels of the linguistic hierarchy and their durations in speech critically undermines the basic mechanism of creating higher-level linguistic units; that is, using their longer oscillatory periods for accumulating signal corresponding to multiple shorter, lower-level units. If this is so, one is left without an account of how higher-level linguistic units are formed. A manifold approach seems necessary, in which word-level and phrase-level representations are not gauged using a single linear measure based on temporal duration, such as the number of syllables.

Syntactic phrases violate strict contiguity owing to nesting

The analogy between chunking for syllables and chunking for syntax discussed earlier breaks down when the property of contiguity is considered; that is, whether units follow one another in a linear and unidirectional succession. Contiguity is generally fulfilled for syllabic representations as they consecutively follow each other: each new syllable begins where a preceding one ends. Such contiguity is essential in enabling entrainment using an oscillation: consecutive syllables fall within consecutive cycles of a theta oscillation. Hence, to the extent that an oscillator can flexibly and promptly adjust to variation in syllable length, a single oscillation is sufficient for segmenting the speech stream into syllables.

The same 'linear' angle has been attempted for syntactic processing, whereby syntactic phrases are chained sequentially: "In the domain of abstract syntactic processing, listeners are biased to group words into implicit phrases with a period that is highly regular across both time and participants"[13]. Yet, famously, syntactic constituents do not present themselves as a linear sequence in which a constituent begins where another constituent ends. Instead, syntactic constituents are part of a hierarchical representation in which constituents are ubiquitously nested within one another (Box 1). In real-time terms, the noun phrase 'a remote village', the prepositional phrase 'from a remote village', the N-bar (an intermediate of a noun phrase) 'man in a tunic from a remote village', the N-bar 'cheerful man in a tunic from a remote village' and several other syntactic phrases in the sentence overlap in time with one another. As syntactic phrases do not succeed each other consecutively, consecutive periods of an oscillation cannot be assigned to the various phrases in the input sequence in the same way as for syllables. In other words, because syntactic phrases are characterized by nesting rather than contiguity, they cannot be fit into successive cycles of an oscillation, even if the oscillation could conceivably handle a rather large variation in constituent length. This holds regardless of whether entrainment proper[11] or intrinsic synchronicity with endogenously computed abstract linguistic units[12,13,44] is assumed. Either way, an intuitively simple claim that a single oscillation can synchronize with syntactic phrases is deceptive.

Syntactic structure building is not just consecutive word grouping

Syntactic structure building often requires building relationships that are more complex than those that can be achieved by delta chunking. Take the question 'What did John say that the grey monkey ate?' Note that the question word 'what' is the object of the verb 'ate' and is separated from it by multiple intervening words. The correct interpretation of the sentence requires relating elements that are non-adjacent in the surface string; that is, 'what' needs to be related to the verb 'ate' as opposed to the closer verb 'say'. Furthermore, the distance between the question word 'what' and the licensing verb 'ate' can be increasingly long; for example, 'What did John say that the grey monkey sitting on the top shelf of the cabinet ate?' Such examples demonstrate that it is insufficient to view syntactic structure building in terms of constructing progressively longer, continuous syntactic constituents. What is also required is the ability to relate elements that are apart, skipping over the linguistic material that separates them. A chunking view in which the only means of creating a linguistic unit is by squeezing sequences of words into a single oscillatory cycle is not fit for purpose.

Delta phrases are not primitive units in syntactic structure building

Our final argument challenges a conceptual point that chunking approaches concur on, namely that the initial step of syntactic processing (usually) yields multiword phrases (Fig. 1). Because acoustically such phrases are considered to fall within one cycle of a delta rhythm, they are often termed 'delta phrases', following ref.[9]. Many authors (notably not Ghitza himself in ref.[41]) consider these delta phrases to be the output of the first step of syntactic processing and the input to further stages of structure building. For example, Meyer[12] writes, "a syntactic processing stream is thought to involve the chunking of multiple words into so-called syntactic phrases to compose complex meaning; multiple phrases form a hierarchy that codes for the relationships within the scenario that the sentence as a whole encodes." But this begs the question of how the syntactic relationships inside a delta phrase get encoded. In syntactic terms, the status of delta phrases is at best unclear: whereas they may be short, or shortish, they are in no way syntactically primitive. Delta phrases such as 'from a remote village' are not syntactically simple: they have an internal structure that must be represented just like the structure of larger phrases. The mechanisms that enable the combining of the delta phrase 'from a remote village' with the preceding 'man in a tunic' into the higher-level unit 'man in a tunic from a remote village' must also have been at play to combine 'from' with 'a remote village', or 'a' with 'remote village' or 'remote' with 'village'. The assumption that syntactic structure building starts at the level of delta phrases is not tenable because syntactic operations of differing complexity must have already occurred to segregate the phrase. It is also not tenable in light of solid psycholinguistic evidence that sentences are processed incrementally, roughly word-by-word[5,45–47]. If syntactic structure is built and updated roughly as each next word becomes available in the input, multiword delta phrases are simply too long to be syntactically simple.

By contrast, delta entrainment may play a role in sensorimotor coupling or working memory during language processing. On the sensorimotor front, an active sensing approach[48,49] proposes that reorganization of delta oscillations in the sensory and motor cortices as a result of a proactive prediction modulates the excitability of task-relevant neural assemblies and thus aids sensory selection[24,26]. In audition, delta motor activity imposes temporal constraints on the sampling of sensory information and (partially) 'guides' auditory attention in speech or music[49–51]. Speech-specific benefits of motor predictions have also been asserted, such as anticipating the end of a speaker's phrase or utterance, including for turn-taking in a conversation[52,53]. Empirical support for such a view comes from the demonstrations of motor delta oscillatory activity modulating the tracking of prosody by the left auditory cortex[54,55] and of motor delta tracking correlating with auditory comprehension[11]. Hence, the motor cortex may contribute to speech processing via analysis of slow acoustic dynamics that provides a contextual temporal frame for parsing[49].

On the working memory front, classic two-stage parsing models[56–58] assume that, owing to working memory capacity limits, the parser shunts out parsed material corresponding to one or two phrases roughly every half a dozen words as the outcome of the first stage (as discussed by Fodor, prosodic factors may interact with the parser's decision to shunt a phrase out[59]). This shunting out thus happens rhythmically and falls within a delta rhythm. In both scenarios, it may be more prudent to refer to the output of delta entrainment unspecifically as 'delta chunks' without attributing a primarily syntactic function to them[41]. In summary, while delta tracking may be a useful step from a sensorimotor or working memory viewpoint (and therefore constitute a factor to which parsing models must pay attention), it explains little in terms of the details of syntactic structure building.

Oscillations for integration

In the first part of this Perspective, we argued against the chunking view, according to which the role of oscillations in syntactic structure building lies in linearly segmenting the input string into syntactically relevant units. We now turn to the integration view, whereby the potential of oscillations for syntactic structure building lies in their capacity to integrate elements across timescales and/or brain regions; that is, to represent relationships between elements in the linguistic input notwithstanding their adjacency or contiguity. In short, oscillatory activity is used to group elements in the linguistic input into larger entities, for which an internal organization is defined by hierarchical principles rather than by linear principles.

The potential of neural oscillations for the task of information integration was originally discussed in the context of the binding problem in vision[60–62]; that is, the question of how individual features of a visual object that are represented by distinct neural populations – such as edges, colour and movement – are bound into a unified representation within a neural circuit. In theoretical terms, it was proposed that synchronized or otherwise temporally correlated neuronal activity between neuronal ensembles coding for individual features can provide the means for the task of information integration. The alignment of neuronal firing among the assemblies that encode features belonging to a unique visual object increases the chance of dissociating it from other objects. The first empirical support for this proposal was provided via Gray and Singer's demonstrations of 'binding by synchrony'[63,64].

They showed that spatially segregated neuronal assemblies in the cat visual cortex that selectively respond to different visual features fired rhythmically and in synchrony with one another and/or with the local field potential when a dynamic visual stimulus was presented to an animal. Crucially, the rhythmic synchronization was internally generated and depended on the stimulus as a whole; that is, it could not be explained solely by the firing patterns of individual neuronal assemblies. This led to the conclusion that oscillatory neural activity enabled binding of individual sensory features into a perceptual gestalt. As Singer put it, "the cerebral cortex might exploit the option to synchronize the discharges of neurons with millisecond precision in order to bind these responses for further joint processing, i.e. to exploit temporal synchrony to encode relations"[65]. Since these early discoveries, the modulatory role of neural oscillations in the firing of neurons constituting an assembly has been upheld. Neuronal firing (the spike code) can be amplified or dampened according to the oscillatory phase (the temporal code). The phenomenon has been demonstrated in different cognitive domains, including action[48], olfaction[66] and attention[67,68], to name a few. Oscillations (also known as 'excitability fluctuations') also provide a functional (as opposed to an anatomical) connectivity structure for communicating signals across brain areas, as they influence the likelihood of spike output and, at the same time, the sensitivity to synaptic input in a downstream level of the network[69,70].

The demonstrations of interacting oscillatory activity across multiple frequency bands as a means of providing a neural code for information are of relevance to the idea that oscillations may play an integrative function in language processing. A well-known phenomenon from spatial navigation referred to as 'phase precession' serves as an illuminating example of how oscillations contribute to the task of information coding and integration across time. When a rat runs along a familiar path, selected neurons in the rat's hippocampus ('place cells') fire only in a specific location on the path ('place field')[71]. The place cell's firing rate is in the gamma range; it is strongest in the centre of the place field and decreases as the animal moves away from it, thus providing a rate code for location. However, the rate code is insufficient to encode the direction of the rat's movement. Crucially, the rat's location and direction can both be retrieved from the precise timing of the place cell's firing relative to the background local field potential theta rhythm[72–74]. In particular, the place cell fires in a late phase of the theta cycle when the animal first enters the place field, with the firing progressively moving to an earlier phase within theta cycles as the animal moves through the place field until it exits it. Such a neural code based on gamma activity nested within the theta rhythm can be used retrospectively during the initial encoding of the memory or prospectively for trajectory planning or predicting future locations[73–76]. Jensen and Lisman have argued that such an encoding of sequences of spatial positions exemplifies how the brain represents sequences in working memory more generally[77]. That is, the neural code for serial order can be based on cross-frequency interaction, with the individual items represented by a high-frequency activity (for example, gamma activity representing different objects, letters and so on) arranged at different phases of a slower frequency (for example, theta) so that an earlier phase represents an earlier serial position.

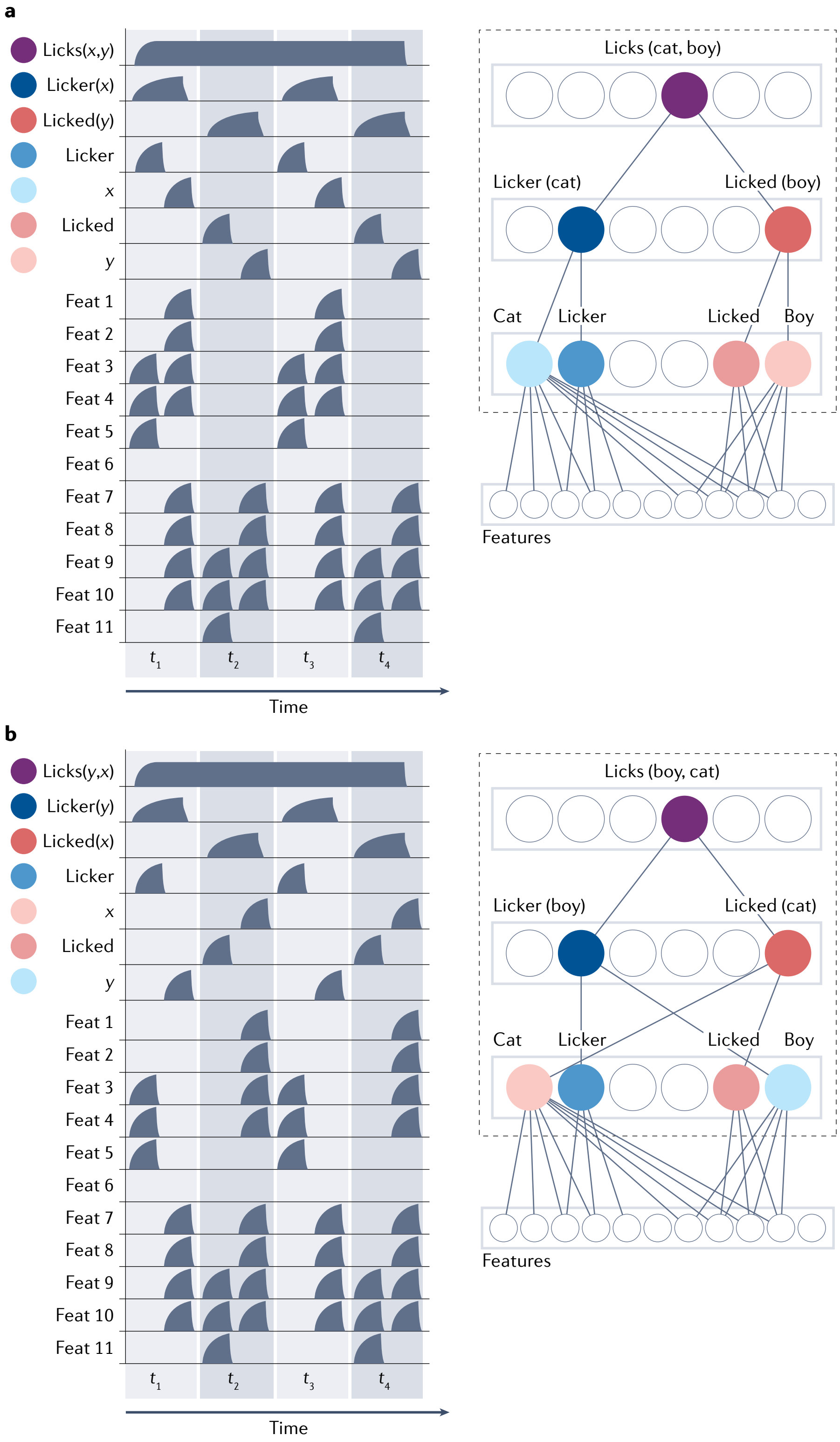
In the past decade, the idea that cross-frequency interaction may underlie structure building in language has been advanced[8,78]. Cross-frequency interaction has been considered as a mechanism for packaging information to encode linguistically relevant categories or structures. Friederici and Singer[78] provided an example of a nested relationship whereby the assembly oscillating in the low-frequency range represents a supraordinate content (for example, a grammatical category or a syntactic phrase), while the nested fast-oscillating assemblies reflect more specific items (for example, words constituting the category or phrase). Ding and colleagues[8] focused on the role of cross-frequency interaction for hierarchical structure building: "the neural representation of smaller linguistic units is embedded at different phases of the neural activity tracking a higher-level structure."

A critical property that distinguishes the 'oscillations for integration' view from the 'oscillations for chunking' view generally, and with regard to models of syntactic structure building particularly, is that in the oscillations for integration approach the oscillation period does not need to match the duration of the unit that is being built. In phase precession, for example, a single theta cycle (5–8 Hz, corresponding to a period of 125–200 ms) encodes a sequence of multiple successive locations on a running track. The amount of time taken by the rat to traverse these locations on the track is one to two orders of magnitude greater (the phase precession slope increases with running speed[79], highlighting the lack of a one-to-one correspondence between the behavioural timescale and the encoding timescale). Similarly, in serial order encoding, the sequence of objects, say A-B-C-D-E, is encoded via the phase of gamma activity representing each object against a theta oscillation. Critically, the theta frequency (that is, its period) is in no way related to the physical duration of the sequence A-B-C-D-E; the factual duration of the sequence can exceed the theta period by one to two orders of magnitude[80,81]. In both cases, the encoding mechanism is such that the time is compressed or, more precisely, recoded. Hence, the criticisms outlined above for the oscillations for chunking view do not apply to oscillations for integration approaches that enjoy a non-isomorphic mapping between the time of the external stimulus and the time of internal oscillation-based representations. For example, a varying length of syntactic phrases is of no consequence if the oscillation period is not commensurate with it, and neither are potentially overlapping acoustic durations of words and phrases. This is because oscillations are recruited not to solve the problem of segmenting the linguistic input but, rather, to represent the structured relationships that are inferred from the input (using long-term grammatical knowledge).

In the following subsections, we consider three proposals that capitalize on the integrative capacity of oscillations for syntactic structure building. In all cases, the key insight is that oscillatory activity is the basis for building a hierarchical representation that is unconstrained by adjacency in the linear input string. Although the proposals may lack mechanistic detail or be restricted in scope, they each feature an interesting idea that can be useful for subsequent work. We discuss key properties of these proposals with the goal of highlighting their strengths and exposing the limitations to be overcome in future research.

Biolinguistics proposals

The first proposal comes from the field of biolinguistics, in which several attempts have been made to present neural oscillations as an essential mechanism for syntactic structure building[14–16,20,82,83]. In these studies, the authors identified subprocesses that are integral to syntactic structure building from the perspective of linguistic theory and aimed to align each of them with a neural oscillatory process. For example, Boeckx and Theofanopoulou[20] distinguished four interacting linguistic subprocesses involved in syntactic structure building: monadicization, which is 'bundling' of lexical feature assemblies into



a single unit, akin to binding of visual features into a coherent visual object; combination, which enables the combining of two elements into a set (also known as a phrase) — for example, 'green' and 'sand' combined into 'green sand'; labelling, by which a phrase is identified — for example, the noun phrase label assigned to the phrase 'green sand' or a verb phrase label assigned to the phrase 'we run'; and consolidation, which makes it possible for a labelled phrase, such as the noun phase 'green sand' to be represented in working memory as a consolidated

Fig. 2 | Binding and symbolic propositions in DORA. a, Left, a diagram demonstrating DORA's time-based binding for the proposition '*x* licked *y*'. Black shapes represent activation patterns of brain activity. The predicate Lick(*x*,*y*) intermittently activates its subordinate unary relations Licker(*x*) and Licked(*y*). At time t_1, while Licker(*x*) is firing, the unit representing the licker role (and relevant semantic feature units among feats 1–11 that encode its semantic content in a distributed fashion) become active; this is followed by the activation of the unit representing *x* and its features. The sequential firing of the licker role and *x* during t_1 marks *x* as bound to the licker role. Similarly, while Licked(*y*) is firing during t_2 the units for the licked role and for *y* become active one after the other and are bound together. The activity pattern during t_1 and t_2 repeats during t_3 and t_4. Right, the proposition 'The cat licked the boy' represented in the DORA architecture (the time dimension is collapsed). Coloured circles correspond to symbolic units representing individual binary predicates (top layer), unary predicates (middle layer) or roles and other concepts (bottom layer). The colour of units links back to the left panel, with *x* and *y* mapping to 'cat' and 'boy', respectively. **b**, The inverse proposition, 'The boy licked the cat', activates the same units but the assignment of fillers to roles is different, with *x* and *y* represented by 'boy' and 'cat', respectively. Note a modified temporal firing sequence in the left panel and modified links between the levels in the right panel. Adapted with permission from ref.[19], Elsevier.

chunk. These subprocesses form the basis for mapping the grammar to the neural operations. Each subprocess is claimed to be carried out by a specific frequency band or via cross-frequency activity: monadicization is carried out via alpha–gamma coupling; combination and labelling are achieved via the alpha rhythm and the beta rhythm, respectively; and consolidation is carried out via theta–gamma coupling. Benitez-Burraco and Murphy's proposal[14] is logically similar, although the linguistic subprocesses and their respective neural correlates differ (see also refs.[15,16,83]). (In a nutshell, feature set construction, combinatorial processing, phrasal memory and linguistic prediction are carried out via theta–gamma coupling, delta–gamma coupling, delta–theta coupling and alpha–gamma coupling, respectively.) Both sets of authors support their approach by citing neurophysiological findings demonstrating a correlation between a linguistic subprocess and its purported neurophysiological correlate.

Although grounding in linguistic theory and an explicit alignment of the steps of linguistic computation with neural oscillations is admirable, the biolinguistics proposals fall short of providing a viable mechanistic explanation of structure building. In particular, it remains unclear how various linguistic and corresponding oscillatory subprocesses are coordinated and interact in time to yield even a most basic higher-level unit such as a two-word phrase (for example, 'blue cat'). This is related to a more global point (discussed later): a direct mapping from syntactic theory to neural processes is of little use. The endeavour of mapping linguistic computations to neural substrates requires an explicit parsing theory that outlines a detailed algorithm for how various computational subprocesses are executed in real time to yield a higher-level representation.

There are two other proposals in the literature in which syntactic structure building is instantiated via oscillatory mechanisms, based on cross-frequency interaction. They each pick a different focus on structure building and scope, but each provides some level of mechanistic detail.

DORA model

The first proposal is represented by the (neuro)computational model DORA (for 'discovery of relations by analogy')[84], which itself derives from an earlier model, LISA (for 'learning and inference with schemas and analogies')[85]. DORA was originally developed as a model of relational reasoning[84], and was later extended into a model of cortical computation of syntactic structure in refs.[18,19] (see also ref.[86]). The model focuses on sentences containing unary or binary predicates such as 'The boy sneezed' or 'The cat licked the boy', respectively. DORA's key feature lies in the use of firing (a)synchrony to keep representations of different arguments distinct from each other. Indeed, representing 'The cat licked the boy' requires an ability to bind 'the cat' to the subject role of 'licked' and 'the boy' to the object role, and to maintain these bindings simultaneously yet distinctly. In DORA, this is achieved by splitting each binary predicate into two subordinate unary predicates; that is, Lick(*x*,*y*) is represented via the combination of Licker(*x*) (that who licks) and Licked(*y*) (that who is being licked) (Fig. 2a). Activating the predicate 'lick' activates its subordinate unary relationships intermittently within an activation cycle of the overarching predicate Lick(*x*,*y*). One of the unary predicates, here Licker(*x*), is activated first, and is followed by the second unary predicate, Licked(*y*). While active, each unary predicate activates its subordinates, also asynchronously; for example, in Fig. 2a, Licker(*x*) activates the role 'licker' first, then the concept 'cat' that is bound to it. Translated into neural oscillatory terms, DORA encodes progressively higher levels of the syntactic hierarchy via progressively slower oscillations; higher-level unit formation relies on cross-frequency interaction, with lower-level units firing at a specific phase within a higher-level unit's cycle. The use of temporal asynchrony in the coding scheme has an important consequence of maintaining independence between the roles and fillers: an inverse containment relation – that is, 'The boy licked the cat' – can be represented simply by rearranging the temporal sequence of firing of the same four units so that 'licker' and 'boy' are active during the Licker(*x*) cycle and 'licked' and 'cat' are active during the Licked(*y*) cycle (Fig. 2b).

Alongside its strengths in its application as a neurocognitive model of syntactic structure, DORA also has weaknesses. First, the model[18] interprets rhythms rigidly in that a phrase is represented by a rhythm proportional to the physical duration of that phrase (for example, given a word rate of 4 Hz, a two-word phrase is represented via a 2-Hz rhythm). This makes the model prone to the concerns discussed in the first part of this Perspective; namely, that phrases come in different lengths and thus building phrases in the model would require an infeasibly wide range of different rhythms. Second, each layer in DORA is dedicated to a specific type of unit: the top layer contains full predicates, the middle layer hosts unary relations such as Licker(*x*) or Licked(*y*) and the bottom layer hosts individual roles ('licker', 'licked') and concepts ('cat', 'boy'). Given this strict layering in the model's architecture, it is unclear how phrases such as 'dusty dry fur' can be built; that is, how the output of Dry(*x*) from the middle layer can serve as input to Dusty(*x*) in the same layer. Last, DORA takes a radically lexicalized approach to syntax, whereby syntactic structure building is fully driven by properties of individual predicates stored in long-term memory, to the exclusion of abstract grammatical knowledge per se. Whereas a high degree of lexicalization of syntax is not uncommon, all lexicalized approaches (but not DORA) also invariably incorporate grammatical knowledge and/or grammatical operations in a form that is not individually tied to lexical items. For example, to interpret the sentence 'A sad flower chatted nobly with a bin' (Box 1), one must use one's grammatical knowledge to find relationships between words, then build a hierarchical structure on the basis of them and derive the sentence interpretation. The grammatical knowledge includes identifying 'sad' as an adjective, 'flowers' as a noun and 'a' as a determiner or article and then accessing the grammatical rules AP + N′ → N′ (an adjective phrase combines with an N-bar-level phrase to produce another N-bar-level phrase) and D + N′ → NP (a determiner combines with an N-bar-level-phrase to produce a noun phrase) that allow one to progressively combine these words into a noun phrase (note that the exact form in which grammatical knowledge is represented is debated in the literature; some of the well-known alternatives are phrase-structure rules similar to the ones above[87,88], auxiliary trees combined with the adjunction or substitution operations in tree-adjoining grammar[89], feature structures and unification in unification-based approaches to grammar[90], and the merge operation in minimalism[91]). Across all variants, grammatical knowledge captures generalizations that go beyond individual lexical words. Without such knowledge, the system cannot handle building adjunct relations that fall beyond saturating the meaning of a lexical item; for example, 'a cat from the attic' or 'a cat with a bow' cannot be represented because there is nothing in the stored lexical entry for 'cat' that enables or requires its combination with the prepositional phrase 'from the attic' or 'with a bow'. Rather, this combination is possible due to a syntactic principle that enables modification of noun phrases with prepositional phrases. Indeed, a prepositional phrase such as 'from the attic' can be attached to any noun phrase and will be interpretable (even though often the resulting interpretation will be implausible or even bizarre, as in 'his ears from the attic'), highlighting that the availability of a prepositional phrase modifier is not predicted by the meaning of a specific lexical item. Consequently, whereas DORA can serve as a neurocognitive model of argument structure (that is, lexical representation of argument-taking lexical items and meaning saturation), it lacks a way of representing syntactic knowledge that is compulsory for any model of syntactic structure building.

VS-BIND model

The other neurocognitive proposal devoted to syntactic structure building, also cast within a symbolic connectionist architecture, is represented by the vector-symbolic sequencing of binding instantiating dependencies (VS-BIND) model[17] (Fig. 3). VS-BIND uses vector symbolic architecture, and explicitly supports combinatorial and symbolic operations over vectors. A key operation is that of binding, indicated by ⊗, which mathematically corresponds to circular convolution (a convolution of two periodic functions[92]) and outputs a vector of the same length as the input vectors. Such dimensionality reduction (or, more precisely, non-proliferation) makes it possible to repeat the binding operation many times to represent complex syntactic structure. The binding can be 'undone' to recover representations similar to the original input by computing a new binding between itself and the (approximate) inverse of one of the inputs; for example, if $A \otimes B$ is bound with the 'key' $\neg B$ ('not B') as in $(A \otimes B) \otimes \neg B$, the output is highly similar to A. Because of the lossy encoding (that is, the fact that constituents retrieved from the output are degraded compared with the original constituents), the system cannot be considered perfectly compositional. Yet Calmus et al. suggest this is a desirable outcome that is a parallel of an ecological limit on the human working memory and of an ensuing limit on the depth of syntactic nesting that can be reasonably handled in natural language[17].

VS-BIND focuses on an inherently syntactic phenomenon of (adjacent or non-adjacent) hierarchical dependencies. The key insight is that hierarchical dependency relationships can be represented using principles of serial encoding via positional tagging, with each item bound to a serial positional tag so that sequence = 1st ⊗ Item 1 + 2nd ⊗ Item 2 + 3rd ⊗ Item 3 .... This basic positional tagging is assumed to involve the dorsolateral prefrontal cortex, the motor and premotor cortices and the hippocampus, following refs.[93–95]. In a simple case, a single dependency such as 'The boy sneezed' can be represented as sentence = 1st ⊗ 'the boy' + 2nd ⊗ 'sneezed' (Fig. 3a). If the sentence contains multiple dependencies nested within each other, as in the case of a centre-embedded 'The boy the cat licked sneezed', a set of positional codes is used for each dependency so that the main clause is represented as 1st ⊗ 'the boy' + 2nd ⊗ 'sneezed' and the embedded clause as 1st ⊗ 'the cat' + 2nd ⊗ 'licked' (Fig. 3b). The representation of two separate clauses is achieved via an assignment of two separate sets of positional codes at the lower level (one per clause) and subsequently binding of each clause to a higher-level set of positional codes (1st_{C}, 2nd_{C}, ..., the index 'C' standing for 'chunk'). This exemplifies a more general point that VS-BIND can define sets of positional tags that are specific to each level of the hierarchy. The networks responsible for extraction and maintenance of non-local hierarchical dependencies are considered to involve, following ref.[96], the dorsal ventrolateral prefrontal cortex, including Brodmann areas 44 and 45. The output is of the form sentence = 1st_{C} ⊗ (1st ⊗ 'the boy' + 2nd ⊗ 'sneezed') + 2nd_{C} ⊗ (1st ⊗ 'the cat' + 2nd ⊗ 'licked'), and within it the clauses are clearly distinguishable and retrievable (for example, binding the output with the key $\neg 1\text{st}_{\mathrm{C}}$ recovers the main clause. The retrieval can be deeper, e.g. the key $\neg(2\text{nd} \otimes 1\text{st}_{\mathrm{C}})$ recovers the second item of the first chunk (that is, 'sneezed').

In neural terms, VS-BIND achieves serial position coding – and thus syntactic structure building – via cross-frequency interaction. As proposed by Lisman and colleagues, serial order in the working memory is encoded by the phase at which fast-frequency representations of individual items are nested within a slower frequency; namely, gamma-within-theta nesting[81,97]. In VS-BIND, individual items (words or grammatical constructs such as subject, verb and object) are represented by gamma activity; the phase of each gamma burst within the theta cycle reflects the word's serial position[17].

VS-BIND's encoding principle makes it possible to create multiple hierarchical dependencies and to package information contained in them while retaining an ability to retrieve – albeit imperfectly – original constituents. VS-BIND supports flexible creation and manipulation of larger chunks, with a valued consequence that with time and/or repetition, such chunks can give rise to a memory trace[17]. A major limitation currently is the lack of discussion of how grammatical knowledge is stored and interacts with the model; for example, how are 'the boy' and 'sneezed' or 'the cat' and 'licked' recognized as members of the same dependency? Furthermore, it is unclear how much hierarchy the model encodes; for example, the final representation 1st_{C} ⊗ (1st ⊗ 'the boy' + 2nd ⊗ 'sneezed') + 2nd_{C} ⊗ (1st ⊗ 'the cat' + 2nd ⊗ 'licked') fails to represent that the first clause is hierarchically higher than the second one. Last, some of the model's behaviour may be at odds with human data. As discussed above, unbinding of the output using various keys yields a welcome possibility to recover information from the outcome, such as individual clauses or individual items in each clause. However, the same process also makes it possible to recover irrelevant information with the same ease. For instance, applying the key ¬2nd to the final sentence representation 1st_{C} ⊗ (1st ⊗ 'the boy' + 2nd ⊗

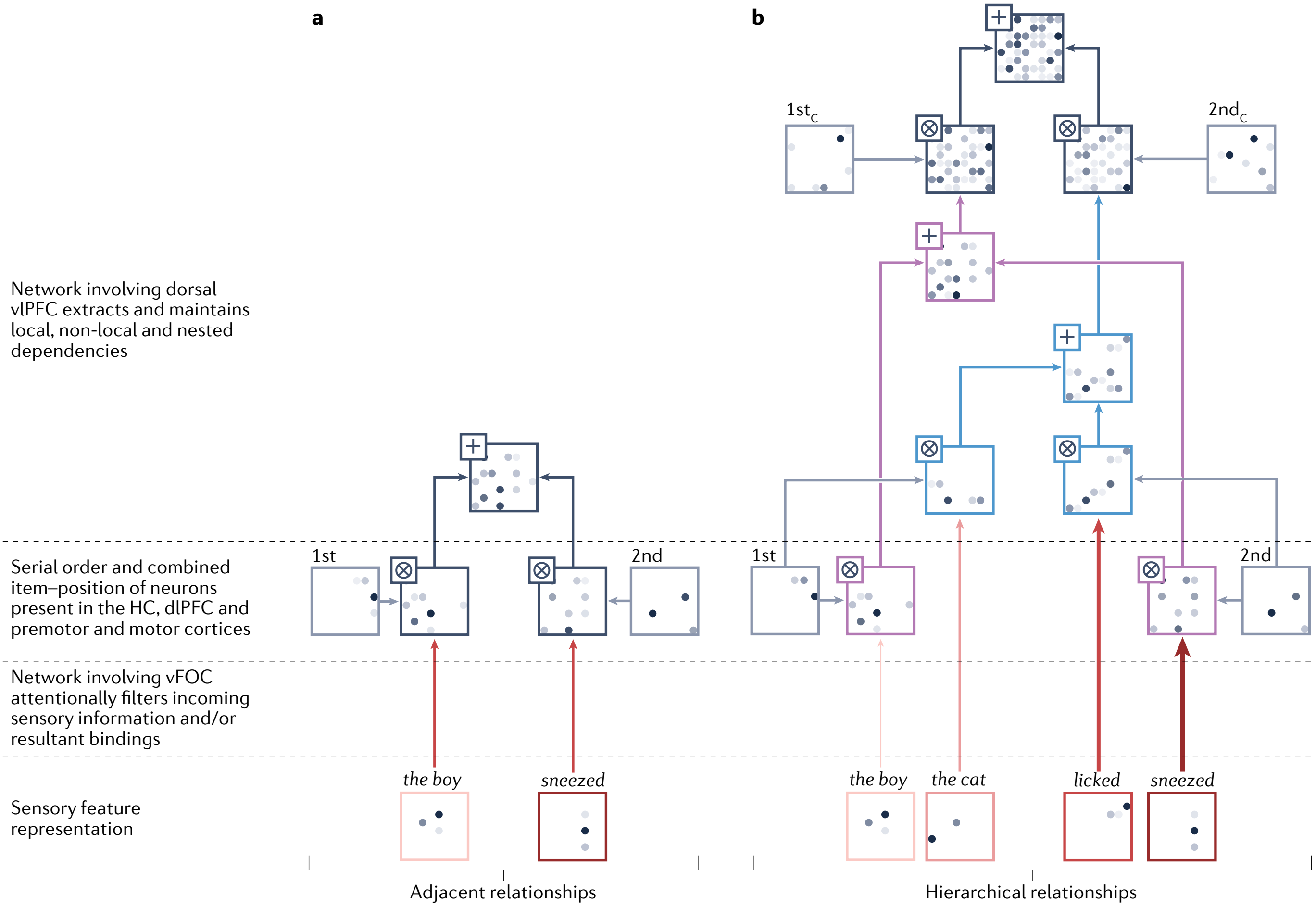


Fig. 3 | Representation of a sentence with a single subject–verb dependency and multiple subject–verb dependencies in VS-BIND. The descriptions on the left represent various levels of representation alongside their anatomical loci. **a**, Representation of a single subject–verb dependency in 'The boy sneezed'. Discrete sensory items ('the boy', 'sneezed'), encoded as latent vector symbolic representations, are bound to distinct positional tags 1st, 2nd, … (grey boxes) to form position–item representations. A linear, ordered representation of the input sequence can thus be maintained in a decaying memory buffer as sequence = 1st ⊗ 'the boy' + 2nd ⊗ 'sneezed' (shown by black boxes and arrows). **b**, Representation of multiple nested dependencies in 'The boy the cat licked sneezed'. Items 1 and 4 ('the boy' and 'sneezed') and items 2 and 3 ('the cat' and 'licked') are chunked through binding and superposition as shown by pink and blue boxes and arrows, respectively. These two chunks form a superchunk representing a nested dependency (shown by black boxes and arrows). Proposed anatomical locations for each level of conceptual and syntactic representation are given in the left column. Feedback connections are permitted in the vector-symbolic sequencing of binding instantiating dependencies (VS-BIND) model (for example, from a cognitively abstract cortical region to sensory cortex) but are not shown here. dlPFC, dorsolateral prefrontal cortex; HC, hippocampus; vFOC, ventral frontal opercular cortex; vlPFC, ventrolateral prefrontal cortex. Adapted with permission from ref.[17], The Royal Society.

'sneezed') + $2nd_C$ ⊗ (1st ⊗ 'the cat' + 2nd ⊗ 'licked') recovers the second item of every chunk, which is unlikely to be easily recoverable by humans. More broadly, while serial encoding principles beyond doubt bring significant potential to the task of syntactic structure building, it is important to keep in sight that in language hierarchy trumps linear order, indicating that mechanisms additional to those enabling serial encoding are at play.

In summary, the mechanisms for building relationships between elements in a sentence in both DORA and VS-BIND fundamentally rely on temporal coordination of activity between neuronal ensembles: neural activity across different frequencies is regulated in time to represent different arguments of a predicate in DORA or to build a subject–verb dependency in VS-BIND. Empirical work, ideally involving high-precision intracortical recordings, that tests these mechanisms will be essential to assess the specific proposals as well as to catalyse the field generally. Importantly, the scope of either proposal is limited and key details are missing; for example, how grammatical knowledge is represented in long-term memory is unclear. As we argue in the next section, in expanding these models and proposing new ones, one should pay attention to the parser, a system that uses long-term grammatical knowledge to convert a string of words into a syntactic representation in real time.

Conclusions and future directions

Neurocognitive models that use oscillations for syntactic structure building do so in two distinct ways: chunking the linguistic input using linear principles versus relating elements in the input to one another using hierarchical principles (integration). Here we have argued against the position that oscillations chunk the input into syntactic phrases by demonstrating that such linear chunking misses the essence of syntactic structure. We have also discussed the potential of oscillations for integration and reviewed several proposals that capitalize on the idea of using oscillatory activity for hierarchical structure building, a major asset in our view, while also noting their limitations.

How can the oscillations for integration approach be developed further? In a search for neural operations for structure building, it is instrumental to set explicit goals for the undertaking. One possibility is that a neurocognitive model of syntactic structure building must align its objectives with that of a grammatical theory; that is, finding a neural representation for a static syntactic structure of a full sentence. The models we discussed in the second part of this Perspective are by and large instances of such an approach. DORA is inherently atemporal in that it does not consider the left-to-right unfolding of linguistic input. VS-BIND's tracking of the linear position makes an important step towards incrementality; however, the core of the model is still atemporal in that all grammatical relations are determined on the basis of a full sentence rather than a sentence fragment available at a given point in time.

Alternatively, the development of a neurocognitive model of syntactic structure building would take as an essential starting point a model of a psychologically plausible parser. In this case, the objective is to translate into neural terms the parsing process as defined in psycholinguistics and/or computational linguistics; that is, incremental conversion of a linear sequence of words into a connected syntactic representation. In a parser, long-term grammatical knowledge interfaces with relevant psychological factors and systems, such as incremental processing of the input, limited working memory capacity and so on. A critical status of these factors has been convincingly demonstrated in psycholinguistics research: linguistic input is processed incrementally (roughly word-by-word) as it unfolds in time, and much of syntactic structure building concerns representing partial syntactic structure for a sentence fragment, rather than a full sentence[46,98–101]. As an important side note, whereas in psycholinguistic and computational literature the parser is typically assumed to receive linguistic input monotonously word-by-word, the research reviewed in the first part of this Perspective highlights that the parser may operate on a buffer that contains several words (that is, 'delta chunks'). If so, then the parsing theory must be adjusted accordingly, for example, as proposed in ref.[102]. Owing to incrementality, other processes, such as prediction, become indispensable and require thorough discussion and implementation[103]. Moreover, a detailed consideration of how grammatical information could be represented over the long term will also be required, both for completeness and to ensure that the neural format of such long-term grammatical representations (for example, based on synaptic weights) can smoothly interface with the format of short-term syntactic representations (for example, based on oscillatory coupling). Finally, because syntactic representation is short-lived and cycled out of memory soon after the sentence is heard or read[104], the question of what syntactic representations are converted into and when also needs to be considered.

Although the claim that one should be mapping neural operations to parsing steps may meet with little disagreement, practical steps towards it are largely non-existent. As a notable exception, Nelson and colleagues[105] demonstrated that the dynamic profile of high-gamma power recorded intracranially from language-related regions correlated with the structure building profile by a left-corner parser. Essentially, the high-gamma power increased during intervals in which the parser needed to maintain a progressively more complex syntactic representation in working memory and decreased whenever a snippet of syntactic structure was completed and cleared out of working memory. Hence, investigating how high-gamma activity is generated may be a step towards elucidating the mechanistic aspects of syntactic structure building by the brain. The time is ripe for the field to take concrete steps in this direction.

Published online: 2 December 2022

References

1. Chomsky, N. *Syntactic Structures* (Mouton, 1957).
2. Adger, D. *Language Unlimited: The Science Behind Our Most Creative Power* (Oxford Univ. Press, 2019).
3. Jackendoff, R. *Foundations of Language* (Oxford Univ. Press, 2002).
4. Adger, D. Syntax. *WIREs Cogn. Sci.* **6**, 131–147 (2015).
5. Crocker, M. W. *Computational Psycholinguistics: An Interdisciplinary Approach to the Study of Language* (Kluwer Academic, 1996).
6. Hale, J. T. What a rational parser would do. *Cogn. Sci.* **35**, 399–443 (2011).
7. Hale, J. T. *Automaton Theories of Human Sentence Comprehension* (CSLI, 2014).
8. Ding, N., Melloni, L., Zhang, H., Tian, X. & Poeppel, D. Cortical tracking of hierarchical linguistic structures in connected speech. *Nat. Neurosci.* **19**, 158–164 (2016).
9. Ghitza, O. Acoustic-driven delta rhythms as prosodic markers. *Lang. Cogn. Neurosci.* **32**, 545–561 (2017).
10. Kaufeld, G. et al. Linguistic structure and meaning organize neural oscillations into a content-specific hierarchy. *J. Neurosci.* **40**, 9467–9475 (2020).
11. Keitel, A., Gross, J. & Kayser, C. Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. *PLoS Biol.* **16**, e2004473 (2018).
12. Meyer, L. The neural oscillations of speech processing and language comprehension: state of the art and emerging mechanisms. *Eur. J. Neurosci.* **48**, 2609–2621 (2017).
13. Meyer, L., Sun, Y. & Martin, A. E. Synchronous, but not entrained: exogenous and endogenous cortical rhythms of speech and language processing. *Lang. Cogn. Neurosci.* **35**, 1089–1099 (2019).
14. Benítez-Burraco, A. & Murphy, E. Why brain oscillations are improving our understanding of language. *Front. Behav. Neurosci.* **13**, 190 (2019).
15. Murphy, E. The brain dynamics of linguistic computation. *Front. Psychol.* **6**, 1515 (2015).
16. Murphy, E. *The Oscillatory Nature of Language* (Cambridge Univ. Press, 2020).
17. Calmus, R., Wilson, B., Kikuchi, Y. & Petkov, C. I. Structured sequence processing and combinatorial binding: neurobiologically and computationally informed hypotheses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **375**, 20190304 (2020).
18. Martin, A. E. & Doumas, L. A. A mechanism for the cortical computation of hierarchical linguistic structure. *PLoS Biol.* **15**, e2000663 (2017).
19. Martin, A. E. & Doumas, L. A. Predicate learning in neural systems: using oscillations to discover latent structure. *Curr. Opin. Behav. Sci.* **29**, 77–83 (2019).
20. Boeckx, C. & Theofanopoulou, C. in *Language, Syntax, and the Natural Sciences* (eds Gallego, A. J. & Martin, R.) 295–315 (Cambridge Univ. Press, 2018).
21. Giraud, A. L. Oscillations for all ¯_(ツ)_/¯? A commentary on Meyer, Sun & Martin (2020). *Lang. Cogn. Neurosci.* **35**, 1106–1113 (2020).
22. Doelling, K. B. & Assaneo, F. M. Neural oscillations are a start toward understanding brain activity rather than the end. *PLoS Biol.* **19**, e3001234 (2021).
23. Obleser, J., Henry, M. J. & Lakatos, P. What do we talk about when we talk about rhythm? *PLoS Biol.* **15**, e2002794 (2017).
24. Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I. & Schroeder, C. E. Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science* **320**, 110–113 (2008).
25. Lakatos, P., Gross, J. & Thut, G. A new unifying account of the roles of neuronal entrainment. *Curr. Biol.* **29**, R890–R905 (2019).
26. Schroeder, C. E. & Lakatos, P. Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci.* **32**, 9–18 (2009).
27. Giraud, A. L. & Poeppel, D. Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* **15**, 511–517 (2012).
28. Ghitza, O. Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. *Front. Psychol.* **2**, 130 (2011).
29. Ding, N. et al. Temporal modulations in speech and music. *Neurosci. Biobehav. Rev.* **81**, 181–187 (2017).
30. Pellegrino, F., Coupé, C. & Marsico, E. Across-language perspective on speech information rate. *Language* **87**, 539–558 (2011).
31. Norcia, A. M., Appelbaum, L. G. G., Ales, J. M. J. M., Cottereau, B. R. B. R. & Rossion, B. The steady-state visual evoked potential in vision research: a review. *J. Vis.* **15**, 1–46 (2015).
32. Glushko, A., Poeppel, D. & Steinhauer, K. Overt and implicit prosody contribute to neurophysiological responses previously attributed to grammatical processing. *Sci. Rep.* **12**, 1459 (2022).
33. Kalenkovich, E., Shestakova, A. & Kazanina, N. Frequency tagging of syntactic structure or lexical properties; a registered MEG study. *Cortex* **146**, 24–38 (2022).

34. Burroughs, A., Kazanina, N. & Houghton, C. Grammatical category and the neural processing of phrases. *Sci. Rep.* **11**, 2446 (2021).
35. Makov, S. et al. Sleep disrupts high-level speech parsing despite significant basic auditory processing. *J. Neurosci.* **37**, 7772–7781 (2017).
36. Ding, N. et al. Characterizing neural entrainment to hierarchical linguistic units using electroencephalography (EEG). *Front. Hum. Neurosci.* **11**, 481 (2017).
37. Marcus, M. P., Santorini, B. & Marcinkiewicz, M. A. Building a large annotated corpus of English: the Penn Treebank. *Comput. Linguist.* **19**, 313–330 (1993).
38. Bird, S., Klein, E. & Loper, E. *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit* (O'Reilly Media, 2009).
39. Gwilliams, L. Hierarchical oscillators in speech comprehension: a commentary on Meyer, Sun, and Martin (2019). *Lang. Cogn. Neurosci.* **35**, 1114–1118 (2020).
40. Ghitza, O. & Greenberg, S. On the possible role of brain rhythms in speech perception: intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica* **66**, 113–126 (2009).
41. Ghitza, O. "Acoustic-driven oscillators as cortical pacemaker": a commentary on Meyer, Sun & Martin (2019). *Lang. Cogn. Neurosci.* **35**, 1100–1105 (2020).
42. Honey, C. J. et al. Slow cortical dynamics and the accumulation of information over long timescales. *Neuron* **76**, 423–434 (2012).
43. Hasson, U., Yang, E., Vallines, I., Heeger, D. J. & Rubin, N. A hierarchy of temporal receptive windows in human cortex. *J. Neurosci.* **28**, 2539–2550 (2008).
44. Meyer, L., Sun, Y. & Martin, A. E. "Entraining" to speech, generating language? *Lang. Cogn. Neurosci.* **35**, 1138–1148 (2020).
45. Crocker, M. W. in *Perspectives on Sentence Processing* (eds Clifton, C., Frazier, L. & Rayner, K.) 245–266 (L. Erlbaum Associates, 1994).
46. Sturt, P. & Lombardo, V. Processing coordinated structures: incrementality and connectedness. *Cogn. Sci.* **29**, 291–305 (2005).
47. Sturt, P. & Crocker, M. W. Monotonic syntactic processing: a cross-linguistic study of attachment and reanalysis. *Lang. Cogn. Process.* **11**, 449–494 (1996).
48. Schroeder, C. E., Wilson, D. A., Radman, T., Scharfman, H. & Lakatos, P. Dynamics of active sensing and perceptual selection. *Curr. Opin. Neurobiol.* **20**, 172–176 (2010).
49. Morillon, B., Arnal, L. H., Schroeder, C. E. & Keitel, A. Prominence of delta oscillatory rhythms in the motor cortex and their relevance for auditory and speech perception. *Neurosci. Biobehav. Rev.* **107**, 136–142 (2019).
50. Morillon, B. & Baillet, S. Motor origin of temporal predictions in auditory attention. *Proc. Natl Acad. Sci. USA* **114**, E8913–E8921 (2017).
51. Zalta, A., Petkoski, S. & Morillon, B. Natural rhythms of periodic temporal attention. *Nat. Commun.* **11**, 1051 (2020).
52. Wilson, M. & Wilson, T. P. An oscillator model of the timing of turn-taking. *Psychon. Bull. Rev.* **12**, 957–968 (2005).
53. Scott, S. K., McGettigan, C. & Eisner, F. A little more conversation, a little less action—candidate roles for the motor cortex in speech perception. *Nat. Rev. Neurosci.* **10**, 295–302 (2009).
54. Keitel, A., Ince, R. A. A., Gross, J. & Kayser, C. Auditory cortical delta-entrainment interacts with oscillatory power in multiple fronto-parietal networks. *Neuroimage* **147**, 32–42 (2017).
55. Park, H., Ince, R. A. A., Schyns, P. G., Thut, G. & Gross, J. Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Curr. Biol.* **25**, 1649–1653 (2015).
56. Kimball, J. Seven principles of surface structure parsing in natural language. *Cognition* **2**, 15–47 (1973).
57. Frazier, L. & Clifton Jr, C. *Construal* (MIT Press, 1996).
58. Frazier, L. & Fodor, J. D. The sausage machine: a new two-stage parsing model. *Cognition* **6**, 291–325 (1978).
59. Fodor, J. D. Learning to parse? *J. Psycholinguist. Res.* **27**, 285–319 (1998).
60. Milner, P. M. A model for visual shape recognition. *Psychol. Rev.* **81**, 521–535 (1974).
61. von der Malsburg, C. Nervous structures with dynamical links. *Ber. Bunsenges.* **89**, 703–710 (1985).
62. von der Malsburg, C. *The Correlation Theory of Brain Function*. Internal report 81–82 (Max Planck Institute for Biophysical Chemistry, 1981).
63. Gray, C. M. & Singer, W. Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proc. Natl Acad. Sci. USA* **86**, 1698–1702 (1989).
64. Gray, C. M., König, P., Engel, A. K. & Singer, W. Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature* **338**, 334–337 (1989).
65. Singer, W. Binding by synchrony. *Scholarpedia* **2**, 1657 (2007).
66. Perez-Orive, J. et al. Oscillations and sparsening of odor representations in the mushroom body. *Science* **297**, 359–365 (2002).
67. Busch, N. A. & VanRullen, R. Spontaneous EEG oscillations reveal periodic sampling of visual attention. *Proc. Natl Acad. Sci. USA* **107**, 16048–16053 (2010).
68. Dugué, L., McLelland, D., Lajous, M. & VanRullen, R. Attention searches nonuniformly in space and in time. *Proc. Natl Acad. Sci. USA* **112**, 15214–15219 (2015).
69. Fries, P. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn. Sci.* **9**, 474–480 (2005).
70. Fries, P., Nikolić, D. & Singer, W. The gamma cycle. *Trends Neurosci.* **30**, 309–316 (2007).
71. O'Keefe, J. & Dostrovsky, J. The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Res.* **34**, 171–175 (1971).
72. O'Keefe, J. & Recce, M. Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus* **3**, 317–330 (1993).
73. Bose, A. & Recce, M. Phase precession and phase-locking of hippocampal pyramidal cells. *Hippocampus* **11**, 204–215 (2001).
74. Skaggs, W. E., McNaughton, B. L., Wilson, M. A. & Barnes, C. A. Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus* **6**, 149–172 (1996).
75. Drieu, C. & Zugaro, M. Hippocampal sequences during exploration: mechanisms and functions. *Front. Cell Neurosci.* **13**, 1–22 (2019).
76. Gupta, A. S., van der Meer, M. A. A., Touretzky, D. S. & Redish, A. D. Segmentation of spatial experience by hippocampal theta sequences. *Nat. Neurosci.* **15**, 1032–1039 (2012).
77. Jensen, O. & Lisman, J. E. Hippocampal sequence-encoding driven by a cortical multi-item working memory buffer. *Trends Neurosci.* **28**, 67–72 (2005).
78. Friederici, A. D. & Singer, W. Grounding language processing on basic neurophysiological principles. *Trends Cogn. Sci.* **19**, 329–338 (2015).
79. King, C., Recce, M. & O'keefe, J. The rhythmicity of cells of the medial septum/diagonal band of Broca in the awake freely moving rat: relationships with behaviour and hippocampal theta. *Eur. J. Neurosci.* **10**, 464–477 (1998).
80. Heusser, A. C., Poeppel, D., Ezzyat, Y. & Davachi, L. Episodic sequence memory is supported by a theta-gamma phase code. *Nat. Neurosci.* **19**, 1374–1380 (2016).
81. Lisman, J. E. & Jensen, O. The theta-gamma neural code. *Neuron* **77**, 1002–1016 (2013).
82. Boeckx, C. & Benítez-Burraco, A. The shape of the human language-ready brain. *Front. Psychol.* **5**, 1–23 (2014).
83. Murphy, E. in *The Talking Species: Perspectives on the Evolutionary, Neuronal and Cultural Foundations of Language* (eds Luef, E. & Manuela, M.) 251–269 (Unipress Graz, 2018).
84. Doumas, L. A. A., Hummel, J. E. & Sandhofer, C. M. A theory of the discovery and predication of relational concepts. *Psychol. Rev.* **115**, 1–43 (2008).
85. Hummel, J. E. & Holyoak, K. J. Distributed representations of structure: a theory of analogical access and mapping. *Psychol. Rev.* **104**, 427–466 (1997).
86. Martin, A. E. A compositional neural architecture for language. *J. Cogn. Neurosci.* **32**, 1407–1427 (2020).
87. Chomsky, N. *Lectures on Government and Binding: The Pisa Lectures* (Foris, 1981).
88. Chomsky, N. *Aspects of the Theory of Syntax* (MIT Press, 1965).
89. Joshi, A. K., Levy, L. S. & Takahashi, M. Tree adjunct grammars. *J. Comput. Syst. Sci.* **10**, 136–163 (1975).
90. Shieber, S. M. *An Introduction to Unification-Based Approaches to Grammar* (Microtome, 2003).
91. Chomsky, N. *The Minimalist Program* (MIT Press, 1995).
92. Plate, T. A. Holographic reduced representations. *IEEE Trans. Neural Netw.* **6**, 623–641 (1995).
93. Carpenter, A. F., Baud-Bovy, G., Georgopoulos, A. P. & Pellizzer, G. Encoding of serial order in working memory: neuronal activity in motor, premotor, and prefrontal cortex during a memory scanning task. *J. Neurosci.* **38**, 4912–4933 (2018).
94. Petrides, M. Functional specialization within the dorsolateral frontal cortex for serial order memory. *Proc. R. Soc. Lond. B Biol. Sci.* **246**, 299–306 (1991).
95. Long, N. M. & Kahana, M. J. Hippocampal contributions to serial-order memory. *Hippocampus* **29**, 252–259 (2019).
96. Friederici, A. D., Fiebach, C. J., Schlesewsky, M., Bornkessel, I. D. & von Cramon, D. Y. Processing linguistic complexity and grammaticality in the left frontal cortex. *Cereb. Cortex* **16**, 1709–1717 (2006).
97. Lisman, J. E. & Idiart, M. A. P. Storage of 7±2 short-term memories in oscillatory subcycles. *Science* **267**, 1512–1515 (1995).
98. Bader, M. & Lasser, I. in *Perspectives on Sentence Processing* (eds Clifton, C., Frazier, L. & Reiner, K.) 225–242 (L. Erlbaum Associates, 1994).
99. Inoue, A. & Fodor, J. D. in *Japanese Sentence Processing* (eds Mazuka, R & Nagai, N.) 9–63 (L. Erlbaum Associates, 1995).
100. Mazuka, R. & Itoh, K. In *Japanese Sentence Processing* (eds Mazuka, R. & Nagai, N.) 295–329 (L. Erlbaum Associates, 1995).
101. Miyamoto, E. T. Case markers as clause boundary inducers in Japanese. *J. Psycholinguist. Res.* **31**, 307–347 (2002).
102. Tabor, W., Galantucci, B. & Richardson, D. Effects of merely local syntactic coherence on sentence processing. *J. Mem. Lang.* **50**, 355–370 (2004).
103. Altmann, G. T. M. & Mirković, J. Incrementality and prediction in human sentence processing. *Cogn. Sci.* **33**, 583–609 (2009).
104. Bransford, J. D. & Johnson, M. K. Contextual prerequisites for understanding: some investigations of comprehension and recall. *J. Verbal Learn. Verbal Behav.* **11**, 717–726 (1972).
105. Nelson, M. J. et al. Neurophysiological dynamics of phrase-structure building during sentence processing. *Proc. Natl Acad. Sci. USA* **114**, E3669–E3678 (2017).
106. Uddén, J., de Jesus Dias Martins, M., Zuidema, W. & Tecumseh Fitch, W. Hierarchical structure in sequence processing: how to measure it and determine its neural implementation. *Top. Cogn. Sci.* **12**, 910–924 (2020).
107. Carnie, A. *Syntax: A Generative Introduction* (Blackwell, 2002).
108. Berger, H. Über das Elektroenkephalogramm des Menschen. *Arch. Psychiatr. Nervenkr.* **87**, 527–570 (1929).
109. Nunez, P. L. & Srinivasan, R. Electroencephalogram. *Scholarpedia* **2**, 1348 (2007).
110. Rodin, E. & Funke, M. Cerebral electromagnetic activity in the subdelta range. *J. Clin. Neurophysiol.* **23**, 238–244 (2006).

111. Buzsaki, G. & Watson, B. O. Brain rhythms and neural syntax: implications for efficient coding of cognitive content and neuropsychiatric disease. *Dialogues Clin. Neurosci.* **14**, 345–367 (2012).
112. Klimesch, W. The frequency architecture of brain and brain body oscillations: an analysis. *Eur. J. Neurosci.* **48**, 2431–2453 (2018).
113. Breska, A. & Deouell, L. Y. Neural mechanisms of rhythm-based temporal prediction: delta phase-locking reflects temporal predictability but not rhythmic entrainment. *PLoS Biol.* **15**, e2001665 (2017).
114. Pikovsky, A., Kurths, J., Rosenblum, M. & Kurths, J. *Synchronization: A Universal Concept in Nonlinear Sciences* (Cambridge Univ. Press, 2003).
115. Strogatz, S. H. *Nonlinear Dynamics and Chaos with Student Solutions Manual: With Applications to Physics, Biology, Chemistry, and Engineering* (CRC, 2018).
116. Kopell, N., Ermentrout, G. B., Whittington, M. A. & Traub, R. D. Gamma rhythms and beta rhythms have different synchronization properties. *Proc. Natl Acad. Sci. USA* **97**, 1867–1872 (2000).
117. Buzsáki, G. & Draguhn, A. Neuronal oscillations in cortical networks. *Science* **304**, 1926–1929 (2004).
118. Herreras, O. Local field potentials: myths and misunderstandings. *Front. Neural Circuits* **10**, 101 (2016).
119. Buzsáki, G., Anastassiou, C. A. & Koch, C. The origin of extracellular fields and currents-EEG, ECoG, LFP and spikes. *Nat. Rev. Neurosci.* **13**, 407–420 (2012).
120. Doelling, K. B., Florencia Assaneo, M., Bevilacqua, D., Pesaran, B. & Poeppel, D. An oscillator model better predicts cortical entrainment to music. *Proc. Natl Acad. Sci. USA* **116**, 10113–10121 (2019).
121. Helfrich, R. F., Breska, A. & Knight, R. T. Neural entrainment and network resonance in support of top-down guided attention. *Curr. Opin. Psychol.* **29**, 82–89 (2019).
122. Obleser, J., Herrmann, B. & Henry, M. J. Neural oscillations in speech: don't be enslaved by the envelope. *Front. Hum. Neurosci.* **6**, 2008–2011 (2012).
123. Doelling, K. B., Arnal, L. H., Ghitza, O. & Poeppel, D. Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *Neuroimage* **85**, 761–768 (2014).
124. van Rullen, R. Perceptual cycles. *Trends Cogn. Sci.* **20**, 723–735 (2016).
125. Shamma, S. A., Elhilali, M. & Micheyl, C. Temporal coherence and attention in auditory scene analysis. *Trends Neurosci.* **34**, 114–123 (2011).

Acknowledgements

The authors are very grateful to E. Lau and M. Yoshida for their vast input and feedback. They also thank J. Mitchell for help with corpus data, J. Bowers for his comments and S. Brendecke for help with illustrations. N.K. acknowledges the support of the International Laboratory of Social Neurobiology, Institute for Cognitive Neuroscience, Higher School of Economics, Russian Federation (grant 075-15-2022-1037). A.T. acknowledges the support of the Max Planck Society.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41583-022-00659-5.

Correspondence should be addressed to Nina Kazanina.

Peer review information *Nature Reviews Neuroscience* thanks M. F. Assaneo and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Related links

Natural Language Toolkit: https://www.nltk.org/

© Springer Nature Limited 2022