scientific reports

OPEN

Structural and sequential regularities modulate phrase-rate neural tracking

Junyuan Zhao[1], Andrea E. Martin[2,3] & Cas W. Coopmans[2,3]✉

Electrophysiological brain activity has been shown to synchronize with the quasi-regular repetition of grammatical phrases in connected speech—so-called phrase-rate neural tracking. Current debate centers around whether this phenomenon is best explained in terms of the syntactic properties of phrases or in terms of syntax-external information, such as the sequential repetition of parts of speech. As these two factors were confounded in previous studies, much of the literature is compatible with both accounts. Here, we used electroencephalography (EEG) to determine if and when the brain is sensitive to both types of information. Twenty native speakers of Mandarin Chinese listened to isochronously presented streams of monosyllabic words, which contained either grammatical two-word phrases (e.g., catch fish, sell house) or non-grammatical word combinations (e.g., full lend, bread far). Within the grammatical conditions, we varied two structural factors: the position of the head of each phrase and the type of attachment. Within the non-grammatical conditions, we varied the consistency with which parts of speech were repeated. Tracking was quantified through evoked power and inter-trial phase coherence, both derived from the frequency-domain representation of EEG responses. As expected, neural tracking at the phrase rate was stronger in grammatical sequences than in non-grammatical sequences without syntactic structure. Moreover, it was modulated by both attachment type and head position, revealing the structure-sensitivity of phrase-rate tracking. We additionally found that the brain tracks the repetition of parts of speech in non-grammatical sequences. These data provide an integrative perspective on the current debate about neural tracking effects, revealing that the brain utilizes regularities computed over multiple levels of linguistic representation in guiding rhythmic computation.

Language comprehension requires the transformation of a continuous physical signal (speech or sign) into discrete structured representations (linguistic units and structured relations between them). Evidence shows that during successful comprehension, low-frequency neural activity becomes aligned with linguistic units such as phonemes and syllables[1,2]. This 'neural tracking' effect was initially shown for these lower-level linguistic units, but more recently it has been observed for higher-level and rather abstract units such as words and phrases as well[3–9]. And while syllables can be physically characterized by salient energy increases[10], these higher-level units have no one-to-one physical correlates in the speech signal, meaning that they require endogenous computations to support neural tracking[11,12].

It has been established that the brain tracks both *sequential* and *structural* linguistic information in spoken language. That is, when a connected speech stream contains a regular sequential pattern of (certain subtypes of) words or syllables, the brain response will show neural tracking at the rate at which these linguistic units occur in the speech[5,13,14]. What is striking is that structurally organized information elicits similar effects. In particular, when phrasal units like verb phrases or noun phrases are regularly repeated in a connected speech stream, the brain response will show neural tracking at the phrase rate[3,6–9]. Since phrase structure is organized hierarchically, these structural regularities are not directly visible in the sequential structure of language—indeed, the notion 'phrase rate' presumes the cognitive capacity to represent hierarchical phrase structure by abstracting away from the sequential input. It thus appears that the brain concurrently tracks linguistic information at multiple levels of representation and at multiple timescales[5]. Throughout our discussion of neural tracking, we will use the terms 'structural information' and 'sequential information' to denote the contrast between linguistic information that is hierarchically organized (in particular, compositional phrase structures) and linguistic information

[1]Department of Linguistics, University of Michigan, Ann Arbor, MI, USA. [2]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands. [3]Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands. ✉email: Cas.Coopmans@mpi.nl

that is sequentially organized in speech (in particular, the sequential repetition of lexical information). While we note that the contrast is not dichotomous, we use this description because it captures one of the fundamental challenges of language processing: as hierarchical structure is not directly encoded in the linear speech input, tracking hierarchical structure requires processes of perceptual inference that go beyond merely following the sequential input[11,15].

Because sequential distribution is (partially) determined by structural properties, structural and sequential regularities in natural language are correlated. For instance, syntactic structure has correlates in the surface statistics of sentences[16–18], and processing syntactic structure can give rise to statistical expectations about sequential properties of sentences (e.g., the likelihood that certain words are adjacent[19,20]). Due to the correlation between structural and sequential information, it is difficult to delineate their functional roles in language processing, in particular when viewed through the lens of neural speech tracking[9,21]. In this electroencephalogram (EEG) study, we address the role of sequential and structural information in neural speech tracking by presenting participants with speech streams that vary in syntactic structure. In particular, we ask (1) whether structural and sequential regularities affect tracking differently, and (2) how syntactic variables modulate the neural tracking of phrases.

Neural speech tracking is modulated by structural information

Because natural language is hierarchically structured, it is rife with properties that separate meaning from the linear or temporal order in which it is produced[22–26]. Processing language therefore requires internal computations that integrate and resolve information over temporal displacement and variation[11,15]. Such internal computations allow basic elements to break free from the linear order they are imposed with and be interpreted with respect to hierarchically related units. In this section, we will discuss a growing body of research that shows a connection between hierarchical structure and the neural tracking of speech.

The first attempts to reveal the role of hierarchical structure in neural tracking are attributed to Ding and colleagues[4,5]. In their "frequency-tagging" studies, participants listened to sequences of monosyllabic words, which were isochronously presented at 4 Hz. In these sequences, two adjacent words could be grouped into two-word phrases (occurring at 2 Hz), and two adjacent phrases could be grouped into four-word sentences (occurring at 1 Hz). As the sequences contained synthesized speech that was presented without prosody or coarticulation, neither phrases nor sentences were auditorily cued in the stimulus, and their rates were frequency-tagged only by virtue of participants' knowledge of grammar. It was found that electrophysiological brain activity closely matched the occurrence rates of these abstract linguistic units: phrases elicited spectral peaks at 2 Hz and sentences elicited a spectral peak at 1 Hz (see also refs.[27,28]). Crucially, this frequency-tagged effect of grammatical structure disappeared when participants listened to structurally identical materials in an unfamiliar language. These findings suggest that structural knowledge of language supports 'chunking' continuous speech into structured, multi-word units[29].

In addition to such abstract grouping patterns, further studies showed that the content of linguistic structure also modulates neural speech tracking. Using naturally produced speech, two studies[6,7] showed that phrases were tracked more strongly in sentences than in prosodically natural jabberwocky sentences and in unstructured word lists. Moreover, studies have found that the strength of phrase-rate tracking is correlated with behavioral measures of comprehension[30,31], thus showing the perceptual relevance of neural tracking. Note that by phrase-rate tracking we mean tracking effects at the frequency that corresponds to the occurrence rate of phrasal units. It can, but need not, refer to the tracking of syntactic phrases.

That neural tracking of speech is indicative of a structural specification extending beyond the contrast between sentences and word lists is also suggested by a recent frequency-tagging study. A recent study by Burroughs and colleagues[3] reported stronger tracking at the phrase rate when participants listened to sequences of structurally consistent adjective-noun phrases than when they listened to sequences that were composed of structurally-variable two-word phrases (e.g., adverb-adjective, verb-noun, preposition-noun, etc.). Their findings are interesting in two respects. First, the fact that phrases were tracked even when they were composed of different parts of speech shows that a purely sequential account of phrase-rate tracking is insufficient, because these stimuli contained no sequential regularities at the phrase rate. Instead, the only phrase-rate regularity in these variable sequences is a structural one, namely the repetition of two-word phrase structures. Second, the fact that phrases in structurally variable sequences were nevertheless tracked less closely than phrases in structurally consistent sequences shows that, even in the context of intact syntactic composition, the structural and sequential details of a phrase influence phrasal tracking. However, as the structurally-variable condition in this study[3] differed from the structurally-consistent adjective-noun condition in various linguistic respects (e.g., type of attachment, sequential position of the head, see Section. "The present study"), it remains unclear which linguistic factors modulate phrasal tracking. As studies that explicitly examine the role of syntactic or semantic information are sparse, a gap remains between our rather coarse-grained understanding of neural speech tracking and our fine-grained knowledge of linguistic structure[32].

Neural speech tracking is modulated by sequential information

Contrasting with these structural accounts, it has been argued that the regular, sequential occurrence of lexical information provides a more parsimonious explanation of phrase-rate tracking[13,14]. According to this view, a tracking effect emerges at the phrase rate because certain word-level features are repeated at the same frequency as phrases. A 2-Hz phrase-rate peak arises when participants listen to isochronously presented 1-s sentences like "dry fur rubs skin" simply because nouns (i.e., "fur", "skin") are repeated every 1/2 second[13]. This account of tracking is based on the sequential occurrence of word-level information (i.e., the semantic features associated with parts of speech, derived from statistical distributions) and therefore does not require hierarchically structured representations.

Another non-hierarchical account of phrase-rate tracking holds that the effect is driven by sequential statistics, that is, the rhythmic temporal variation in transitional probabilities between words. In the sentence "dry fur rubs skin", word-to-word transitional probabilities are higher within the phrases "dry fur" and "rubs skin" than between these two phrases (i.e., at the boundary between "fur" and "rubs"), leading to an alternating high-low pattern that coincides with phrase occurrence (see ref.[18]). Given that the human brain is sensitive to transitional probabilities during speech segmentation (to extract word-like units[33]), it might employ similar mechanisms to extract phrases in speech sequences. Indeed, similar tracking effects are found when participants listen to connected sequences of unfamiliar words that are cued only by high transitional probabilities between syllables[34–36]. For example, in a recent study by Bai and colleagues[34], Dutch native speakers were presented with sequences of disyllabic Chinese words. As these participants had no prior knowledge of Chinese, the only cues available to extract multi-syllabic units were the differences in syllable-to-syllable transitional probabilities, which were high within words and low between words. Over the course of the experiment, tracking at the word rate emerged for these statistically indexed words. In order to show that it is linguistic structure that drives neural tracking in a frequency-tagging design, it is therefore important to control for the differences between conditions in terms of the sequential transitional probability between linguistic units.

The present study

The previous two sections show that both structural and sequential regularities correlate with rhythmic neural activation. In the present study, we seek to enrich our understanding of these relationships by presenting participants with speech streams that vary in syntactic structure and sequential-statistical information. By measuring phrase-rate tracking in these different speech streams, we investigate (1) how structural (syntactic) and sequential regularities together affect the neural tracking of speech, and (2) which aspects of syntactic structure are reflected in tracking effects. To answer the first question, we used a balanced design in which structural and sequential information were orthogonalized as much as possible (we say as much as possible because a full orthogonalization between structural and sequential information can hardly be achieved since these variables are highly correlated in natural language). Specifically, half our conditions contain grammatical word combinations, the other half contain non-grammatical combinations. Moreover, half of our conditions contain a sequential regularity at the phrase rate (i.e., the sequential repetition of lexical information), the other half do not. The structural and sequential factors were fully crossed, such that our design contained all combinations of +/- syntactic structure and +/- lexical repetition.

In the three grammatical conditions, we probed the effects of structural information on the neural tracking phrases. The grammatical conditions consisted of repetitions of two-word phrases, in which we varied two properties that play a prominent role in syntactic theory. In particular, we varied the position of the head of the two-word phrase (Head position) and the syntactic function of the word attached to the head (Attachment type).

Syntactic structures are endocentric, which means that the syntactic status of a phrase is determined by one of its elements, the 'head'. Thus, the verb is the head of a verb phrase (VP; e.g., *win* in *win the game*), the preposition is the head of a prepositional phrase (PP; e.g., *on* in *on Monday*), and so on. Moreover, the syntactic function of the other element in a phrase is dependent on the selectional properties of the head. Elements that are required to complete the meaning of the head and receive a thematic role are called complements (or arguments). Adjuncts, instead, do not receive a thematic role; they modify the grammatical head and typically convey optional information[37,38]. Consider the VP *win the game on Monday*. The head of the phrase, the verb *win*, assigns a thematic role to *the game*, which is the verb's complement (*the game* completes the meaning of *win*). The PP *on Monday* is not inherently required by the verb and does not receive a thematic role, so it is considered an adjunct. Certain syntactic theories additionally propose a difference between complement and adjunct attachments in terms of their underlying syntactic structure[38,39].

Typologically, languages differ in whether the head occupies the left or right edge of the phrase. In a head-initial language like English, the head consistently precedes its complements. Adjuncts, on the other hand, are freer and can often be placed on either side of the head. These syntactic differences in attachment and head position and are visually illustrated in Fig. 1.

We included the factors Attachment type and Head position because the structurally variable 'mixed-phrase' condition in the study by Burroughs and colleagues[3] contained phrases that varied in these respects. Recall that they found stronger phrase-rate tracking in sequences containing structurally consistent adjective-noun (A-N) phrases than in sequences containing mixed phrases. An example of their mixed-phrase condition is given below:

<u>too good</u> <u>want fame</u> <u>from rice</u> <u>sit more</u>
Adv   A   V   N   P   N   V Adv

The position of the head of these two-word phrases is not consistent: the adjective *good* in the adjectival phrase (AP) *too good* is the second word, the verb *want* in the VP *want fame* is the first word, the preposition *from* in the PP *from rice* is the first word, and so on. Moreover, the word attached to the head in each phrase does not consistently fulfill the same syntactic function; it can be a complement or an adjunct. The adverb *too* is an adjunct to *good*, *fame* is the complement of *want*, *rice* is the complement of *from*, etcetera. Inconsistency in either of these variables (Head position and Attachment type) might have led mixed phrases to be tracked less closely than structurally consistent A-N phrases. We consider this a reasonable possibility because both variables affect languages processing.
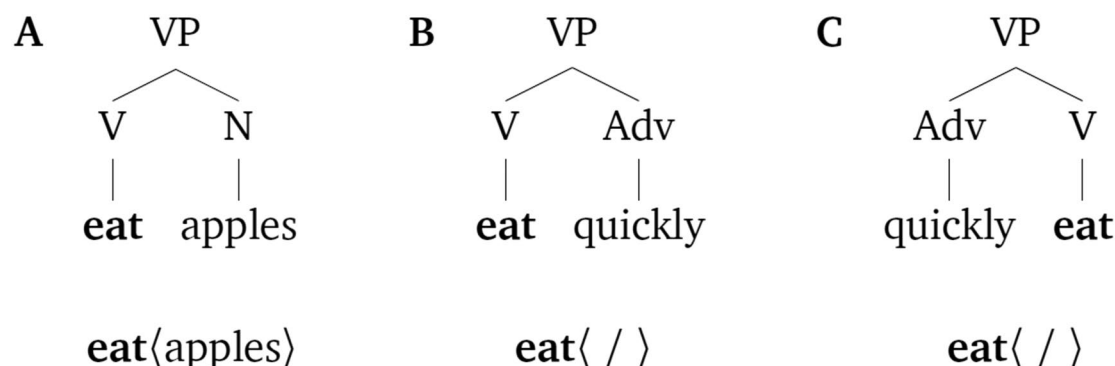
Figure 1. Syntactic representations of different verb phrases. In (**A**), the noun *apples* is the complement of the verb *eat*, which is the head of the VP (in bold). In (**B**), the adverb *quickly* is attached to the verb phrase as an adjunct. Because the phrase contains no complement, the thematic role associated with the verb is not assigned, represented as ⟨/⟩. The verb phrase in (**C**) is the same as the one in (**B**) except that the adverb precedes the verb, which appears in final position.

In head-initial structures, complements are easier to process than adjuncts. For instance, the PP *for a raise* is read faster than the PP *for a month* when preceded by *The company lawyers considered employee demands…*[40–44]. The noun *demands* takes a PP complement. Reading *demands* therefore facilitates the processing of *for a raise*, which can fulfill that complement role (i.e., *to demand a raise*), compared to *for a month*, which is a temporal adjunct to the verb *considered*. This example also illustrates the role of the head (here, *demands*) in syntactic processing, which allows syntactic structure to be projected (as in head-driven parsing models[40,45]) and can trigger expectations for upcoming linguistic material[41,44,46]. Moreover, cross-linguistic differences in head position modulate sentence-processing effects. In head-initial structures it is commonly found that processing costs are increased when the distance between the head and its dependents is increased[47,48]. Studies investigating head-final structures, instead, have found the exact opposite, i.e., a facilitation in processing of words integrating longer dependencies, which is typically attributed to an increase in the prediction for the upcoming head[49,50]. Given the robust effects of Attachment type and Head position on sentence processing, we incorporated both factors in our design.

Constructing minimal contrasts with these two factors resulted in three grammatical conditions, whose labels refer to the parts of speech they are composed of: V–N, V-Adv and V-Adv (alt.). The V–N condition contains sequences of Verb-Noun combinations, which are consistently head-initial and contain a noun as the complement of the verb (like in Fig. 1A). V-Adv sequences contain Verb-Adverb phrases that are consistently head-initial as well, but the adverb is an adjunct to the verb phrase (like in Fig. 1B). In a third condition, V-Adv (alt.), sequences are composed of Verb-Adverb (adjunct) combinations whose order was shuffled, creating a random alternation between head-initial and head-final phrases (Figs. 1B and C, respectively). Our stimuli were in Mandarin Chinese because this language naturally allows phrases of both the V-Adv and the Adv-V order. In sum, the effect of Attachment type (complement vs. adjunct) is tested in the comparison between V–N and V-Adv. The effect of Head position (consistent vs. varying head position) is tested in the comparison between V-Adv and V-Adv (alt.).

In three non-grammatical conditions, we probed the effects of non-structural information, i.e., the sequential repetition of lexical information. The N-R condition contains sequences composed of two-word pairs with a Noun and a Random word, yielding sequential repetition of the noun at the phrase rate in the absence of phrasal structure. The V-V condition consists of random Verb sequences, thus having sequential repetition of verbs at the word rate. The R-R condition contains random word sequences and therefore contains no regular repetition of lexical information at all. All non-grammatical conditions were constructed such that there were no grammatical phrases in the concatenated sequence. An overview of the experimental design with example materials is given in Table 1.

We recorded the EEG of participants while they listened to Mandarin Chinese speech streams that vary in structural and sequential regularities (see Fig. 2). The streams consisted of connected syllables isochronously presented at 2 Hz, yielding a 1-Hz rate for two-word phrases. Following previous work[3,4,8], neural tracking was quantified through measures of evoked power and inter-trial phase coherence (ITPC) of the frequency-domain representation of the EEG signal.

We consider three theoretical possibilities, which make different predictions about the presence of a tracking effect at the 1-Hz phrase rate. First, a *structure*-only hypothesis posits that the brain only tracks syntactic structure. Following this hypothesis, a phrase-rate tracking effect is expected only in the conditions that contain syntactic structure (i.e., the grammatical conditions V–N, V-Adv, and V-Adv (alt.)). We also hypothesize that our structural manipulations will yield different magnitudes of tracking[3,6]. However, we cannot make specific predictions about the directions of these effects, because in existing data several factors co-occur (e.g., in the study by Burroughs et al.[3], the adjective-noun and mixed-phrase conditions differed in multiple respects), and because mechanistic linking hypotheses between frequency representations and syntactic-structure building are missing[32]. Second, a *sequence*-only hypothesis posits that the brain only tracks the sequential repetition of information. It therefore predicts phrase-rate tracking effects in conditions with regular repetition at the word level (i.e., V–N, V-Adv, and N-R), regardless of whether these conditions contain syntactic structure as well. The

| Condition	Structural regularity	Syntactic specification	Sequential regularity	Examples	
V–N	✓	Head-Initial Complement	✓	看天 kan4 tian1 **look** sky	喂鱼 wei4 yu2 **feed** fish
V-Adv	✓	Head-Initial Adjunct	✓	放下 fang4 xia4 **put** downwards	派出 pai4 chu1 **send** outwards
V-Adv (alt.)	✓	Head-random Adjunct	X	重做 chong2 zuo4 again **do**	说明 shuo1 ming2 **state** clearly
N-R	X	–	✓	电伪 dian4 wei3 electricity fake	发再 fa4 zai4 hair again
R-R	X	–	X	下讲 xia4 jiang3 down talk	变降 bian4 jiang4 change descend
V-V	X	–	X	取听 qu3 ting1 get listen	剥搬 bo1 ban1 peel move

Table 1. Overview of the experimental design, with examples of each condition. The three lines in the Examples column represent Chinese characters, Chinese phonological transcription with tones, and word-by-word translation in English, respectively. In the grammatical conditions, the head of each verb phrase is indicated in bold.
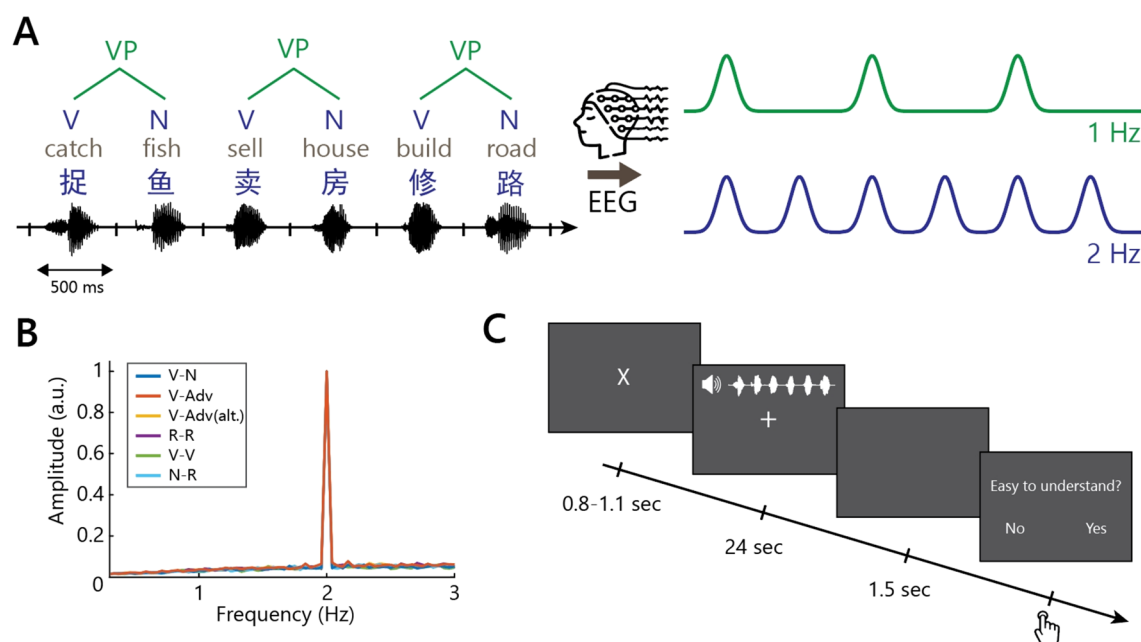
Figure 2. Schematic representation of the materials and experimental procedure. (**A**) Multi-level linguistic units present in connected speech sequences (left) and hypothetical neural responses to different levels of linguistic representation (right). This is a toy illustration of the temporal unfolding of monosyllabic words (every 500 ms) and phrases (every 1000 ms), with hypothetical neural responses reflecting the processing of words and phrases at the corresponding rates of presentation. (**B**) The spectrum of stimulus intensity reflects the average spectral amplitude of synthesized speech signals in the frequency domain. A spectral peak was present in all conditions only at 2 Hz, corresponding to the word rate. The six conditions were not acoustically different. (**C**) Illustration of an experimental trial in one condition. (This figure includes icons from Flaticon.com).

structure-only and the *sequence*-only hypotheses may appear too simplistic, because it is unlikely that the brain only tracks one type of information. We nevertheless present the hypotheses this way because existing accounts of neural tracking effects lean towards one of the two explanations[8,13,14]. A third, *hybrid* hypothesis holds that the brain tracks both structural and sequential regularities. It predicts phrase-rate tracking effects in all conditions that contain regularity at the phrase rate, whether its source is structural or sequential (i.e., the grammatical conditions V–N, V-Adv, V-Adv (alt.), and the non-grammatical condition N-R), but does not specify the exact source of a tracking effect in cases where both types of regularity co-exist (i.e., in V–N and V-Adv). We consider the hybrid hypothesis most likely a priori, but nonetheless aim to assess the degree to which these three competing hypotheses are supported by our results.
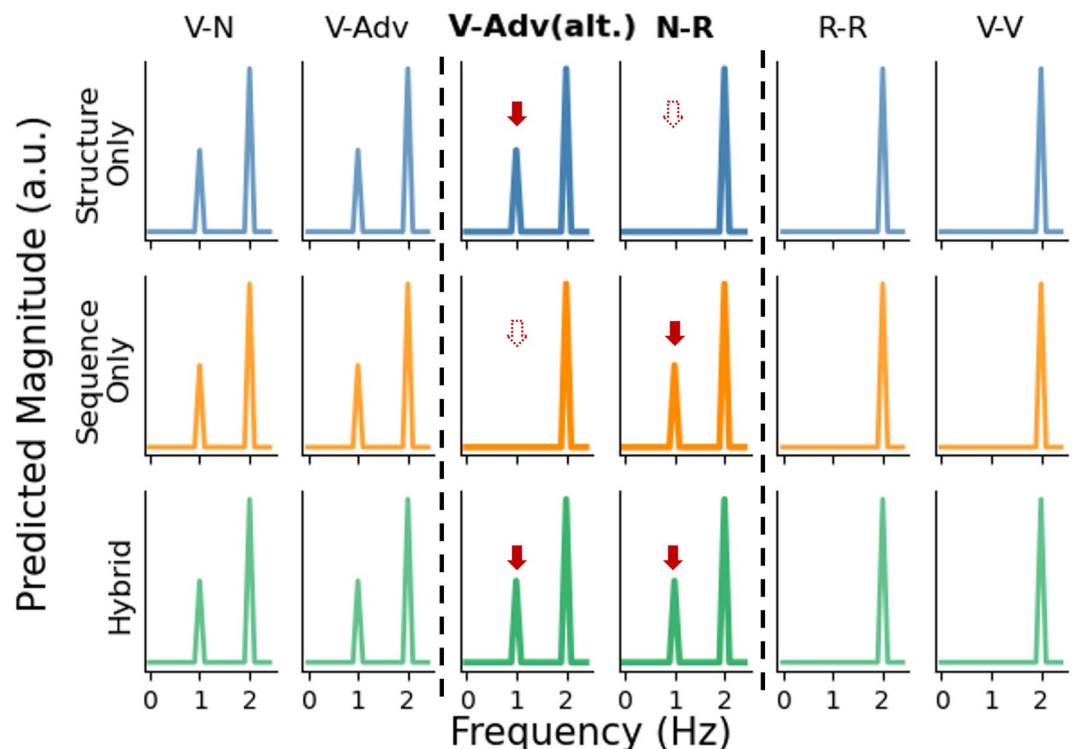
Figure 3. Predicted outcomes in the frequency domain following the three hypotheses. The predictions only indicate the presence or absence of peaks and do not reflect effect sizes. The columns for V-Adv (alt.) and N-R are emphasized because these are the critical conditions for which the three accounts make different predictions. The solid red arrows in these two columns indicate phrasal peaks that are expected under the different hypotheses. Dashed arrows indicate the absence of spectral peaks at the phrasal frequency.

Figure 3 presents the predicted effects following the three hypotheses. As can be seen, the three hypotheses converge for conditions V–N, V-Adv, R-R, and V-V regarding the presence or absence of tracking effects. Given that they diverge only for V-Adv (alt.) and N-R, these conditions will be the main focus of this study, in accordance with the research questions outlined above.

Results

At the 2-Hz syllable rate, all conditions elicited peaks in both power and ITPC (all $p < 0.001$). A further comparison between conditions did not show any reliable difference in power or ITPC of the 2-Hz syllable peaks (all $p > 0.4$). Figures 4A and C present the power and ITPC spectra of all conditions.

At the 1-Hz phrase rate, the results for evoked power and ITPC were slightly different. For evoked power (Fig. 4A), significant phrase-rate peaks were present in all grammatical conditions (V–N: $p < 0.001$; V-Adv: $p < 0.001$; V-Adv (alt.): $p = 0.047$). We did not find a phrasal peak in the power spectrum of the N-R condition ($p = 0.46$). For ITPC (Fig. 4C), phrase-rate peaks were again found in the grammatical conditions V–N ($p < 0.001$), V-Adv ($p = 0.002$), V-Adv (alt.) ($p = 0.044$), but also in the non-grammatical N-R condition ($p = 0.043$). Besides this difference in frequency spectra, power and ITPC at 1 Hz also had different scalp distributions (see Figs. 4B and D). Such differences between power and ITPC are not entirely unexpected, since ITPC has been shown mathematically to be more sensitive to neural responses that are synchronized to external stimuli[51]. This, combined with the fact that four out of six conditions showed a 1-Hz peak in ITPC (compared to three out of six conditions showing a 1-Hz peak in evoked power), reveals the higher sensitivity of ITPC to tracking effects. Our pairwise comparisons will therefore be focused on ITPC.

To examine whether neural tracking at the phrase rate was different across conditions, we performed paired samples t-tests on all conditions that showed a 1-Hz peak in the ITPC spectrum (i.e., V–N, V-Adv, V-Adv (alt.), and N-R). The comparisons are shown in Fig. 5A. Two theoretically informed planned contrasts (i.e., V–N vs. V-Adv for the effect of Attachment and V-Adv vs. V-Adv (alt.) for the effect of Head position) revealed that phrases with complement attachments were tracked more strongly than phrases with adjunct attachments (V–N > V-Adv, $\Delta = 0.057$, $p = 0.034$) and that phrases with a consistent head position were tracked more strongly than phrases with a varying head position (V-Adv > V-Adv (alt.), $\Delta = 0.042$, $p = 0.012$). We also found stronger tracking for V–N than for both V-Adv (alt.) ($\Delta = 0.010$, $p < 0.001$) and N-R ($\Delta = 0.103$, $p < 0.001$), and stronger tracking for V-Adv than for N-R ($\Delta = 0.046$, $p = 0.008$). There was no difference between V-Adv (alt.) and N-R ($\Delta = 0.004$, $p = 0.772$).
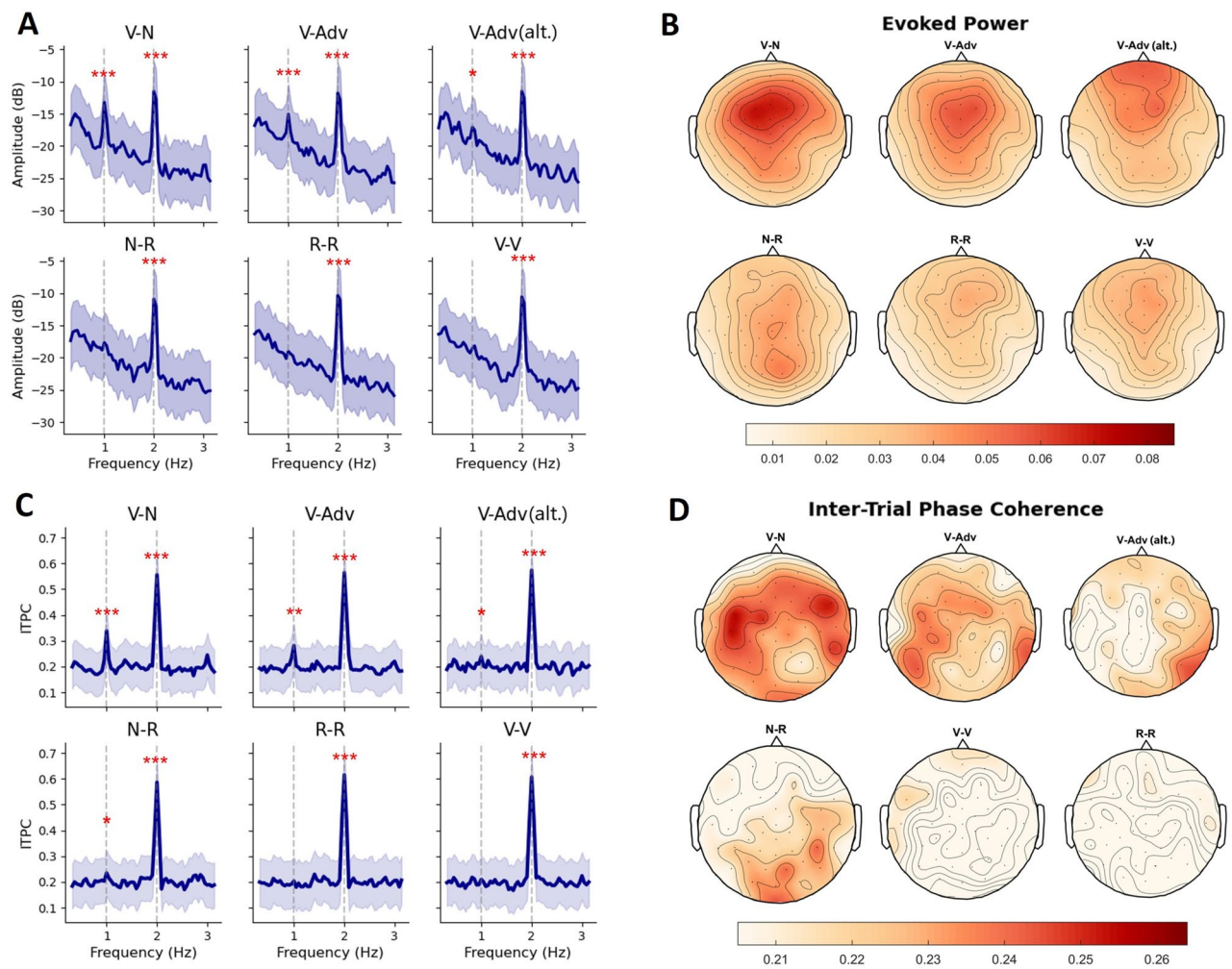
Figure 4. EEG responses in all conditions. (**A**) Spectra of evoked power. (**B**) Topographies of 1-Hz evoked power. (**C**) Spectra of ITPC. (**D**) Topographies of 1-Hz ITPC. In (**A**) and (**C**), the shaded areas indicate the between-subjects standard error of the mean. Significance: *, $p < 0.05$; **, p < 0.01; ***, $p < 0.001$.

The ITPC results show that the brain is sensitive to both sequential and structural regularities. The next question is whether these effects are independently additive or whether they interact with one another. To test this, we ran a two-way repeated measures analysis of variance (ANOVA) on a subset of the data (see Fig. 5B). In this subset, the two binary factors STRUCTURAL and SEQUENTIAL regularities were fully orthogonalized, yielding the following four groups: V-Adv ([+ SEQ, + STRUCT], V-Adv (alt.) [-SEQ, + STRUCT], N-R [+ SEQ, -STRUCT], and R-R [-SEQ, -STRUCT]). As expected, the ANOVA yielded significant main effects of both structural (F_{STRUCT} (1, 19) = 13.13, $p = 0.002$) and sequential regularities (F_{SEQ} (1, 19) = 6.86, $p = 0.017$). The interaction was not significant ($F_{interaction}$(1, 19) = 0.33, $p = 0.572$). Yet, the results in Fig. 5B nevertheless suggest that the effect of SEQUENTIAL regularity is not the same for conditions with versus without STRUCTURAL regularity. A post-hoc analysis on estimated marginal means[52] tentatively confirmed this suspicion. Regarding the role of sequential regularity, when the stimuli contained grammatical structure, adding sequential regularity increased ITPC ([+ SEQ, + STRUCT] > [-SEQ, + STRUCT], $\beta = 0.042$, $SE = 0.014$, $t(19) = 2.85$, $p = 0.012$). However, this effect of sequential regularity vanished when the stimuli contained no grammatical structure ([+ SEQ, -STRUCT] = [-SEQ, -STRUCT], $\beta = 0.032$, $SE = 0.019$, $t(19) = 1.68$, $p = 0.140$). The role of structural regularity, instead, appears to be more consistent. Adding structural regularity increased ITPC, both when there was sequential regularity in the stimuli ([+ SEQ, + STRUCT] > [+ SEQ, -STRUCT], $\beta = 0.046$, $SE = 0.016$, $t(19) = 2.96$, $p = 0.008$) and when there was no sequential regularity in the stimuli ([-SEQ, + STRUCT] > [-SEQ, -STRUCT], $\beta = 0.036$, $SE = 0.014$, $t(19) = 2.61$, $p = 0.017$). All p-values reported for the post-hoc analysis were adjusted with Tukey's method. Thus, it appears that the effect of STRUCTURAL regularity is more robust than the effect of SEQUENTIAL regularity: the effect of structural regularity shows up even in the face of sequential inconsistency, but the effect of sequential regularity is dependent on the presence of grammatical structure. We do note that these post-hoc pairwise comparisons should be interpreted with caution, because the interaction in the two-way ANOVA was not significant.

As an additional control analysis, we used a linear regression analysis to test if the 1-Hz ITPC effects can be explained by differences in the transitional probabilities (TPs) of the two-word combinations in the different conditions. This analysis included the conditions V–N, V-Adv, and V-Adv (alt.). The condition N–R was not
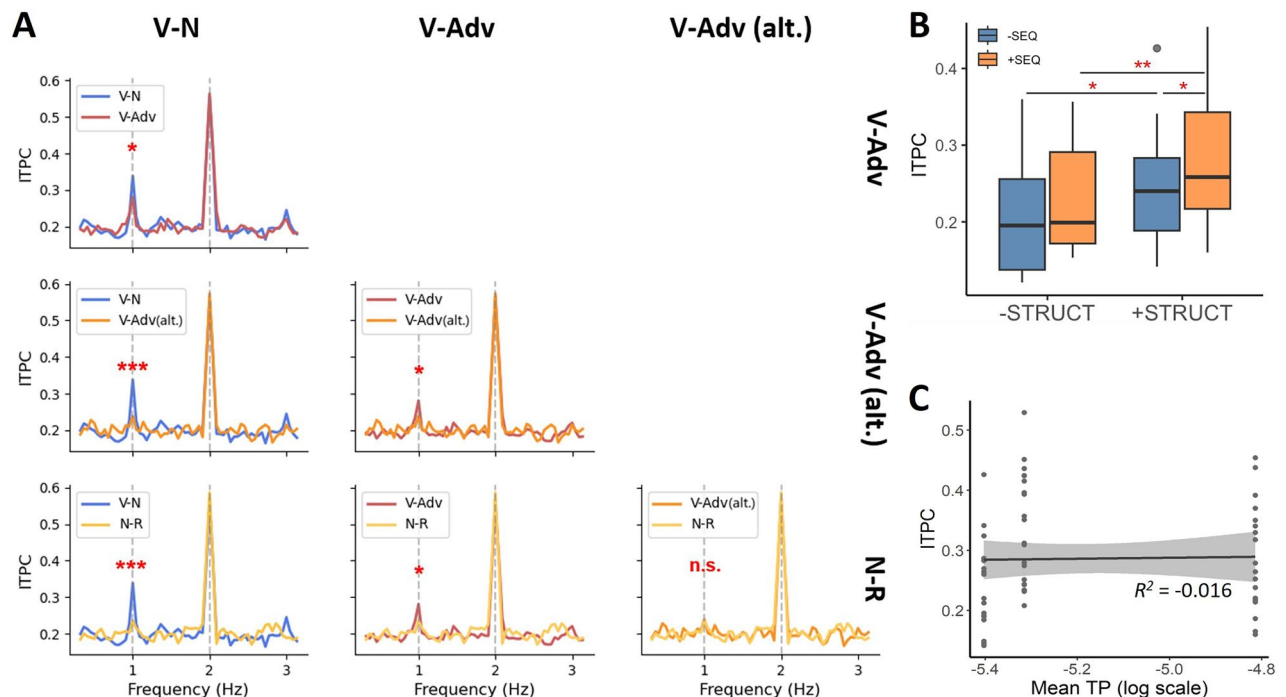
Figure 5. (**A**) Pairwise comparisons of 1-Hz peaks in ITPC. (**B**) Boxplots of ITPC grouped by STRUCTURAL and SEQUENTIAL regularities, where [+ SEQ, + STRUCT] = V-Adv, [-SEQ, + STRUCT] = V-Adv (alt.), [+ SEQ, -STRUCT] = N-R, and [-SEQ, -STRUCT] = R-R. Asterisks indicate the significant results of post-hoc comparisons following a two-way repeated measures ANOVA. (**C**) Scatter plot and least-squares linear regression fitted between ITPC and TP in all grammatical conditions. Mean TP values of each condition were used as the predictor. The shaded area indicates 95% confidence interval. Significance: *, $p < .05$; **, $p < .01$; ***, $p < .001$.

included because there were not enough N-R bigrams in the corpus to compute a sufficiently representative average TP for these non-grammatical combinations. The bigram TP of a phrase w_1w_2 was computed as the probability of encountering that phrase divided by the probability of encountering all other two-word phrases starting with w_1, or $P(w2|w1) = \frac{P(w1w2)}{P(w1\#)}$. Word frequencies were extracted from the Chinese Wikipedia corpus used in a previous study[13]. ITPC on individual trials and average TP in the V–N, V-Adv, and V-Adv (alt.) conditions were then entered into a linear regression model, which showed that TP did not reliably predict ITPC (Fig. 5C; $R^2 = - 0.016$, $\beta = 0.22$, $SE = 2.3$, $t(58) = 0.097$, $p = 0.923$). Thus, while there might be differences between the conditions in terms of sequential statistics, this cannot explain the pattern of results observed here.

Discussion

In this EEG study, we investigated whether and how neural tracking of phrases in connected speech is modulated by regularities stemming from syntactic structure and from sequential lexical information. Besides robust syllable tracking in all conditions[1], we observed tracking effects at the phrase rate in all conditions that contained a structural and/or sequential regularity at that rate. In line with the literature[3–8,30], all grammatical conditions that can be characterized with a regular structural pattern elicited phrase-rate tracking, whether they contained sequential regularities (V–N, V-Adv) or not (V-Adv (alt.)). These grammatical tracking effects were furthermore modulated by the syntactic properties of the phrases. Moreover, we found phrase-rate tracking in conditions with sequential lexical regularities, even in the absence of grammatical structure (N-R). In all, these findings suggest that the phrase-rate tracking effect is a general neural readout of regularity tracking, which is sensitive to the presence of different levels of linguistic representation.

Tracking structural regularities

In both ITPC and evoked power, we observed 1-Hz phrase-rate tracking for all three conditions that contained grammatical structure (i.e., Verb-Noun, Verb-Adverb, and Verb-Adverb (alternation)), which had a central-anterior scalp distribution in evoked power. Pairwise comparisons revealed effects of syntactic properties: verb phrases with complements were tracked more strongly than verb phrases with adjuncts (V–N > V-Adv), and verb phrases with a consistent head position were tracked more strongly than verb phrases with varying head positions (V-Adv > V-Adv (alt.)).

The overall pattern of phrase-rate tracking in grammatical conditions is not predicted by existing non-syntactic accounts. For instance, the tracking effects cannot be explained by prosody[53], because there is no difference between V–N and V-Adv sequences in either explicit prosodic structure (see Fig. 2B) or implicit prosodic grouping (i.e., all grammatical conditions contained repetitions of verb phrases). Neither can the effects be captured

in terms of differences in transitional probabilities[34,36], because there was no reliable relationship between the magnitude of the tracking effects and bigram transitional probabilities across conditions (see Fig. 5C). Alternatively, one might attribute the effect to the fact that V–N and V-Adv, as constructions, differ in frequency (e.g., the frequencies of the combinations V–N and V-Adv might be different). However, this presumes the representational capacity for syntactic abstraction in the first place. If people can compute statistics based on grammatical category combinations, they are able to represent syntactic structure extending beyond lexical representations. Frequency-based explanations of this kind must therefore be elaborated with structural language representations.

Verb-noun phrases were tracked more closely than verb-adverb phrases. This effect of attachment type is consistent both with the theoretical distinction between complements and adjuncts and with their different processing characteristics. According to certain linguistic theories, complements and adjuncts are different in both syntax[38,39] and semantics[37]. A potential processing-based explanation of increased phrase-rate tracking of complement phrases is that complements immediately satisfy the selectional restriction of the verb and saturate its argument structure, and therefore reduce uncertainty in processing. Adjuncts, by contrast, only modify the verb without fulfilling a grammatical requirement. They therefore do not significantly reduce uncertainty. The resulting uncertainty around adjunct attachment might yield increased variance in a stream of V-Adv structures, reducing phase coherence and hence phrase-rate tracking. An alternative explanation is that the effect of attachment type simply reflects the (semantic and/or syntactic) difference between nouns and adverbs. While we cannot rule this out, we consider it unlikely given the extensive psycholinguistic literature that shows a difference in the processing of complements and adjuncts, even when their parts of speech are controlled[41–46]. To directly test this alternative explanation of the neural effect of attachment type, future studies could use prepositional constructions that differ in attachment type but are matched in parts of speech, like "waiting for the boat" (complement attachment) versus "waiting on the boat" (adjunct attachment).

Consistently head-initial phrases were tracked more closely than sequences with varying head position. The tracking difference between these conditions can be explained in terms of different positions of the grammatical head, which determines the syntactic and semantic status of the phrase and guides language processing[40,41,44–46]. We find the head-position effect intriguing because it shows that the brain accomplishes a seemingly dichotomous task, namely tracking *abstract* linguistic structure while preserving structural *detail*. That is, it is unlikely that the head only serves as a salient landmark that repeatedly evokes a neural response or re-aligns neural oscillations[1,10]. If this were so, we should not have obtained consistent 1-Hz phrasal tracking in the V-Adv (alt.) condition, given the temporal fluctuation in the position of the head on a scale of several hundred milliseconds (i.e., phrase-initial and phrase-final heads randomly lagged by 500 ms). Instead, the head position effect calls for a mechanism that resolves temporal inconsistency by inferring linguistic structure from the signal (see ref.[9]). A candidate mechanism of such kind is the time-based binding mechanism described in a recent computational model of neural oscillations[54]. This model infers structure from temporally unfolding signals to generate pulses, eventually rendering a consistent readout in the frequency domain. Under this framework, the effect of head position can be explained in terms of the temporal inconsistency associated with temporal binding: the 1-Hz peak in the V-Adv (alt.) condition was lower than the 1-Hz peak in the V-Adv condition, because in the former case there was more temporal jitter in the brain's generation of a periodic endogenous pulse (reflecting the composition of a verb phrase) in the absence of reliable physical cues. This proposal is compositional in the sense that higher-level units are hierarchically composed of low-level elements and decouples from constituent-level variance (e.g., head-position inconsistency). Such a compositional theory aligns with neurobiological theories of language processing that include a combinatorial operation that hierarchically combines smaller elements into larger elements[11,15,55–59]. Our findings so far suggest that the process underlying phrasal tracking is likely sensitive to order (i.e., the timepoint at which structure building is initiated differs depending on the position of the head of the phrase) as well as syntactic and semantic information (i.e., different types of linguistic structures elicit different degrees of tracking). These findings therefore enrich the neural tracking literature by showing how the tracking of phrases in connected speech is modulated by the specific properties of these phrases (cf. [3,5,7,8]).

Tracking sequential regularities

In parallel with structural tracking effects, we observed phrase-rate tracking in conditions that exhibited a sequential lexical regularity at 1 Hz, both in grammatical conditions (V–N, V-Adv) as well as in a condition without any grammatical structure (N-R). We interpret the 1-Hz effect in N-R sequences as indicating a sequential process external to grammatical processing. However, as we will discuss below, the sequential nature of the N-R effect does not reduce all grammatical tracking effects to sequential regularities.

One might observe that, besides a difference in grammaticality, N-R also differs from V–N and V-Adv in the amount of sequential regularity. That is, the two grammatical conditions regularly repeat both parts of speech, whereas the N-R condition only regularly repeats the noun. A better control condition would have been a non-grammatical condition in which both parts of speech are repeated, like in N-Adv sequences. However, the results of Burroughs et al.[3] suggest that this factor does not fully explain the difference between grammatical and non-grammatical conditions. They found stronger phrase-rate tracking in A-N sequences than in A-V sequences despite a match in sequential regularity, showing that this contrast does not reduce to a mere difference in the amount of sequential regularity.

Effects of sequential statistics are omnipresent in language processing. During sentence processing, word-level statistical information that is activated in tandem with higher-level structural information modulates neural activity[60,61]. Statistical information can also be exploited in the absence of grammatical structure, e.g., during speech segmentation[33]. Likewise, a recent frequency-tagging study[62] reported multi-word tracking effects for sequences without any grammatical structure, when participants were instructed to perform a task tuned to specific semantic categories of the individual words (i.e., "detect two-word chunks consisting of living or non-living

nouns"). Any pattern-recognition strategy operating over sequential, word-level information will yield a phrase-rate tracking readout if the presentation rate of the word-level features coincides with the frequency of phrases. In N-R sequences, participants heard a noun every 500 ms, hence this condition elicits a tracking effect at the 1-Hz phrase rate.

This sequential strategy[13] suffices to explain our findings in the N-R condition. However, it seems at odds with a recent study by Lo and colleagues, which showed no phrase-rate tracking in a 'reversed phrase' condition (e.g., 'fur dry skin rubs' vs. 'dry fur rubs skin') that, like our N-R condition, contained lexical regularities but no syntactic structure[8]. A potential reason behind this difference is that in the Chinese stimuli of Lo et al. the reversal of certain sentences affected the grammatical categories of the words. For instance, they used derivational nouns involving a non-noun constituent and a suffix (e.g., the noun "teach-er"), of which a syllabic reversal breaks the desired part-of-speech pattern (e.g., "er-teach" is not a noun). Moreover, their sentences contained a number of disyllabic Chinese compound verbs, such as in the A-N-V-V sentence "red bird fly_1 fly_2", which roughly means "the red bird flies". Here, "fly_1" and "fly_2" are two distinct verbal morphemes that together form a verbal compound. As such, these sentences contain no lexical repetition at 2 Hz, so the observed 2-Hz tracking effect likely stems from a combination of phrasal composition (the NP "red bird") and morphological composition (the verbal compound "fly_1 fly_2"). In the reversed version of this type of sentence (i.e., a N-A-V-V sequence "bird red fly_2 fly_1"), there is again no lexical repetition. Reversing the sentence also breaks the NP and the compound, which can explain why they did not observe a phrase-rate tracking effect for reversed sentences even though the brain tracks sequential lexical regularities.

While the sequential strategy is consistent with the N-R effect, it fails to explain the effects in other conditions. First of all, the 1-Hz effect in N-R had a posterior scalp distribution in evoked power, quite different from the central-anterior distribution of the effects in V–N and V-Adv (see Fig. 4B). Thus, although phrase-rate tracking effects converged in the frequency spectra, different processing strategies underlying tracking effects appear to be distinguishable in terms of spatial topography (see also ref.[62]). Second, a word-level tracking mechanism cannot explain the presence of phrase-rate tracking in V-Adv (alt.). Here, there is no consistent word-level regularity at the phrase rate at all, so the 1-Hz peak cannot come from the tracking of word-level features alone. And third, an exploratory analysis of the potential interaction between structural and sequential regularities showed that the magnitude of sequential tracking was boosted by the addition of structure (Fig. 5B), suggesting that structural and sequential effects are not identical in nature and that the presence of grammatical structure increases the neural sensitivity to sequential regularity. In all, this pattern suggests that structural and sequential information go hand in hand in the service of extracting structured meaning from spoken language.

A regularity-based hybrid account of phrase-rate tracking

In this section, we motivate and sketch a hybrid account of phrase-rate tracking, where regularities at multiple levels are required to explain neural patterns. This account receives support from a comparison between the observed pattern of phrase-rate tracking (see Fig. 4C) and the patterns predicted under different theoretical hypotheses (see Fig. 3). We consider a hybrid account most plausible, given a discrepancy between the current data and existing accounts that place their explanatory emphasis on a single (linguistic) factor. That is, the difference between V–N and V-Adv raises a challenge for prosodic accounts[53], because prosodic grouping was identical across conditions (see also the discussion in ref.[21]). Phrasal tracking in V-Adv (alt.) contradicts lexical accounts[13,14], since there was no reliable lexical repetition in this condition. And although the gradience within syntactic effects is compatible with a syntactic explanation[3,5,8], N-R does not contain grammatical structure at all, indicating that a purely syntactic account is also inadequate. To integrate our findings with the abovementioned accounts, we provide a rough sketch of a hybrid account of phrase-rate tracking, in which the brain tracks regularities stemming from (syntactic) structure and from sequential (lexical) information. Given that language processing is situated in a biological organ sensitive to multiple levels of information[11,15,63], neither structural nor sequential regularities are unique to the neural tracking of language. Instead, we suggest a rigorous dissection[9] of, for example, auditory[14], lexical[13], and task-modulated aspects[62] of this frequency-tagged readout. A dissociation between some of these factors has been found with neuro-computational language models in the temporal-spatial domain[19,63–65]. We believe that it will be helpful to extend this line of research into the frequency domain.

To summarize, this study shows that phrase-rate neural tracking is sensitive to regularities at multiple representational levels. As regularities can be computed over both the sequential pattern of lexical information and the hierarchical structure of phrasal units, our findings call for a neurobiological account of language processing wherein the brain leverages regularities computed over multiple levels of linguistic representation to guide rhythmic computation, and which seeks to integrate the contributions of structural and sequential information to language processing, and to behavior more generally.

Methods
Participants
Twenty-two native speakers of Mandarin Chinese participated in this study (15 females, mean age = 27.3, SD = 4.19) at the Max Planck Institute for Psycholinguistics (MPI). All subjects were reimbursed for their time (27 euros for 2.5 h). Written informed consent was obtained prior to the experiment. This study was approved by the Ethics Committee of the Faculty of Social Sciences at Radboud University Nijmegen. The entire procedure was performed in accordance with relevant guidelines and regulations.

Stimuli and design
Participants listened to streams of monosyllabic Mandarin Chinese words that were isochronously presented at 2 Hz (see Fig. 2A; the full stimulus list with English translations can be found at https://osf.io/fw7bc/). There

were six conditions in total (see Table 1 for examples), consisting of repetitions of: verb-noun phrases (V–N), verb-adverb phrases (V-Adv), verb-adverb phrases with varying order (V-Adv (alt.)), combinations of a noun and a pseudorandom word (N-R), pseudorandom words (R-R), and verb-verb combinations (V-V). In the N-R and R-R conditions, R words were chosen pseudorandomly rather than randomly because we had to make sure that no grammatical combinations were present. For instance, the R word in the N-R condition could never be a verb, as that could generate a subject-verb sequence. In the first three conditions (i.e., V–N, V-Adv, and V-Adv (alt.)), the combination of two adjacent words yields two-word phrases, which occur at 1 Hz. These are the grammatical conditions. In the latter three conditions (i.e., N-R, R-R, and V-V), no grammatical combinations could be formed by combining adjacent words, which makes them non-grammatical conditions. Furthermore, the conditions V–N, V-Adv and N-R contain a sequential regularity at 1 Hz, because one or more of their parts of speech are repeated every 1 s. The conditions V-Adv (alt.), R-R, and V-V contain no sequential regularities at 1 Hz.

Monosyllabic words were first synthesized using Google Text-to-Speech (Mainland Standard Chinese, WaveNet-C, male voice, speech rate = 0.75). Their duration was adjusted to exactly 500 ms by either padding zeros to the edges or truncating exceeding signals. After length normalization, a sinewave ramping window was applied to the first and last 10% of each signal to make the speech envelope more natural. Then, 48 words were concatenated into streams of 24 s. Figure 2B shows the power spectrum of the auditory streams of each condition, which was computed by first applying a Hilbert transform to the half-wave rectified speech signal to extract the temporal envelopes, and then applying a discrete Fourier transform to the down-sampled (200 Hz) envelope of the stimuli. The power spectrum of all conditions contains a clear peak at 2 Hz only, showing that the acoustic signals only contain information at the word rate and that the conditions cannot be distinguished acoustically.

For each of the six conditions, we created 288 unique two-word combinations. These two-word combinations formed syntactic phrases in the grammatical conditions and non-syntactic 'phrases' (two-word combinations) in the non-grammatical conditions. Each phrase was repeated twice in different halves of the experiment, so there were 576 phrases per condition in total. These were divided over 24 streams per condition, each of which contained 24 phrases (i.e., 48 monosyllabic words per condition). These streams were then distributed over 24 blocks, with each block containing one stream from each of the six conditions. The order of conditions within each block was pseudo-randomized, and the order of the blocks across participants was counterbalanced. Participants were allowed to take short breaks after each block. The interval between trials ranged randomly from 800 to 1100 ms. The structure of each trial is illustrated in Fig. 2C.

We note that the auditory presentation, combined with the one-to-many mapping between syllables and morphemes in Chinese, could pose a challenge to the listener in the recognition of individual words in syllable streams. However, the frequency-tagging design mitigates this issue. That is, while the first syllable in a two-word combination might be ambiguous, the auditory input is rendered unambiguous upon the presentation of the second syllable, meaning that unambiguous two-word phrases could be constructed. This is the case for all grammatical two-word combinations, and so ambiguity about the first word is unlikely to affect neural processes happening at the phrasal timescale.

We also note that the conditions V-Adv and V-Adv (alt.) differ in the semantic relation between the verb and the adverb. Indeed, the V-Adv order in Chinese poses stronger restrictions on the semantic relation between the two constituents than the Adv-V order does[66,67]. Hence, the intermixing of V-Adv and Adv-V phrases in the V-Adv (alt.) condition introduces not only syntactic variability (in head position) but also semantic variability. We believe that this is an interesting avenue for future research, which should attempt to determine to what extent phrase-rate tracking is modulated by semantic relations at the phrasal level.

Procedure

Participants listened to speech streams played through loudspeakers in a sound-proof and electromagnetically shielded room. They were instructed to sit still, blink normally and pay attention to the audio. After audio playback, they had to indicate by button press whether they thought the stream was "easy to understand". Participants were implicitly informed through feedback in a practice block that grammatical sequences are easier to understand than non-grammatical sequences. This behavioral task served to keep participants' attention focused on the stimuli in a natural way without invoking artificial and task-specific grouping strategies. In contrast to previous studies[3,4], we chose not to use a phrase- or syllable-detection/monitoring task after observing in pilot runs that such tasks make participants inclined to strategically group every two or four syllables, even in non-grammatical conditions. On average, participants answered 97.9% of the questions correctly (SD = 2.17%, range of average accuracy across conditions = 87.5—100%).

EEG recording

EEG signals were recorded using an MPI custom ActiCAP 64-electrode montage (Brain Products, Munich, Germany), of which 59 electrodes were mounted in the electrode cap. Horizontal eye movements were recorded by two electrodes placed on the outer canthi of the left and right eyes. Eye blinks were recorded by an electrode placed below the left eye. One electrode was placed on the right mastoid (RM), the reference electrode was placed on the left mastoid (LM) and the ground electrode was placed on the forehead. The EEG signal was amplified through BrainAmp DC amplifiers, referenced online to LM, sampled at 500 Hz and filtered with a passband of 0.016–249 Hz. The impedance of each electrode was kept below 25 kΩ by applying electrolyte gel prior to data recording.

EEG preprocessing

EEG signals were preprocessed and analyzed with the Fieldtrip toolbox[68] in the Matlab environment (Mathworks Inc., version 2021b). The EEG data were filtered with a 0.3–25 Hz bandpass filter and re-referenced to the

average of the left and right mastoids (LM/RM). Channels that were malfunctioning or showed excessive drifts across the entire experiment were removed and then interpolated using the weighted average of their neighboring channels. Then, the data were epoched into single trials, from the onset of the third phrase (t = 2s; to avoid transient auditory evoked responses at stimulus onset) to the end of the sequence (t = 24s). Whole trials were rejected only if there was a consistent electrode-level artifact, such as excessive muscle movements or jumps. We used independent component analysis to regress out artifacts resulting from eye blinks and eye movements. Data from two participants were rejected during preprocessing. The data of one of them was very noisy due to excessive head movements (over 60% of trials rejected). The other participant did not show reliable tracking at the syllable rate, which is at odds with robust findings that acoustic input reliably evokes spectral peaks at the corresponding frequency of presentation[3–5,8,62]. We excluded this participant because we consider syllable tracking to be a prerequisite for phrase tracking.

Spectral analysis

The preprocessed data were converted into the frequency domain by applying a discrete Fourier transformation with a frequency resolution of 1/22 Hz (after removing the first two phrases, the duration of each trial was 22 s). We computed evoked power and ITPC between 0.3 and 5 Hz. Evoked power was computed for each condition by first averaging over all trials in each condition and then applying a Fourier transform. ITPC at frequency *f* was computed as the mean of the squared sum of Fourier coefficients at *f*, which can be calculated with the formula:

$$R(f) = \frac{1}{k}\left[\left(\sum_k \cos\theta_k\right)^2 + \left(\sum_k \sin\theta_k\right)^2\right]$$

where *k* denotes the number of trials for each condition and θ_k is the phase angle of the complex-valued Fourier coefficient.

Statistical analysis

We ran statistical tests on both power and ITPC at 1 and 2 Hz. For power, we tested if the amplitude at the target frequency was higher than the average of three neighboring frequency bins to its left. By only including left-neighboring bins, this analysis takes into account the 1/*f* trend in power spectra. For ITPC, we tested if the ITPC at the target frequency was higher than the average of three neighboring frequency bins on each side.

Since power and ITPC in different frequency bins were not normally distributed, we ran non-parametric random permutation tests for peak detection. In each permutation, group labels (e.g., "target", "neighbor") were randomly assigned to observations from a merged sample containing all groups, resulting in permuted groups. Then, an arbitrary test statistic (here, we used the difference in mean between groups, denoted as Δ) was calculated for the permutation groups. Via 10,000 permutations, we created a sampling distribution of the values of the test statistic and calculated the probability of observing the actual experimental value under this sampling distribution.

After detecting peaks in each individual condition, we performed planned pairwise comparisons (paired samples t-tests) on ITPC between all conditions that showed a 1- or 2-Hz spectral peak to evaluate how structural and/or sequential regularities contribute to phrase-rate tracking. A false discovery rate (FDR) correction was performed after all pairwise comparisons to correct for multiple comparisons[69].

Data availability

The full stimulus list, the raw EEG data, and the analysis scripts of this study are available at https://osf.io/fw7bc/.

References

1. Giraud, A.-L. & Poeppel, D. Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* **15**, 511–517 (2012).
2. Peelle, J. & Davis, M. Neural oscillations carry speech rhythm through to comprehension. *Front. Psychol.* https://doi.org/10.3389/fpsyg.2012.00320 (2012).
3. Burroughs, A., Kazanina, N. & Houghton, C. Grammatical category and the neural processing of phrases. *Sci. Rep.* **11**, 2446 (2021).
4. Ding, N. *et al.* Characterizing neural entrainment to hierarchical linguistic units using electroencephalography (EEG). *Front. Hum. Neurosci.* https://doi.org/10.3389/fnhum.2017.00481 (2017).
5. 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).
6. Coopmans, C. W., de Hoop, H., Hagoort, P. & Martin, A. E. Effects of structure and meaning on cortical tracking of linguistic units in naturalistic speech. *Neurobiol. Lang.* https://doi.org/10.1162/nol_a_00070 (2022).
7. Kaufeld, G. *et al.* Linguistic structure and meaning organize neural oscillations into a content-specific hierarchy. *J. Neurosci.* **40**, 9467–9475 (2020).
8. Lo, C.-W., Tung, T.-Y., Ke, A. H. & Brennan, J. R. Hierarchy, not lexical regularity, modulates low-frequency neural synchrony during language comprehension. *Neurobiol. Lang.* **3**, 538–555 (2022).
9. Oever, S. T., Kaushik, K. & Martin, A. E. Inferring the nature of linguistic computations in the brain. *PLOS Comput. Biol.* **18**, e1010269 (2022).
10. 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).
11. Martin, A. E. A compositional neural architecture for language. *J. Cogn. Neurosci.* **32**, 1407–1427 (2020).

12. 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 (2020).
13. Frank, S. L. & Yang, J. Lexical representation explains cortical entrainment during speech comprehension. *PLOS One* **13**, e0197304 (2018).
14. Kalenkovich, E., Shestakova, A. & Kazanina, N. Frequency tagging of syntactic structure or lexical properties; A registered MEG study. *Cortex* https://doi.org/10.1016/j.cortex.2021.09.012 (2021).
15. Martin, A. E. Language processing as cue integration: grounding the psychology of language in perception and neurophysiology. *Front. Psychol.* https://doi.org/10.3389/fpsyg.2016.00120 (2016).
16. Slaats, S. & Martin, A. E. What's surprising about surprisal. Preprint at https://doi.org/10.31234/osf.io/7pvau (2023).
17. Takahashi, E. & Lidz, J. Beyond statistical learning in syntax. in *Language acquisition and development: Proceedings of GALA* 446–456 (Cambridge Scholars Publishing).
18. Thompson, S. P. & Newport, E. L. Statistical learning of syntax: The role of transitional probability. *Lang. Learn. Dev.* **3**, 1–42 (2007).
19. Brennan, J. R., Dyer, C., Kuncoro, A. & Hale, J. T. Localizing syntactic predictions using recurrent neural network grammars. *Neuropsychologia* **146**, 107479 (2020).
20. Brennan, J. R. & Hale, J. T. Hierarchical structure guides rapid linguistic predictions during naturalistic listening. *PLOS One* **14**, e0207741 (2019).
21. Coopmans, C. W. & Martin, A. E. Prosody vs. syntax, or prosody and syntax? Evaluating accounts of delta-band tracking. In *Rhythms of Speech and Language: Culture, Cognition, and the Brain* (eds Meyer, L. & Strauss, A.) (Cambridge University Press, in press).
22. Coopmans, C. W., de Hoop, H., Kaushik, K., Hagoort, P. & Martin, A. E. Hierarchy in language interpretation: Evidence from behavioural experiments and computational modelling. *Lang. Cogn. Neurosci.* **37**, 420–439 (2022).
23. Chomsky, N. *Aspects of the Theory of Syntax* (MIT Press, 1965).
24. Everaert, M. B. H., Huybregts, M. A. C., Chomsky, N., Berwick, R. C. & Bolhuis, J. J. Structures, not strings: Linguistics as part of the cognitive sciences. *Trends Cogn. Sci.* **19**, 729–743 (2015).
25. Sprouse, J. & Hornstein, N. *Experimental Syntax and Island Effects* (Cambridge University Press, 2013).
26. Martin, A. E., Nieuwland, M. S. & Carreiras, M. Agreement attraction during comprehension of grammatical sentences: ERP evidence from ellipsis. *Brain Lang.* **135**, 42–51 (2014).
27. Blanco-Elorrieta, E., Ding, N., Pylkkänen, L. & Poeppel, D. Understanding requires tracking: Noise and knowledge interact in bilingual comprehension. *J. Cogn. Neurosci.* **32**, 1975–1983 (2020).
28. Sheng, J. *et al.* The cortical maps of hierarchical linguistic structures during speech perception. *Cereb. Cortex* **29**, 3232–3240 (2019).
29. Ding, N., Melloni, L., Tian, X. & Poeppel, D. Rule-based and word-level statistics-based processing of language: Insights from neuroscience. *Lang. Cogn. Neurosci.* **32**, 570–575 (2017).
30. ten Oever, S., Carta, S., Kaufeld, G. & Martin, A. E. Neural tracking of phrases in spoken language comprehension is automatic and task-dependent. *eLife* **11**, e77468 (2022).
31. Keitel, A., Gross, J. & Kayser, C. Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. *PLOS Biol.* **16**, e2004473 (2018).
32. Poeppel, D. & Embick, D. Defining the Relation Between Linguistics and Neuroscience. In *Twenty-first century psycholinguistics: Four cornerstones* (eds Poeppel, D. & Embick, D.) 103–118 (Lawrence Erlbaum Associates Publishers, 2005).
33. Saffran, J. R., Aslin, R. N. & Newport, E. L. Statistical learning by 8-month-old infants. *Science* **274**, 1926–1928 (1996).
34. Bai, F. *Neural representation of speech segmentation and syntactic structure discrimination*. PhD Dissertation, Radboud University Nijmegen (2022).
35. Getz, H., Ding, N., Newport, E. L. & Poeppel, D. Cortical tracking of constituent structure in language acquisition. *Cognition* **181**, 135–140 (2018).
36. Henin, S. *et al.* Learning hierarchical sequence representations across human cortex and hippocampus. *Sci. Adv.* **7**, eabc4530 (2021).
37. Heim, I. & Kratzer, A. *Semantics in Generative Grammar* (Wiley-Blackwell, 1998).
38. Jackendoff, R. S. *X Syntax: A Study of Phrase Structure* (MIT Press, 1977).
39. Carnie, A. *Syntax: A Generative Introduction* (John Wiley & Sons, 2021).
40. Abney, S. P. A computational model of human parsing. *J. Psycholinguist. Res.* **18**, 129–144 (1989).
41. Boland, J. E. & Blodgett, A. Argument status and PP-attachment. *J. Psycholinguist. Res.* **35**, 385–403 (2006).
42. Frazier, L. Sentence processing: A tutorial review. In *Attention and performance 12: The psychology of reading* (ed. Frazier, L.) 559–586 (Lawrence Erlbaum Associates Inc, 1987).
43. Frazier, L. & Clifton, C. Construal: Overview, motivation, and some new evidence. *J. Psycholinguist. Res.* **26**, 277–295 (1997).
44. Schütze, C. T. & Gibson, E. Argumenthood and english prepositional phrase attachment. *J. Mem. Lang.* **40**, 409–431 (1999).
45. Pritchett, B. L. *Grammatical Competence and Parsing Performance* (University of Chicago Press, 1992).
46. Coopmans, C. W. & Schoenmakers, G.-J. Incremental structure building of preverbal PPs in Dutch. *Linguist. Neth.* **37**, 38–52 (2020).
47. Bartek, B., Lewis, R. L., Vasishth, S. & Smith, M. R. In search of on-line locality effects in sentence comprehension. *J. Exp. Psychol. Learn. Mem. Cogn.* **37**, 1178–1198 (2011).
48. Gibson, E. Linguistic complexity: Locality of syntactic dependencies. *Cognition* **68**, 1–76 (1998).
49. Husain, S., Vasishth, S. & Srinivasan, N. Strong expectations cancel locality effects: Evidence from Hindi. *PLOS One* **9**, e100986 (2014).
50. Konieczny, L. Locality and parsing complexity. *J. Psycholinguist. Res.* **29**, 627–645 (2000).
51. Ding, N. & Simon, J. Z. Power and phase properties of oscillatory neural responses in the presence of background activity. *J. Comput. Neurosci.* **34**, 337–343 (2013).
52. Lenth, R., Singmann, H., Love, J., Buerkner, P. & Herve, M. emmeans: Estimated Marginal Means, aka Least-Squares Means. (2020).
53. Glushko, A., Poeppel, D. & Steinhauer, K. Overt and implicit prosody contribute to neurophysiological responses previously attributed to grammatical processing. *Sci. Rep.* **12**, 14759 (2022).
54. Martin, A. E. & Doumas, L. A. A. A mechanism for the cortical computation of hierarchical linguistic structure. *PLOS Biol.* **15**, e2000663 (2017).
55. Hagoort, P. On Broca, brain, and binding: A new framework. *Trends Cogn. Sci.* **9**, 416–423 (2005).
56. Hagoort, P. The core and beyond in the language-ready brain. *Neurosci. Biobehav. Rev.* **81**, 194–204 (2017).
57. Friederici, A. D. The brain basis of language processing: from structure to function. *Physiol. Rev.* **91**, 1357–1392 (2011).
58. Matchin, W. & Hickok, G. The cortical organization of syntax. *Cereb. Cortex* **30**, 1481–1498 (2020).
59. Pylkkänen, L. Neural basis of basic composition: What we have learned from the red–boat studies and their extensions. *Philos. Trans. R. Soc. B* **375**, 20190299 (2020).
60. Slaats, S., Weissbart, H., Schoffelen, J.-M., Meyer, A. S. & Martin, A. E. Delta-band neural responses to individual words are modulated by sentence processing. *J. Neurosci.* **43**, 4867–4883 (2023).
61. Heilbron, M., Armeni, K., Schoffelen, J.-M., Hagoort, P. & de Lange, F. P. A hierarchy of linguistic predictions during natural language comprehension. *Proc. Natl. Acad. Sci.* **119**, e2201968119 (2022).
62. Lu, Y., Jin, P., Ding, N. & Tian, X. Delta-band neural tracking primarily reflects rule-based chunking instead of semantic relatedness between words. *Cereb. Cortex* **33**, 4448–4458 (2023).

63. Stanojević, M., Brennan, J. R., Dunagan, D., Steedman, M. & Hale, J. T. Modeling structure-building in the brain with CCG parsing and large language models. *Cogn. Sci.* **47**, e13312 (2023).
64. Coopmans, C. W., de Hoop, H., Tezcan, F., Hagoort, P. & Martin, A. E. Neural dynamics express syntax in the time domain during natural story listening. *bioRxiv* https://doi.org/10.1101/2024.03.19.585683 (2024).
65. Nelson, M. J. *et al.* Neurophysiological dynamics of phrase-structure building during sentence processing. *Proc. Natl. Acad. Sci.* **114**, E3669–E3678 (2017).
66. Huang, CT James, YH Audrey Li, and Andrew Simpson, eds. The handbook of Chinese linguistics. John Wiley & Sons (2018).
67. Guiyu Wang (汪贵玉). Investigating "AV" Disyllabic Verb Phrases in Chinese ("AV"式双音动词研究), Master Thesis, Huazhong University of Science and Technology (2020).
68. Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J.-M. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* **2011**, e156869 (2010).
69. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).

Acknowledgements

Author contributions

J.Z., A.E.M., and C.W.C. conceptualized the work. J.Z. and C.W.C. designed the experiment. J.Z. carried out the experiment. J.Z. and C.W.C. performed data analysis, prepared the figures, and wrote the first draft of the manuscript. J.Z., A.E.M., and C.W.C. revised the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to C.W.C.

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.