scientific reports

OPEN

The child the apple eats: processing of argument structure in Mandarin verb-final sentences

Max Wolpert[1,2,3✉], Jiarui Ao[2], Hui Zhang[4], Shari Baum[3,5] & Karsten Steinhauer[3,5]

Mandarin Chinese is typologically unusual among the world's languages in having flexible word order despite a near absence of inflectional morphology. These features of Mandarin challenge conventional linguistic notions such as subject and object and the divide between syntax and semantics. In the present study, we tested monolingual processing of argument structure in Mandarin verb-final sentences, where word order alone is not a reliable cue. We collected participants' responses to a forced agent-assignment task while measuring their electroencephalography data to capture real-time processing throughout each sentence. We found that sentence interpretation was not informed by word order in the absence of other cues, and while the coverbs BA and BEI were strong signals for agent selection, comprehension was a result of multiple cues. These results challenge previous reports of a linear ranking of cue strength. Event-related potentials showed that BA and BEI impacted participants' processing even before the verb was read and that role reversal anomalies elicited an N400 effect without a subsequent semantic P600. This study demonstrates that Mandarin sentence comprehension requires online interaction among cues in a language-specific manner, consistent with models that predict crosslinguistic differences in core sentence processing mechanisms.

Languages differ in how they express meaning, but these differences are constrained by the biology of the human brain. Thus, many accounts of processing propose that every language is processed in fundamentally the same ways. For instance, all languages appear to communicate information at a similar rate[1] and have a similar distance between hierarchically dependent linguistic units[2]. Likewise, all languages utilize the same brain networks for comprehension and production[3], with the same amount of time required to access and retrieve word meaning[4–6]. Language users also rely on common processing heuristics to build sentence structure in real time[7,8], including assumptions about which sentence elements are probable agents[9,10]. It follows, then, that behavioral and neural metrics of sentence processing should yield similar results no matter the language.

Nonetheless, models of language processing must account for the full extent of crosslinguistic variability, and many so-called universals do not hold for all languages[11]. For example, processing of verb-final sentences relies on distinct profiles of working memory and predictive parsing[12,13], and comprehension of these structures varies across languages[14]. Additionally, different linguistic features lead to comprehenders prioritizing different units of information[15], such as languages with flexible word order enlisting different processing resources from languages with fixed order[16]. Suffice it to say, studying diverse languages is part of appreciating the extent of variation in sentence processing.

In the present study, we considered verb-final sentences in Mandarin, a language unusual for having flexible word order—a feature normally associated with highly inflected languages—despite virtually no morphological inflection[17,18], making Mandarin a "typologically hybrid" language[19]. For example, the two sentences 孩子苹果吃掉了 "child apple ate" and 苹果孩子吃掉了 "apple child ate" both mean that the child ate the apple, despite the word order being reversed. Even basic terms like "subject" and "object" are not always appropriate for describing

[1]Key Laboratory for Biomedical Engineering of Ministry of Education, College of Biomedical Engineering and Instrument Sciences, Zhejiang University, Hangzhou 310027, China. [2]Integrated Program in Neuroscience, McGill University, Room 302 Irving Ludmer Building, 1033 Pine Avenue West, Montreal, QC H3A 1A1, Canada. [3]Centre for Research on Brain, Language and Music, 2001 Av. McGill College #6, Montreal, QC H3A 1G1, Canada. [4]School of Foreign Languages and Cultures, Nanjing Normal University, Suiyuan Campus, Building 500, Nanjing 210024, Jiangsu, China. [5]School of Communication Sciences and Disorders, McGill University, 2001 Av. McGill College #8, Montreal, QC H3A 1G1, Canada. ✉email: max.wolpert@zju.edu.cn

Mandarin sentences[20,21]. These grammatical properties make it challenging to create an "unambiguously ungrammatical" Mandarin sentence[22], leading some to describe Mandarin as a "semantics-based" as opposed to a "syntax-based" language[23]. These features make Mandarin an ideal language for testing assumptions about meaning and structure in sentence processing models.

The divide between syntax and semantics has long been a focus of sentence processing research[4,24,25]. At the syntax-semantics interface is argument structure, or the grammatical roles that verbs assign to sentence elements[26,27]. Although many categories of roles have been described, there is consensus that the human mind distinguishes between the doer, or agent, and the receiver, or patient, of a verb's action[28–30]. We use the terms agent and patient to represent the concept of proto-agent and proto-patient, two roles argued to be psychologically real across languages[28]. Languages differ in how agent and patient roles are expressed, and grammatical information is often redundant with semantic knowledge of stereotypical argument roles[31]. It is thus productive to pit semantics against syntax to characterize how they are prioritized in different languages for argument structure interpretation.

Mandarin has another feature important for argument structure: the coverbs BA and BEI. These coverbs can occur in verb-final sentences to disambiguate agent and patient roles. BA assigns agent status to its preceding noun, as in 孩子把苹果吃掉了 "child BA apple ate", resulting in subject-object-verb word order. BEI assigns patient status to its preceding noun, as in 苹果被孩子吃掉了 "apple BEI child ate", resulting in object-subject-verb word order. These word orders are also possible without the coverbs BA and BEI, as already indicated, but may be pragmatically restricted[32]. There has been debate about the syntactic categories of BA and BEI[33,34], but for simplicity we refer to them as coverbs[35]. Although BA and BEI are both common in verb-final clauses, they have differences in structure and usage. BA must be followed by a noun phrase and is limited in which verbs it can be used with, and BEI is typically analyzed as a passive construction and can be followed directly by a verb[36,37]. Despite these differences, BA and BEI are powerful cues for argument structure assignment and each assigns a different interpretation in verb-final sentences.

In the present study of Mandarin sentence processing, we used electroencephalography (EEG) and behavioral measures to systematically compare the impact of syntactic (coverbs and word order) and semantic information (animacy and event knowledge). Below, we first introduce an impactful model of crosslinguistic sentence processing, the Competition Model, and then turn to the specific case of role reversals and existing accounts that consider Mandarin data. We then summarize the critical elements of our experimental design and relevant predictions.

Competition Model

For experimental comparison of crosslinguistic differences, an impactful framework is the Competition Model[38], a key reference for language processing researchers for the last 40 years. As its name suggests, the Competition Model describes processing as a competitive arena where fundamental units of information called "cues" compete to shape decisions in comprehension and production[39]. Much of the research in the framework of the Competition Model has targeted argument structure processing, employing a forced-choice task for agent selection with orthogonal cue comparison across languages to assess the relative strength and validity of specific cues in offline judgments[38]. Agent selection differs among languages depending on how often certain cues are present (cue availability) and how often these cues correctly indicate the agent (cue reliability); in the case of competition among cues, the primary cue driving argument structure assignment is said to have greater cue strength[40]. Mandarin speakers have been shown to rely on the following cues in order of decreasing strength: BEI, animacy, word order, and BA, although BA may be as strong as word order in verb-final sentences[32].

Role Reversals and EEG Experiments

Sentences where the stereotypical arguments of agent and patient are swapped without grammatical violations are known as role reversals, such as "the child the apple ate." Over the past twenty years, researchers have studied role reversals with event-related potentials (ERPs), with specific focus on the N400 and P600 components. Traditionally, N400s were linked to semantic processing[4] and P600s to syntactic processing[25,41]. However, this functional distinction has become more nuanced, especially when implausible role reversals were shown to elicit "semantic P600 effects" without modulation of the N400[42–44].

Different explanations for semantic P600 effects have been proposed, including parallel syntactic and semantic processing streams[45–47] and classifying the N400 and P600 as indexing retrieval and integration instead of semantics and syntax[48,49]. However, few explanations have considered the potential for different languages to require different processing profiles for sentence comprehension. Given its unusual linguistic properties, Mandarin is a fruitful target for testing crosslinguistic validity of sentence processing models.

We are aware of two processing models that have explicitly considered Mandarin to study argument structure in role reversals. The first is the extended Argument Dependency Model (eADM), which, like the Competition Model, sets crosslinguistic diversity at the center of its model architecture[16,50,51]. According to the eADM, processing of sentence elements is divided in separate streams for nouns and verbs, and nouns are assigned a proto-agent or proto-patient role as argument structure is built iteratively. When parsers encounter difficulty in initial stages of relational and plausibility processing, an N400 effect is elicited; P600 effects, however, are limited to subsequent well-formedness and repair computations. Crucially, the computations for processing nouns and verbs depend on language-specific patterns of cue weighting, and the model predicts that role reversals only elicit semantic P600 effects without N400 modulation in languages that have rigid word order[16,52]. Conversely, role reversals for a language with flexible word order like Mandarin will elicit N400 effects with or without P600 modulation[16]. These predictions arise from revisions of earlier models to better account for conflict between syntactic and thematic processing streams[52,53].

Reversibility	Agent Animacy	Example sentence
Reversible	Animate	老板技工举报了。 *boss technician denounced*
	Inanimate	温度气压预测了。 *temperature pressure predicted*
Irreversible	Animate	仆人镜子擦亮了。 *servant mirror polished*
	Inanimate	鸟笼喜鹊困住了。 *birdcage magpie trapped*

Table 1. Examples of sentence stimuli with semantic manipulations. All reversible sentences had two nouns with shared animacy status, while all irreversible sentences had two nouns with opposing animacy status. For each of the four conditions in the table, we used a unique set of verbs and noun pairs. All nouns and verbs were two syllables and all sentences ended with the aspect particle LE.

The second model that has considered Mandarin is the Bag of Arguments account, which encompasses multiple studies over the past decade[54–57] that together inform an overarching processing model of argument structure[58]. The Bag of Arguments model centers around the N400 as an index of prediction during real-time sentence processing, and role reversals are expected to modulate N400 amplitude only if (1) the verb and its arguments are combinable and highly predictable[54], (2) arguments are in the same clause as the verb[55,57], and (3) parsers have more than 800 ms to utilize structural role information to predict the verb[56]. If these conditions are met, then role reversals should elicit N400 effects, with or without modulation of the P600[56,58]. The eADM and Bag of Arguments accounts stemmed from different motivations, with the eADM focused on crosslinguistic differences in argument structure and plausibility computations, while the Bag of Arguments model is concerned with the timing of verb prediction based on preceding arguments. The eADM also makes predictions for ERPs prior to the verb[59,60]. Accordingly, the models are not incompatible, and both can inform interpretation of the present study.

Present Study

The present study had two primary aims. First, we sought to improve characterization of cue weighting in Mandarin using an experiment with a balanced design and monolingual Mandarin-speaking participants. While prior descriptions of the pattern of Mandarin cue weighting have been impactful[23,32,61], there has been a recent call to use updated methodological approaches to evaluate these findings' replicability[62]. To appreciate the iterative nature of argument structure assignment, we also compared ERPs at pre-verb sentence positions. Second, we used ERPs to test processing of role reversals, which have been widely reported to elicit semantic P600 effects[48]. While most accounts assume that role reversals are processed similarly across languages[48,54], there is evidence that different language process these structures differently[16]. Prior experiments have tested Mandarin role reversals, but there are conflicting reports of N400s[16], semantic P600s[54,56], or no ERP modulation[58]. These studies have not reported controlling for participants' bilingual language knowledge, despite findings showing sentence comprehension can be impacted by second language knowledge[23,61].

Accordingly, we designed stimuli by systematically manipulating semantic and syntactic cues: Reversibility, Agent Animacy, Order, and Structure. For Reversibility, there were reversible sentences where either noun was equally plausible as agent and irreversible sentences where only one noun was a plausible agent. To maximize the plausibility difference between reversible and irreversible sentences, we manipulated Agent Animacy such

Order	Structure	Example sentence	Plausibility
Plausible agent in first position	NNV	仆人镜子擦亮了。 *servant mirror polished*	Plausible
	BA	仆人把镜子擦亮了。 *servant BA mirror polished*	Plausible
	BEI	仆人被镜子擦亮了。 *servant BEI mirror polished*	Implausible role reversal
Plausible agent in second position	NNV	镜子仆人擦亮了。 *mirror servant polished*	Plausible
	BA	镜子把仆人擦亮了。 *mirror BA servant polished*	Implausible role reversal
	BEI	镜子被仆人擦亮了。 *mirror BEI servant polished*	Plausible

Table 2. Examples of sentence stimuli with syntactic manipulations. Each sentence created according to the semantic cues of Reversibility and Agent Animacy was paired with the cues of Order and Structure to display to participants in the experiment. Due to Mandarin's flexible word order in NNV structures without a coverb, either Order of the plausible noun should result in the plausible interpretation. The example given is for a sentence with a plausible animate agent, but the same procedure was applied for all sentence materials. Note that Order and Structure were manipulated for reversible sentences as well, but in those cases there was no semantic cue for informing agent selection, and as such all arrangements were potentially plausible.

that irreversible sentences had contrasting animacy between the two nouns and reversible sentences had two nouns with shared animacy status. Unlike many prior studies[23,32,61], we included plausible inanimate agents, thus dissociating animacy and plausibility. The stimuli designed with the above semantic restraints were then crossed with Order, whether a given plausible agent was in first or second position in the sentence, and Structure, whether a sentence used BA, BEI, or noun-noun–verb (NNV) structure without a coverb. The experimental conditions are summarized in Table 1 for the semantic variables and Table 2 for the syntactic variables.

Predictions for behavioral and ERP results were informed by three sentence processing models considering Mandarin data, as previously summarized. In contrast to previous Competition Model reports of a linear ranking of cue strength[23,32,61], we expected that by dissociating animacy from plausibility, animacy would not be the strongest cue. For role reversal sentences, we expected that if Mandarin plausibility comprehension comprises language-specific mechanisms, then there would be an N400 response, with or without P600 modulation, as predicted by the eADM[16]. Alternatively, if Mandarin comprehension relies on mechanisms identical to other languages, then there should be a semantic P600 response without modulation of the N400, as found in prior studies[42–44,54,56]. Crucially, we note that each of the three models has different aims and proposed mechanisms underlying their proposed processing architecture, and thus the corresponding predictions in the context of the current study (overviewed in Table 3) are not strict tests of the validity of any single model over another. In fact, the models are sufficiently distinct in their targeted explanations that direct comparison may be inappropriate. Instead, the present study can give support, nuance, or reconsideration for specific aspects of each model.

Note that towards our first aim of characterizing cue weighting, we further considered ERPs prior to the verb at sentence-initial noun and coverbs. Previous studies have reported greater N400 components for inanimate nouns in comparison to animate nouns at sentence-initial position[59,60]. According to the eADM, this is attributed to a preference for sentence-initial arguments to be agents (or subjects)[10], and thus non-ideal inanimate agents elicit greater N400 amplitude than their animate counterparts[59,60]. Per this explanation, we should observe an N400 effect for sentence-initial inanimate nouns. These predictions are summarized in Table 3.

Results
Behavioral Results
Reversible Sentences
Participants' interpretation of reversible sentences showed two key effects related to our predictions. First, BA and BEI were the strongest cues for agent assignment (BA: β = 16.96, SE = 5.83, Z = 8.23, p < 0.001; BEI: β = 0.06, SE = 0.02, Z = − 7.73, p < 0.001), which confirmed previous behavioral results for verb-final sentences[23,32]. In contrast to prior reports, however, our participants had no inherent word order preference for agent selection in NNV sentences (β = 1.25, SE = 0.24, Z = 1.20, p = 0.232). Alongside these simple effects, BA and BEI interacted with Agent Animacy, with the coverb cue being slightly weaker when indicating an inanimate agent (BA: β = 1.24, SE = 0.12, Z = 2.18, p = 0.03; BEI: β = 0.65, SE = 0.06, Z = − 4.66, p < 0.001). Model results are visualized in Fig. 1a.
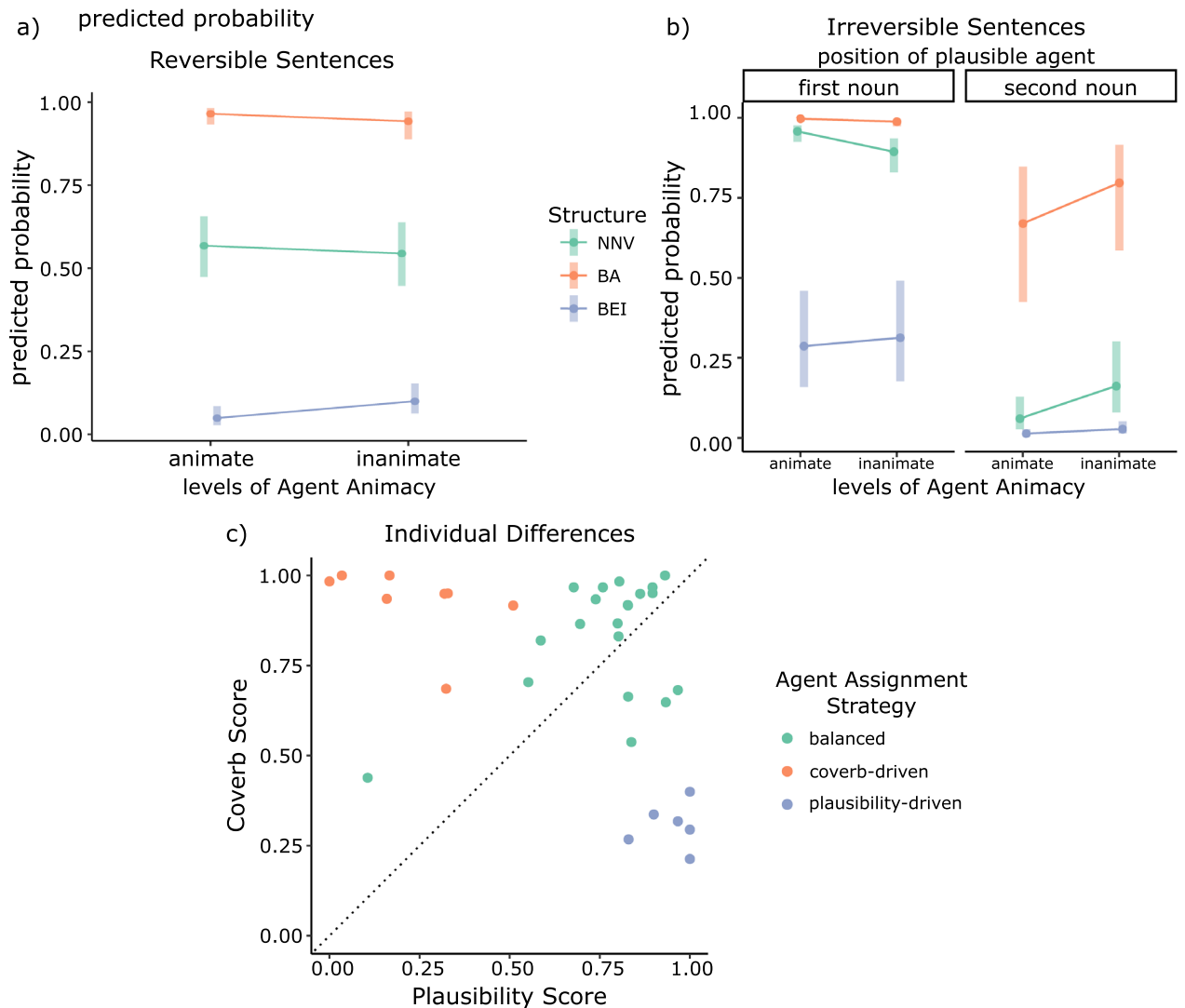
Irreversible Sentences
Just as in reversible sentences, BA and BEI were the strongest cues for agent assignment in irreversible sentences (BA: β = 18.21, SE = 7.00, Z = 7.55, p < 0.001; BEI: β = 0.07, SE = 0.03, Z = − 6.21, p < 0.001). Unlike reversible sentences, Order had a strong effect on agent assignment in the absence of coverbs, demonstrating that irreversible sentences had only one plausible interpretation (β = 11.16, SE = 2.67, Z = 10.09, p < 0.001). Structure and Order further interacted (BA: β = 0.72, SE = 0.13, Z = − 1.86, p = 0.063; BEI: β = 0.42, SE = 0.06, Z = − 6.52, p < 0.001), such that the effect of Order was considerably weaker for BA and BEI structures than for NNV sentences but not entirely absent. Essentially, when the positioning of the plausible agent agreed with the coverb cue, participants' agent assignment was influenced more than by either cue alone (BA: β = 64.44, SE = 34.90, Z = 7.69, p < 0.001 ; BEI: β = 0.05, SE = 0.09, Z = 6.24, p < 0.001).

The model further revealed a two-way interaction between Agent Animacy and Order and a three-way interaction among the cues of Structure, Order, and Agent Animacy. In the two-way interaction, the effect of Order was stronger for plausible animate agents than for plausible inanimate agents (β = 1.69, SE = 0.13, Z = 6.73, p < 0.001). We note that this effect is similar in magnitude to the effect of Agent Animacy for reversible sentences. The three-way interaction was only significant for BEI sentences (β = 0.69, SE = 0.08, Z = − 3.12, p = 0.002), where, unlike for BA or NNV, there was minimal difference in the effect of Order between plausible animate agents and plausible inanimate agents (animate: β = 29.5, SE = 15.8, Z = 6.30, p < 0.001; inanimate: β = 15.9, SE = 8.13, Z = 5.41,

Competition Model	• Agent assignment driven by the following cues in order of cue strength: BEI > animacy > BA > word order • In the absence of other cues, Mandarin verb-final sentence interpretation will follow object-subject-verb word order
eADM	• Role reversals elicit N400 effects for Mandarin due to language-specific features of cue prominence • Role reversals may elicit semantic P600 effects depending on task • At sentence-initial position, inanimate nouns elicit larger N400 effect than animate nouns
Bag of Arguments	• Role reversals only modulate the N400 provided there is enough time (> 800 ms) for structural roles to inform predictions and sufficient relatedness between nouns and verb • Role reversals may elicit semantic P600 effects

Table 3. Predictions from three sentence processing models. Refer to the text for clarification on where our predictions diverged from the models.

Figure 1. Results of agent assignment task. (**a**) Model predictions for interaction between Agent Animacy and Structure for first agent noun selection in reversible sentences. Y-axis shows probability of selecting the first noun as agent. Error bars show 95% confidence intervals; (**b**) model predictions for interaction among Order, Agent Animacy, and Structure for first agent noun selection in irreversible sentences. Y-axis shows probability of selecting the first noun as agent. Error bars show 95% confidence intervals. (**c**) Individual differences in Coverb and Plausibility scores. Dotted line shows the line y = x where Coverb and Plausibility Scores are equal in value. Discrete labels for agent assignment strategy are shown here, but subsequent use of scores for analysis was done with Difference Score as a continuous variable.

$p < 0.001$). This effect was most apparent when comparing BA and BEI role reversal sentences. While BA was more successful at driving role reversal interpretations when the plausible agent was inanimate (animate: 67% (SE = 11.4) probability first noun agent selection; inanimate: 80% (SE = 8.4)): BEI role reversal interpretations were unaffected by Agent Animacy (animate: 29% (SE = 7.9); inanimate: 31% (SE = 8.3)). Model results are visualized in Fig. 1b.

Individual Differences
While the behavioral models showed group-level effects, further data exploration showed that not all participants used the same comprehension strategies. To quantify individual differences, we calculated individual scores for reliance on coverbs in reversible sentences (Coverb Score) and Order in the absence of a coverb in irreversible sentences (Plausibility Score). These scores respectively represent individuals' reliance on coverbs and plausibility when there are no competing cues for agent assignment. The maximum score value of 1 indicates a given participant always used the corresponding cue, whereas a score of 0 indicates the cue was always disregarded.

From visual inspection of individuals' Coverb and Plausibility Scores, we identified three discrete strategies: using only plausibility, only coverbs, or both plausibility and coverbs. To create a continuous metric, we calculated a Difference Score by subtracting the Coverb Score from the Plausibility Score, where individual scores ranged from − 0.98 (most coverb-driven) to 0.79 (most plausibility-driven). For visualization, we labeled participants
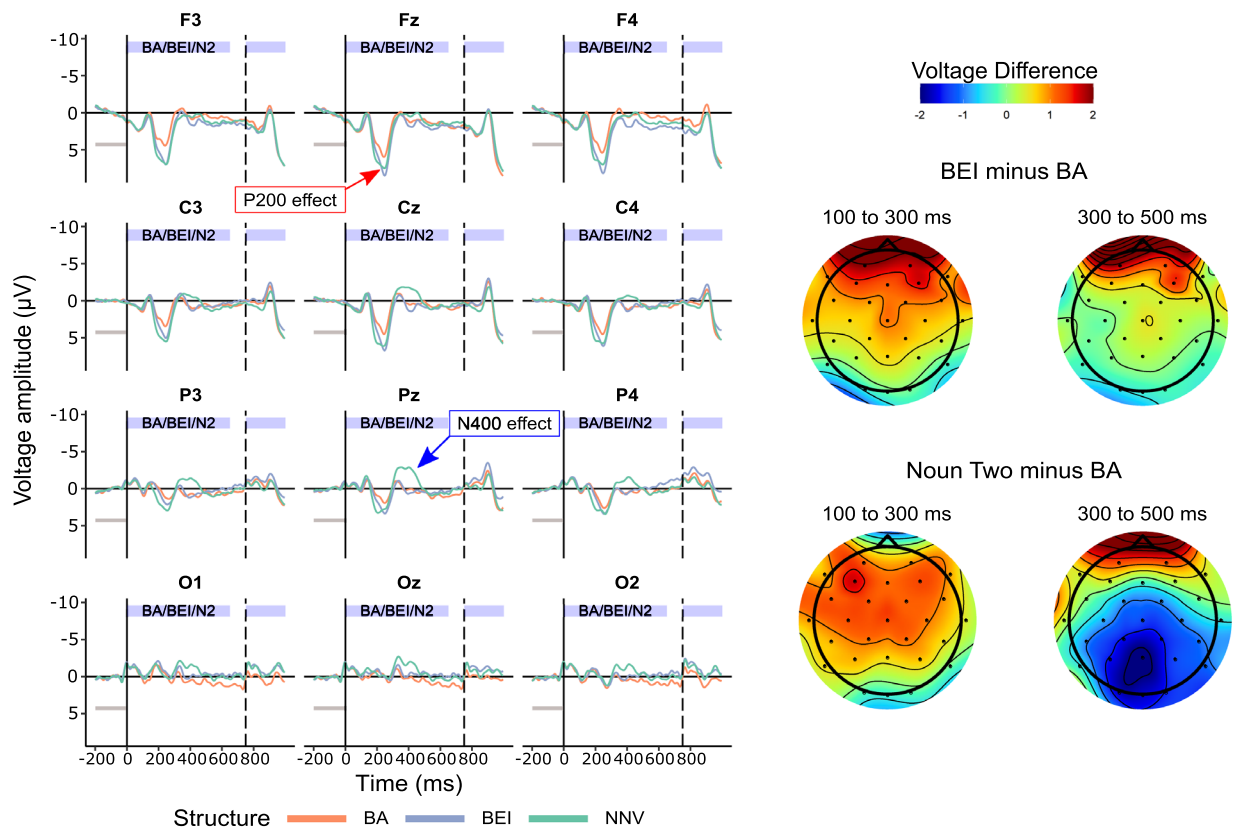
Figure 2. ERPs for the effect of Structure at the second word position. The 200-ms pre-onset baseline interval is indicated with a gray rectangle. Scalp maps show BEI minus BA (top) and noun two minus BA (bottom) for the P200 and N400 time windows.

with a Difference Score between – 0.33 and 0.33 as having a balanced strategy (i.e., the middle third of the values range), where they gave approximately equal weight to the two cues of plausibility and coverb, as depicted in Fig. 1c. This Difference Score further predicted participant reaction times, with more plausibility-driven participants taking longer to respond to reversible NNV sentences (see Supplementary Materials).

ERP Results

Noun One
Visual inspection of ERPs between animate and inanimate nouns in initial sentence position did not reveal a substantial difference in N400 amplitude. A mixed effects model for midline electrodes showed no significant effect of noun one animacy ($\beta = 0.08$ μV, SE $= 0.20$, Z $= 0.40$, p $= 0.69$, model results reported in Supplementary Materials).

Coverb
ERPs at the respective second word in NNV, BA and BEI sentences (i.e., contrasting noun two in NNV and the coverbs) showed strong variation in the P200 amplitude of the coverbs BA and BEI, as can be seen in Fig. 2. Direct comparison between ERPs showed that BA elicited a smaller P200 than BEI and noun two in NNV sentences, as can be appreciated in Fig. 2. Noun two further elicited a sizeable N400 component, consistent with word class effects[63]. The smaller P200 for BA was not predicted a priori, but given the striking visual difference in the ERP waveforms, we analyzed single trial amplitudes in the P200 and N400 time windows.

The average P200 amplitude for BA sentences was significantly smaller than for NNV sentences at electrodes Fz and Cz (ps < 0.01). With correction for multiple comparisons, P200 amplitude for BEI did not differ significantly from that for BA or Noun Two in NNV sentences. Average N400 amplitude for noun two in NNV sentences was larger than for BA and BEI (ps < = 0.01).

Role Reversals
Visual inspection of role reversal results showed that implausible sentences elicited a broadly distributed, centroparietal N400 between 300 and 500 ms and a sustained, localized frontal positivity around 800 ms, as depicted in Fig. 3. We further compared ERPs broken down by Structure and Agent Animacy (see Supplementary Materials). BA role reversals showed a larger and broader N400 effect with a sustained frontal positivity, and a later, broad positivity beginning around 700 ms in both posterior and frontal locations. BEI role reversals showed a smaller,
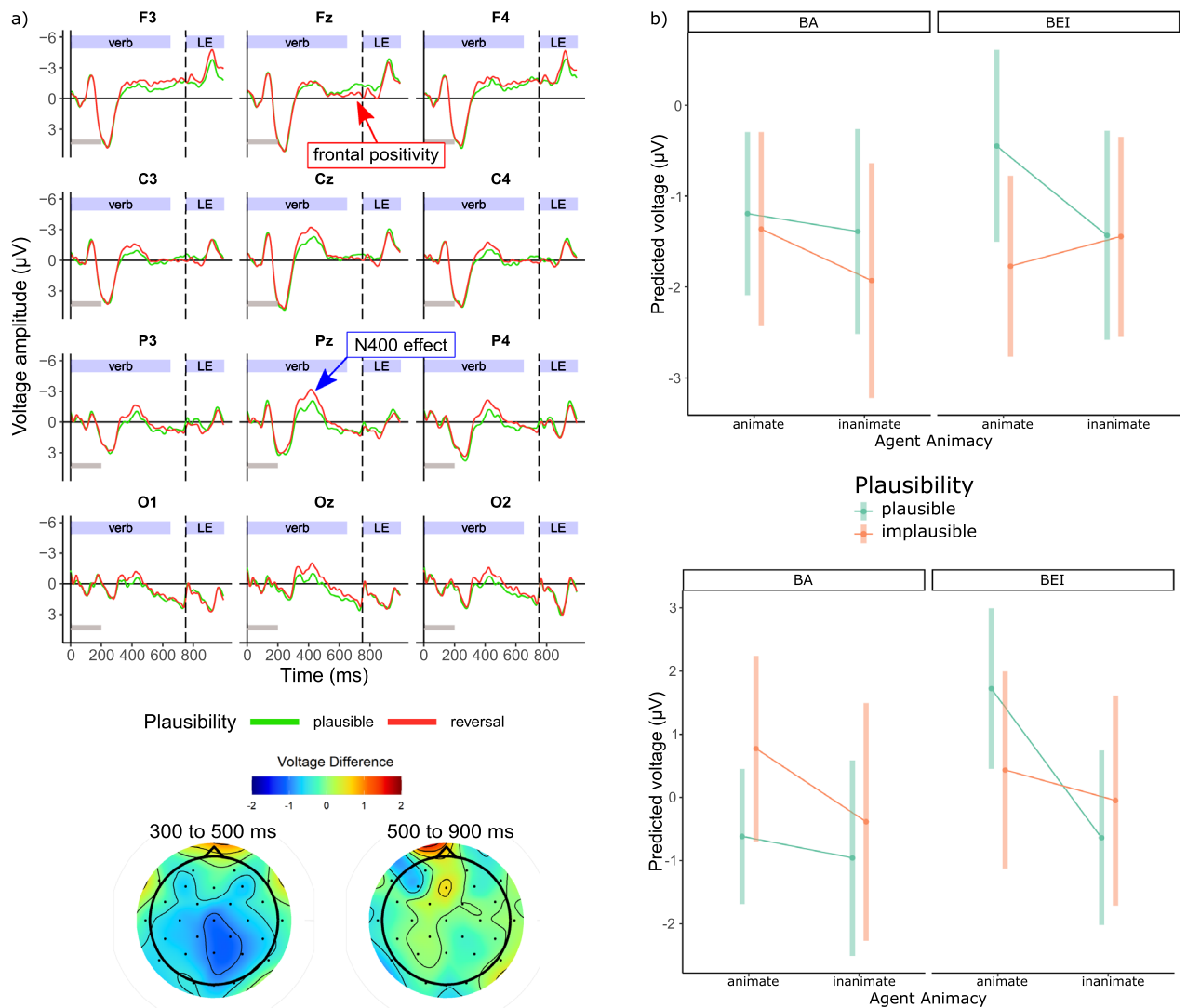
Figure 3. ERPs in response to role reversals. (**a**) ERPs are averaged across Agent Animacy and Structure. Scalp maps show reversal minus plausibility (averaged across other factors) for the N400 time window from 300 to 500 ms and the P600 time window from 500 to 900 ms. The 200-ms post-onset baseline interval is indicated with a gray rectangle. (**b**) Model predictions for three-way interaction among Structure, Agent Animacy, and Plausibility. Note that while the three-way interactions were significant for both time windows, not all post-hoc pairwise comparisons were significant. The simple effect of Plausibility was significant for the N400 time window and not for the P600 time window.

more localized N400 effect with a sustained frontal negativity, and a central-posterior right-lateralized positivity in the late P600 time window also beginning around 700 ms. For plausible animate agents, role reversals showed a broad N400 effect, and role reversals with plausible inanimate agents appeared to show a localized N400 effect followed by a broadly distributed positivity. However, these interactions should be interpreted cautiously due to relatively low power for statistical inference between subconditions.

N400.   Despite the appearance of an N400 effect for Plausibility in the ERP waveforms, our initial model results indicated this effect was only marginally significant ($\beta = -0.51$ μV, SE $= 0.32$, $Z = -1.61$, $p = 0.106$). Based on our contrast coding, the coefficient for Plausibility represented the effect at the reference level of Pz; however, the model including only Pz showed a significant effect of Plausibility ($\beta = -0.92$ μV, SE $= 0.35$, $Z = -2.65$, $p = 0.008$). We determined that the discrepancy between these model values was due to our limiting the output to three-way interactions (Voltage ~ Structure*Plausibility*Agent Animacy + Structure*Plausibility*Electrode + Structure*Agent Animacy*Electrode + Plausibility*Agent Animacy*Electrode). For a full model including all possible interactions (Voltage ~ Structure*Plausibility*Agent Animacy*Electrode), the corresponding coefficients were identical and the effect of Plausibility was significant. These results are challenging to reconcile and could lead to opposite conclusions based on sometimes arbitrary decisions. We bring up this difficulty because such challenges are faced by many, if not all, users of mixed effects models and the field continues to develop standards for best practice[64].

To address these issues, we report here the results of the midline model with a reduced random structure for item variability excluding a random slope for Plausibility to increase power[65], with other models reported in Supplementary Materials. The present model revealed a main effect of Plausibility (β = − 0.46 μV, SE = 0.18, Z = 2.55, p = 0.011), demonstrating that role reversals elicited an N400 effect averaged across Structure and Agent Animacy. There was a further three-way interaction among Structure, Agent Animacy, and Plausibility (β = 0.39 μV, SE = 0.18, Z = 2.16, p = 0.03). Post-hoc pairwise comparisons showed that the largest N400 effect was for implausible BEI sentences with animate agents (e.g., the reversal sentence 仆人被镜子擦亮了 "servant BEI mirror polished"; β = − 1.39 μV, SE = 0.35, Z = 3.99, p < 0.001). Described qualitatively from the model predictions in Fig. 3b, implausible BA role reversals elicited a numerically greater N400 than plausible BA sentences regardless of Agent Animacy status, while implausible role reversals with BEI sentences elicited a greater N400 for animate agents but not for inanimate agents.

P600. Although the ERP visualization suggested a small frontal positivity in the P600 time window, the statistical model did not reveal a significant main effect of Plausibility. Note that for consistency with the N400 analysis, we also excluded the random slope for Plausibility in the item random structure. The only significant model coefficient was a three-way interaction among Agent Animacy, Structure, and Plausibility (β = 0.62 μV, SE = 0.23, Z = 2.66, p = 0.008). Post-hoc pairwise comparison showed that for BEI sentences, ERP amplitudes to reversals with plausible animate agents were significantly more positive than their congruent counterparts. (β = 1.67 μV, SE = 0.46, Z = 3.63, p < 0.01). Nonetheless, we note that this frontal positivity is not a typical semantic P600 distribution[56,66].

Discussion

We investigated argument structure comprehension in verb-final sentences in Mandarin, a language with flexible word order and virtually no inflection. We analyzed monolingual, Mandarin native speakers' behavioral and EEG data to capture a cohesive picture of real-time cue competition. Our behavioral results demonstrated that 1) word order was not used to assign argument structure in the absence of other cues; 2) the coverbs BA and BEI were the strongest cues for agent assignment but were differently impacted by Agent Animacy in the case of role reversals; and 3) participants showed individual differences in their reliance on semantic and syntactic cues. Our EEG results showed that 1) sentence-initial noun animacy did not impact N400 amplitude; 2) BA elicited a reduced P200 amplitude relative to BEI and nouns; and 3) the disambiguating verb in role reversal sentences elicited an N400 effect without a subsequent semantic P600. To our knowledge, this is the first time a forced agent-assignment task has been used in an EEG experiment. A key advantage of this task is that it directly reveals a reader's sentence interpretation and resembles aspects of natural language processing, where individuals must understand the relation of a verb to its arguments. Additionally, all experimental sentences were, in principle, grammatical Mandarin structures, thus minimizing acceptability judgments.

With respect to prior findings, our behavioral results showed a distinct cue weighting profile for Mandarin. First, Agent Animacy was not the most important cue, with participants accepting both animate and inanimate plausible agents. We note that prior experiments often confounded animacy with plausibility[23,32,61,67], thus overestimating the role of animacy in driving sentence interpretation. Second, in reversible sentences without coverbs, word order did not drive agent assignment, challenging the idea that there is an inherent preference for object-subject-verb[32,68] or subject-object-verb[67] word order. Our findings support the idea that verb-final Mandarin sentences where word order is the only cue may be ambiguous[69]. In contrast to English, where pre-verb and post-verb positioning of arguments reliably signals structural roles[15,61], we suggest that only post-verb positioning is reliable in Mandarin[10,17].

Our behavioral results further challenge previous descriptions of cue weighting as mere linear ranking. Instead, multiple cues were weighted to varying extents in different contexts. Consequently, a particular profile of cues could result in different interpretations, meaning any given interpretation is best described probabilistically in terms of the available cues, as in Fig. 1. Animacy is a good example: while it had no simple effect for driving agent assignment, Agent Animacy interacted with other cues to subtly affect comprehension. While the coverb BA was consistently a stronger cue for agent assignment when its preceding noun was animate, this was not the case for the coverb BEI, for which interpretations or role reversal sentences were not impacted by Agent Animacy. This divergence hints at nuanced processing demands between BA and BEI[70]. We comment further on animacy below in conjunction with the ERP results.

While the group-level findings demonstrate an overarching pattern for Mandarin cue weighting, we further found individual differences in interpretation strategy. Although most participants used both plausibility (i.e., Order in irreversible sentences) and coverbs to make their interpretations, a subset of participants relied on one cue while largely ignoring the other. This response pattern further impacted reaction times (see Supplementary Materials), with plausibility-driven participants taking longer to respond to reversible NNV sentences than their coverb-driven counterparts. While individuals varied in comprehension strategy, group averages reflected core characteristics of the cue weighting profile of Mandarin, with individual variation occurring within the confines of these characteristics. For instance, while individuals disregarded the coverb cue in favor of plausibility, no participant interpreted BA as if it were BEI, or vice versa.

In conjunction with the behavioral results, our ERP findings provide insights about online, incremental parsing decisions. We considered three sentence time windows: the first noun, the second word, and the verb. At the first noun, we found no effect of animacy on N400 amplitude, unlike previous reports in English[59,60], which alongside the behavioral findings indicates that the cue of animacy alone is not sufficient to drive assignment of argument structure. Instead, animacy can interact with other cues, and inanimate nouns can be preferentially interpreted as agents given certain semantic contexts. Importantly, the potential of inanimate nouns to be agents

is likely impacted by multiple factors, including concreteness[71], situational relationships[72], motor and social cognition[73], and whether the nouns correspond to places[74] or natural forces[75]. These factors were not controlled in the present experiment, so it is possible that the perceived "agentiveness" of the inanimate nouns in the present study was relatively high, thus leading to the particular profile of agent interpretation and the lack of a sentence-initial N400 amplitude difference from animate nouns.

At the second word position, we observed a striking decrease in P200 amplitude for BA in comparison to BEI and nouns. Although not predicted, the P200 effect is compelling evidence for differences between BA and BEI in cognitive demand prior to the verb[76]. Upon encountering BA, most participants likely assigned agent status to the previous noun; when reading BEI, participants had to wait for the upcoming noun to complete agent assignment. We note that this P200 effect cannot be due to the visual simplicity of BA in relation to other characters[77,78]; if this were the case, then BA (把) and BEI (被) both should have P200 amplitudes smaller than the more visually complex, two-character nouns. There has also been a report of a similar P200 difference in the auditory domain[79]. This effect may stem from a difference between assigning agent or patient arguments to an iteratively constructed sentence structure during online processing. To test this hypothesis, future studies should test single-argument sentences with BA and BEI (e.g., 把苹果吃掉了 "BA apple ate") and the alternation of the experimental task to patient instead of agent identification. We note that this manipulation is only possible with flexible word order for pre-verbal arguments and could thus represent processing specific to languages with frequent verb-final structures, although it may be limited to the current experimental task.

At the position of the verb, we found that role reversals elicited an N400 effect followed by a local frontal positivity. In contrast to reports of semantic P600 effects to role reversals[45,46,48,54], our findings indicate that Mandarin role reversal anomalies were detected via relatively early, automatic semantic processing and meaning retrieval mechanisms. Our finding of a frontal positivity contrasts with predictions for semantic P600s, given that frontal positivities are often dissociated from typically posterior P600 effects[56,66]. This lack of a semantic P600 effect is not the first such report[16,58], suggesting it may be prudent to reconsider the label "semantic P600", just as early labels of the N400 and P600 components as indexing processing of semantics[5,80] and syntax[81] were mis-leading. If P600 responses to role reversals are primarily driven by task[16,58], this component may be understood better as a member of the P300 family[82], in which case prior reports of semantic P600 effects were likely driven by acceptability judgment tasks. To appreciate the interplay of the N400 and P600 components, computational models like the retrieval-integration account[49,83,84] and noisy channel models[85,86] are well equipped to explain and predict ERP responses to role reversals, although the present study suggests that crosslinguistic differences cannot be discounted. One way to integrate crosslinguistic differences may be via a language-specific filter based on cue reliability and prominence, with quantitative and qualitative differences among languages and language experience[39,87].

Our ERP results may be partially explained by existing models, as outlined in Table 3. The N400 effect can be, in principle, consistent with both the eADM and the Bag of Arguments account, although there are some discrepancies with prior reports. According to the eADM, and consistent with the present study's findings, Mandarin role reversals modulate the N400 because of the language's flexible word order, which requires greater weighting of plausibility cues at early processing stages[16]. A prior eADM study of Mandarin found N400 modulation only for BEI reversals and not BA reversals[16], which could conceivably stem from task or modality differences, as well as the present experiment's systematic comparison of multiple cues. For the Bag of Arguments account, the model can be adjusted to explain the present observation of N400 modulation with 750 ms stimulus onset asynchrony (SOA). While experiments have demonstrated that 600 ms is insufficient[56,58] and 800 ms is sufficient[58] for argument role assignment to constrain verb prediction as reflected in modulation of the N400 component, there is not a functional explanation for the necessary time to complete the associated computations. One possibility is a mechanism proposed by the Memory, Unification, and Control model[88], where semantic information is integrated in processing cycles. Combining the structural roles of two sentence elements to aid in verb prediction could require two complete processing cycles, perhaps corresponding to twice the N400 latency, which could also be consistent with the present study's SOA of 750 ms.

Neither the eADM nor the Bag of Arguments account specifically predicted the present experiment's lack of a P600 effect, although both research groups have highlighted task as driving P600 modulation[16,58]. While eADM experiments have not reported P600 effects for Mandarin role reversals, the model does not explain why role reversals for some sequence-independent languages elicit just an N400 effect (Mandarin and Turkish) and others elicit a biphasic N400-P600 effect (German and Icelandic)[16]. For the Bag of Arguments account, most studies have reported a semantic P600 effect[54–56], although recent experiments with fewer task demands found no P600 modulation[58], and the model itself does not make explicit predictions for P600 modulation. Additionally, in contrast to the present experiment, Bag of Arguments studies of Mandarin have manipulated cloze probability and combinability in specifically BA verb-final sentences with animate agents[54,56,58]. If our nouns and verbs were less related to each other than stimuli in previous studies, or the presence of an explicit task with multiple competing cues in BA, NNV, and BEI sentences impacted participants' parsing strategies, these differences could explain the inconsistent findings. Task and stimuli differences have been shown to affect N400 and P600 responses to role reversals[59,60], and acceptability tasks are especially linked with P600 effects[82,87,89]. While some researchers describe the N400 as less sensitive to task modulation than the P600[60,89–91], it should be noted that the N400 can show nuanced variation depending on task and context, potentially because of component overlap[92]. Because the present study is the first ERP experiment using a forced-choice agent assignment task, we cannot discount the potential for task to play a role in driving our effects.

The present study updates our understanding of Mandarin sentence processing and the incremental, online processing of argument structure assignment. In contrast to previous studies describing Mandarin cue weighting as a simple linear ranking[23,32,61], we used logit models to more accurately portray a probability-based profile of Mandarin[62], capturing the gradient nature of cues for sentence comprehension[93]. Our results are consistent with

Variable	Mean	SD	Range
Self-reported Mandarin proficiency (scale of 1–10)	8.6	0.9	6.7–10
Self-reported English proficiency (scale of 1 to 10)	3.3	2.0	1–7.3
Self-reported Dialect proficiency (scale of 1 to 10)	6.9	2.5	0–10
English age of acquisition (self-reported)	8.9	2.2	5–13
Mandarin LexTALE score (out of 50)	45 out of 50 (90%)	3.8	32–50
English LexTALE score (out of 100)	54 out of 100 (54%)	9.2	40–73
Mandarin exposure percent	46.3	29.0	2.7–98.7
English exposure percent	5.4	5.1	0–17.5
Dialect exposure percent	48.3	30.2	0–97
Mandarin usage percent	92.6	10.2	60.3–100
English usage percent	7.0	9.4	0–35.8
Dialect and other languages usage percent	0.4	1.8	0–9.6

Table 4. Participants' language experience and proficiency.

models proposing crosslinguistic differences in core processing steps[16,87,94] and provide new information for the timing of argument role information in comprehension of verb-final sentences[58]. We cannot discount the impact of task and stimuli differences in driving some of our findings, especially for the ERP effects that have smaller effect sizes than the behavioral results. Verifying the extent of crosslinguistic differences will require systematic comparison of diverse languages and sentence types beyond role reversals. Even so, inconsistent ERP findings for Mandarin role reversals, including N400 effects[16], P600 and N400-P600 effects[54,56], and null effects[58] suggest that this language may merit special consideration in neurocognitive models of sentence processing. Of broader implication, recent advances in machine learning have led to successful decoding of sentences from blood-oxygen-level-dependent data[95,96], indicating the potential for rapid advancement in analyzing neuroimaging data. As the current study shows, basic tenets of syntactic and semantic processing diverge among languages, and thus decoding nuanced sentence meaning from brain data may require precise targeting to language-specific features.

Methods
Participants
In total, 39 Mandarin native speakers participated in the study. Of these 39, four were excluded from analysis due to technical problems during experiment delivery and one was excluded due to failure to stay attentive during the experimental session, resulting in 34 (19 to 25 years old, mean age = 22, SD = 1.9, 19 female) datasets. Additionally, there were six subjects who did not see the BA reversal subcondition (approximately 30 sentences in total) due to error in experiment delivery. We opted to include their data in all analysis, as mixed effects models (see Data Analysis section below) can appropriately handle missing data[97].

All participants were recruited via online advertisement and word of mouth in Nanjing and tested at Nanjing Normal University. All participants were right-handed based on the Edinburgh Handedness Inventory[98] (average score = 83) had normal vision or wore corrective lenses, and did not have any history of neurological disorders. Participants gave written informed consent and were compensated 150 RMB for their time.

To ensure that Mandarin processing was not influenced by other language experience, we limited recruitment to participants who primarily communicated in Mandarin and had limited knowledge of English and other languages, including Chinese languages and dialects. Because English is a required subject in Chinese primary, secondary, and tertiary schools[99], all participants had some previous exposure to English. To minimize the influence of English on processing, we restricted recruitment to only those who self-reported an English level of 3 or below on a scale from 1 to 6 (1 being no knowledge of English, 6 being nativelike), who did not use English on a regular basis, and who were at or below the College English Test Level 4, which is typically below communicative competence[100]. If participants had exposure to a dialect other than standard Mandarin, this was restricted to Northern dialects (e.g., Nanjing, Xuzhou, Nantong, Shandong, Hebei) which are classified as belonging to the Mandarin dialect family and are mutually intelligible[17]. Note that there were exceptionally three participants in the present study who had knowledge of a Chinese language outside of the Mandarin dialect family (Wu, Gan, and Xiang), but they had minimal exposure to these languages in their adult life and primarily used Mandarin.

Participants further completed a detailed language background and usage questionnaire, from which we report summary values in Table 4. To further evaluate their language knowledge, participants also completed a LexTALE lexical decision task in English[101] and Mandarin[102]. Self-reported proficiency values represent a mean of three separate values for reading, writing, and listening. Exposure percentages represent the self-reported average percent of exposure time from birth to the present. Participants reported percentages in approximately three-year increments throughout their lives, which we then averaged to create an aggregate estimate of lifetime language exposure. Note that the dialect exposure numbers primarily reflect Mandarin dialects (e.g., Nanjing, Nantong, and Xuzhou dialects), which are mutually intelligible with standard Mandarin. Usage percentages represent the average of self-reports of percent of time a language is used in different social contexts, including at school, at the workplace, speaking with friends, and general reading.

Reversibility	Agent Animacy	Variable	Pair one—mean of noun 1 (upper value) noun 2 (lower value) ± SD	Pair two—mean of noun 1 (upper value) noun 2 (lower value) ± SD	Mean of verb ± SD
Reversible	Animate	log(Frequency)	2.36 ± 0.84	2.30 ± 0.87	2.88 ± 0.51
			2.59 ± 0.58	2.66 ± 0.66	
		Number of strokes	13.0 ± 4.16	15.0. ± 4.31	16.5 ± 3.08
			14.9 ± 4.35	15.4 ± 4.20	
	Inanimate	log(Frequency)	2.73 ± 0.60	2.46 ± 0.82	2.63 ± 0.57
			2.61 ± 0.67	2.59 ± 0.64	
		Number of strokes	15.6 ± 3.89	15.8 ± 3.85	16.8 ± 3.88
			16.6 ± 3.96	15.4 ± 3.65	
Irreversible	Animate	log(Frequency)	2.36 ± 0.80	2.48 ± 0.61	2.62 ± 0.56
			2.40 ± 0.81	2.60 ± 0.67	
		Number of strokes	14.0 ± 4.30	14.1 ± 4.88	17.4 ± 3.94
			16.8 ± 4.74	17.3 ± 3.59	
	Inanimate	log(Frequency)	2.29 ± 0.90	2.36 ± 0.67	2.49 ± 0.72
			2.50 ± 0.78	2.38 ± 0.83	
		Number of strokes	15.5 ± 3.99	14.7 ± 4.39	17.67 ± 4.44
			15.2 ± 4.38	14.7 ± 5.08	

Table 5. Controlled variables for sentence materials.

Materials

We created verb-final sentences with two noun arguments across the two levels of Reversibility (reversible, irreversible) and Agent Animacy (animate, inanimate). Crossing these two factors resulted in four conditions: reversible animate agent, reversible inanimate agent, irreversible animate agent, and irreversible inanimate agent (as summarized in Table 1). To maximize ambiguity in reversible sentences, we chose nouns that shared the same animacy status. We selected 30 transitive verbs for reversible inanimate and irreversible inanimate sentences and 31 transitive verbs for reversible animate and irreversible animate sentences, resulting in 122 unique verbs. To minimize repetitions of sentence materials during the experiment, we selected two noun pairs (noun pair one and noun pair two) for each verb, such that each pair combined with the verb to meet the requirements of the corresponding condition (e.g., reversible with animate agent: 老板技工举报了 "boss technician denounced"; 证人被告举报了 "witness defendant denounced"). These steps resulted in a total of 244 unique noun pairs. Within these parameters, we further controlled for frequency (using subtitle frequencies[103]) and number of strokes. Frequency values and number of strokes are reported in Table 5. The full sentence materials are reported in the Supplementary Materials.

In designing the sentence materials, we endeavored to select ideal agent-patient verbs and nouns, but a minority of verbs were closer to the experiencer verbs category (e.g., 强化 "strengthen", 冻死 "freeze", 安慰 "comfort"). This variability in our stimuli may affect our results, as experiencer verbs used in role reversal sentences have been shown to elicit biphasic N400-P600 effects in English[59,60]. However, we note that the majority of our stimuli use agentive verbs, and the designation of experiencer or agentive verb does not neatly extend across languages. In the case of Mandarin, the structure of verb complements means that there can be a mismatch between the functional and structural roles of arguments[104]. For these cases, the framework of proto-agent and proto-patient still applies[28].

Prior to running the EEG experiment, we created an offline questionnaire with our sentence materials to receive information for agent assignment and acceptability ratings from native Mandarin speakers. Note that these questionnaires did not include coverbs, which naturally resulted in decreased acceptability in the absence of a conversational context. Although including BA or BEI in these sentences would increase the naturalness, we wanted to understand how our sentences were comprehended at a purely semantic level and that they would meet a minimum level of acceptability in the NNV structure without a coverb, without systematic differences

Reversibility	Agent Animacy	Acceptability rating (mean ± SD)	Agent preference (mean ± SD)
Reversible	Animate	2.12 ± 1.40	0.44 ± 0.50
	Inanimate	2.50 ± 1.58	0.50 ± 0.50
Irreversible	Animate	2.21 ± 1.48	0.83 ± 0.38
	Inanimate	2.13 ± 1.45	0.83 ± 0.38

Table 6. Pretest results for sentence materials. Acceptability Rating is based on a scale from 1 to 5, where 1 indicated "completely unacceptable" and 5 indicated "completely acceptable". Agent Preference is an average value where each sentence item was presented to participants in both possible orders. A value closer to 1 or 0 indicates a strong preference for one of the nouns, while a value closer to 0.5 indicates no preference for either noun.

between conditions in acceptability, as well as ensuring there was a clear semantic direction for our irreversible sentences. These results are summarized in Table 6.

To make our list of stimuli for running the experiment, we next crossed our factors Reversibility and Agent Animacy with Structure (NNV, BA, and BEI) and Order (first and second, representing position of the plausible noun). Note that for reversible sentences, one of the orders was arbitrarily assigned as first so that Order could still be tested and controlled for these items. We assembled ordered lists for presenting sentences to participants. Each of the 122 verbs was used three times, once for each level of Structure, resulting in 366 total sentences. To minimize the effects of repetition, we used the two noun pairs for each verb, so that a given noun only repeated a maximum of once. For example, the two noun pairs 喜鹊鸟笼/老鼠箱子 "magpie birdcage/mouse box"and verb 困住 "trap" might appear in the experiment as follows: 喜鹊被鸟笼困住了。 "magpie BEI birdcage trapped"; 喜鹊把鸟笼困住了。 "magpie BA birdcage trapped"; 箱子老鼠困住了。 "box mouse trapped". Note that each sentence ended with the aspect particle LE and a period.

To pseudorandomize our stimuli, we used the program Mix[105], constraining the randomization such that each level of Structure could repeat a maximum of two times consecutively and a given verb occurred a minimum of 90 trials before or after its previous occurrence. Due to the design of the sentence materials, there was an equal probability of the first or second noun being animate or inanimate and actor or undergoer, so there was no way for participants to develop strategies to predict the role of the nouns until they saw BEI or BA and the final verb. Stimuli were pseudorandomized to maximize distance between repeated verbs (at least 90 items between repetitions) and minimize repetitions of same structure condition to two.

Procedure

All parts of the experiment were approved by the McGill Faculty of Medicine Institutional Review Board following the guidelines of the Canadian Tri-Council Policy Statement and by the School of Foreign Languages and Cultures at Nanjing Normal University (南京师范大学外国语学院). After reviewing and signing the consent form, participants sat in a sound-attenuated booth. All stimuli were presented with Presentation® software (Version 17.2, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com) using the Windows XP operating system.

Sentences were presented visually word-by-word, for 650 ms per word and each word followed by a 100 ms inter-stimulus interval (i.e., stimulus onset asynchrony (SOA) = 750 ms). Each trial began with a cue for the participants to blink ("(−)") for 2000 ms, then a fixation cross ("+") for 500 ms. Each sentence ended with the particle LE appearing for 650, before displaying a prompt for the agent assignment task. For the task prompt, the two nouns from the preceding sentence appeared on the screen, with the first noun on the left and the second noun on the right, with " !!!!! " between them. Participants were asked to choose which of the two nouns was the agent (施事) by pressing the A or L key, corresponding to the left or right noun, respectively. The experiment was divided into four blocks of approximately 90 sentences each, with a scheduled break between each block. Participants could also pause at any response prompt to rest before continuing. The total experiment time, including preparation of the cap and cleanup, lasted from two to three hours.

EEG Recording and Preprocessing

Participants' EEG was recorded from 32 Ag/AgCl electrodes mounted on an elastic cap according to the international 10–20 system (EASYCAP, Herrsching, Germany). To monitor vertical and horizontal eye movement, one electrode was positioned below the right eye and another to the left of the left eye. Electrode impedance was kept below 10 kΩ. The recordings were amplified online with a bandpass filter of 0.05–100 Hz, referenced online to electrode FCz, and digitized at a sampling rate of 500 Hz.

We used EEGLAB (v2019.1) and ERPLAB (v7.0.0) to preprocess the data. The EEG signal was downsampled to 250 Hz and re-referenced to the average of the linked mastoids (TP9 and TP10). We then used a high-pass filter at a cutoff of 0.1 Hz (FIR filter, Kaiser window, Kaiser beta = 4.89856, transition bandwidth = 0.2, filter order = 3934) and a low-pass filter at a cutoff of 30 Hz (FIR filter, Kaiser window, transition bandwidth = 10, Kaiser beta = 4.89856, filter order = 80). To correct eye movement artifacts, we decomposed the data using independent component analysis (ICA, runica algorithm in EEGLAB, with the option 'extended', 1). Note that exclusively for the ICA decomposition, we used data that was high-pass filtered at a cutoff of 0.5 Hz (FIR filter, Kaiser window, Kaiser beta = 4.89856, transition bandwidth = 0.2, filter order = 3934) because using a higher high-pass filter on data for ICA decomposition improves the signal-to-noise ratio[106]. The data were cleaned automatically prior to ICA using the pop_rejcont function (epochlength 2, overlap 1, freqlimit 1–25, threshold 10, taper hamming). The final ICs were then copied to the data filtered at 0.1 and 30 Hz for analysis. We removed a maximum of two ICs per participant, one IC each for vertical and horizontal eye movement.

The signal was then segmented into epochs from −200 to 1000 ms around each critical word (first noun, coverb, second noun, and verb), with pre-stimulus 200 ms baseline correction. To appreciate the ERP changes across the entire sentence, we additionally created whole-sentence epochs with 200 ms pre-onset baselines; including the baseline interval, these epochs spanned 3100 ms for NNV sentences and 3850 ms for BA and BEI sentences. Based on visual inspection of role-reversal sentences in the whole-sentence epochs, we determined that there were important ERP differences occurring before verb onset, which could make a pre-onset baseline problematic[107]. To minimize the impact of baseline differences at the verb, we re-epoched the verb time window with a post-stimulus-onset 200 ms baseline correction (baseline interval of 0 to 200 ms), with the assumption that early components in this interval should have minimal difference between conditions. Results from this post-stimulus-onset 200 ms baseline correction are reported in the text because we believe this more accurately reflects verb-linked activity; results from the original pre-stimulus 200 ms baseline correction are reported in Supplementary Materials.

Artifact rejection was performed across epochs with a moving window threshold of 80 μV (window size = 500 ms). During review of the artifact rejection process, we determined that electrodes Fp1 and Fp2 were exceptionally noisy across participants and excluded them from analysis. For select subjects whose automatic rejection resulted in greater than 15% of trials being rejected, epoched data were manually inspected to include additional trials and the overall quality of individual datasets. This review resulted in all 34 subjects being included for final analysis (i.e., all individuals had fewer than 15% rejected trials).

Data analysis

We analyzed responses and reaction times for the agent assignment task and ERPs time-locked to the onset of target words in sentences. The four factors manipulated in the sentence materials were included in the analysis of each of these measures. The factor Structure was comprised of three levels: NNV, BA, and BEI. Structure was treatment coded such that NNV was the reference level to evaluate the effect of coverbs in relation to sentences with no coverb. The factors Reversibility (reversible and irreversible), Agent Animacy (animate and inanimate) and Order (first and second, denoting position of the plausible noun in irreversible sentences) were sum coded. Unless otherwise noted, these were the factors and contrast coding.

All data analysis was done using *R* version 4.02[108]. To account for variability across items and within participants, we computed mixed effects models using the *glmer* function from package *lme4* version 1.1-23[109], including the optimizer = 'bobyqa' parameter. Model coefficients were calculated by maximum likelihood estimates using the Laplace approximation. In the case of the binary data from the agent assignment task, we added the argument family = 'binomial' to fit a logistic mixed effects model. To ensure that model effects were interpretable, we limited fixed and random effects to a maximum of three-way interactions, even if there were possible higher order interactions. For random effects structures, all factors with possible variability within items or participants were included in the maximal possible structure. Note that because Agent Animacy and Reversibility did not vary across individual items, these factors were not included in the random structure for item.

All *p*-values were calculated with the Satterthwaite approximation calculated based on Wald *Z*-scores in *lmerTest* package version 3.1-2[110]. To construct the maximum possible model and random structure, we used the *buildmer* package version 1.8[111] using the direction = 'order' parameter, which adds effects to the model in order of their contribution to log-likelihood. We then again used *buildmer* to do stepwise removal of model variables with the direction = 'backward' parameter to maximize log-likelihood score. These optimized models are reported in the text to maximize power and minimize overfitting[112], but the maximal models are reported for reference in the supplementary materials[113].

Significant interactions were followed up with post-hoc tests for pairwise comparisons using the *emmeans* package version 1.4.8[114] and Tukey method for adjustment of *p*-values to correct for multiple comparisons. Interactions were visualized with the emmip function from the *emmeans* package. Note that while *p*-values are reported for model results, our inferences and interpretations were not limited to significance testing; instead, we further considered our hypotheses and predictions, effect sizes, and the limitations of data quantity and quality[115].

For ease of interpretability, model results are reported with graphical depictions of coefficients and confidence intervals generated by the plot_model function from the *sjPlot* package version 2.8.9[116]; full model outputs, including random effects, are reported in the supplementary materials in tables generated from the tab_model function from the *sjPlot* package. For simplicity, model results reported in the text are limited to significant effects or effects that were related to initial predictions. For data arrangement and general plotting, we used the *tidyverse* package[117], with final figure adjustment performed using the software Inkscape version 0.92[118].

Agent Assignment

Binary agent assignment responses were analyzed with logistic mixed effects models. Because irreversible sentences had a single plausible interpretation, while reversible sentences had two plausible interpretations, we analyzed reversible and irreversible sentences separately. This allowed us to better understand the effect of plausibility on agent assignment, while limiting model coefficients to a maximum of three-way interactions. For both reversible and irreversible sentences, the maximum specified model included the fixed effects of Structure, Agent Animacy, and Order, random slopes and intercepts for Structure, Agent Animacy, and Order by participant, and random slopes and intercepts for Structure and Order by item. We report coefficients, confidence intervals, and *p*-values on the odds ratios scale, but original tests were performed on the log odds scale. For interpretability, interactions are illustrated on the probability scale.

Reaction Times

Reaction times for agent assignment responses were analyzed with linear mixed effects models. Reaction times were first cleaned to exclude response times above 10 s or below 100 ms. We then cleaned reaction times by condition, limiting to those values within 1.5 standard deviations for each subcondition of Structure, Reversibility, and Animacy[119]. These steps resulted in excluding 12.4% of trials from further analysis; we note that some of the excluded trials included instances when participants took breaks before responding. Reaction times were then natural log transformed to ensure that we met assumptions of distribution normality for analysis[120]. Note that we also analyzed the raw reaction time values[121] and results were similar to those found for the log-transformed data; these results are reported in supplementary materials for transparency[122]. The maximum specified model included the fixed effects of Structure, Reversibility, Agent Animacy, and Order, random slopes and intercepts for Structure, Reversibility, Agent Animacy, and Order by participants, and random slopes and intercepts for Structure and Order by item. As an additional step, we also ran a model with the additional factor of Difference Score (the difference between a participant's reliance on plausibility cues and their reliance on coverb cues), which is introduced in the section Individual Differences in Cue Weighting for Agent Assignment. Recent

work has demonstrated the importance of individual differences in psychology and language research[123,124], and including Difference Score in the model explained additional variability in the data. We report coefficients, confidence intervals, and *p*-values on the log-transformed scale, but model predictions were back-transformed to milliseconds for interpretability. Note that all reaction time results are reported in Supplementary Materials.

ERPs

As noted above, ERPs were analyzed at the first noun, coverb, second noun, and verb position of the sentence. Condition averages were calculated for each subject and then grand average ERPs were calculated for each condition. These grand average ERPs by condition were used for visual inspection and are represented in all ERP figures in the present study. Statistical models, however, were all based on average amplitudes for specific time windows in single trial epochs.

For the first noun of the sentence, referred to hereafter as noun one, we analyzed average amplitude in the N400 time window from 300 to 500 ms. This time window analysis was planned a priori based on reports of greater N400 effects for inanimate nouns than for animate nouns[59,60]. At the second-word position of the sentence, there was either another noun (noun two, in the case of NNV sentences), the coverb BA, or the coverb BEI. At this sentence position, we analyzed average amplitudes in the P200 (100 to 300 ms) and N400 (300 to 500 ms) time windows. The P200 time window was selected for analysis after visual observation of large differences in the ERPs between sentence structure types. The N400 time window was analyzed as a validation step to confirm expectations that nouns elicited larger N400 amplitudes than coverbs, thus giving more weight to the unexpected differences in P200 amplitude. At the verb position of the sentence, we analyzed the N400 (300 to 500 ms) and P600 (700 to 900 ms).

For each time window analyzed, we used single trial average amplitude to calculate linear mixed effects models. We first ran models on midline electrodes, including the factor Electrode (Fz, Cz, Pz, Oz), to confirm the presence of effects, and then over all other electrodes on the scalp excluding the midline, with the additional levels of Anteriority (frontal, central, and posterior) and Laterality (right, left). N400 and P600 effects, the primary components investigated in the present study, typically present with a posterior distribution on the scalp[125]; with this in mind, we treatment coded the factors Electrode and Anteriority with the reference levels of Pz and posterior, respectively. In contrast, P200 effects typically have a frontal distribution[78], so for models in the P200 time window, we exceptionally used the reference levels of Fz and frontal for Electrode and Anteriority, respectively. Note that for the model specifications below, the factor Electrode was substituted by Anteriority and Laterality for the models over non-midline electrodes.

For noun one, the maximum specified model included the fixed effects of noun one Animacy (animate, inanimate) and Electrode, with random slopes and intercepts included for noun one Animacy and Electrode by participant and by item. For noun two and coverb, the maximum specified model included the fixed effects of Structure (NNV, BA, BEI) and Electrode, with random slopes and intercepts included for Structure and Electrode by participant and by item.

At the verb, only unambiguous role reversal sentences were analyzed (contrasting plausible vs implausible sentences), which limited trials to irreversible BA and BEI sentences (see Table 2). For clarity with respect to predictions about role reversal effects, the factor Order was recoded in terms of Plausibility; BA sentences with the plausible noun in first position were coded as plausible, while BEI sentences with the plausible noun in first position were coded as implausible, with the same logic applied for sentences with the plausible noun in second position. Plausibility was treatment coded with plausible as the reference level. Because there were only two levels of Structure (BA and BEI) with neither level more suited as a reference, Structure was sum coded for this analysis. The maximum specified model for the verb included the fixed effects of Structure, Agent Animacy, Plausibility, and Electrode. Random slopes and intercepts were included for Structure, Agent Animacy, Plausibility, and Electrode by participant, and random slopes and intercepts for Structure, Plausibility, and Electrode by item. Additionally at the verb, we ran models at individual midline electrodes as confirmation for the effects across midline electrodes. These models are reported in supplementary materials.

Because the components of interest in the present study (N400, P600, P200) are typically maximal at or near midline electrodes[125], we primarily report in the text results from models at the midline; results from lateral electrodes are reported in the text if they show effects beyond the models at midline electrodes. Note that full models for lateral electrodes excluding the midline are reported in the Supplementary Materials. Additionally, simple effects of the topographical factors Electrode, Anteriority, and Laterality, or interactions involving only these factors, are not reported or discussed in the text because they are not related to the experimental manipulations. Lastly, for models at the verb, we discuss in the text only those effects that included Plausibility because this is the only factor for which we had predictions.

All final analyses were performed on single trial average amplitudes, but average ERPs were calculated by condition for plotting purposes. All figures showing ERP voltage against time and scalp maps reflect these average ERPs and were plotted using the R package *ERPscope*[126].

Data availability

The datasets generated during and/or analyzed during the current study will be available upon request in the Borealis Dataverse.

References

1. Coupé, C., Oh, Y. M., Dediu, D. & Pellegrino, F. Different languages, similar encoding efficiency: Comparable information rates across the human communicative niche. *Sci. Adv.* **5**, 66 (2019).
2. Sainburg, T., Theilman, B., Thielk, M. & Gentner, T. Q. Parallels in the sequential organization of birdsong and human speech. *Nat. Commun.* **10**, 1–11 (2019).
3. Hickok, G. Chapter 4—The dual stream model of speech and language processing. In *Aphasia* (eds. Hillis, A. E. & Fridriksson, J.) 185, 57–69 (Elsevier, 2022).
4. Kutas, M. & Hillyard, S. Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science (80)* **207**, 203–205 (1980).
5. Lau, E. F., Phillips, C. & Poeppel, D. A cortical network for semantics: (De)constructing the N400. *Nat. Rev. Neurosci.* **9**, 920–933 (2008).
6. Costa, A., Strijkers, K., Martin, C. & Thierry, G. The time course of word retrieval revealed by event-related brain potentials during overt speech. *Proc. Natl. Acad. Sci.* **106**, 21442 (2009).
7. Ferreira, F. & Patson, N. D. The 'Good Enough' approach to language comprehension. *Lang. Linguist. Compass* **1**, 71–83 (2007).
8. Christiansen, M. H. & Chater, N. The Now-or-Never bottleneck: A fundamental constraint on language. *Behav. Brain Sci.* **39**, 66 (2015).
9. Krebs, J., Malaia, E., Wilbur, R. B. & Roehm, D. Subject preference emerges as cross-modal strategy for linguistic processing. *Brain Res.* **1691**, 105–117 (2018).
10. Wang, L., Schlesewsky, M., Bickel, B. & Bornkessel-Schlesewsky, I. Exploring the nature of the 'subject'-preference: Evidence from the online comprehension of simple sentences in Mandarin Chinese. *Lang. Cogn. Process.* **24**, 1180–1226 (2009).
11. Evans, N. & Levinson, S. C. The myth of language universals: Language diversity and its importance for cognitive science. *Behav. Brain Sci.* **32**, 429–448 (2009).
12. Friederici, A. D. & Frisch, S. Verb argument structure processing: The role of verb-specific and argument-specific information. *J. Mem. Lang.* **43**, 476–507 (2000).
13. Levy, R. P. & Keller, F. Expectation and locality effects in German verb-final structures. *J. Mem. Lang.* **68**, 199–222 (2013).
14. Vasishth, S., Suckow, K., Lewis, R. L. & Kern, S. Short-term forgetting in sentence comprehension: Crosslinguistic evidence from verb-final structures. *Lang. Cogn. Process.* **25**, 533–567 (2010).
15. Bates, E., McNew, S., MacWhinney, B., Devescovi, A. & Smith, S. Functional constraints on sentence processing: A cross-linguistic study. *Cognition* **11**, 245–299 (1982).
16. Bornkessel-Schlesewsky, I. *et al.* Think globally: Cross-linguistic variation in electrophysiological activity during sentence comprehension. *Brain Lang.* **117**, 133–152 (2011).
17. Li, A., & Thompson, S. A. *Mandarin Chinese: A Functional Reference Grammar*. (University of California Press, 1989).
18. Wu, F. & He, Y. Some typological characteristics of Mandarin Chinese Syntax. In *The Oxford Handbook of Chinese Linguistics* (Oxford University Press, 2015). https://doi.org/10.1093/oxfordhb/9780199856336.013.0020
19. Chappell, H., Ming, L. & Peyraube, A. Chinese linguistics and typology: The state of the art. *Linguist. Typol.* **11**, 187–211 (2007).
20. LaPolla, R. *Arguments Against 'Subject' and 'Direct Object' as Viable Concepts* **(in Chinese)** (1993).
21. Xu, L. Topic prominence. In *The Oxford Handbook of Chinese Linguistics* (Oxford University Press, 2015). https://doi.org/10.1093/oxfordhb/9780199856336.013.0072
22. Lu, C.-C. *et al.* Judgements of grammaticality in aphasia: The special case of Chinese. *Aphasiology* **14**, 1021–1054 (2000).
23. Su, I. R. Transfer of sentence processing strategies: A comparison of L2 learners of Chinese and English. *Appl. Psycholinguist.* **22**, 83–112 (2001).
24. Pylkkänen, L. The neural basis of combinatory syntax and semantics. *Science (80)* **366**, 62–66 (2019).
25. Neville, H., Nicol, J. L., Barss, A., Forster, K. I. & Garrett, M. F. Syntactically based sentence processing classes: Evidence from event-related brain potentials. *J. Cogn. Neurosci.* **3**, 151–165 (1991).
26. Chomsky, N. Chapter 2. Subsystems of core grammar. 2.6. LF-representation and θ-theory (2). In *Lectures on Government and Binding* 101–117 (DE GRUYTER MOUTON, 1993). https://doi.org/10.1515/9783110884166.101
27. Druks, J. Verbs and nouns—A review of the literature. *J. Neurolinguistics* **15**, 289–315 (2002).
28. Dowty, D. Thematic proto-roles and argument selection. *Language* **67**, 547 (1991).
29. Bisang, W. From meaning to syntax—Semantic roles and beyond. In *Semantic Role Universals and Argument Linking* 191–236 (Mouton de Gruyter, 2006). https://doi.org/10.1515/9783110219272.191
30. Rissman, L. & Majid, A. Thematic roles: Core knowledge or linguistic construct?. *Psychon. Bull. Rev.* **26**, 1850–1869 (2019).
31. Mahowald, K., Diachek, E., Gibson, E., Fedorenko, E. & Futrell, R. Grammatical cues are largely, but not completely, redundant with word meanings in natural language. CoRR abs/2201.1 (2022).
32. Li, P., Bates, E., Liu, H. & MacWhinney, B. Cues as functional constraints on sentence processing in Chinese. In *Language Processing in Chinese* (eds. Chen, H. C. & Tzeng, O.) 207–234 (North-Holland, 1992). https://doi.org/10.1016/S0166-4115(08)61893-2
33. Bender, E. The syntax of Mandarin Ba: Reconsidering the verbal analysis. *J. East Asian Ling.* **9**, 105–145 (2000).
34. Ting, J. Deriving the bei-construction in Mandarin Chinese. *J. East Asian Ling.* **7**, 319–354 (1998).
35. Li, C. & Thompson, S. Co-verbs in Mandarin Chinese: Verbs of prepositions?. *J. Chin. Linguist.* **2**, 257–278 (1974).
36. Deng, X., Mai, Z. & Yip, V. An aspectual account of ba and bei constructions in child Mandarin. *First Lang.* **38**, 243–262 (2018).
37. Huang, C.-T. J., Li, Y.-H. A. & Li, Y. *The Syntax of Chinese* (Cambridge University Press, 2009). https://doi.org/10.1017/CBO9781139166935
38. MacWhinney, B. The competition model: Past and future. In *Mechanisms of Language Acquisition* (ed. MacWhinney, B.) 3–16 (Psychology Press, 2022). https://doi.org/10.1007/978-3-030-66175-5_1
39. Martin, A. E. Language processing as cue integration: Grounding the psychology of language in perception and neurophysiology. *Front. Psychol.* **7**, 1–17 (2016).
40. MacWhinney, B. Chapter 14. A Unified Model of First and Second Language Learning. In 287–312 (2018). https://doi.org/10.1075/tilar.22.15mac
41. Osterhout, L. & Holcomb, P. J. Event-related brain potentials elicited by syntactic anomaly. *J. Mem. Lang.* **31**, 785–806 (1992).
42. Hoeks, J. C., Stowe, L. A. & Doedens, G. Seeing words in context: The interaction of lexical and sentence level information during reading. *Cogn. Brain Res.* **19**, 59–73 (2004).
43. Kolk, H. H. J., Chwilla, D. J., Van Herten, M. & Oor, P. J. W. Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain Lang.* **85**, 1–36 (2003).
44. Kim, A. & Osterhout, L. The independence of combinatory semantic processing: Evidence from event-related potentials. *J. Mem. Lang.* **52**, 205–225 (2005).
45. Kolk, H. H. J. & Chwilla, D. Late positivities in unusual situations. *Brain Lang.* **100**, 257–261 (2007).
46. Kuperberg, G. R. Neural mechanisms of language comprehension: Challenges to syntax. *Brain Res.* **1146**, 23–49 (2007).
47. Van Herten, M., Kolk, H. H. J. & Chwilla, D. J. An ERP study of P600 effects elicited by semantic anomalies. *Cogn. Brain Res.* **22**, 241–255 (2005).

48. Brouwer, H., Fitz, H. & Hoeks, J. Getting real about Semantic Illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Res.* **1446**, 127–143 (2012).
49. Brouwer, H., Delogu, F., Venhuizen, N. J. & Crocker, M. W. Neurobehavioral correlates of surprisal in language comprehension: A neurocomputational model. *Front. Psychol.* **12**, 1–16 (2021).
50. Bornkessel, I. & Schlesewsky, M. The extended argument dependency model: A neurocognitive approach to sentence comprehension across languages. *Psychol. Rev.* **113**, 787–821 (2006).
51. Bornkessel-Schlesewsky, I. & Schlesewsky, M. The argument dependency model. In *Neurobiology of Language* 357–369 (Elsevier, 2016). https://doi.org/10.1016/B978-0-12-407794-2.00030-4
52. Bornkessel-Schlesewsky, I. & Schlesewsky, M. An alternative perspective on 'semantic P600' effects in language comprehension. *Brain Res. Rev.* **59**, 55–73 (2008).
53. Bornkessel, I. & Schlesewsky, M. The extended argument dependency model: A neurocognitive approach to sentence comprehension across languages. *Psychol. Rev.* **113**, 787–821 (2006).
54. Chow, W.-Y. & Phillips, C. No semantic illusions in the "Semantic P600" phenomenon: ERP evidence from Mandarin Chinese. *Brain Res.* **1506**, 76–93 (2013).
55. Chow, W.-Y., Smith, C., Lau, E. & Phillips, C. A "bag-of-arguments" mechanism for initial verb predictions. *Lang. Cogn. Neurosci.* **31**, 577–596 (2016).
56. Chow, W.-Y., Lau, E., Wang, S. & Phillips, C. Wait a second! Delayed impact of argument roles on on-line verb prediction. *Lang. Cogn. Neurosci.* **33**, 803–828 (2018).
57. Chow, W.-Y., Momma, S., Smith, C., Lau, E. & Phillips, C. Prediction as memory retrieval: Timing and mechanisms. *Lang. Cogn. Neurosci.* **31**, 617–627 (2016).
58. Liao, C. H., Lau, E. & Chow, W.-Y. Towards a processing model for argument-verb computations in online sentence comprehension. *J. Mem. Lang.* **126**, 104350 (2022).
59. Bourguignon, N., Drury, J. E., Valois, D. & Steinhauer, K. Decomposing animacy reversals between agents and experiencers: An ERP study. *Brain Lang.* **122**, 179–189 (2012).
60. Kyriaki, L., Schlesewsky, M. & Bornkessel-Schlesewsky, I. Semantic reversal anomalies under the microscope: Task and modality influences on language-associated event-related potentials. *Eur. J. Neurosci.* **52**, 3803–3827 (2020).
61. Liu, H., Bates, E. & Li, P. Sentence interpretation in bilingual speakers of English and Chinese. *Appl. Psycholinguist.* **13**, 451–484 (1992).
62. Skalicky, S. & Chen, V. Forward and backward transfer of sentence processing cues in English and Mandarin Chinese: A call for approximate replication of Liu, Bates, and Li (1992) and Su (2001). *Lang. Teach.* https://doi.org/10.1017/S0261444820000464 (2020).
63. Steinhauer, K., Pancheva, R., Newman, A. J., Gennari, S. & Ullman, M. T. How the mass counts: An electrophysiological approach to the processing of lexical features. *Neuroreport* **12**, 66 (2001).
64. Meteyard, L. & Davies, R. A. I. Best practice guidance for linear mixed-effects models in psychological science. *J. Mem. Lang.* **112**, 104092 (2020).
65. Brysbaert, M. & Stevens, M. Power analysis and effect size in mixed effects models: A tutorial. *J. Cogn.* **1**, 66 (2018).
66. Kuperberg, G. R., Brothers, T. & Wlotko, E. W. A tale of two positivities and the N400: Distinct neural signatures are evoked by confirmed and violated predictions at different levels of representation. *J. Cogn. Neurosci.* **32**, 12–35 (2020).
67. Miao, X. Word order and semantic strategies in Chinese sentence comprehension. *Int. J. Psycholinguist.* **8**, 109–122 (1981).
68. Wang, L., Schlesewsky, M., Philipp, M. & Bornkessel-Schlesewsky, I. The role of Animacy in online argument interpretation in Mandarin Chinese. *Case Word Order Promin.* **40**, 91–119 (2012).
69. Yu, S. & Tamaoka, K. Age-related differences in the acceptability of non-canonical word orders in Mandarin Chinese. *Ling. Sin.* **4**, 66 (2018).
70. Hsien-Yi, G. W. The Syntax-Semantics Interface: The BA and BEI Constructions in Mandarin (National University of Singapore, 1998).
71. Scorolli, C. *et al.* Abstract and concrete sentences, embodiment, and languages. *Front. Psychol.* **2**, 66 (2011).
72. Barsalou, L. W., Dutriaux, L. & Scheepers, C. Moving beyond the distinction between concrete and abstract concepts. *Philos. Trans. R. Soc. B Biol. Sci.* **373**, 20170144 (2018).
73. Borghi, A. M. *et al.* Words as social tools: Language, sociality and inner grounding in abstract concepts. *Phys. Life Rev.* **29**, 120–153 (2019).
74. Lowder, M. W., Zhou, A. & Gordon, P. C. The lab discovered: Place-for-institution metonyms appearing in subject position are processed as agents. *J. Exp. Psychol. Learn. Mem. Cogn.* https://doi.org/10.1037/xlm0001314 (2023).
75. Lowder, M. W. & Gordon, P. C. Natural forces as agents: Reconceptualizing the animate–inanimate distinction. *Cognition* **136**, 85–90 (2015).
76. Schendan, H. E. & Kutas, M. Neurophysiological evidence for transfer appropriate processing of memory: Processing versus feature similarity. *Psychon. Bull. Rev.* **14**, 612–619 (2007).
77. Potts, G. F. An ERP index of task relevance evaluation of visual stimuli. *Brain Cogn.* **56**, 5–13 (2004).
78. Potts, G. F. & Tucker, D. M. Frontal evaluation and posterior representation in target detection. *Cogn. Brain Res.* **11**, 147–156 (2001).
79. Philipp, M., Bornkessel-Schlesewsky, I., Bisang, W. & Schlesewsky, M. The role of animacy in the real time comprehension of Mandarin Chinese: Evidence from auditory event-related brain potentials. *Brain Lang.* **105**, 112–133 (2008).
80. Kutas, M. & Federmeier, K. D. Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.* **62**, 621–647 (2011).
81. Friederici, A. D. The time course of syntactic activation during language processing: A model based on neuropsychological and neurophysiological data. *Brain Lang.* **50**, 259–281 (1995).
82. Sassenhagen, J., Schlesewsky, M. & Bornkessel-Schlesewsky, I. The P600-as-P3 hypothesis revisited: Single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. *Brain Lang.* **137**, 29–39 (2014).
83. Brouwer, H., Crocker, M. W., Venhuizen, N. J. & Hoeks, J. C. J. A neurocomputational model of the N400 and the P600 in language processing. *Cogn. Sci.* **41**, 1318–1352 (2017).
84. Brouwer, H. & Crocker, M. W. On the proper treatment of the N400 and P600 in language comprehension. *Front. Psychol.* **8**, 66 (2017).
85. Li, J. & Ettinger, A. Heuristic interpretation as rational inference: A computational model of the N400 and P600 in language processing. *Cognition* **233**, 105359 (2023).
86. Ryskin, R. *et al.* An ERP index of real-time error correction within a noisy-channel framework of human communication. *Neuropsychologia* **158**, 107855 (2021).
87. Bornkessel-Schlesewsky, I. & Schlesewsky, M. Toward a neurobiologically plausible model of language-related, negative event-related potentials. *Front. Psychol.* **10**, 1–17 (2019).
88. Baggio, G. & Hagoort, P. The balance between memory and unification in semantics: A dynamic account of the N400. *Lang. Cogn. Process.* **26**, 1338–1367 (2011).
89. Royle, P., Drury, J. E. & Steinhauer, K. ERPs and task effects in the auditory processing of gender agreement and semantics in French. *Ment. Lex.* **8**, 216–244 (2013).

90. Schacht, A., Sommer, W., Shmuilovich, O., Martínez, P. C. & Martín-Loeches, M. Differential task effects on N400 and P600 elicited by semantic and syntactic violations. *PLoS ONE* **9**, 1–7 (2014).
91. Deacon, D. & Shelley-Tremblay, J. How automatically is meaning accessed: A review of the effects of attention on semantic processing. *Front. Biosci.* **5**, e82 (2000).
92. Lau, E. F., Holcomb, P. J. & Kuperberg, G. R. Dissociating N400 effects of prediction from association in single-word contexts. *J. Cogn. Neurosci.* **25**, 484–502 (2013).
93. Levshina, N. *et al.* Why we need a gradient approach to word order. *Linguistics* **61**, 825–883 (2023).
94. MacWhinney, B. The competition model: Past and future. In *A Life in Cognition* (eds. Gervain, J., Csibra, G. & Kovács, K.) 3–16 (2022). https://doi.org/10.1007/978-3-030-66175-5_1
95. Tang, J., LeBel, A., Jain, S. & Huth, A. G. Semantic reconstruction of continuous language from non-invasive brain recordings. *bioRxiv* **26**, 66 (2022).
96. Cao, L., Huang, D., Zhang, Y., Jiang, X. & Chen, Y. Brain decoding using fNIRS. In *35th AAAI Conference Artificial Intelligence AAAI 2021*, vol. 14A 12602–12611 (2021).
97. Gueorguieva, R. & Krystal, J. H. Move over ANOVA. *Arch. Gen. Psychiatry* **61**, 310 (2004).
98. Yang, N., Waddington, G., Adams, R. & Han, J. Translation, cultural adaption, and test–retest reliability of Chinese versions of the Edinburgh Handedness Inventory and Waterloo Footedness Questionnaire. *Laterality Asymm. Body Brain Cogn.* **23**, 255–273 (2018).
99. Yang, J. Learners and users of English in China. *English Today* **22**, 3–10 (2006).
100. Yan, J. & Huizhong, Y. The English proficiency of college and university students in China: As reflected in the CET. *Lang. Cult. Curric.* **19**, 21–36 (2006).
101. Lemhöfer, K. & Broersma, M. Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behav. Res. Methods* **44**, 325–343 (2012).
102. Chan, I. L. & Chang, C. B. LEXTALE_CH: A quick, character-based proficiency test for Mandarin Chinese. *Proc. Annu. Bost. Univ. Conf. Lang. Dev.* **42**, 114–130 (2018).
103. Cai, Q. & Brysbaert, M. SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS ONE* **5**, e10729 (2010).
104. Her, O.-S. Argument-function mismatches in Mandarin resultatives: A lexical mapping account. *Lingua* **117**, 221–246 (2007).
105. Van Casteren, M. & Davis, M. H. Mix, a program for pseudorandomization. *Behav. Res. Methods* **38**, 584–589 (2006).
106. Winkler, I., Debener, S., Muller, K. R. & Tangermann, M. On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP. In *Proceedings of the Annual International Conference on IEEE Engineering in Medicine and Biology Society EMBS* 2015-Novem 4101–4105 (2015).
107. Steinhauer, K. & Drury, J. E. On the early left-anterior negativity (ELAN) in syntax studies. *Brain Lang.* **120**, 135–162 (2012).
108. R Core Team. *R: A Language and Environment for Statistical Computing* (2017).
109. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 66 (2015).
110. Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. lmerTest package: Tests in linear mixed effects models. *J. Stat. Softw.* **82**, 66 (2017).
111. Voeten, C. C. *buildmer: Stepwise Elimination and Term Reordering for Mixed-Effects Regression* (2021).
112. Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H. & Bates, D. Balancing Type I error and power in linear mixed models. *J. Mem. Lang.* **94**, 305–315 (2017).
113. Barr, D. J., Levy, R., Scheepers, C. & Tily, H. J. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J. Mem. Lang.* **68**, 255–278 (2013).
114. Russell, L. *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.4.2 (2019).
115. Loken, E. & Gelman, A. Measurement error and the replication crisis—The assumption that measurement error always reduces effect sizes is fals. *Science (80)* **355**, 582–584 (2017).
116. Lüdecke, D. *sjPlot: Data Visualization for Statistics in Social Science* (2021).
117. Wickham, H. *et al.* Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
118. Inkscape Project. *Inkscape* (2020).
119. Ratcliff, R. Methods for dealing with reaction time outliers. *Psychol. Bull.* **114**, 510–532 (1993).
120. Gelman, A. & Hill, J. *Data Analysis Using Regression and Multilevel/Hierarchical Models* (Cambridge University Press, 2006). https://doi.org/10.1017/CBO9780511790942
121. Lo, S. & Andrews, S. To transform or not to transform: using generalized linear mixed models to analyse reaction time data. *Front. Psychol.* **6**, 66 (2015).
122. Morís Fernández, L. & Vadillo, M. A. Flexibility in reaction time analysis: Many roads to a false positive?. *R. Soc. Open Sci.* **7**, 190831 (2020).
123. Cunnings, I. & Fujita, H. Quantifying individual differences in native and nonnative sentence processing. *Appl. Psycholinguist.* **42**, 579–599 (2021).
124. Kidd, E., Donnelly, S. & Christiansen, M. H. Individual differences in language acquisition and processing. *Trends Cogn. Sci.* **22**, 154–169 (2018).
125. Kaan, E. Event-related potentials and language processing: A brief overview. *Lang. Linguist. Compass* **1**, 571–591 (2007).
126. Herbay, A. *ERPscope* (2022).

Author contributions
M.W. and K.S. designed the experiment with input from S.B. and H.Z. K.S. provided the funding for participants and H.Z. provided the lab space and testing materials. M.W. collected the data, conducted the analysis, and wrote the manuscript. J.A. participated in coding and analyzing questionnaire and reaction time data and assisted in optimizing the protocol for preprocessing EEG data. All authors gave feedback on analysis and the manuscript.

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/s41598-024-70318-5.

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

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