Applied Psycholinguistics (2025), **46**, e51, 1–27
doi:10.1017/S0142716425100404

CAMBRIDGE UNIVERSITY PRESS

ORIGINAL ARTICLE

Comprehending semantic and syntactic anomalies in text attributed to an LLM versus a human: An ERP study

Xiaohui Rao[1], Hanlin Wu[1] and Zhenguang G. Cai[1,2]

[1]Department of Linguistics and Modern Languages, The Chinese University of Hong Kong, Hong Kong, China and [2]Brain and Mind Institute, The Chinese University of Hong Kong, Hong Kong, China
Corresponding author: Zhenguang G. Cai; Email: zhenguangcai@cuhk.edu.hk

(Received 6 June 2025; revised 11 October 2025; accepted 18 November 2025)

Abstract

As people increasingly interact with large language models (LLMs), a critical question emerges: do humans process language differently when communicating with an LLM versus another human? While there is good evidence that people adapt comprehension based on their expectations toward their interlocutor in human–human interaction, human–computer interaction research suggests the adaptation to machines is often suspended until expectation violation occurs. We conducted two event-related potential experiments examining Chinese sentence comprehension, measuring neural responses to semantic and syntactic anomalies attributed to an LLM or a human. Experiment 1 revealed reduced N400 but larger P600 responses to semantic anomalies in LLM-attributed text than human-attributed one, suggesting participants anticipated semantic errors yet required increased composition/integration efforts. Experiment 2 showed enhanced P600 responses to LLM-attributed than human-attributed syntactic anomalies, reflecting greater reanalysis or integration difficulty in the former than in the latter. Notably, neural responses to LLM-attributed semantic anomalies (but not syntactic anomalies) were further modulated by participants' belief about humanlike knowledge in LLMs, with a larger N400 and a smaller P600 in participants with stronger belief of humanlike knowledge in LLMs. These findings provide the first neurocognitive evidence that people develop mental models of LLM capabilities and adapt neural processing accordingly, offering theoretical insights aligned with multidisciplinary frameworks and practical implications for designing effective human–AI communication systems.

Keywords: ERP; human–computer interaction; language comprehension; large language models; semantics; syntax

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0142716425100404 Published online by Cambridge University Press

Introduction

Communication is extending beyond human interactions. Computers have long contributed to content generation and message exchange (Sundar & Lee, 2022); yet they have previously been constrained by limitations in world knowledge (Broussard, 2018) and linguistic competence (Kennedy et al., 1988; Yngve, 1967), which have hindered the potential of human–computer interaction (HCI). The advent of large language models (LLMs) represents a pivotal shift, allowing for authentic, natural language conversations with artificial intelligence (AI; Kim et al., 2024; Ni & Li, 2024). The popularity of interactions with LLMs is on the rise, with these models increasingly integrated into communicative agents such as virtual assistants (Hoy, 2018), companion robots (Irfan et al., 2023; Khoo et al., 2023), language teachers (Schmidt & Strassner, 2022), and therapists (Fiske et al., 2019). This shift highlights the importance of understanding how people perceive and interact with LLMs. While research has shown that people react differently to images and voices generated by LLMs versus those generated by humans (Cohn et al., 2019; Miller et al., 2023; Nightingale & Farid, 2022; Zellou et al., 2021), whether they also differentially comprehend language from LLMs versus humans remains under-investigated.

Long before the rise of LLMs, the *Computers Are Social Actors* framework postulates that people increasingly recognized computers as "social actors" (Nass & Moon, 2000). In HCI, people treat computers similarly to their human peers by applying expectations, such as demographic stereotypes (Nass et al., 1997, 2000), and social norms including politeness (Kopp et al., 2005; Nass et al., 1999), reciprocity (Fogg & Nass, 1997; Moon, 2000), and cooperation (Krämer, 2005; Parise et al., 1999). Although computers are understood to not to be human, these machines can still elicit in their human interlocutors communicative behaviors and emotional responses typically observed in human–human interaction. The reasons for these seemingly irrational responses—treating inanimate machines as if they possess gender, personality, feelings, and intentions—remain unclear. One explanation is mindlessness (Lee, 2008, 2010; Nass & Moon, 2000; Sundar & Nass, 2000): in the context of HCI, people often do not engage in deep, effortful processing regarding the computer's nonhuman nature and instead fall back on automatic social reactions learned from human–human interaction.

While the literature on human language comprehension demonstrates that listeners attend to the source of a message, using speaker identity and context to inform their expectations and pragmatic inferences (e.g., Noveck, 2018; Rueschemeyer et al., 2015; van Berkum et al., 2008), the mindlessness default in HCI suggests that people may not pay attention to the nonhuman source of language unless they have a specific reason to do so, particularly when a message somewhat contradicts their expectations (Lee, 2024). This contradiction leads to a reassessment of the source, influenced by their beliefs (also known as machine heuristics) about what machines are like and how they perform (Molina & Sundar, 2022; Sundar & Kim, 2019). Previous research in HCI has shown that these beliefs affect how people respond to machine agents compared to human counterparts (Lee, 2024; Liu & Wei, 2018; Sundar & Kim, 2019; Wang, 2021). To maximize the

https://doi.org/10.1017/S0142716425100404 Published online by Cambridge University Press

potential for machine heuristics to modulate language comprehension, we utilized semantic and syntactic anomalies. This approach is grounded in two theoretical lines: first, in human studies, anomalies elicit robust event-related potential (ERP) components for comparison; and second, in HCI theory, anomalies serve as the expectation violation predicted to shift participants from the default "mindless" state to an effortful use of source-specific beliefs.

How do people process computer-attributed language versus human language?

Research observed differences between HCI and human–human interaction, particularly regarding linguistic features and the complexity of interactions. People use shorter sentences with less diverse lexical expressions when addressing computers than human peers (Amalberti et al., 1993; Hara & Iqbal, 2015; Kennedy et al., 1988; Shechtman & Horowitz, 2003). Such differences suggest that people might perceive computers as linguistically less competent than humans, thus favoring clarity and simplicity over richness and complexity (e.g., Branigan et al., 2011). Additionally, people exhibited less politeness toward robots (Kanda et al., 2008) and reported feeling less social stress in the presence of computers (Vollmer et al., 2018).

Studies have further shown that, in HCI, people adapt their language use in light of their beliefs about computer systems. People are more likely to re-use their interlocutor's prior lexical expressions when addressing computer interlocutors than human interlocutors (Bergmann et al., 2015; Branigan et al., 2004, 2011; Shen & Wang, 2023). They also more often re-use a computer's lexical expressions when communicating with a more basic than more advanced computer, presumably because doing so can help to maximize the likelihood of being understood by a linguistically limited computer (Branigan et al., 2011; Pearson et al., 2006). It is worth noting that people's tendency to mimic computer-attributed sentence structures has decreased over time. Earlier studies found that people sometimes repeat the sentence structures used by computers more often than those by humans (Branigan et al., 2003). However, more recent research has revealed no such differences (Cowan et al., 2015) and even suggested a reversal of this trend (Heyselaar et al., 2017). These findings indicate a possible decline in aligning syntactic structures with computers as technology has advanced, perhaps reflecting a growing perception of computers as more competent than in the past. The findings that people use their knowledge about their interlocutor to adapt their language production is consistent with the theory that language communication is goal-oriented, with interlocutors comprehending and producing language in a way to maximize communicative success (e.g., Cai et al., 2017; Goodman & Frank, 2016). For example, people construct a model of important demographics of their interlocutor and use such a model to adjust their language output (Cai et al., 2021) and constrain their language understanding (Cai, 2022; Wu, Duan, & Cai, 2024), a process known as interlocutor modeling. Extended to HCI, this account suggests that, when interacting with computers, people may modify their language based on their beliefs about computers; in particular, they may accordingly adjust language comprehension and production if they believe that computers tend to be more limited in their linguistic capabilities. Whether

https://doi.org/10.1017/S0142716425100404 Published online by Cambridge University Press

and how people also consider the properties of LLMs in human–AI interaction remains an open question.

Do people treat LLMs differently from their human peers?

Previously, computers were regarded as distinct from humans, resulting in people treating them differently while still broadly recognizing them as social actors. Recent developments in generative AI have resulted in AI agents that are "more intimate, proximate, embodied, and humanlike" (Jones-Jang & Park, 2023, p. 1). It is thus interesting to see whether people still distinguish between LLMs and humans.

Research on face science revealed that AI-synthesized faces are indistinguishable from real faces and more trustworthy (Nightingale & Farid, 2022). Cabrero-Daniel and Sanagustín Cabrero (2023) showed that, while most participants acknowledged that LLM-attributed text exhibits gender bias, half of them argued that these AI systems perform as well as or even better than humans in mitigating gender bias. In addition, there is also research showing that LLM-generated narratives could reduce hostile media bias by engaging the machine heuristic, a mental shortcut that causes people to perceive machines as more impartial, systematic, and accurate than humans (Cloudy et al., 2023). People also have stereotypical expectations regarding LLMs' consistent performance (i.e., AI systems will perform tasks accurately and reliably every time), leading to frustration when these expectations are unmet. For instance, people often exhibit heightened sensitivity to LLMs' inconsistencies, resulting in more punitive evaluations of the fairness of their decisions (Jones-Jang & Park, 2023). Additionally, people are still skeptical about LLMs' ability to fully understand human language and to align with human values, especially on complex social issues (Wu, Xu, et al., 2024).

Beyond general attitudes, emotional responses to LLMs differ from those to humans. Yin et al. (2024) showed that, after expressing their feelings, participants reported being heard more often if they received responses generated by LLMs than by humans (with response source unknown to participants); however, when the source of response was known, participants reported being heard more often with human responses than LLM-generated responses. These findings suggest that people may perceive LLMs emotionally differently from human interlocutors, even when LLMs can generate well-written responses. Nevertheless, most relevant studies rely on qualitative data, such as ratings, with few directly comparing the real-time mechanisms underlying human–human interaction and human–AI interaction from a neurocognitive perspective. Furthermore, research often overlooks the role of language in human–AI interaction, limiting our understanding of how people comprehend LLM-attributed compared to human-attributed language.

Neural responses to semantic and syntactic anomalies in LLM-attributed text

In this study, we examined people's brain responses to language attributed to an LLM versus to a human peer, especially when the message or the syntax of the language deviates from people's expectation (i.e., when sentences contain semantic and syntactic anomalies). This represents a novel area of investigation with limited

https://doi.org/10.1017/S0142716425100404 Published online by Cambridge University Press

prior research to guide specific predictions about how people's neural responses might differ when processing LLM-attributed versus human-attributed anomalies.

Extensive research has employed ERPs to investigate how people respond to semantic and syntactic anomalies in human language. In particular, a semantic anomaly evokes an N400 effect in comprehenders, a negative component peaking around 400 ms after encountering the problematic word. For instance, at reading a sentence such as *The Dutch trains are yellow/sour and very crowded*, comprehenders show an N400 effect with the word *sour* (which results in a semantic anomaly) compared to *yellow*, indicating more cognitive efforts in integrating the meaning of *sour* into the context (Federmeier & Kutas, 1999; Hagoort et al., 2004; Kutas & Hillyard, 1980; van Berkum et al., 1999). A syntactic anomaly, in contrast, evokes a P600 effect, a positive shift starting at around 600 ms after presentation of a syntax-violating word (Hagoort & Brown, 2000; Molinaro et al., 2011). For instance, in reading *The spoilt child throws/throw the toys on the floor*, there is a P600 effect at *throw* compared to *throws*, indicating more cognitive efforts in re-analyzing or repairing the sentence's syntactic structure (Hagoort & Brown, 2000).

Importantly, these core neural signatures of language processing are modulated by the comprehender's mental model of their interlocutor. This model encompasses beliefs and expectations regarding the interlocutor's sociodemographic attributes (van Berkum et al., 2008), lexical choices (Ryskin et al., 2020), syntactic preferences (Xu et al., 2022), communicative styles (Regel et al., 2010), social connections (Barr et al., 2014), and reliability (Brothers et al., 2019). Such modulation results from the integration of linguistic and interlocutor context, which collectively shapes language expectations and interpretation (Hanulíková & Carreiras, 2015; van Berkum et al., 2008; Wu & Cai, 2024, 2025; Wu, Duan, & Cai, 2024; Wu et al., 2025). For instance, van Berkum et al. (2008) showed that comprehenders use their demographic knowledge of the speaker (such as gender, age, and socioeconomic backgrounds) to interpret their sentences; they showed an N400 effect at hearing the sentence *If only I looked like Britney Spears* spoken in a male voice relative to a female voice. Wu, Duan & Cai (2024) further showed that the perceived linguistic capacity of the speaker modulates listeners' neural signatures in speech comprehension. They found an N400 effect in response to label switching (i.e., when a speaker deviates from an earlier linguistic label in referring to the same thing); more interestingly, they showed that the label switching effect, reflected as N400, was larger when the speaker was a child than an adult, suggesting that people expect children to be linguistically less flexible than adults, thus feeling more "surprised" at label switching by children. Collectively, these findings show that lexical-semantic processing, indexed by the N400, is modulated by the interlocutor context (see also Bornkessel-Schlesewsky et al., 2013; Brothers et al., 2019; Foucart et al., 2019; Foucart & Hartsuiker, 2021; Grant et al., 2021). While the N400 may reflect low-level processes like semantic retrieval or integration (Baggio & Hagoort, 2011; Kutas & Federmeier, 2011), extensive research has established that this component is sensitive not just to local word-level features (like frequency or association) but also to global contextual constraints. Higher-level information, like the interlocutor context, often overrides the low-level processes (Federmeier & Kutas, 1999; Kutas & Federmeier, 2011; Kutas & Hillyard, 1980). The interlocutor context is also known

https://doi.org/10.1017/S0142716425100404 Published online by Cambridge University Press

to modulate the P600 effect associated with both syntactic reanalysis (Caffarra et al., 2020; Hanulíková et al., 2012; Hanulíková & Carreiras, 2015; Zhou et al., 2019) and composition/integration processes (Foucart et al., 2019; Regel et al., 2010). When the interlocutor is an LLM, we assume that people's heuristic expectations of LLMs also modulate these neural signatures.

The present study

The present study aims to use ERPs (mainly N400 and P600) to investigate whether people differentially comprehend semantic and syntactic anomalies in text attributed to an LLM versus a human. We report two experiments in Mandarin Chinese examining these questions. In Experiment 1, we examined whether participants exhibited different N400 effects when reading semantically anomalous sentences (relative to semantically coherent counterparts) as LLM-attributed versus human-attributed. In Experiment 2, we examined whether perceived LLM- and human-attributed syntactic anomalies result in different P600 effects. Previous studies have demonstrated that people's beliefs about machine agents' properties can modulate how people respond to them (Lee, 2024; Liu & Wei, 2018; Sundar & Kim, 2019; Wang, 2021). Based on this, we predicted that people would show different ERP responses to language anomalies when attributed to LLMs versus humans. However, the specific direction of these differences remains exploratory, given the novel context of LLM-attributed language processing. Additionally, previous research has shown that the way people use language with a computer is influenced by the perceived level of "knowledge" the computer possesses (e.g., participants more often lexically aligned with a basic computer than with an advanced computer; Branigan et al., 2011). Hence, in both experiments, we also measured participants' beliefs about the extent to which LLMs possess humanlike knowledge and explored whether these beliefs modulate neural responses to LLM-attributed anomalies.

Experiment 1

We adopted a factorial design of 2 (Anomaly: anomalous vs. control, manipulated within participants and within items) × 2 (Attribution: LLM vs. human, manipulated between participants and within items). Participants read Chinese sentences, one phrase at a time at a rate of 300 ms, that were either semantically anomalous or their control versions (see Table 1); they then decided whether a sentence was sensible or not. In addition, we further attributed sentences to either an LLM or a human being by telling participants, at the start of the experiment, that the sentences they were to read had been produced by either an LLM or a human being.

Method

Participants

Sixty-four neurologically healthy native speakers of Mandarin Chinese (32 males, mean age = 23 years; $SD = 1.98$) took part in the experiment. Six of them were later excluded from data analyses (see below), leaving 58 participants. All participants gave

https://doi.org/10.1017/S0142716425100404 Published online by Cambridge University Press

Table 1. An example of the Chinese sentences used in Experiment 1, with glossary and translation. The bolded phrases were the critical region. The spaces in the Chinese text indicate the segmentation of phrases that were presented one at a time during the experiment

小玲	刚	调好	**闹钟/月亮**	就 立马	睡觉 了。
Xiǎolíng	gāng	tiáohǎo	**nàozhōng/yuèliàng**	jiù lìmǎ	shuìjiào le.
Xiaoling	just	set	**clock/moon**	immediately	sleep.
After setting the clock/moon, Xiaoling immediately went to sleep.					

their written consent before participating in the experiment. The research methodology and procedures were approved by the Joint Chinese University of Hong Kong-New Territories East Cluster Clinical Research Ethics Committee.

Materials

We created 80 target sentences in Mandarin Chinese, each with a semantically coherent version and a semantically anomalous version (see Table 1), supplemented by 80 fillers (40 containing semantic anomalies and 40 without). The anomalous version was created by replacing the verb-following noun phrase with another noun phrase that is semantically incompatible with the verb (e.g., changing *set the clock* into *set the moon*; see Table 1). Both the verb and the critical noun phrases always consisted of two characters. As the anomaly-causing critical noun phrase normally appeared around the middle of the sentence, to avoid attracting participants' attention to this error location, following Zhu et al. (2012), we also included 40 filler sentences with semantic anomalies caused by a word located in various other positions within the sentences. Also, in filler sentences, there were other types of semantic anomalies (e.g., in the prepositional phrase, the adverbial phrase, the adjective phrase) to prevent participants from developing reading strategies. We created two lists of stimuli, each containing 40 semantically coherent sentences, 40 semantically anomalous sentences, and 80 fillers, according to Latin Square Design, such that each experimental sentence only appeared once in a list.

Procedure

Participants were seated in a soundproof booth specially built for EEG experiments. After giving their written consent, half of the participants were assigned to the LLM condition and the other half to the human condition. In the LLM condition, participants were told that they were to read sentences that had been produced by an LLM. To enhance their belief that they would be reading LLM-attributed text, prior to the experiment, participants were asked to first have an interaction in Mandarin Chinese via typing for about three minutes with the LLM (GPT-4). The live interaction was conducted in an in-house interface connected to the API of GPT-4. We prompted the LLM to produce short sentences (without any type of errors) during the interaction to resemble the experimental materials that we used. Participants in the human condition were told that they would be reading sentences that other participants had previously produced for a corpus of human language

https://doi.org/10.1017/S0142716425100404 Published online by Cambridge University Press

that we were building. The experiment was implemented using E-Prime 3.0 (Psychology Software Tools, Pittsburgh, PA). A trial began with a visual fixation mark for 500 ms, followed by a sentence presented phrase by phrase. Each phrase remained at the center of the screen for 300 ms, followed by a blank screen for 300 ms. All participants were asked to read the sentence carefully and decide whether the sentence was semantically sensible by pressing a key (*F* or *J* on the keyboard) within three seconds. The pairing between the two keys and participant responses (semantically sensible or not) was counterbalanced across participants. The next sentence was presented after a response was made or after the time window expired. As earlier studies have shown that people would adjust their language use depending on their perceived knowledge of computer systems (e.g., Branigan et al., 2011), at the end of the experiment, we asked participants to rate the extent to which they perceived LLMs to resemble humans in their knowledge on a 5-point scale (1 = "not similar at all," 5 = "very similar").

Data recording and preprocessing

Six participants were excluded from the data analyses either due to an excessively high error rate (over 30%) or more than 30% of the trials being artifact-contaminated, leaving 58 participants for subsequent data analyses (with 29 participants in the LLM and human condition, respectively). The EEG signals were recorded using 128 active sintered Ag/AgCl electrodes positioned according to an extended 10–20 system. All electrodes were referred online to the left earlobe. Signals were recorded using a g.HIamp amplifier and digitalized at a sampling rate of 1200 Hz. All electrodes were kept below the impedance of 30 kΩ throughout the task. The EEG data were preprocessed offline using customized scripts and the FieldTrip toolbox (Oostenveld et al., 2011) in MATLAB. The raw data were bandpass-filtered between 0.1 and 30 Hz (Luck, 2014; Tanner et al., 2015) and then re-referenced to the left and right earlobes (A1 and A2). For our high-density EEG data, Principal Component Analysis was first applied to the 1–30 Hz filtered continuous data (Luck, 2022) to reduce dimensionality to 30 components. Subsequently, Independent Component Analysis was used to identify and remove ocular artifacts. This two-stage approach is common practice for high-density data and artifact correction (Ghandeharion & Erfanian, 2010; Winkler et al., 2011; Wu & Cai; 2025). Then, the data were epoched from 200 ms before the stimulus onset to 1000 ms after the onset. Trials with inaccurate behavioral responses were excluded (4.07%). The epochs containing amplitudes exceeding ±150 μV were excluded as artifacts (3.88%). Finally, the remaining epochs were baseline-corrected by subtracting the mean amplitude from 200 ms to 0 ms prior to the critical word onset. Seventy-five electrodes in the central-parietal region (FCC3h, PPO1h, PPO2h, P1, P2, PPO3h, PPO4h, CPz, Pz, Cz, CCP1h, CCP2h, CPP1h, CPP2h, C1, C2, CP2, CP1, CCP3h, CCP4h, CPP3h, CPP4h, CP3, CP4, T7, C5, C3, C4, C6, T8, TP7, CP5, CP6, TP8, P7, P5, P3, P4, P6, P8, PO7, PO3, POz, PO4, PO8, O1, Oz, O2, FTT7h, FCC5h, FCC1h, FCC2h, FCC4h, FCC6h, FTT8h, TTP7h, CCP5h, CCP6h, TTP8h, TPP7h, CPP5h, CPP6h, TPP8h, PPO7h, PPO5h, PPO6h, PPO8h, POO5, POO1, POO2, POO6, POO9h, OI1h, OI2h, POO10h) were selected for statistical analysis, covering central and parietal scalp regions related to the observation of the N400 component in previous studies (Hahne & Friederici, 2002; Kutas & Van Petten, 1994).

https://doi.org/10.1017/S0142716425100404 Published online by Cambridge University Press

Table 2. LME models for by-trial amplitude analysis across the centro-parietal sites

Fixed effects	*β*	*SE*	*t*	*p*
N400 (300–500 ms)				
Intercept	−0.19	0.24	−0.80	0.429
Anomaly	**−0.85**	**0.03**	**−29.04**	**<.001**
Attribution	0.24	0.41	0.60	0.553
Anomaly: Attribution	**0.66**	**0.06**	**11.25**	**<.001**
P600 (600–800 ms)				
Intercept	**0.76**	**0.24**	**3.23**	**0.002**
Anomaly	**1.36**	**0.04**	**38.76**	**<.001**
Attribution	0.27	0.35	0.76	0.449
Anomaly: Attribution	**0.92**	**0.44**	**2.09**	**0.040**

Notes: The model for N400 analysis: Amplitude ~ Anomaly * Attribution + (1|Participant) + (1|Item) + (Attribution + 1|Channel); the model for P600 analysis: Amplitude ~ Anomaly * Attribution + (1|Participant) + (Anomaly : Attribution +1|Item) + (Anomaly : Attribution + Attribution + 1|Channel). Significant and marginally significant results are marked bold.

We preserved the data of each channel (instead of collapsing the mean amplitudes within the ROI) for each trial to conduct channel-level analyses. This allowed us to include Channel as a random effect in the subsequent linear mixed-effects (LME) models, following the methodology of recent studies (Lenzoni et al., 2024; Ryskin et al., 2021; Volpert-Esmond et al., 2018; Wu & Cai, 2025).

Results and discussion

We focused on the classic time window of the N400 effect, namely, 300–500 ms (see also van Berkum et al., 2005). We fit LME model to the mean amplitudes extracted over the target time windows for each channel of each trial, with Anomaly (control = −0.5, anomalous = 0.5), Attribution (human = −0.5, LLM = 0.5), and the interaction between them as fixed effects. We included Participant, Item, and Channel as random effects (See Table 2 for the full model structure). For all LME analyses, we used the maximal random-effect structure justified by the data and determined through forward model comparison ($\alpha = 0.2$; Matuschek et al., 2017).

Table 2 presents the LME results for by-trial amplitude analysis in the time window of 300–500 ms. There was a significant main effect of Anomaly, with more negative amplitudes in the semantically anomalous condition ($M = -0.63$ μV) compared to the semantically coherent condition ($M = 0.23$ μV) during the time window of 300–500 ms (see also Figure 1). The interaction between Anomaly and Attribution was significant. Simple-effect analysis showed that in the LLM condition, the amplitudes were significantly more negative for semantically anomalous sentences ($M = -0.35$ μV) than for semantically coherent sentences ($M = 0.22$ μV, $\beta = -0.49$, $t = -11.73$, $p < .001$). In the human condition, the amplitudes were significantly more negative for semantically anomalous sentences ($M = -0.90$ μV) than for semantically coherent sentences ($M = 0.23$ μV,

https://doi.org/10.1017/S0142716425100404 Published online by Cambridge University Press

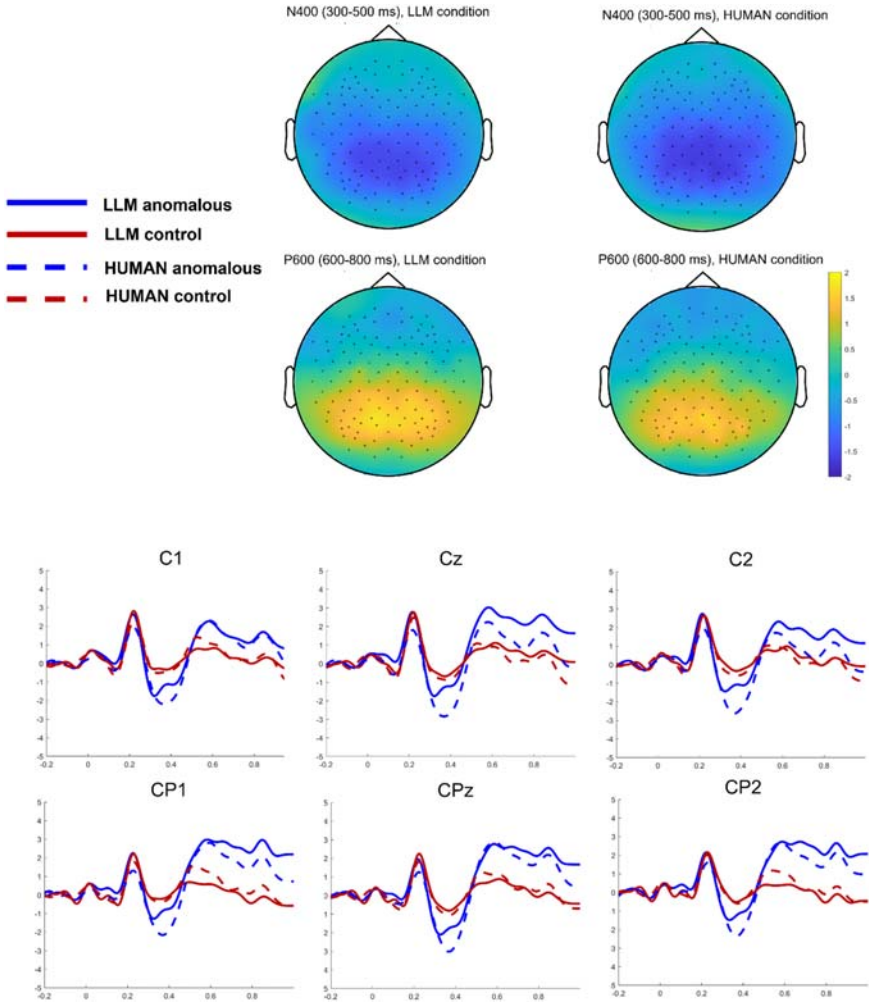


Figure 1. Grand average ERPs elicited by the semantically anomalous sentences and the semantically coherent sentences in the LLM and human conditions. Waveforms are illustrated with six centro-parietal sites C1, Cz, C2, CP1, CPz, and CP2.

$\beta = -1.16$, $t = -28.77$, $p < .001$). These results suggest that the N400 effect was weaker in the LLM condition (−0.57 μV) than in the human condition (−1.13 μV).

Additionally, our analysis revealed an unexpected pattern in the 600–800 ms time window. As shown in Table 2 and Figure 1, there was a significant main effect of Anomaly, with more positive deflections in the semantically anomalous condition ($M = 1.43$ μV) than in the semantically coherent condition ($M = 0.09$ μV), suggesting the presence of a P600 component in response to semantic anomalies. The interaction between Anomaly and Attribution was significant. Simple-effect analysis showed that in the LLM condition, the amplitudes were significantly more positive for semantically anomalous sentences ($M = 1.77$ μV) than for semantically

https://doi.org/10.1017/S0142716425100404 Published online by Cambridge University Press

Table 3. LME models for by-trial amplitude analysis of knowledge resemblance in the LLM condition

Fixed effects	*β*	*SE*	*t*	*p*
N400 (300–500 ms)				
Intercept	−0.03	0.32	−0.10	0.925
Anomaly	**−0.42**	**0.04**	**−10.14**	**<.001**
Knowledge Resemblance	−0.52	0.28	−1.82	0.080
Anomaly: Knowledge Resemblance	**−0.80**	**0.04**	**−18.49**	**<.001**
N600 (600–800 ms)				
Intercept	**0.93**	**0.31**	**2.96**	**0.005**
Anomaly	**1.87**	**0.05**	**36.82**	**<.001**
Knowledge Resemblance	−0.31	0.26	−1.20	0.242
Anomaly: Knowledge Resemblance	**−0.64**	**0.05**	**−12.11**	**<.001**

Notes: The model for N400 analysis in the LLM condition: Amplitude ~ Anomaly * Knowledge Resemblance + (1|Participant) + (1|Item) + (1|Channel); the model for P600 analysis: Amplitude ~ Anomaly * Knowledge Resemblance + (1|Participant) + (1|Item) + (1|Channel). Significant and marginally significant results are marked bold.

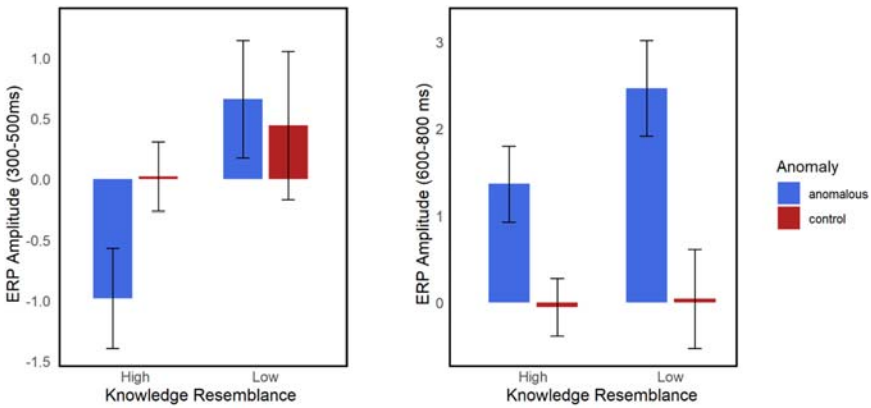


Figure 2. Mean ERP Amplitude in the LLM condition for high versus low knowledge resemblance across semantically anomalous and control conditions.

coherent sentences ($M = 0.02$ μV, $\beta = 1.82$, $t = 35.91$, $p < .001$). In the human condition, the amplitudes were significantly more positive for semantically anomalous sentences ($M = 1.09$ μV) than for semantically coherent sentences ($M = 0.15$ μV, $\beta = 0.91$, $t = 18.55$, $p < .001$). These results indicate that the P600 effect was larger in the LLM condition (1.75 μV) than in the human condition (0.94 μV).

As previous research has shown that the way people use language with a computer is influenced by the perceived level of "knowledge" the computer possesses (e.g., Branigan et al., 2011), we fit LME models on data from the LLM condition with Anomaly and Knowledge Resemblance as interacting fixed effects (see Table 3 for model structures) to investigate whether the perceived resemblance

https://doi.org/10.1017/S0142716425100404 Published online by Cambridge University Press

in knowledge between LLMs and humans could predict the N400 effect and the P600 effect in the LLM condition. During 300–500 ms, the results revealed a significant interaction between Anomaly and Knowledge Resemblance, such that higher perceived knowledge resemblance between LLMs and humans predicted a larger N400 effect when people believed they were reading semantic anomalies from an LLM (see Figure 2). Additionally, during 600–800 ms, the interaction between Anomaly and Knowledge Resemblance was significant, such that higher perceived knowledge resemblance between LLMs and humans predicted a smaller P600 effect when people believed they were reading semantic anomalies from an LLM (see Figure 2).

Experiment 1 showed that semantically anomalous sentences, relative to semantically coherent ones, elicited an N400 effect. Notably, the N400 effect was reduced when the semantic anomaly was believed to be produced by an LLM compared to by a human being, which suggests that participants seemed to be more accommodating with LLM-attributed semantic anomalies than with human-attributed ones. The difference in the N400 effects largely reflects the modulation of participants' expectations about the attributed interlocutor (LLM vs. human). While our stimuli used strong semantic incoherence, which differs from prototypical LLM hallucinations (i.e., coherent but factually false propositions), we interpret the reduced N400 amplitude as stemming from a broader expectation about the source rather than a direct expectation of this specific type of failure. This modulation may be driven by two related factors. First, the widely publicized factual unreliability of LLMs, of which hallucination is a prime example (Maynez et al., 2020; Rawte et al., 2023), may prepare people to anticipate a wider range of unpredictable semantic outputs. Second, participants are likely to have some understanding that LLMs, unlike humans, do not possess genuine world knowledge. This general awareness, gained through media coverage or personal experience with AI limitations, could lead them to view any semantic failure, including the strong incoherence in our materials, as a more predictable outcome of the system. Consequently, LLM-attributed semantic anomalies become more predictable and are thus better tolerated. This is further supported by the observation that the N400 effect to LLM-attributed semantic anomalies increased as a function of the extent of perceived humanlike knowledge of LLMs. This suggests that LLM-attributed semantic anomalies are less predictable or worse tolerated by participants if they believed to a greater extent that LLMs have humanlike knowledge.

Moreover, semantically anomalous sentences elicited a follow-up P600, compared to semantically coherent ones. Nevertheless, the P600 effect to LLM-attributed semantic anomalies was larger compared to that to human-attributed anomalies. Recent research has linked the P600 effect to composition or integration processes (Brouwer et al., 2012; Baggio, 2018, 2021; Fritz & Baggio, 2020, 2022). Baggio (2018) proposes that the P600 reflects formal operations at the syntax–semantics interface, which can be driven by two distinct mechanisms. First, it can represent a top-down influence where a compelling semantic interpretation affects the perception of syntactic form. Alternatively, it can index bottom-up, syntax-driven meaning composition when a rich semantic context is absent, thus forcing the system to engage in resource-intensive structural computation. Therefore, this framework predicts that the amplitude of the P600 component will be modulated by

https://doi.org/10.1017/S0142716425100404 Published online by Cambridge University Press

Table 4. An example of the Chinese sentences used in Experiment 2, with glossary and translation. The bolded phrase was the critical region. The spaces in the Chinese text indicate the segmentation of phrases that were presented one at a time during the experiment

昨天/明天,	小明	**吃过**	汉堡 了,	今天	就 不 了。
zuótiān/míngtiān	xiǎomíng	chī-guò	hànbǎo le,	jīntiān	jiù bù le.
Yesterday/tomorrow,	Xiaoming	eat-guo	hamburger,	today	then not.
Yesterday/tomorrow, Xiaoming ate a hamburger and he will not eat one today.					

the difficulty of composition/integration. Our results thus indicate that LLM-attributed semantic anomalies were harder to compose or integrate compared to human-attributed anomalies. Moreover, our results showed the P600 effect to LLM-attributed semantic anomalies decreased as a function of the extent of humanlike knowledge participants attributed to LLMs. This suggests that LLM-attributed semantic anomalies are easier to compose or integrate if participants believed to a greater extent that LLMs have humanlike knowledge. The complementary pattern of larger N400 and smaller P600 effects in participants with stronger beliefs in LLM humanlike knowledge demonstrates that people attributing more humanlike knowledge to LLMs employ cognitive mechanisms more similar to human–human communication when processing LLM semantic anomalies.

Experiment 2

We switched to syntactic anomalies, with a similar design as in Experiment 1: 2 (Anomaly: anomalous vs. control, manipulated within participants and within items) × 2 (Attribution: LLM vs. human, manipulated between participants and within items). In particular, we manipulated whether a sentence is syntactically anomalous or well-formed (in the control condition).

Method

Participants

Another 64 neurologically healthy native speakers of Mandarin Chinese (27 males, mean age = 24 years; *SD* = 2.50), who had not taken part in Experiment 1, participated in the experiment. Eight of them were later excluded from data analyses (see below), leaving 56 participants. Participants gave their written consent before participating in the experiment.

Materials

We created 80 target sentences in Mandarin Chinese, each with a syntactically well-formed version and a syntactically anomalous version (see Table 4), following the previous research (Hao et al., 2024; Qiu & Zhou, 2012). Mandarin Chinese has a limited range of grammatical inflections compared to Indo-European languages, lacking grammatical morphology that denotes syntactic categories or features (Zhang et al., 2010). Nevertheless, people can determine grammatical relations

https://doi.org/10.1017/S0142716425100404 Published online by Cambridge University Press

among sentence constituents through nuanced interactions of lexical semantics, grammatical particles, and discourse principles. In terms of tense, Chinese verbs are not morphologically marked but can be paired with temporal adverbs and a few aspectual particles to convey temporal references (Dai, 1997). We made use of the syntactic agreement between the temporal phrase and the aspectual particle *guo* (indicating a discontinued prior experience). In simplex sentences (e.g., Table 4), *guo* typically indicates completed past events, functioning similarly to a past tense marker (Lin, 2003). When an event (indicated by the temporal phrase) occurs in the past, then one can use the verb with the aspectual particle *guo* (e.g., *zuótiān, xiǎomíng chī-guò hànbǎo le, jīntiān jiù bù le*, meaning "Yesterday, Xiaoming **ate** a hamburger and he will not eat one today"), but if the event is yet to occur, then the aspectual particle cannot be used with the verb (e.g., *míngtiān, xiǎomíng chī-guò hànbǎo le, jīntiān jiù bù le*, meaning "Tomorrow, Xiaoming **ate** a hamburger and he will not eat one today"). The P600 effect observed by Qiu and Zhou (2012) and Hao et al. (2024) using this paradigm confirms that this temporal disagreement constitutes a syntactic violation. To ensure economical and incremental parsing (Frazier & Fodor, 1978; Konieczny, 2000), our sentences followed a high-frequency sequence: temporal phrase + subject + verb-particle structure. This word order allows immediate resolution to a simplex main clause upon encountering the subject and verb, minimizing sustained local ambiguity and isolating processing costs to the temporal agreement violation. We created 80 syntactically well-formed sentences expressing an event in the past (as indicated by the use of temporal phrases such as *zuótiān*, meaning "yesterday"); then, we modified these sentences into syntactically anomalous ones by changing the temporal phrase into a future one (e.g., *míngtiān* meaning "tomorrow"). To prevent participants from relying on the initial temporal phrase as a deterministic cue for the *guo* violation, we implemented a balanced filler design similar to Liang et al. (2022). Specifically, we included 80 filler sentences (40 syntactically well-formed, 40 with different nontemporal syntactic anomalies such as improper word order, missing components, or incorrect demonstratives). No fillers contained the target temporal agreement violation. Crucially, to decouple the initial temporal phrase from the target syntactic violation, these 80 fillers were balanced to begin with future-tense, present-tense, and past-tense temporal phrases. This balanced filler design ensured that a future-tense phrase had no statistical relationship to the subsequent *guo* violation, thereby requiring participants to process the full sentence structure for grammaticality. We created two lists of stimuli, each containing 40 syntactically well-formed sentences, 40 syntactically anomalous sentences, and 80 fillers, according to Latin Square Design, such that each experimental sentence only appeared once in a list. The presentation of stimuli was the same as Experiment 1.

Procedure

Participants were asked to read the sentence carefully and decide whether the sentence was syntactically well-formed within three seconds. Other procedures were the same as Experiment 1.

https://doi.org/10.1017/S0142716425100404 Published online by Cambridge University Press

Table 5. LME models for amplitude analysis across the centro-parietal sites

Fixed effects	*β*	*SE*	*t*	*p*
P600 (600–800 ms)				
Intercept	**0.49**	**0.20**	**2.50**	**0.013**
Anomaly	0.40	0.26	1.58	0.120
Attribution	−0.20	0.28	−0.71	0.481
Anomaly: Attribution	**0.91**	**0.51**	**1.79**	**0.080**

Notes: The model for P600 analysis: Amplitude ~ Anomaly * Attribution + (Anomaly + 1|Participant) + (1|Item) + (Attribution + 1|Channel). Significant and marginally significant results are marked bold.

Data recording and preprocessing

Eight participants were excluded from the data analyses either due to an excessively high error rate (over 37.5%) or more than 30 trials being artifact-contaminated, comprising approximately 37.5% of the total trials, leaving 56 participants for subsequent data analyses (with 29 in the LLM condition and 27 in the human condition). Data recording and preprocessing were the same as Experiment 1. Trials with inaccurate behavioral responses were excluded (9.55%). The epochs containing amplitudes exceeding ±150 μV were excluded as artifacts (6.06%). We selected 75 electrodes in the central-parietal region, consistent with Experiment 1, for statistical analysis, covering central and parietal scalp regions related to the observation of the P600 component in previous studies (e.g., Hagoort & Brown, 2000; Hao et al., 2024; Hinojosa et al., 2003; Qiu & Zhou, 2012).

Results and discussion

We focused on the classic time window of the P600 effect, namely, 600–800 ms (see also Nevins et al., 2007; Vissers et al., 2010). We fit LME models to the mean amplitudes extracted over the target time windows for each channel of each trial, with Anomaly (control = −0.5, anomalous = 0.5), Attribution (human = −0.5, LLM = 0.5), and the interaction between them as fixed effects. We included Participant, Item, and Channel as random effects (see Table 5 for the full model structures).

Table 5 presents the LME models for by-trial amplitude analysis during 600–800 ms. There was a marginally significant interaction between Anomaly and Attribution (see also Figure 3). Simple-effect analysis showed that in the LLM condition, the amplitudes were significantly more positive for syntactically anomalous sentences ($M = 0.88$ μV) than for syntactically well-formed sentences ($M = 0.01$ μV, $\beta = 0.97$, $t = 2.41$, $p = .023$). In the human condition, the amplitudes were comparable ($\beta = -0.08$, $t = -0.15$, $p = .884$) for syntactically anomalous sentences ($M = 0.61$ μV) and syntactically well-formed sentences ($M = 0.57$ μV).

Moreover, to investigate whether perceived knowledge resemblance between LLMs and humans modulates the P600 effect to LLM-attributed syntactic anomalies, we fit LME models on data from the LLM condition with Anomaly

https://doi.org/10.1017/S0142716425100404 Published online by Cambridge University Press

Table 6. LME models for by-trial amplitude analysis of knowledge resemblance in the LLM condition

Fixed effects	*β*	*SE*	*t*	*p*
P600 (600–800 ms)				
Intercept	**0.40**	**0.23**	**1.70**	**0.093**
Anomaly	**1.00**	**0.41**	**2.46**	**0.021**
Knowledge Resemblance	−0.18	0.18	−1.00	0.327
Anomaly: Knowledge Resemblance	−0.34	0.45	−0.77	0.449

Notes: The model for P600 analysis in the LLM condition: Amplitude ~ Anomaly * Knowledge Resemblance + (Anomaly + 1|Participant) + (1|Item) + (1|Channel). Significant and marginally significant results are marked bold.

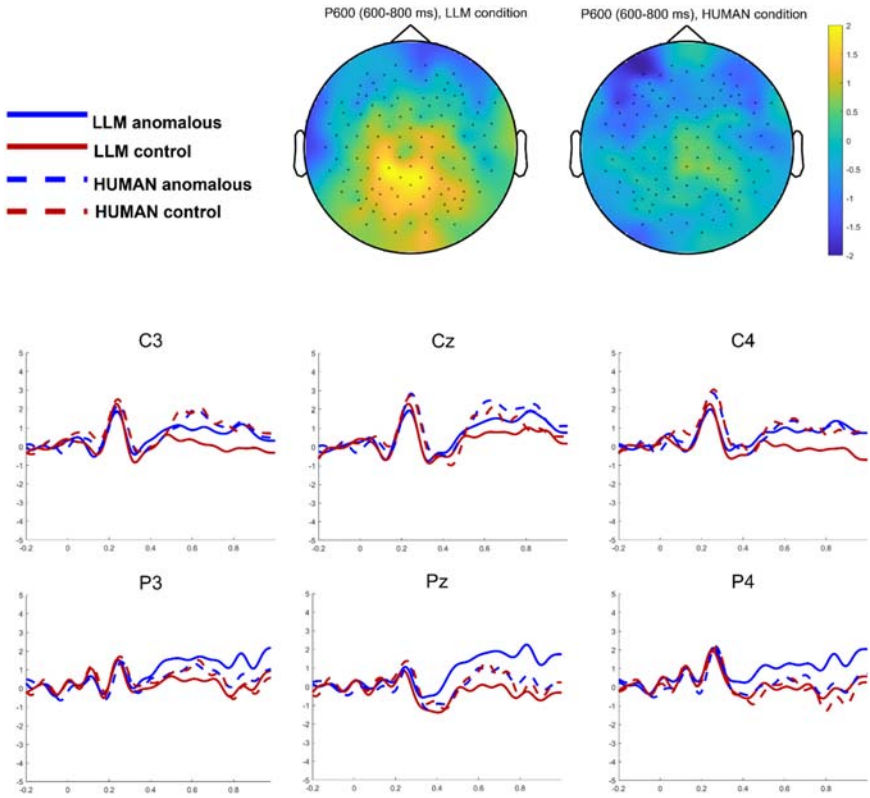


Figure 3. Grand average ERPs elicited by the syntactically anomalous sentences and syntactically well-formed sentences in the LLM and human conditions. Waveforms are illustrated with six centro-parietal sites C3, Cz, C4, P3, Pz, and P4.

and Knowledge Resemblance as interacting fixed effects (see Table 6 for model structures). The results showed the interaction between Anomaly and Knowledge Resemblance did not reach significance, indicating that perceived knowledge resemblance does not modulate the P600 effect to syntactic anomalies from LLMs.

https://doi.org/10.1017/S0142716425100404 Published online by Cambridge University Press

The results of Experiment 2 demonstrate that syntactic anomalies in LLM-attributed text elicited a robust P600 effect, whereas those in human-attributed text did not. The P600 effect is an ERP component associated with increased cognitive efforts in detecting syntactic errors and performing syntactic integration (Hagoort & Brown, 2000; Molinaro et al., 2011). This indicates that participants exerted more cognitive efforts when processing syntactic anomalies in language attributed to LLMs, likely due to an expectation that LLMs should have high syntactic capabilities. In contrast, when they realized that those syntactic anomalies were attributed to humans, they processed them more effortlessly, as they did not hold the same high syntactic expectations for humans.

More recent research suggests that the P600 effect reflects domain-general composition or integration difficulty at the syntax–semantics interface (Baggio, 2018, 2021; Brouwer et al., 2012; Fritz & Baggio, 2020, 2022). Accordingly, there is another interpretation of this result. Experiment 2 reveals that integration processes are modulated by the interlocutor context. The enhanced P600 for LLM-attributed syntactic anomalies reflects additional efforts to assign an interpretation to syntactically anomalous sentences when participants believed that these were generated by an LLM. Critically, the consistent P600 enhancement for LLM-attributed anomalies across both experiments suggests a unified account: participants may engage additional composition or integration efforts when assigning an interpretation to LLM-attributed sentences, regardless of anomaly type, as evidenced by the P600 effects observed in both studies.

The P600 effect to syntactic anomalies from LLMs was not predicted by perceived knowledge resemblance between LLMs and humans, which contrasts with our findings on the N400 and P600 effects to semantic anomalies from LLMs. This suggests that the "knowledge" in question is part of world knowledge or factual reliability, irrelevant to syntactic capabilities. Cognitive efforts in processing syntactic anomalies appear to be influenced by specific expectations about LLMs' syntactic capabilities rather than their general knowledge. Further investigation is necessary to clarify this issue.

General discussion

The present study provides an interdisciplinary investigation integrating linguistic, psychological, and neuroscientific perspectives within HCI research. In two experiments, we investigated the brain responses when people encounter semantic and syntactic anomalies believed to be produced by an LLM versus by a human being. For semantic anomalies, we observed a smaller N400 effect and a larger P600 effect for LLM-attributed anomalies compared to human-attributed ones. This finding indicates that LLM-attributed semantic anomalies were better predictable and more tolerant, but still harder to compose or integrate. Moreover, we observed a larger P600 effect for LLM-attributed syntactically anomalous sentences than human-attributed ones. This indicates that LLM syntactic anomalies were worse predictable or harder to compose or integrate. The consistently enhanced P600 to LLM anomalies across two experiments suggests a unified interpretation: participants may engage additional composition or integration processes when assigning an interpretation to LLM-attributed sentences, regardless of anomaly type.

https://doi.org/10.1017/S0142716425100404 Published online by Cambridge University Press

Notably, individual differences in perceived LLM humanlike knowledge selectively modulated semantic but not syntactic processing. Participants with stronger beliefs in LLM humanlike capabilities showed larger N400 and smaller P600 effects for semantic anomalies, indicating they employed cognitive mechanisms more similar to human–human communication when processing LLM semantic anomalies. The absence of knowledge resemblance effects on syntactic processing indicates that perceived "knowledge" reflects world knowledge or factual reliability rather than syntactic capabilities, with syntactic processing influenced by distinct expectations about LLMs' syntactic capabilities. These findings illuminate differences in the cognitive mechanisms deployed when comprehending LLM-attributed versus human-attributed language.

It should be noted that we observed only a nonsignificant trend toward a P600 effect when participants read syntactically anomalous sentences relative to well-form ones in the human condition, a finding that is in contrast with that in Qiu and Zhou (2012) and Hao et al. (2024) on the violation of the same syntactic agreement in Chinese. It is possible that as we explicitly emphasized to participants in the human condition that they were to read sentences previously produced by another person, participants might have expected "human errors" in these sentences, leading to increased tolerance with syntactic anomalies. Nevertheless, we must also acknowledge that participants might have developed strategies to predict upcoming syntactic violations through the initial temporal phrase. While our design included fillers with balanced initial temporal phrases to reduce this predictability, following a similar filler design to Liang et al. (2022) and representing an approach not employed in previous *guo* violation studies (e.g., Hao et al., 2024; Qiu & Zhou, 2012), we recognize that this may not have completely eliminated the potential for strategic processing. Although our filler distribution was not extensive, it provided more control over this predictability issue than previous studies and was sufficient for drawing our main conclusions. This represents a limitation inherent to this experimental paradigm that future research could address by incorporating additional balanced fillers or developing experimental designs.

Mental model construction of LLM properties during language processing

In general, the findings from the current study align with earlier research demonstrating that people take into account various attributes of their interlocutors (e.g., language background, gender, and socioeconomic status) during language comprehension and use such information to interpret speech and adjust expectations about the content and form of the interlocutor's utterances (e.g., Cai, 2022; Cai et al., 2017, 2021; Corps et al., 2022). For instance, people use their speaker's accent to infer their dialectic background and understand cross-dialectally ambiguous words (Cai, 2022; Cai et al., 2017) and predict the likelihood of specific words or phrases based on the speaker's language background (Martin et al., 2016). People also consider the speaker's gender, anticipating male and female speakers to discuss stereotypically masculine and feminine topics, respectively (Corps et al., 2022). This gender-based expectation has been demonstrated in ERP studies, where comprehenders experience processing difficulties (reflected as enhanced N400 effects) when the message of a sentence does not align with the

https://doi.org/10.1017/S0142716425100404 Published online by Cambridge University Press

stereotypical expectations associated with the speaker's age or gender (van Berkum et al., 2008; Wu, Duan, & Cai., 2024; Wu & Cai, 2025).

Our findings suggest that people can still apply interlocutor modeling when comprehending LLM-attributed language. That is, people seem to take into account an LLM's factual unreliability and lack of genuine world knowledge, as well as its relatively high linguistic competence. This leads to the seemingly contradictory pattern: people are less "surprised" with semantic anomalies but are more "surprised" with syntactic anomalies in LLM language comprehension than human language comprehension. These findings are thus consistent with earlier findings that people treat AI systems with certain human traits. For instance, people perceive AI systems as having certain gender, age, and language competence. They assign gender to robots based on their appearance or voice (Eyssel & Kuchenbrandt, 2012) and make assumptions about their knowledge based on stereotypical gender roles (Powers et al., 2005; Powers & Kiesler, 2006). People also interact differently with robots depending on their perceived age, being more likely to follow requests from baby-faced robots than adult-faced ones (Powers & Kiesler, 2006) and perceiving child-voice robots as more acceptable for home companions and adult-voice robots as more suitable for educational tutors (Dou et al., 2021). Furthermore, people are sensitive to a computer's language background and adjust their interaction accordingly. When collaborating with a computer speaking in an Irish or American accent, people tend to produce dialect-appropriate lexical expressions (Cowan et al., 2019). They also adapt their interaction based on the perceived language competence of a computer, aligning more with a computer than with a human (Branigan et al., 2011; Shen & Wang, 2023) and more with a basic computer than a more advanced one (Branigan et al., 2011; Pearson et al., 2006). Our findings are thus in line with these previous findings, showing that people treat LLMs as having certain humanlike knowledge bases and adjust their language accordingly.

The conclusion that people construct a mental model of an LLM is further substantiated by the modulation of ERP components by the perceived knowledge resemblance between LLMs and humans. For semantic anomalies, attributing higher humanlike knowledge to LLMs predicted a larger N400 effect and a smaller P600 effect (Experiment 1). This indicates that as the perceived gap between human and LLM knowledge narrows, the cognitive mechanisms underlying semantic processing of LLM language is more humanlike. Nevertheless, no such modulation was observed for syntactic processing, revealing a dissociation that suggests people's mental models distinguish between LLMs' semantic and syntactic capabilities. In other words, people calibrate their expectations of LLMs across different linguistic domains. Further research is necessary to fully explore how domain-specific perceptions of LLM capabilities affect human–LLM interaction dynamics.

Theoretical insights into the divergent responses to anomalies in HCI

Overall, our results suggest that people do indeed treat LLMs similarly to their human peers by applying expectations, particularly in their language behaviors, in line with the *Computers Are Social Actors* paradigm (Nass & Moon, 2000). However, these behaviors at times are not identical; people exhibit varying neural responses toward LLMs in certain contexts due to differing expectations. Our findings can also

https://doi.org/10.1017/S0142716425100404 Published online by Cambridge University Press

be interpreted within the framework of machine heuristics (Lee, 2024). This framework posits that people attend to the source of information (machines vs. humans) primarily when linguistic output contradicts their expectations, and people's prior beliefs about machines modulate their reaction to the unexpected performance of machines. Specifically, people with positive perceptions of machines may exhibit heightened negativity toward their underperformance due to unmet expectations, while those with negative beliefs may demonstrate greater tolerance for such shortcomings. Our results indicate that semantic anomalies elicited a reduced N400 response in LLM condition, consistent with the idea that people with negative beliefs about LLMs are more tolerant of their underperformance. As these negative beliefs decrease, tolerance also diminishes, as evidenced by a more pronounced N400 effect in LLM-attributed text among participants who held stronger convictions regarding the humanlike knowledge of LLMs. Additionally, the observation that syntactic anomalies perceived to be LLM-attributed elicited a larger P600, likely due to people's belief in LLMs' high syntactic capabilities, aligns with the account as well. Collectively, these findings highlight the significant moderating effect of task-relevant beliefs regarding LLMs' specific properties on the neuro-cognitive processes involved in understanding LLM-attributed language.

Implications for advancing AI systems

Our findings have practical implications for designing systems that align with human cognitive needs. Our study demonstrates that core language comprehension processes—semantic and syntactic processing—can be influenced by people's beliefs about LLMs. These beliefs may induce cognitive biases: people who believe that LLMs are flawless might overlook mistakes in their responses, while those who are skeptical may disengage or distrust the outputs, even if the information is accurate. Given the growing integration of LLMs into educational settings (Shahzad et al., 2025), addressing cognitive biases and subconscious influences is urgent.

One approach is to improve explainable AI frameworks, which focus on making the decision-making processes of AI systems transparent and easily understandable for users (Shamsuddin et al., 2025). By designing systems that actively counteract subconscious biases through transparency, training, and human–AI collaboration, it is possible to foster environments where LLMs enhance rather than hinder the quality of interaction. This may also enhance user trust and engagement with AI technologies. Furthermore, understanding how cognitive biases vary across different contexts, such as education, healthcare, and entertainment, is an important direction for future research.

Conclusion

Our research reveals that people's brains respond differently to LLM-attributed language compared to human-attributed language. In two experiments measuring brain activity, we found that people show more expectation for LLM-attributed semantic anomalies (a smaller N400 effect) yet still require more efforts to compose or integrate them (a larger P600 effect). In contrast, people show less expectation or have more integration difficulty when processing LLM syntactically

https://doi.org/10.1017/S0142716425100404 Published online by Cambridge University Press

anomalous sentences compared to human ones (a larger P600 effect). These findings suggest people develop expectations about LLMs, anticipating potential factual unreliability or semantic errors while expecting high syntactic capabilities. Notably, perceived humanlike knowledge of LLMs selectively modulated semantic processing: as perceived humanlike knowledge of LLMs increased, participants showed enhanced sensitivity to semantic anomalies from an LLM (larger N400) and reduced composition or integration efforts (smaller P600), resembling humanlike cognitive mechanisms. Nevertheless, perceived knowledge resemblance between LLMs and humans did not influence the syntactic processing of LLM language, indicating that people's mental models distinguish between LLMs' factual reliability and syntactic capabilities. As LLMs continue to evolve into more sophisticated communication partners, understanding these cognitive adaptations becomes increasingly important for designing effective human–AI interaction. Our study provides the first neuroscientific evidence that people's perceptions of language source (human vs. LLMs) alter how their brains process information, highlighting the unique cognitive dynamics at play in the emerging landscape of human–AI communication.

Replication package. All experimental stimuli, data preprocessing scripts, analytic scripts, and raw data have been made publicly available on the Open Science Framework at https://osf.io/kfmzu, with supplementary raw data available at https://osf.io/7y9ht and https://osf.io/jr2ys.

Funding Statement. This work was supported by a General Research Fund (14600220) from University Grants Committee, Hong Kong.

Competing interests. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Amalberti, R., Carbonell, N., & Falzon, P. (1993). User representations of computer systems in human-computer speech interaction. *International Journal of Man-Machine Studies*, **38**(4), 547–566. https://doi.org/10.1006/imms.1993.1026

Baggio, G. (2018). *Meaning in the brain*. MIT Press.

Baggio, G. (2021). Compositionality in a parallel architecture for language processing. *Cognitive Science*, **45**(5), e12949. https://doi.org/10.1111/cogs.12949

Baggio, G., & Hagoort, P. (2011). The balance between memory and unification in semantics: A dynamic account of the N400. *Language and Cognitive Processes*, **26**(9), 1338–1367. https://doi.org/10.1080/01690965.2010.542671

Barr, D. J., Jackson, L., & Phillips, I. (2014). Using a voice to put a name to a face: The psycholinguistics of proper name comprehension. *Journal of Experimental Psychology: General*, **143**(1), 404. https://doi.org/10.1037/a0031813

Bergmann, K., Branigan, H. P., & Kopp, S. (2015). Exploring the alignment space–lexical and gestural alignment with real and virtual humans. *Frontiers in ICT*, **2**, 7. https://doi.org/10.3389/fict.2015.00007

Bornkessel-Schlesewsky, I., Krauspenhaar, S., & Schlesewsky, M. (2013). Yes, you can? A speaker's potency to act upon his words orchestrates early neural responses to message-level meaning. *PloS one*, **8**(7), e69173. https://doi.org/10.1371/journal.pone.0069173

Branigan, H. P., Pickering, M. J., Pearson, J., McLean, J. F., & Brown, A. (2011). The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition*, **121**(1), 41–57. https://doi.org/10.1016/j.cognition.2011.05.011

https://doi.org/10.1017/S0142716425100404 Published online by Cambridge University Press

Branigan, H. P., Pickering, M. J., Pearson, J., McLean, J. F., & Nass, C. I. (2003, July). Syntactic alignment between computers and people: The role of belief about mental states. *Proceedings of the 25th annual conference of the cognitive science society* (Vol. 31). Hillsdale, NJ, USA: Lawrence Erlbaum Associates.

Branigan, H. P., Pickering, M. J., Pearson, J., McLean, J. F., Nass, C. I., & Hu, J. (2004). Beliefs about mental states in lexical and syntactic alignment: Evidence from human-computer dialogs. *Proceedings of the 17th CUNY Conference on Human Sentence Processing*.

Brothers, T., Dave, S., Hoversten, L. J., Traxler, M. J., & Swaab, T. Y. (2019). Flexible predictions during listening comprehension: Speaker reliability affects anticipatory processes. *Neuropsychologia*, **135**, 107225. https://doi.org/10.1016/j.neuropsychologia.2019.107225

Broussard, M. (2018). *Artificial unintelligence: How computers misunderstand the world*. MIT Press.

Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about semantic illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research*, **1446**, 127–143. https://doi.org/10.1016/j.brainres.2012.01.055

Cabrero-Daniel, B., & Sanagustín Cabrero, A. (2023, July). Perceived trustworthiness of natural language generators. *Proceedings of the First International Symposium on Trustworthy Autonomous Systems* (pp. 1–9). https://doi.org/10.1145/3597512.3599715

Cai, Z. G. (2022). Interlocutor modelling in comprehending speech from interleaved interlocutors of different dialectic backgrounds. *Psychonomic Bulletin & Review*, **29**, 1026–1034. https://doi.org/10.3758/s13423-022-02055-7

Cai, Z. G., Gilbert, R. A., Davis, M. H., Gaskell, M. G., Farrar, L., Adler, S., & Rodd, J. M. (2017). Accent modulates access to word meaning: Evidence for a speaker model account of spoken word recognition. *Cognitive Psychology*, **98**, 73–101. https://doi.org/10.1016/j.cogpsych.2017.08.003

Cai, Z. G., Sun, Z., & Zhao, N. (2021). Interlocutor modelling in lexical alignment: The role of linguistic competence. *Journal of Memory and Language*, **121**(December 2020), 104278. https://doi.org/10.1016/j.jml.2021.104278

Caffarra, S., Wolpert, M., Scarinci, D., & Mancini, S. (2020). Who are you talking to? The role of addressee identity in utterance comprehension. *Psychophysiology*, **57**(4), e13527. https://doi.org/10.1111/psyp.13527

Cloudy, J., Banks, J., & Bowman, N. D. (2023). The str (AI) ght scoop: Artificial intelligence cues reduce perceptions of hostile media bias. *Digital Journalism*, **11**(9), 1577–1596. https://doi.org/10.1080/21670811.2021.1969974

Cohn, M., Segedin, B. F., & Zellou, G. (2019). Imitating Siri: Socially-mediated vocal alignment to device and human voices. *Proceedings of the 19th International Congress of Phonetic Sciences (ICPhS)* (Vol. 17).

Corps, R. E., Brooke, C., & Pickering, M. J. (2022). Prediction involves two stages: Evidence from visual-world eye-tracking. *Journal of Memory and Language*, **122**, 104298. https://doi.org/10.1016/j.jml.2021.104298

Cowan, B. R., Branigan, H. P., Obregón, M., Bugis, E., & Beale, R. (2015). Voice anthropomorphism, interlocutor modelling and alignment effects on syntactic choices in human- computer dialogue. *International Journal of Human-Computer Studies*, **83**, 27–42. https://doi.org/10.1016/j.ijhcs.2015.05.008

Cowan, B. R., Doyle, P., Edwards, J., ... **& Clark, L.** (2019). What's in an accent? The impact of accented synthetic speech on lexical choice in human-machine dialogue. *Proceedings of the 1st International Conference on Conversational User Interfaces* (pp. 1–8). https://doi.org/10.1145/3342775.3342786

Dai, Y. (1997). Studies on Aspectual Systems of Modern Chinese (Xiandai Hanyu Shiti Xitong Yanjiu).

Dou, X., Wu, C. F., Lin, K. C., Gan, S., & Tseng, T. M. (2021). Effects of different types of social robot voices on affective evaluations in different application fields. *International Journal of Social Robotics*, **13**, 615–628. https://doi.org/10.1007/s12369-020-00654-9

Eyssel, F., & Kuchenbrandt, D. (2012). Social categorization of social robots: Anthropomorphism as a function of robot group membership. *British Journal of Social Psychology*, **51**(4), 724–731. https://doi.org/10.1111/j.2044-8309.2011.02082.x

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, **41**(4), 469–495. https://doi.org/10.1006/jmla.1999.2660

Fiske, A., Henningsen, P., & Buyx, A. (2019). Your robot therapist will see you now: Ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. *Journal of Medical Internet Research*, **21**(5), 1–12. https://doi.org/10.2196/13216

https://doi.org/10.1017/S0142716425100404 Published online by Cambridge University Press

Fogg, B. J., & Nass, C. (1997). How users reciprocate to computers: An experiment that demonstrates behavior change. *CHI'97 Extended Abstracts on Human Factors in Computing Systems* (pp. 331–332), USA. https://doi.org/10.1145/1120212.1120419

Foucart, A., Santamaría-García, H., & Hartsuiker, R. J. (2019). Short exposure to a foreign accent impacts subsequent cognitive processes. *Neuropsychologia*, **129**, 1–9. https://doi.org/10.1016/j.neuropsychologia.2019.02.021

Foucart, A., & Hartsuiker, R. J. (2021). Are foreign-accented speakers that 'incredible'? The impact of the speaker's indexical properties on sentence processing. *Neuropsychologia*, **158**, 107902. https://doi.org/10.1016/j.neuropsychologia.2021.107902. Get rights and content

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, **6**(4), 291–325. https://doi.org/10.1016/0010-0277(78)90002-1

Fritz, I., & Baggio, G. (2020). Meaning composition in minimal phrasal contexts: Distinct ERP effects of intensionality and denotation. *Language, Cognition and Neuroscience*, **35**(10), 1295–1313. https://doi.org/10.1080/23273798.2020.1749678

Fritz, I., & Baggio, G. (2022). Neural and behavioural effects of typicality, denotation and composition in an adjective–noun combination task. *Language, Cognition and Neuroscience*, **37**(5), 537–559. https://doi.org/10.1080/23273798.2021.2004176

Ghandeharion, H., & Erfanian, A. (2010). A fully automatic ocular artifact suppression from EEG data using higher order statistics: Improved performance by wavelet analysis. *Medical Engineering & Physics*, **32**(7), 720–729. https://doi.org/10.1016/j.medengphy.2010.04.010

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, **20**(11), 818–829. https://doi.org/10.1016/j.tics.2016.08.005

Grant, A. M., Brisson-McKenna, M., & Phillips, N. A. (2021). Electrophysiology reflects the influence of discourse context on auditory semantic processing in bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **47**(2), 343. https://doi.org/10.1037/xlm0000813

Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech compared to reading: The P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia*, **38**(11), 1531–1549. https://doi.org/10.1016/S0028-3932(00)00053-1

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, **304**(5669), 438–441. https://doi.org/10.1126/science.1095455

Hahne, A., & Friederici, A. D. (2002). Differential task effects on semantic and syntactic processes as revealed by ERPs. *Cognitive Brain Research*, **13**(3), 339–356. https://doi.org/10.1016/S0926-6410(01)00127-6

Hanulíková, A., Van Alphen, P. M., Van Goch, M. M., & Weber, A. (2012). When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience*, **24**(4), 878–887. https://doi.org/10.1162/jocn_a_00103

Hanulíková, A., & Carreiras, M. (2015). Electrophysiology of subject-verb agreement mediated by speakers' gender. *Frontiers in Psychology*, **6**(September), 1–12. https://doi.org/10.3389/fpsyg.2015.01396

Hao, Y., Duan, X., Zha, S., & Xu, T. (2024). How L2 learners process different means of time encoding in a tenseless language: An ERP study of Mandarin. *Journal of Psycholinguistic Research*, **53**(4), 60. https://doi.org/10.1007/s10936-024-10097-2

Hara, K., & Iqbal, S. T. (2015, April). Effect of machine translation in interlingual conversation: Lessons from a formative study. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 3473–3482).

Heyselaar, E., Hagoort, P., & Segaert, K. (2017). In dialogue with an avatar, language behavior is identical to dialogue with a human partner. *Behavior Research Methods*, **49**, 46–60. https://doi.org/10.3758/s13428-015-0688-7

Hinojosa, J., Martín-Loeches, M., Casado, P., Munoz, F., & Rubia, F. (2003). Similarities and differences between phrase structure and morphosyntactic violations in Spanish: An event-related potentials study. *Language and Cognitive Processes*, **18**(2), 113–142. https://doi.org/10.1080/01690960143000489

Hoy, M. B. (2018). Alexa, Siri, Cortana, and more: An introduction to voice assistants. *Medical Reference Services Quarterly*, **37**(1), 81–88. https://doi.org/10.1080/02763869.2018.1404391

Irfan, B., Kuoppamäki, S. M., & Skantze, G. (2023). Between reality and delusion: Challenges of applying large language models to companion robots for open-domain dialogues with older adults. *Autonomous Robots*, **49**(1), 9. https://doi.org/10.1007/s10514-025-10190-y

https://doi.org/10.1017/S0142716425100404 Published online by Cambridge University Press

Jones-Jang, S. M., & Park, Y. J. (2023). How do people react to AI failure? Automation bias, algorithmic aversion, and perceived controllability. *Journal of Computer-Mediated Communication*, **28**(1), 1–8. https://doi.org/10.1093/jcmc/zmac029

Kanda, T., Miyashita, T., Osada, T., Haikawa, Y., & Ishiguro, H. (2008). Analysis of humanoid appearances in human–robot interaction. *IEEE Transactions on Robotics*, **24**(3), 725–735. https://doi.org/10.1109/TRO.2008.921566.

Kennedy, A., Wilkes, A., Elder, L., & Murray, W. S. (1988). Dialogue with machines. *Cognition*, **30**(1), 37–72. https://doi.org/10.1016/0010-0277(88)90003-0

Khoo, W., Hsu, L. J., Amon, K. J., Chakilam, P. V., Chen, W. C., Kaufman, Z., … **& Sabanović, S.** (2023, March). Spill the tea: When robot conversation agents support well-being for older adults. *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 178–182).

Kim, Y., Kim, D., Choi, J., Park, J., Oh, N., & Park, D. (2024). A Survey on Integration of Large Language Models with Intelligent Robots. *arXiv preprint arXiv:2404.09228*.

Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research*, **29**(6), 627–645. https://doi.org/10.1023/A:1026528912821

Kopp, S., Gesellensetter, L., Krämer, N. C., & Wachsmuth, I. (2005). A conversational agent as museum guide–design and evaluation of a real-world application. *Intelligent Virtual Agents: 5th International Working Conference, IVA 2005, Kos, Greece, September 12-14, 2005. Proceedings 5* (pp. 329–343). Springer Berlin Heidelberg. https://doi.org/10.1007/11550617_28

Krämer, N. C. (2005). Social communicative effects of a virtual program guide. *Intelligent Virtual Agents: 5th International Working Conference, IVA 2005, Kos, Greece, September 12-14, 2005. Proceedings 5* (pp. 442–453). Springer Berlin Heidelberg. https://doi.org/10.1007/11550617_37

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, **62**(1), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, **207**(4427), 203–205. https://doi.org/10.1126/science.7350657

Kutas, M., & Van Petten. C. (1994). Psycholinguistics Electrified: Event-Related Brain Potential Investigations. In M. Gemsbacher (Ed.), *Handbook of psycholinguistics* (pp. 83–144). Academic Press.

Lee, E. J. (2008). Gender stereotyping of computers: Resource depletion or reduced attention? *Journal of Communication*, **58**(2), 301–320. https://doi.org/10.1111/j.1460-2466.2008.00386.x

Lee, E. J. (2010). What triggers social responses to flattering computers? Experimental tests of anthropomorphism and mindlessness explanations. *Communication Research*, **37**(2), 191–214. https://doi.org/10.1177/0093650209356389

Lee, E. J. (2024). Minding the source: Toward an integrative theory of human–machine communication. *Human Communication Research*, **50**(2), 184–193. https://doi.org/10.1093/hcr/hqad034

Lenzoni, S., Sumich, A. L., & Mograbi, D. C. (2024). Domain specificity of error monitoring: An ERP study in young and older adults. *Psychophysiology*, **61**(8), e14579. https://doi.org/10.1111/psyp.14579

Liang, L., Chondrogianni, V., & Chen, B. (2022). Online processing of the grammatical aspect marker by L2 Chinese learners. *Second Language Research*, **38**(4), 765–786. https://doi.org/10.1177/0267658321996423

Lin, J. W. (2003). Temporal reference in mandarin Chinese. *Journal of East Asian Linguistics*, **12**(3), 259–311. https://doi.org/10.1023/A:1023665301095

Liu, B., & Wei, L. (2018). Reading machine-written news: Effect of machine heuristic and novelty on hostile media perception. *Human-Computer Interaction. Theories, Methods, and Human Issues: 20th International Conference, HCI International 2018, Las Vegas, NV, USA, July 15–20, 2018, Proceedings, Part I 20* (pp. 307–324). Springer International Publishing. https://doi.org/10.1007/978-3-319-91238-7_26

Luck, S. J. (2014). *An introduction to the ERP technique*, second edition. In MIT Press. https://books.google.com.hk/books?id = y4-uAwAAQBAJ

Luck, S. J. (2022). Applied Event-Related Potential Data Analysis. https://socialsci.libretexts.org/Bookshelves/Psychology/Book%3A_Applied_Event-Related_Potential_Data_Analysis_(Luck)

Martin, C. D., Garcia, X., Potter, D., Melinger, A., & Costa, A. (2016). Holiday or vacation? The processing of variation in vocabulary across dialects. *Language, Cognition and Neuroscience*, **31**, 375–390. https://doi.org/10.1080/23273798.2015.1100750

https://doi.org/10.1017/S0142716425100404 Published online by Cambridge University Press

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, **94,** 305–315. https://doi.org/10.1016/j.jml.2017.01.001

Maynez, J., Narayan, S., Bohnet, B., & McDonald, R. (2020). On Faithfulness and Factuality in Abstractive Summarization. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (p. 1906). Association for Computational Linguistics.

Miller, E. J., Foo, Y. Z., Mewton, P., & Dawel, A. (2023). How do people respond to computer-generated versus human faces? A systematic review and meta-analyses. *Computers in Human Behavior Reports*, **10**, 100283. https://doi.org/10.1016/j.chbr.2023.100283

Molina, M. D., & Sundar, S. S. (2022). Does distrust in humans predict greater trust in AI? Role of individual differences in user responses to content moderation. *New Media & Society,* **26**(6). https://doi.org/10.1177/14614448221103534

Molinaro, N., Barber, H. A., & Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex*, **47**(8), 908–930. https://doi.org/10.1016/j.cortex.2011.02.019

Moon, Y. (2000). Intimate exchanges: Using computers to elicit self-disclosure from consumers. *Journal of Consumer Research*, **26**(4), 323–339. https://doi.org/10.1086/209566

Nass, C., Isbister, K., & Lee, E. J. (2000). Truth is beauty: Researching embodied conversational agents. *Embodied Conversational Agents*, **2000**, 374–402.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, **56**(1), 81–103. https://doi.org/10.1111/0022-4537.00153

Nass, C., Moon, Y., & Carney, P. (1999). Are people polite to computers? Responses to computer-based interviewing systems 1. *Journal of Applied Social Psychology*, **29**(5), 1093–1109. https://doi.org/10.1111/j.1559-1816.1999.tb00142.x

Nass, C. I., Moon, Y., & Green, N. (1997). Are machines gender neutral? Gender-stereotypic responses to computers with voices. *Journal of Applied Social Psychology*, **27**(10), 864–876. https://doi.org/10.1111/j.1559-1816.1997.tb00275.x

Nevins, A., Dillon, B., Malhotra, S., & Phillips, C. (2007). The role of feature-number and feature-type in processing Hindi verb agreement violations. *Brain Research*, **1164**, 81–94. https://doi.org/10.1016/j.brainres.2007.05.058

Ni, X., & Li, P. (2024). A Systematic Evaluation of Large Language Models for Natural Language Generation Tasks. *Proceedings of the 22nd Chinese National Conference on Computational Linguistics* (Volume 2: Frontier Forum) (pp. 40–56).

Nightingale, S. J., & Farid, H. (2022). AI-synthesized faces are indistinguishable from real faces and more trustworthy. *Proceedings of the National Academy of Sciences*, **119**(8), e2120481119. https://doi.org/10.1073/pnas.2120481119

Noveck, I. (2018). *Experimental pragmatics: The making of a cognitive science*. Cambridge University Press.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, **2011**(1). https://doi.org/10.1155/2011/156869

Parise, S., Kiesler, S., Sproull, L., & Waters, K. (1999). Cooperating with life-like interface agents. *Computers in Human Behavior*, **15**(2), 123–142. https://doi.org/10.1016/S0747-5632(98)00035-1

Pearson, J., Hu, J., Branigan, H. P., Pickering, M. J., & Nass, C. I. (2006, April). Adaptive language behavior in HCI: how expectations and beliefs about a system affect users' word choice. *Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 1177–1180). https://doi.org/10.1145/1124772.1124948

Powers, A., & Kiesler, S. (2006). The advisor robot: tracing people's mental model from a robot's physical attributes. *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction* (pp. 218–225). https://doi.org/10.1145/1121241.1121280

Powers, A., Kramer, A. D., Lim, S., Kuo, J., Lee, S. L., & Kiesler, S. (2005). Eliciting information from people with a gendered humanoid robot. In *2005 IEEE International Workshop on Robot and Human Interactive Communication* (pp. 158–163). IEEE. https://doi.org/10.1109/ROMAN.2005.1513773

Qiu, Y., & Zhou, X. (2012). Processing temporal agreement in a tenseless language: An ERP study of Mandarin Chinese. *Brain Research*, **1446**, 91–108. https://doi.org/10.1016/j.brainres.2012.01.051

Rawte, V., Sheth, A., & Das, A. (2023). *A survey of hallucination in large foundation models.* arXiv preprint arXiv:2309.05922.

https://doi.org/10.1017/S0142716425100404 Published online by Cambridge University Press

Regel, S., Coulson, S., & Gunter, T. C. (2010). The communicative style of a speaker can affect language comprehension? ERP evidence from the comprehension of irony. *Brain Research*, **1311**, 121–135. https://doi.org/10.1016/j.brainres.2009.10.077

Rueschemeyer, S. A., Gardner, T., & Stoner, C. (2015). The Social N400 effect: How the presence of other listeners affects language comprehension. *Psychonomic Bulletin & Review*, **22**(1), 128–134. https://doi.org/10.3758/s13423-014-0654-x

Ryskin, R., Ng, S., Mimnaugh, K., Brown-Schmidt, S., & Federmeier, K. D. (2020). Talker-specific predictions during language processing. *Language, Cognition and Neuroscience*, **35**(6), 797–812. https://doi.org/10.1080/23273798.2019.1630654

Ryskin, R., Stearns, L., Bergen, L., Eddy, M., Fedorenko, E., & Gibson, E. (2021). An ERP index of real-time error correction within a noisy-channel framework of human communication. *Neuropsychologia*, **158**, 107855. https://doi.org/10.1016/j.neuropsychologia.2021.107855

Schmidt, T., & Strassner, T. (2022). Artificial intelligence in foreign language learning and teaching. *Anglistik*, **33**(1), 165–184. https://doi.org/10.33675/angl/2022/1/14

Shahzad, T., Mazhar, T., Tariq, M. U., Ahmad, W., Ouahada, K., & Hamam, H. (2025). A comprehensive review of large language models: Issues and solutions in learning environments. *Discover Sustainability*, **6**(1), 27.

Shamsuddin, R., Tabrizi, H.B. & Gottimukkula, P.R. (2025). Towards responsible AI: An implementable blueprint for integrating explainability and social-cognitive frameworks in AI systems. *AI Perspectives & Advances*, **7**, 1. https://doi.org/10.1186/s42467-024-00016-5

Shechtman, N., & Horowitz, L. M. (2003, April). Media inequality in conversation: how people behave differently when interacting with computers and people. *Proceedings of the SIGCHI Conference on Human factors in Computing Systems* (pp. 281–288). https://doi.org/10.1145/642611.642661

Shen, H., & Wang, M. (2023). Effects of social skills on lexical alignment in human-human interaction and human-computer interaction. *Computers in Human Behavior*, **143**, 107718. https://doi.org/10.1016/j.chb.2023.107718

Sundar, S. S., & Kim, J. (2019). Machine heuristic: When we trust computers more than humans with our personal information. *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1–9), UK. https://doi.org/10.1145/3290605.300768

Sundar, S. S., & Lee, E.-J. (2022). Rethinking communication in the era of artificial intelligence. *Human Communication Research*, **48**(3), 379–385. https://doi.org/10.1093/hcr/hqac014

Sundar, S. S., & Nass, C. (2000). Source orientation in human–computer interaction: Programmer, networker, or independent social actor. *Communication Research*, **27**(6), 683–703. https://doi.org/10.1177/009365000027006001

Tanner, D., Morgan-Short, K., & Luck, S. J. (2015). How inappropriate high-pass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and cognition. *Psychophysiology*, **52**(8), 997. https://doi.org/10.1111/PSYP.12437

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **31**(3), 443–467. https://doi.org/10.1037/0278-7393.31.3.443

Van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, **11**(6), 657–671. https://doi.org/10.1162/089892999563724

Van Berkum, J. J., Van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, **20**(4), 580–591. https://doi.org/10.1162/jocn.2008.20054

Vissers, C. T. W., Virgillito, D., Fitzgerald, D. A., Speckens, A. E., Tendolkar, I., van Oostrom, I., & Chwilla, D. J. (2010). The influence of mood on the processing of syntactic anomalies: evidence from P600. *Neuropsychologia*, **48**(12), 3521–3531. https://doi.org/10.1016/j.neuropsychologia.2010.08.001

Vollmer, A. L., Read, R., Trippas, D., & Belpaeme, T. (2018). Children conform, adults resist: A robot group induced peer pressure on normative social conformity. *Science Robotics*, **3**(21), eaat7111. https://doi.org/10.1126/scirobotics.aat7111

https://doi.org/10.1017/S0142716425100404 Published online by Cambridge University Press

Volpert-Esmond, H. I., Merkle, E. C., Levsen, M. P., Ito, T. A., & Bartholow, B. D. (2018). Using trial-level data and multilevel modeling to investigate within-task change in event-related potentials. *Psychophysiology*, **55**(5), e13044. https://doi.org/10.1111/psyp.13044

Wang, S. (2021). Moderating uncivil user comments by humans or machines? The effects of moderation agent on perceptions of bias and credibility in news content. *Digital Journalism*, **9**(1), 64–83. https://doi.org/10.1080/21670811.2020.1851279

Winkler, I., Haufe, S., & Tangermann, M. (2011). Automatic classification of artifactual ICA-components for artifact removal in EEG signals. *Behavioral and Brain Functions*, **7**(1), 30. https://doi.org/10.1186/1744-9081-7-30

Wu, H., & Cai, Z. G. (2024). Speaker effects in spoken language comprehension. *arXiv preprint arXiv:2412.07238*.

Wu, H., & Cai, Z. G. (2025). When a man says he is pregnant: Event-related potential evidence for a rational account of speaker-contextualized language comprehension. *Journal of Cognitive Neuroscience*, 1–16. https://doi.org/10.1162/JOCN.a.102

Wu, H., Duan, X., & Cai, Z. G. (2024). Speaker demographics modulate listeners' neural correlates of spoken word processing. *Journal of Cognitive Neuroscience*, **36**(10), 2208–2226. https://doi.org/10.1162/jocn_a_02225

Wu, H., Rao, X., & Cai, Z. G. (2025). Probabilistic adaptation of language comprehension for individual speakers: Evidence from neural oscillations. *Social Cognitive and Affective Neuroscience*, **20**(1), nsaf085. https://doi.org/10.1093/scan/nsaf085

Wu, Q., Xu, Y., Xiao, T., Xiao, Y., Li, Y., Wang, T., … & Yang, Y. (2024). Surveying attitudinal alignment between large language models vs. humans towards 17 sustainable development goals. *arXiv preprint arXiv:2404.13885*.

Xu, J., Abdel Rahman, R., & Sommer, W. (2022). Who speaks next? Adaptations to speaker identity in processing spoken sentences. *Psychophysiology*, **59**(1), e13948. https://doi.org/10.1111/psyp.13948

Yin, Y., Jia, N., & Wakslak, C. J. (2024). AI can help people feel heard, but an AI label diminishes this impact. *Proceedings of the National Academy of Sciences of the United States of America*, **121**(14), e2319112121. https://doi.org/10.1073/pnas.2319112121

Yngve, V. H. (1967). Some applications of computers in linguistics. *Reprinted from: Languages and areas: studies presented to George V.* (pp. 172–179). Bobrinskoy.

Zellou, G., Cohn, M., & Ferenc Segedin, B. (2021). Age-and gender-related differences in speech alignment toward humans and voice-AI. *Frontiers in Communication*, **5**, 600361. https://doi.org/10.3389/fcomm.2020.600361

Zhang, Y., Yu, J., & Boland, J. E. (2010). Semantics does not need a processing license from syntax in reading Chinese. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **36**(3), 765. https://doi.org/10.1037/a0019254

Zhu, Z., Hagoort, P., Zhang, J. X., Feng, G., Chen, H. C., Bastiaansen, M., & Wang, S. (2012). The anterior left inferior frontal gyrus contributes to semantic unification. *Neuroimage*, **60**(4), 2230–2237. https://doi.org/10.1016/j.neuroimage.2012.02.036

Zhou, P., Garnsey, S., & Christianson, K. (2019). Is imagining a voice like listening to it? evidence from ERPs. *Cognition*, **182**, 227–241. https://doi.org/10.1016/j.cognition.2018.10.014

Cite this article: Rao, X., Wu, H., & Cai, Z. G. (2025). Comprehending semantic and syntactic anomalies in text attributed to an LLM versus a human: An ERP study. *Applied Psycholinguistics*. https://doi.org/10.1017/S0142716425100404

https://doi.org/10.1017/S0142716425100404 Published online by Cambridge University Press