Reviewed Preprint
v1 • April 27, 2026
Not revised

Reviewed Preprint
v2 • June 25, 2026
Revised by authors

# Neural categorization of visual words of alphabetic and non-alphabetic languages

**Guo Zheng, Shihui Han**

School of Psychological and Cognitive Sciences, PKU-IDG/McGovern Institute for Brain Research, Peking University, Beijing, China

**For correspondence:** shan@pku.edu.cn

**Competing interests:** No competing interests declared

**Funding:** See page 24

**Reviewing editor:** Nai Ding, Zhejiang University, China



## eLife Assessment

This **important** study investigates how the brain categorizes written words from different writing systems (e.g., alphabetic vs. non-alphabetic). The evidence supporting the authors' claims is **solid** and sheds light on the neural basis of language's social-categorization function.

https://doi.org/10.7554/eLife.110320.2.sa2

## Abstract

Languages provide social-category markers that tag people as one or another social group. How does the brain sort words into different language categories as a basis of the social-categorization function of language? We addressed this issue by testing neural categorization of visual words of different writing systems in nine studies using electroencephalography, magnetoencephalography, and a repetition suppression paradigm. We showed that a neural network, including the anterior temporal, insular, orbital frontal, and ventral occipito-temporal cortices in both hemispheres, was engaged in computations of correlation distances between two words to represent intra-language similarity and inter-language difference during categorization of visual words of alphabetic and non-alphabetic languages. These processes occurred as early as 150 ms post-stimulus, recruited within-hemisphere functional connections, operated independently of words' semantic meanings and pronunciations, and exhibited consistently across individuals with diverse language backgrounds. These findings highlight the neural mechanisms of language-based spontaneous neural categorization of visual words as a basis of the social-categorization function of language.

## Introduction

Language is a sophisticated system for transmission of information (e.g., thoughts, knowledge, and feelings) among individuals (Fedorenko et al., 2024b). This linguistic communication function of language relies on the processing of semantic meanings and pronunciations of spoken or written words. Languages also provide social-category markers that tag individuals who use different languages as separate social groups (Bucholtz and Hall, 2010; DeJesus et al., 2018; Kinzler, 2021; Kinzler and Dautel, 2012; Kinzler et al., 2007; Pietraszewski and Schwartz, 2014; Roberts, 2013). This social-categorization function of language has notable consequences in human societies. For example, language serves as a key dimension of ethnic group identity which in turn influences social behaviors (Sachdev and Bourhis, 1990; Scherer and Giles, 1979). In extreme cases, individuals who spoke a different language were classified into an out-group and persecuted (Greenberg, 2008; Shell, 2001). The social-categorization function of language emerges early during human development (Kinzler, 2021; Rhodes and Baron, 2019). Infants expect two individuals who speak the same language to be affiliated (Liberman et al., 2017b) and favor native over non-native speakers during social learning (Howard et al., 2015) and acts of giving (Kinzler et al., 2012). Studies of adults also revealed evidence that language is used as a social cue for categorization of perceived faces (Baus et al., 2021; Champoux-Larsson et al.,

2022) and social categorization of faces based on language may occur automatically (Lorenzoni et al., 2022). The social-categorization function of language revealed in these behavioral studies implicates that rapid categorization of words of different languages may occur in the human brain. Furthermore, the findings of infant studies (e. g., Liberman et al., 2017b) suggest that the neural process involved in categorization of words of different languages may develop even prior to the processing of linguistic properties (e.g. semantic meanings) of words.

Nevertheless, up to date, there has been little neuroimaging research examining the neural mechanisms underlying automatic and fast categorization of words of different languages.

Language-based categorization of written or spoken words with different pronunciations or semantic meanings lays the foundation for knowing who use the same or different languages and thus belong to the same or diverse social groups. The neural mechanisms of language-based categorization of words have been overlooked in previous research that usually focused on the linguistic communication function of language. Linguistic studies have revealed a core neural network that is dominated by the left hemisphere (including the inferior/middle frontal gyri and superior/middle temporal gyri) and enables the processing of linguistic properties (e.g., pronunciations and semantic meanings) of words (Fedorenko et al., 2024a; Friederici and Gierhan, 2013). This network responds to diverse languages (Malik-Moraleda et al., 2022; Siok et al., 2004), activates similarly to native and non-native languages (Li et al., 2021; Malik-Moraleda et al., 2024), and exhibits similar patterns of responses to spoken and written words/sentences (Fedorenko et al., 2024a). Electroencephalography (EEG) and magnetoencephalography (MEG) studies have revealed that visual words initially activate the left ventral occipito-temporal cortex at approximately 170 ms after stimulus onset, reflecting pre-lexical word form processing (Marinkovic et al., 2003; Nan et al., 2022; Pylkkänen and Marantz, 2003). The activation then spreads to the left superior temporal sulcus (LSTS)/inferolateral temporal area at ~230 ms and anterior temporal lobe (ATL) at ~350 ms, and then to the bilateral inferior prefrontal cortices and orbitofrontal cortices (OFC) at ~400 ms (Marinkovic et al., 2003). The left temporal and frontal activities are particularly important for semantic and phonological processing of words and sentences (Hodgson et al., 2021) around 400 ms after word presentation (Zhu et al., 2022) as well as social-semantic working-memory (Zhang et al., 2023a). Because even words of an unlearned language can serve as a social-category marker of a 'not-us group', the neural dynamics of language-based categorization of words may be different from that involved in the processing of linguistic properties of words. The brain regions underlying social concept/knowledge and social categorization (Arioli et al., 2021; Olson et al., 2013; Pang et al., 2025; Pobric et al., 2016; Zahn et al., 2007; Zhou et al., 2020) may play important roles in categorization of words of different languages.

The present study investigated neural dynamics of categorization of visual words of two different (an alphabetic versus a non-alphabetic, or two different alphabetic) languages by combining EEG/MEG with a repetition suppression (RS) paradigm adopted from previous studies of social categorization of faces (Zhang et al., 2023b; Zhou et al., 2020). RS refers to the attenuation in neural responses to a repeated occurrence of stimuli that engage common neuronal populations or processes due to habituation. The RS paradigm consisted of an alternating condition (Alt-Cond), in which visual words of two different languages were presented alternately, and a repetition condition (Rep-Cond), in which words of one language were presented repeatedly (Fig. 1a). Neural responses to stimuli of the same category were attenuated in the Rep-Cond compared to Alt-Cond due to habituation and this RS effect has been examined to disentangle the neural activities underlying categorization of faces and body silhouettes of a specific social group (Pu and Han, 2025; Zhang and Han, 2021; Zhang et al., 2023b; Zhou et al., 2020).

In nine experiments we recorded EEG/MEG signals from Chinese, English, and German speakers when viewing words of an alphabetic language and a non-alphabetic language (English and Chinese words, or Italian and Korean words) or of two alphabetic languages (English and German) in the Rep-Cond and Alt-Cond. We recorded EEG signals from Chinese participants to examine temporal neural dynamics of spontaneous language-based word categorization in Experiment 1. The similar paradigm was employed in Experiments 2 and 3 to investigate whether perceptual

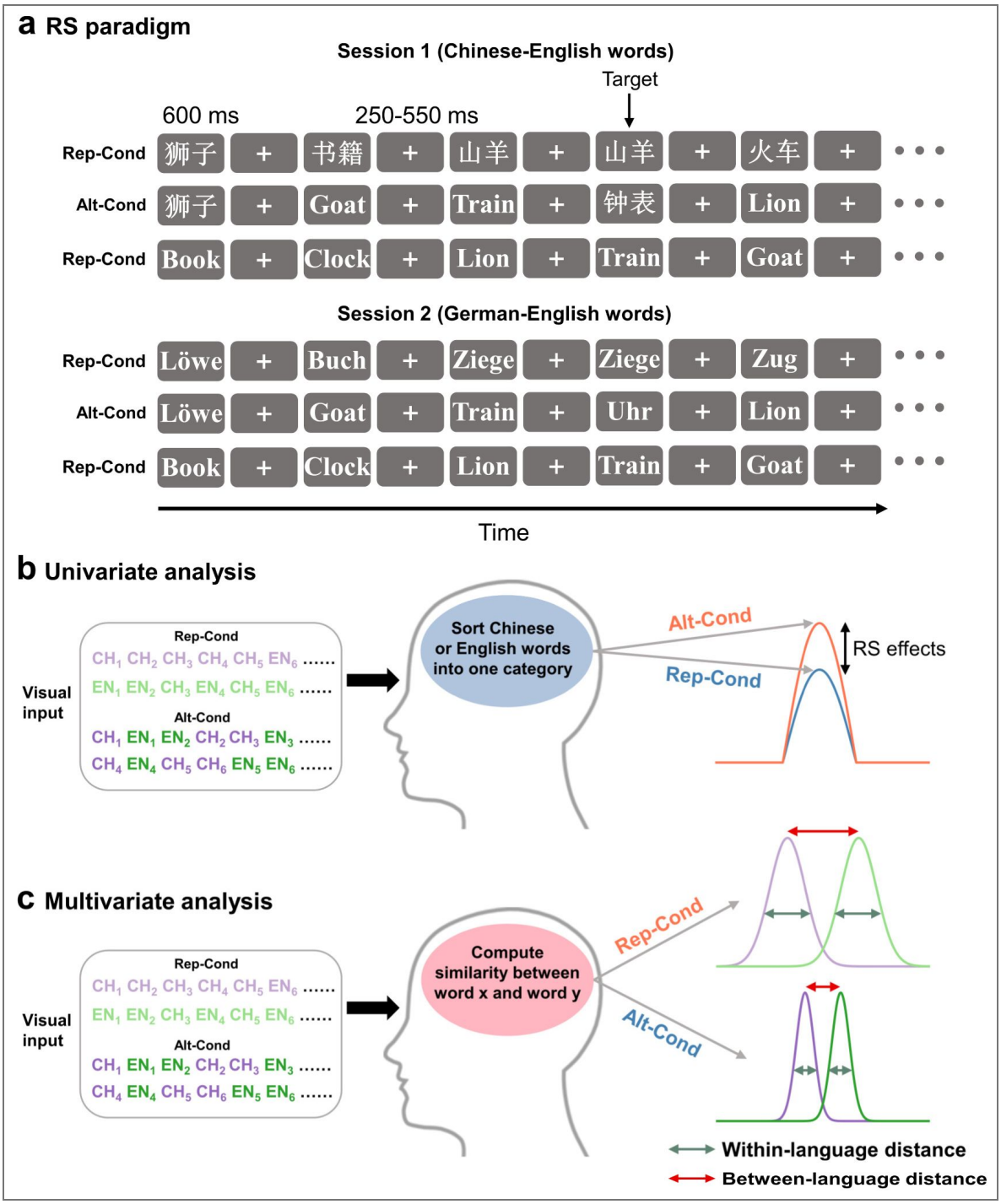


**Figure 1. Illustrations of the RS paradigm and univariate/multivariate analyses of brain activities in response to words.**

(a) The RS paradigm. Words of two different languages (e.g., Chinese vs. English, or English vs. German) were presented alternately in the Alt-Cond, and words of one language were presented repeatedly in the Rep-Cond. (b) Univariate analyses. The RS effect was quantified as the decrease of averaged neural responses to words of one language in the Rep-Cond vs. Alt-Cond.

features or radical/letters of words are sufficient to generate spontaneous language-based categorization of visual words. The results in Experiment 1 were replicated in native English and German speakers in Experiments 4 and 5, respectively. Neural dynamics of categorization of words of two unlearned languages were further investigated in Chinese participants in Experiment 6. Finally, the neural networks supporting the spontaneous categorization of words of two learned or unlearned languages were localized using MEG in Chinese and English speakers in Experiments 7-9, respectively.

We performed univariate analyses of the RS effects on EEG and MEG signals in response to words of the same language to examine both timing and architecture of the integrated neural activities related to language-based word categorization (Fig. 1b). Since visual categorization of objects or faces depends on the processing of both within-category similarity (or intra-group) and between-category (or inter-group) difference between perceived stimuli (Freedman et al., 2001; Ito and Bartholow, 2009; Kriegeskorte et al., 2008; Zhou et al., 2020), we further conducted multivariate representational similarity analyses (Kriegeskorte et al., 2006) of correlation distances between two words to disentangle neural processes of intra-language similarity and inter-language difference between words that are fundamental to language-based word categorization. The processing of intra-language similarity occurs when two words of the same language are perceived repeatedly with short interstimulus intervals.

Because words of the same language were repeatedly presented in the Rep-Cond and words of two different languages were displayed in the Alt-Cond, the processing of intra-language similarity occurred more frequently and would be inhibited in the Rep-Cond (vs. Alt-Cond) due to habituation (Fig. 1c). By contrast, the processing of inter-language difference takes place when two words of different languages are perceived with short interstimulus intervals. Since words of different languages appeared more frequently in the Alt-Cond (vs. Rep-Cond), we would expect RS of the processing of inter-language difference in the Alt-Cond (vs. Rep-Cond). The neural processing of intra-language similarity was quantified as correlation distances between neural responses to two words of the same language whereas the neural processing of inter-language difference was assessed as correlation distances between neural responses to two words of two different languages. The correlation distances from the multivariate analyses were further employed to assess how words of one language are clustered and how far words of two languages are separated in a two-dimensional (2D) space during language-based word categorization. Enhanced language-based word categorization is associated with smaller intra-language correlation distances, which reflect more densely clustered words of the same language, and larger inter-language correlation distances, which manifest further separated words of two different languages. This approach is different from previous research that focused on distinct neural responses to visual words of two different languages in the brain regions such as the fusiform/lingual gyri that are specific to word-form processing (Zhan et al., 2023).

This RS effect manifested habituation of the integrated neural activities that supported classification words into one or another category and occurred more frequently in the Rep-Cond (vs. Alt-Cond). (c) Multivariate analyses. Correlation distances between neural responses to two words were calculated to estimate how words of one language were clustered (i.e., intra-language similarity) and how words of two languages were separated (i.e., inter-language difference) during language-based word categorization.

# Results

## Temporal dynamics of spontaneous language-based word categorization

To examine temporal characteristics of the neural processes involved in spontaneous language-based word categorization, in Experiment 1, we recorded EEG in two sessions from native Chinese speakers who learned English but not German (N=34, see Table S1 for information about all participants in our work). In Session 1 Chinese words (animal and tool names) and English words (translated from the Chinese words, see Table S2 for all the stimuli used in our study) were

presented in the Alt-Cond and Rep-Cond, respectively (Fig. 1a). The same design was applied to English words and German words (translated from the Chinese words) in Session 2. Participants responded to a casual target word that was presented repeatedly in two consecutive trials by pressing a button during EEG recording. This one-back task required neither processing of linguistic properties nor intentional classification of words and thus allowed us to examine neural activities underlying spontaneous language-based word categorization. Behavioral performances in the one-back task did not differ significantly between words of different languages (see Table S3 for details of behavioral results in all studies), indicating comparable attentional demand and task difficulty in detection of target words of different languages.

Event-related brain potentials (ERPs) in response to non-target words were characterized by an early negative activity at 90–140 ms (N1), a positive activity at 140–280 ms (P2), and a late negative activity at 280–350 ms (N2) over the frontal/central electrodes, a long-latency positivity at 400–600 ms (LPP) at the centro-parietal electrodes, and a negative activity at 150–210 ms (N170) at the occipito-temporal electrodes (Fig. 2a and 2b, Fig. S1). We conducted univariate whole-brain cluster-based permutation *t*-tests to examine the RS effect on neural responses to non-target words (i.e., decreased amplitudes in the Rep-Cond vs. Alt-Cond). The results in Session 1 revealed significant clusters showing the RS effects over the middle frontal/central/parietal regions at 134–320 ms for Chinese words and at 172–300 ms for English words (a pre-defined threshold of $P < 0.05$, cluster-level $P < 0.05$, two-tailed, 10,000 iterations). However, similar analyses of brain activities in Session 2 did not show reliable RS of neural responses to English and German words (Fig. S2). These results provided electrophysiological evidence for spontaneous language-based categorization of words as early as 150 ms after stimulus onset between an alphabetic language and a nonalphabetic language (i.e., English and Chinese) but not between two alphabetic languages (i.e., English and German).

Next, we conducted multivariate representational similarity analyses of the neural responses to words to examine the neural processes of intra-language similarity and inter-language difference between words, respectively. A 160 × 160 neural representation dissimilarity matrix (RDM), including 40 words from each language in the Alt-Cond and Rep-Cond, was constructed using the EEG data at the electrodes at which the ERP amplitudes showed the significant RS effects in the univariate analyses (Fig. 2c and 2d). This RDM consisted of a matrix of correlation distances (calculated as 1–Pearson correlation coefficients) between neural responses to two words of the same language, which represents the neural response pattern of the processing of intra-language similarity (a smaller intra-language correlation distance corresponds to a more clustered representation of words of the same language). The RDM also had a matrix of correlation distances between neural responses to two words of different languages, which represents a neural response pattern of the processing of inter-language difference (a larger inter-language correlation distance corresponds to greater separation of representations of words of two languages). As predicted, permutation *t*-tests showed that the mean correlation distance of the RDMs corresponding to intra-language similarity was significantly increased in the Rep-Cond (vs. Alt-Cond) at 180–206 ms for Chinese words and at 168–226 ms for English words (Fig. 2e-g, see Fig. S3 for these RS effects at the individual level), indicating weakened clustered representations of Chinese (or English) words in the Rep-Cond (vs. Alt-Cond) due to habituation. By contrast, the mean correlation distance of the RDMs corresponding to inter-language difference was significantly reduced in the Alt-Cond (vs. Rep-Cond) at 162–218 ms, indicating more closed representations of Chinese and English words in the Alt-Cond (vs. Rep-Cond) due to habituation.

Because these neural RS effects were consistently observed within 300 ms, the following multivariate analyses focused on the results in this time window.

To further assess the neural categorical representations of Chinese and English words in the Alt-Cond and Rep-Cond, we conducted multidimensional scaling analyses of the 160 × 160 RDMs averaged in the time window of the significant RS effects. The first two components of the results of these analyses were then used to construct 2D word spaces in which words of the same language or of the two different languages were plotted. As can be seen in Fig. 2e-g, words of the same language are clustered more densely in the Alt-Cond (vs. Rep-Cond) whereas words of the

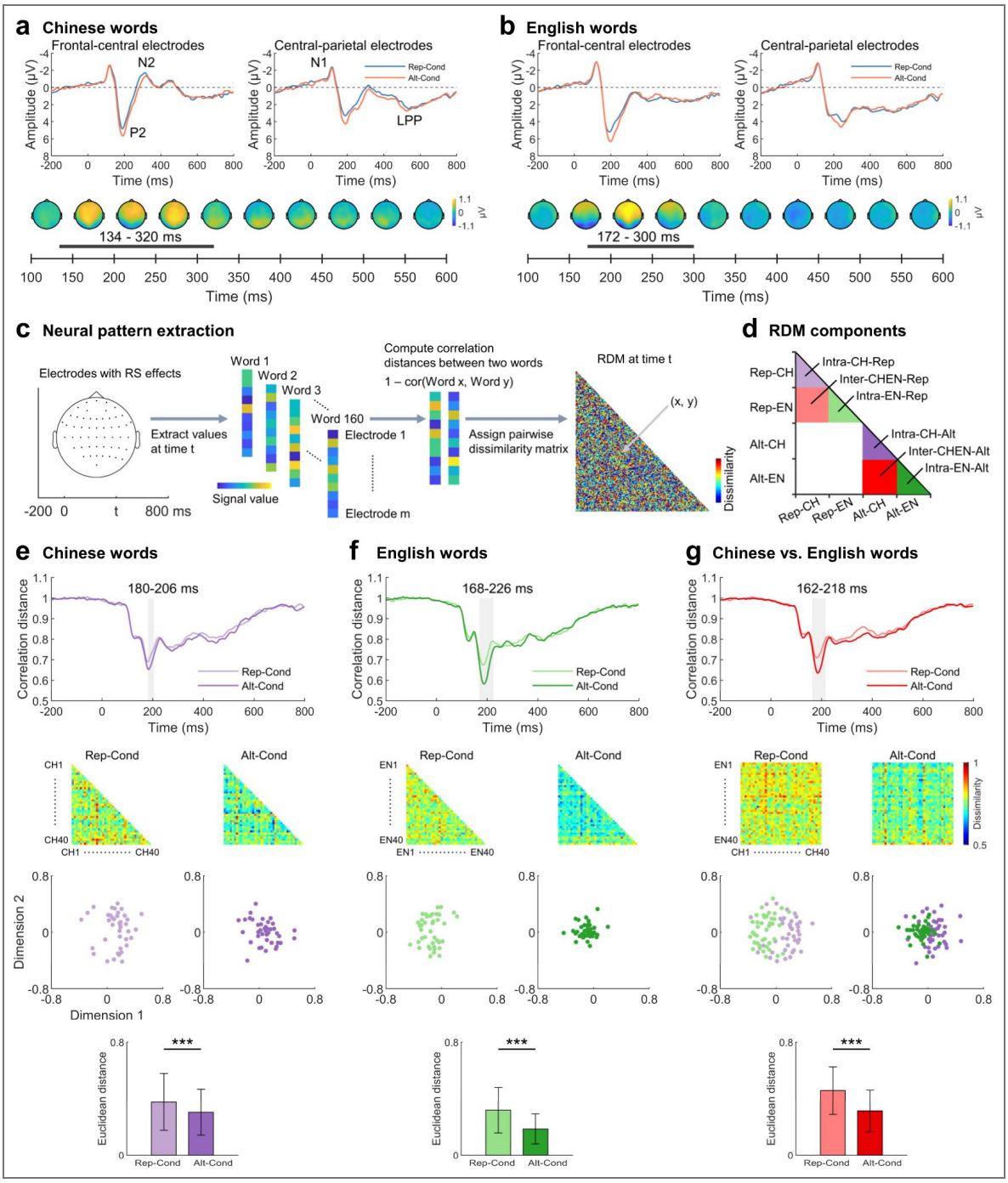


**Figure 2. EEG results of Chinese speakers in Experiment 1.**

(a) and (b) Results of univariate analyses. Top panels illustrate electrophysiological responses to Chinese and English words in the Alt-Cond and Rep-Cond, respectively. Bottom panels show scalp distributions of significant RS effects on neural responses to words of each language. (c) Illustration of the procedure of computing neural RDMs in the multivariate analyses of correlation distances between words. (d) Illustration of the 160 × 160 neural RDM. Triangles represent the neural RDMs corresponding to intra-language similarity. Squares represent the RDMs corresponding to inter-language difference. (e), (f), and (g) Results of multivariate and multidimensional scaling analyses. The top two panels show the time courses of significant differences in correlation distances between words corresponding to intra-language similarity and inter-language difference between the Alt-Cond and Rep-Cond and the neural RDM in the two conditions, respectively. The bottom two panels illustrate clustered representations of the words in the 2D word space built based on the first two dimensions of multidimensional scaling analyses of neural RDMs corresponding to intra-language similarity and inter-language difference, respectively, and the mean Euclidean distances in the 2D word space between two words of the same language and between two words of different languages. *** $P < 0.001$.

two different languages separate more distantly in the Rep-Cond (vs. Alt-Cond). These differences were further quantified by comparing the mean Euclidean distances in the word spaces between two words of the same language and between two words of different languages in the Alt-Cond and Rep-Cond. Together, the results of our multivariate analyses established two neural processes of intra-language similarity and inter-language difference that took place spontaneously during categorization of Chinese and English words in Chinese speakers.

## Perceptual features or radical/letters of words are not sufficient to generate spontaneous language-based categorization of words

Did the neural RS effects observed in Experiment 1 manifest habituation of the processing of non-specific perceptual shape features of words? We clarified this issue in Experiment 2 by creating two sets of scrambled stimuli from the Chinese and English words used in Experiment 1. The scrambled stimuli possess both global and local shape features of the words but lack word-specific information (e.g., semantic meanings and language categories). We recorded EEG signals in response to the scrambled stimuli from an independent sample of Chinese speakers (N=34) using the same experimental procedure as that in Experiment 1. Whole-brain cluster-based permutation *t*-tests did not find any significant RS effect on neural responses to the scrambled stimuli (Fig. S4), indicating that perceptual features of visual words contribute little to the neural RS effects related to spontaneous categorization of Chinese and English words.

In Experiment 3 we further tested whether categorization of radicals of Chinese words and letters of English words engages similar neural processes as those involved in categorization of Chinese and English words. We recorded EEG signals in response to Chinese radicals and English letters from an independent sample of Chinese speakers (N=34) using the same experimental procedure as that in Experiment 1.

Univariate whole brain cluster-based permutation *t*-tests showed significant RS effects on the ERP amplitudes to Chinese radicals at 172–286 ms over the central region and to English letters at 214–364 ms over the occipital regions (Fig. S5a and S5b).

Multivariate analyses of the correlation distances corresponding to intra-radical similarity did not show any significant difference between the Rep-Cond and Alt-Cond. The mean correlation distances corresponding to intra-letter similarity and inter-radical-letter difference were significantly decreased in the Rep-Cond than Alt-Cond but in time windows delayed (after 230 ms) compared to those observed for Chinese and English words in Experiment 1 (Fig. S5c-e). These results provide no evidence that perception of the middle-level units of Chinese and English words (i.e., radicals and letters) employs the same early fronto-central neural processes as those involved in spontaneous categorization of Chinese and English words.

## Neural dynamics of language-based word categorization is independent of people's language backgrounds

To generalize the findings in Experiment 1 to populations of different language backgrounds, we recorded EEG signals to words from native English speakers (Experiment 4, N=34) who learned Chinese (but not German) and native German speakers (Experiment 5, N=34) who learned both Chinese and English. The stimuli and procedure were the same as those in Experiment 1. The results of both univariate and multivariate analyses replicated those observed in Experiment 1 (see Fig. S6–S9 for details), indicating similar neural processes involved in language-based words categorization regardless of speakers' language proficiency and learning experiences.

## Neural dynamics of categorization of words of two unlearned languages

So far, the neural RS effects in Experiments 1, 4, and 5 were observed for words of two learned languages. In Experiment 6 we further investigated to what degree semantic meanings and pronunciations of words contributed to the neural processes of intra-language similarity and

inter-language difference during spontaneous language-based word categorization. We recorded EEG signals in response to Korean and Italian words (see Table S2) from an independent sample of Chinese speakers (N=34) who had not learned Korean and Italian when being tested. The experimental procedure was the same as that in Experiment 1. The participants were informed of viewing Korean and Italian words during EEG recording but did not know semantic meanings of these words and were unable to pronounce the Korean words (though might be able to pronounce Italian words in the way to spell English words since they had learned English). If semantic meanings or pronunciations of words are necessary for spontaneous language-based categorization of words, the neural RS effects would not occur to Korean or Italian words.

Cluster-based permutation *t*-tests of the ERP amplitudes to non-target revealed significant RS effects over the middle frontal/central/parietal regions at 136–310 ms for Korean words and at 156–514 ms for Italian non-target words (Fig. S10a and S10b). Multivariate analyses of neural responses to words showed a significantly increased mean correlation distance corresponding to intra-language similarity in the Rep-Cond (vs. Alt-Cond) at 164–216 ms for Korean words and at 166–246 ms for Italian words (Fig. S10c-e). Moreover, the mean correlation distance corresponding to inter-language difference was significantly reduced in the Alt-Cond (vs. Rep-Cond) at 164–236 ms. Similarly, the multidimensional scaling analyses of the RDMs revealed more densely clustered representations of words of the same language in the Alt-Cond (vs. Rep-Cond) and more distantly separated representations of words of the two different languages in the Rep-Cond (vs. Alt-Cond) in the 2D word space. These results demonstrated that spontaneous language-based categorization also occurred to words of two unlearned languages. Furthermore, the time courses of neural processes of intra-language similarity and inter-language difference involved in categorization of Korean and Italian words were akin to those of categorization of words of two learned languages (i.e., Chinese and English). These results indicate that semantic meanings or pronunciations of words contribute little to spontaneous categorization of words of an alphabetic language and a non-alphabetic language.

## A neural network underlying language-based categorization of words

Next, we sought to localize the neural network underlying spontaneous language-based categorization of words in Experiments 7a and 8a. We recorded 306-channel, whole-head anatomically constrained MEG signals in response to Chinese and English words from two independent samples of native Chinese and English speakers (N=34 in each group). High-resolution structural MRI was combined with temporally precise whole-head high-density MEG to localize brain regions in which activities showed RS effects and to examine the functional roles of this network in processing intra-language similarity and inter-language difference between words. The stimuli and procedures were the same as those used in Experiment 1. Because our EEG data analyses showed consistent neural RS effects in Chinese and English speakers, we combined Chinese and English speakers' MEG data to examine the neural RS effects. We first assessed time courses of the RS effect on sensor-space MEG signals to words by pooling across Chinese and English words. A whole-brain cluster-based permutation *t*-test of sensor-space MEG signals in the Rep-Cond and Alt-Cond revealed three significant clusters (magnetometer signals:136–278 ms and 151–268 ms; gradiometer signals: 154–257 ms, using a predefined threshold of $P < 0.01$, two-tailed, 10,000 iterations, and a cluster-level threshold of $P < 0.05$, Fig. 3a). These results provide MEG replication of our EEG findings of the temporal dynamics of spontaneous language-based word categorization.

To probe the neural network involved in language-based categorization of words, we performed source reconstruction of MEG signals and then combined source-space MEG signals to Chinese and English words across Chinese and English speakers to search for brain regions in which responses to words of the same language were suppressed in the Rep-Cond (vs. Alt-Cond). The results of this univariate analysis revealed significant RS effects on activities in the bilateral ATLs, insula, OFC, occipito-temporal cortices, and the LSTS at a predefined threshold of $P < 0.001$, 10,000 iterations, and a cluster-level threshold of $P < 0.05$, one-tailed (Fig. 3b). To further assess the time courses

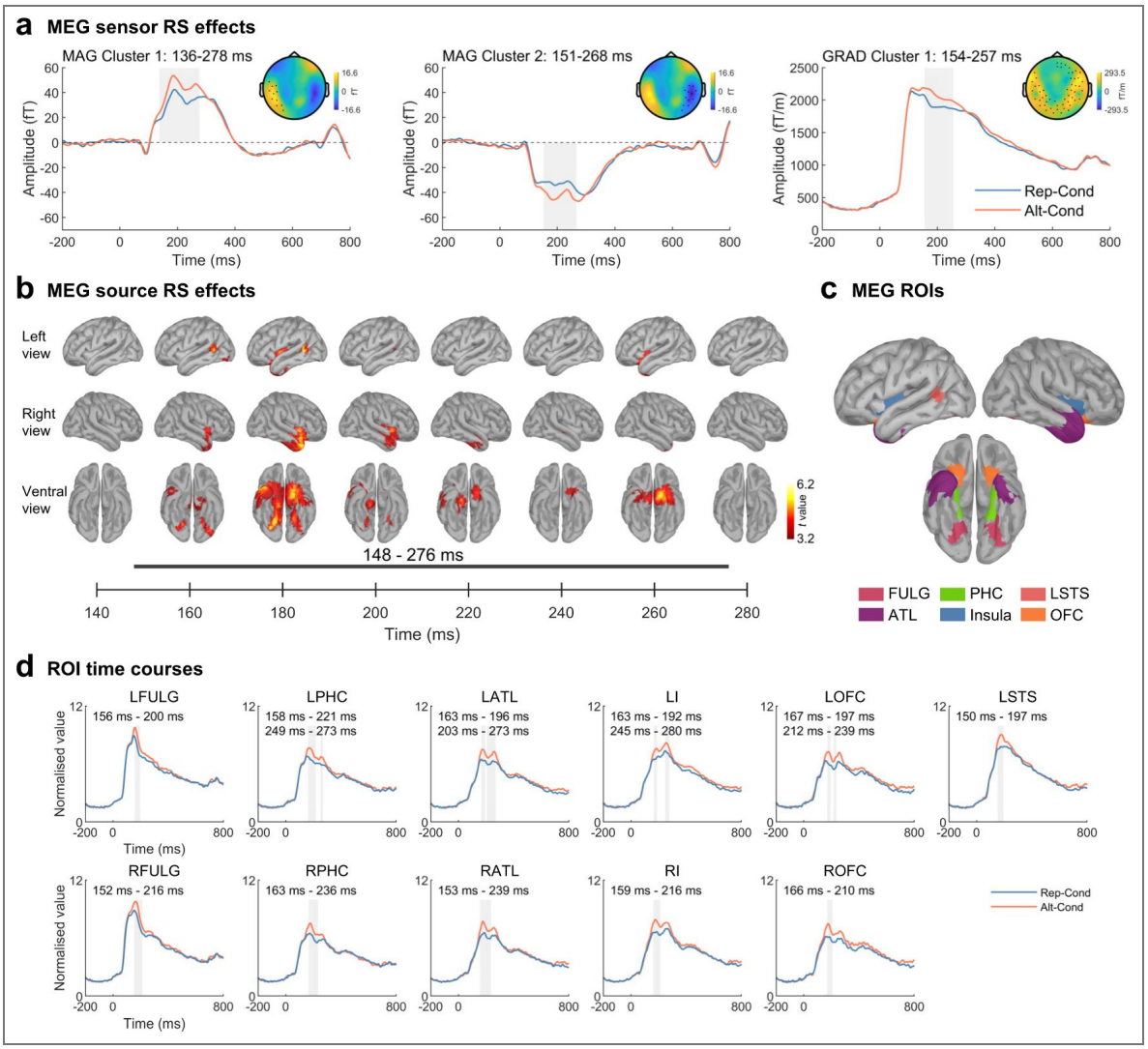


**Figure 3. MEG results combined Chinese and English words across Chinese and English speakers in Experiments 7a and 8a.**

(a) Significant clusters of RS effects on magnetometer signals (the left two panels) and gradiometer signals (the right panel). (b) Results of the whole-brain source analyses. Shown are left, right, and ventral views of brain regions in which MEG signals in response to words of the same language decreased significantly in the Rep-Cond (vs. Alt-Cond). (c) Illustration of the regions of interest used in the following MEG data analyses. (d) Source-space MEG signals in the brain regions showing significant RS effects. Time windows of the significant clusters are shown for each brain region. LFULG=left fusiform and lingual gyrus; RFULG= right fusiform and lingual gyrus; LPHC=left parahippocampal cortex; RPHC=right parahippocampal cortex; LATL=left anterior temporal lobe; RATL=right anterior temporal lobe; LI=left insula; RI=right insula; LOFC=left orbital frontal cortex; ROFC=right orbital frontal cortex; LSTS=left superior temporal sulcus;

and hemispheric asymmetry of the neural RS effects in different nodes of this network, we defined eleven regions of interest (ROIs) based on the intersection of brain regions showing significant RS effects and the Desikan-Killiany-Tourville atlas (Klein and Tourville, 2012), including the ATL, insula, OFC, fusiform and lingual gyrus (FULG), parahippocampal cortex (PHC) in both hemispheres, and LSTS (Fig. 3c). Cluster-based permutation *t*-tests of the activities in these brain regions within 300 ms after word onsets identified significant RS effects at 150–280 ms (Fig. 3d). The neural RS effects started slightly earlier in the right than left hemispheres (except the PHC), though a repeated measures analysis of variance (ANOVA) of the peak latency of the RS effect with Hemisphere (left vs. right) and Brain regions (OFC, insula, ATL, FULG, PHC) as independent within-subjects factors did not find a significant difference between the two hemispheres ($P > 0.9$). We also compared the magnitude of the RS effects in the two hemispheres by conducting ANOVA of the peak RS effects with Hemisphere (left vs. right) and Brain regions (OFC, insula, ATL, FULG, PHC) as independent within-subjects factors but failed to find a significant difference in the neural RS effects between the two hemispheres ($P > 0.4$). These results provide no evidence for dominance of one over the other hemisphere during language-based word categorization.

To further explore the functional role of this network in processing of intra-language similarity and inter-language difference between words, we constructed global network RDMs based on the correlation distances calculated from 313 subregions (each subregion had 5 vertices on average) from the eleven ROIs shown in Fig. 5c (using the mean source-space MEG signals in each subregion, see Methods for details, Fig. 4a). If this neural network serves the processing of intra-language similarity during language-based categorization of words, the network RDMs corresponding to intra-language similarity would be attenuated in the Rep-Cond (vs. Alt-Cond) whereas the network RDMs corresponding to inter-language difference would show a reverse pattern due to habituation. We tested these predictions by performing permutation *t*-tests of the time courses of the RDMs. The results revealed that the correlation distance corresponding to intra-language similarity was significantly increased in the Rep-Cond (vs. Alt-Cond) at 159–200 ms for Chinese words and at 151–198 ms for English words. Moreover, the correlation distance corresponding to inter-language difference was significantly reduced in the Alt-Cond (vs. Rep-Cond) at 150–232 ms and 236–294 ms (Fig. 4b-d, see Fig. S11 for these RS effects at the individual level). The multidimensional scaling analyses of the network RDM in the time windows of these significant clusters further unraveled more densely clustered representations of words of the same language in the Alt-Cond (vs. Rep-Cond) and more distantly separated representations of words of the two different languages in the Rep-Cond (vs. Alt-Cond) in the 2D word space. Similar results were obtained in separate analyses of MEG data in Experiments 7a and 8a, respectively (see Fig. S12 and S13). These results identified the word-categorization network that supported the processing of both intra-language similarity and inter-language difference during categorization of Chinese and English words.

We further assessed dynamic contributions of each brain region to the processing of intra-language similarity and inter-language difference during language-based word categorization. We computed local RDMs using the patterns of activity in each of the 11 ROIs, respectively (see Methods for details). The mean correlation distance values specific to intra-language similarity and inter-language difference were calculated and assigned to all vertices in each ROI, as illustrated in Fig. 5a. We then performed whole-network cluster-based permutation *t*-tests to examine the dynamic RS effects on the mean correlation distance values (a predefined threshold of $P < 0.025$, 10,000 iterations, and a cluster-level threshold of $P < 0.05$, one-tailed). The results showed that the RS effects related to intra-language similarity initiated at 148 ms in the FULG for Chinese words and at 156 ms in the RATL for English words, and these early RS effects spread to other brain regions in the network (Fig. 5b and 5c). The RS effects pertaining to inter-language difference started at 140 ms in the LATL and then spread to other brain regions in the network (Fig. 5d). These results suggested distinct patterns of neural dynamics linked to the computations of intra-language similarity and inter-language difference.

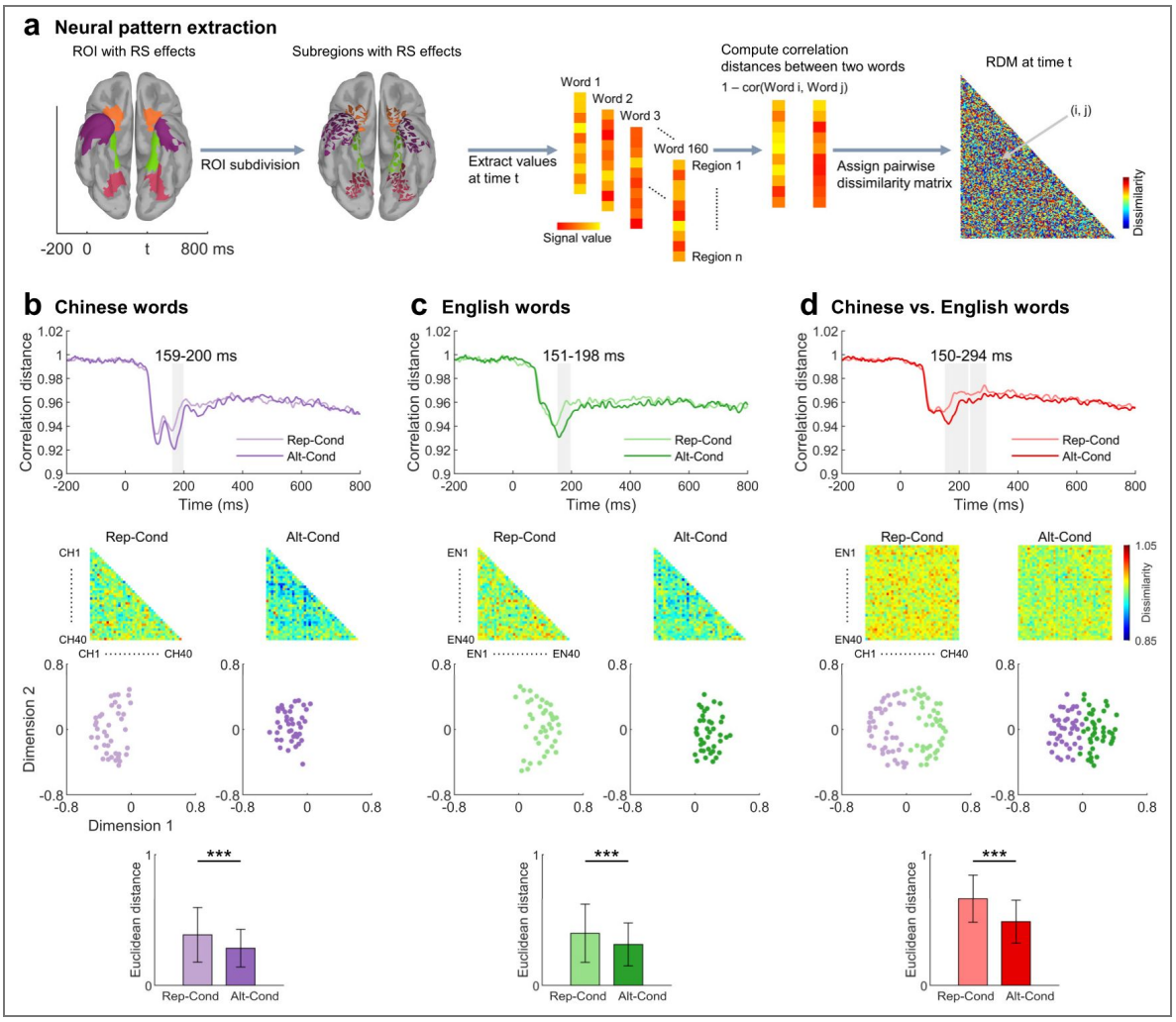


**Figure 4. Results of the analyses of the global network RDM across Chinese and English speakers in Experiments 7a and 8a.**

(a) Illustration of the procedure to calculate network RDMs based on source-space MEG signals. (b), (c), and (d) Results of multivariate analyses. The top two panels show the time courses of significant differences in the correlation distances corresponding to intra-language similarity and inter-language difference between the Alt-Cond and Rep-Cond and the neural RDMs in the two conditions, respectively. The bottom two panels illustrate clustered representations of words in the 2D word space built based on the first two dimensions of multidimensional scaling analyses of neural RDMs corresponding to intra-language similarity and inter-language difference, respectively, and the mean Euclidean distances in the 2D word space between two words of the same language and between two words of different languages. ***$P < 0.001$.

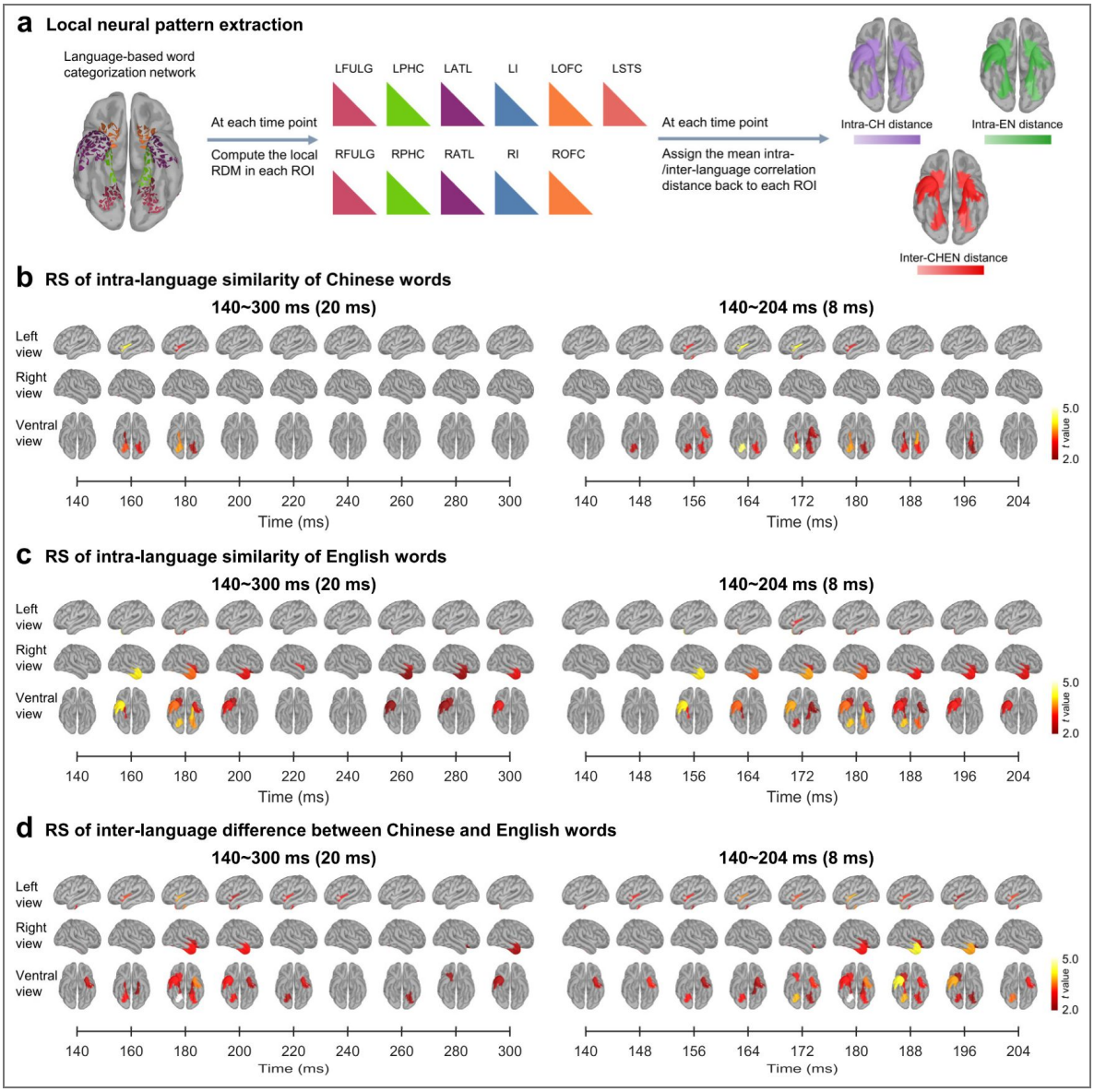


**Figure 5. Results of the analyses of local RDMs across Chinese and English speakers in Experiments 7a and 8a.**

(a) Illustration of the procedure to calculate local RDMs based on source-space MEG signals. (b) and (c) Dynamic RS effects related to intra-language similarity of Chinese and English words in different ROIs. (d) Dynamic RS effects related to inter-language difference in different ROIs.

We further performed a Granger causality analysis (GCA) (Barnett and Seth, 2014) to examine functional connectivity characteristics of the word-categorization network during the processing of intra-language similarity and inter-language difference between words. To this end, we calculated the time series of the RS effects on correlation distances between two words during 100–300 ms in each cell of the network RDM corresponding to intra-language similarity or inter-language difference in the eleven ROIs (Fig. 4c). These time series were then subject to GCA to estimate information flow among the ROIs during language-based word categorization (see Methods for details). The results revealed three connectivity characteristics of the network (Fig. 6a-c). First, functional connections occurred dominantly within the same hemisphere and no cross-hemisphere functional connection passed the same threshold. Second, forward and backward connections existed between adjacent brain regions (e.g., FULG and PHC, or ATL and OFC) whereas long-distance connections were rare. Third, while similar patterns of functional connections were observed for the processing of intra-language similarity and inter-language difference between words, the connectivity strength in the whole network was slightly but significantly greater during processing of intra-language similarity relative to inter-language difference. Finally, the connections were significantly stronger in the left than right hemispheres in connections from FULG to PHC, ATL to PHC, insula to ATL, and OFC to ATL, whereas the connection from ATL to insula was significantly stronger in the right than left hemispheres (*Ps* < 0.05, false discovery rate (FDR) corrected).

To further test the functional roles of neural computations of intra-language similarity and inter-language difference between words in language-based word categorization, we asked the participants in Experiments 7a and 8a to perform an explicit categorization task that required behavioral classification of Chinese and English words (Experiments 7b and 8b, see Methods for details). The participants sorted Chinese and English words with high accuracies and fast responses by pressing one of two buttons (Experiments 7b and 8b: accuracy = 96.04 ± 0.03% and 96.20 ± 0.03%, reaction time = 521 ± 64.3 and 481 ± 45.3 ms). If neural computations of intra-language similarity and inter-language difference are conducted spontaneously during both implicit and explicit language-based word categorization, intra-language similarity and inter-language difference computed in the Alt-Cond in an implicit task (i.e., the one-back task) would predict those computed in the explicit categorization task. In addition, participants with better behavioral classification performances in the explicit categorization task would be associated with smaller intra-language similarity (i.e., stronger clustered neural representations of words of one languages) but a larger inter-language difference (i.e., larger separation of neural representations of words of two languages).

To test these predictions, we extracted the neural activities from the word-categorization network identified in Experiments 7a and 8a during the explicit word categorization task in Experiments 7b and 8b. The correlation distances corresponding to inter-language difference and intra-language similarity were then calculated based on these activities in the time windows showing significant RS effects in the multivariate analyses in Experiments 7a and 8a. We then calculated a neural index (**δ**) that combined the contributions of inter-language difference and intra-language similarity to language-based word categorization:

**δ** = Inter_LD – (Intra_$LS_{CH}$ + Intra_$LS_{EN}$)/2 in which Inter_LD = inter-language difference; Intra_$LS_{CH}$ = intra-language similarity of Chinese words; Intra_$LS_{EN}$ = intra-language similarity of English words. A larger **δ** indicates more clustered neural representations of words of the same language and more separated neural representations of words of different languages. Behavioral performances during the explicit word categorization task were quantified by the mean response accuracy, mean reaction time, and the inverse efficiency score (IES, i.e., the ratio of the mean reaction time to response accuracy) that considered both the response accuracy and reaction time.

Linear regression analyses including Group (Chinese or English speakers) as a covariate showed that the correlation distances corresponding to intra-language similarity, inter-language difference and **δ** in the Alt-Cond in the one-back task in Experiments 7a and 8a positively predicted those in the explicit word categorization task in Experiments 7b and 8b (*Ps* < 0.001, Fig. 6d). These results suggest similar neural computations of intra-language similarity and inter-language

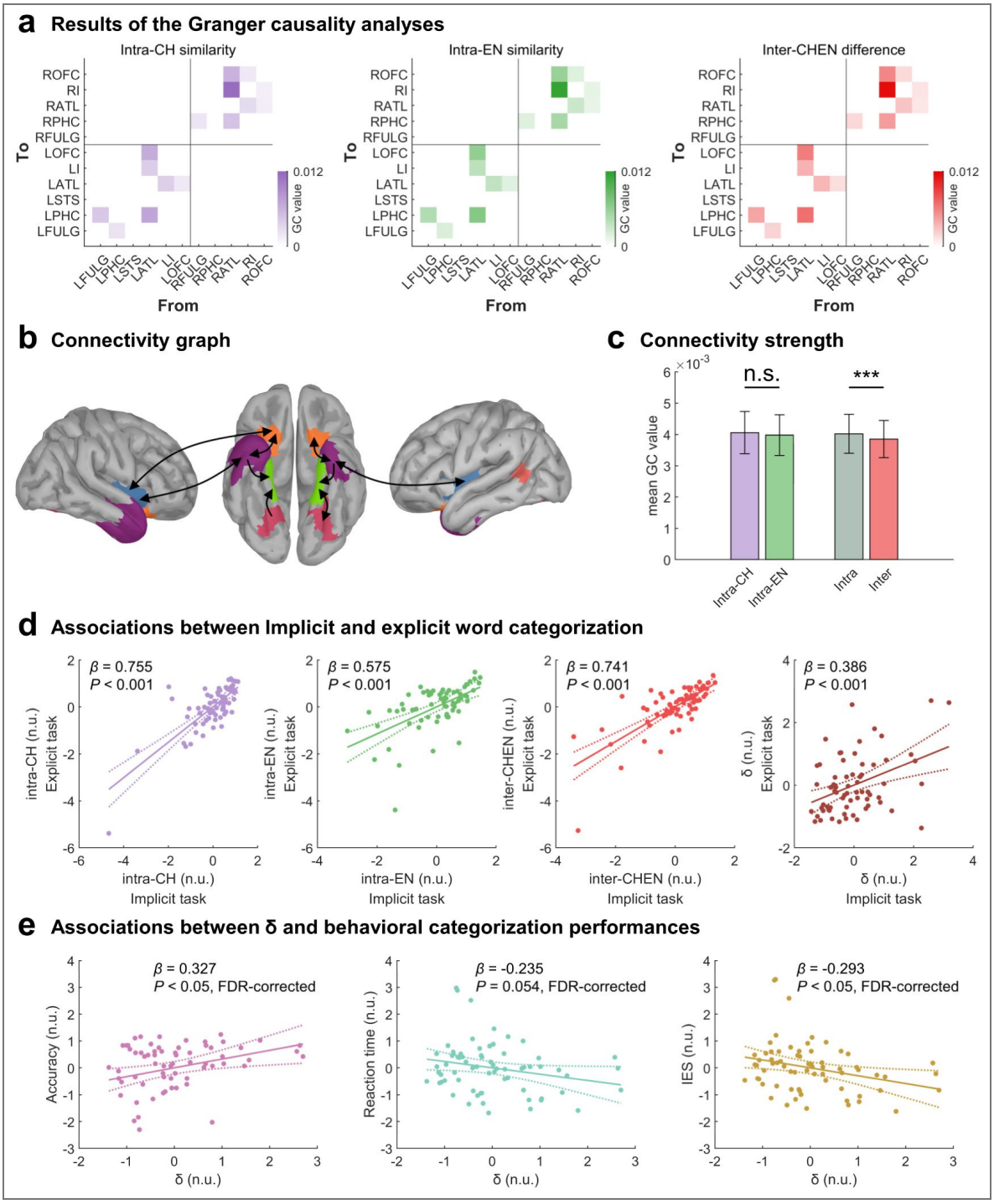


**Figure 6. Results of GCA and linear regression analyses across Chinese and English speakers in Experiments 7a and 8a.**

(a) Connectivity values above a threshold (GC values > 0.001) corresponding to intra-language similarity and inter-language difference. Each square represents connectivity from one brain region indicated by the x-axis to another region indicated by the y-axis. (b) Patterns of functional connectivity between two ROIs. Arrows indicate directions of information flow from one to another brain region. Lines with two arrows indicate information flow with mutual directions. (c) The mean connectivity strength in the whole word-categorization network related to the processing of intra-language similarity and inter-language difference. (d) Associations between the correlation distances corresponding to inter-language difference and intra-language similarity, and the neural index of language-based word categorization (**δ**) during the implicit and explicit language-based word categorization tasks. Shown are partial regression plots (e) Associations between the neural index of language-based word categorization (**δ**) and behavioral performances (response accuracies, reaction time and inverse efficiency scores (IES)) during the explicit language-based word categorization tasks. *** $P < 0.001$; n.s. = no significance.

difference in the same neural network during implicit and explicit language-based word categorization. More importantly, a larger **δ** calculated using the neural network responses to words in the explicit language-based word categorization significantly and positively predicted better behavioral categorization performances (i.e., higher accuracies: $P < 0.05$; shorter reaction times: $P = 0.054$, and smaller IES: $P < 0.05$, FDR-corrected, Fig. 6e), in Experiments 7b and 8b. However, the **δ** calculated using the source-space MEG signals in each of the eleven ROIs of the word-categorization network failed to predict behavioral performances during explicit language-based word categorization. These results indicate that the integrated computations of intra-language similarity and inter-language difference supported by the whole word-categorization network (rather than by a single region of the word-categorization network) provide a neural mechanism of behavioral categorization of words of an alphabetic language and a non-alphabetic language.

We also conducted a disruption analysis to examine which node of the word-categorization network was necessary for predicting behavioral categorization of words of two languages. To this end, we computed the **δ** of the word-categorization network by removing the bilateral FULG, PHC, ATL, insula, OFC, and LSTS, respectively. Thereafter, we conducted linear regression analyses to test whether **δ** of the disrupted network predicted behavioral categorization of words. The results showed that the association between **δ** and behavioral categorization of words was significant except when the bilateral FULGs were removed from the network (see Fig. S14), suggesting that the FULG may be more critical than the other nodes of the network in supporting behavioral performances during language-based categorization of words.

### The word-categorization network functions independently of words' linguistic properties

Finally, we tested whether the neural network identified in Experiments 7a and 8a also serves language-based categorization of words of two unlearned languages in Experiment 9. We recorded MEG signals from an independent sample of Chinese speakers (N=34) who had not learned Korean and Italian when being tested. The stimuli and procedure were the same as those in Experiment 7 except that Korean and Italian words were used. The results in Experiment 9 replicated the results in Experiments 7a and 8a (Fig. S15–S17). These results provided further evidence that the word-categorization neural network appears to support the processing of intra-language similarity and inter-language difference during spontaneous categorization of words of an alphabetic language and a non-alphabetic language independently of the processing of words' linguistic features (e.g., pronunciations or semantic meanings).

## Discussion

Together, our EEG/MEG findings uncovered the dynamic neurocognitive mechanisms of spontaneous language-based categorization of words. Importantly, our findings established a neural network of which dynamic activities underlie the processing of intra-language similarity and inter-language difference between words during language-based categorization of visual words. This network does not include the left middle and inferior frontal cortices which are fundamental for the processing of semantic, syntactic, and phonological information of words (Fedorenko et al., 2024a; Hodgson et al., 2021; Tan et al., 2005) but did not show neural RS effect related to language-based word categorization (see Fig. S18 for results). The word-categorization network is different from the language switching network that consists of the caudate (Crinion et al., 2006), lateral frontal cortices (Rodriguez-Fornells et al., 2005; Wang et al., 2007; Zhu et al., 2020), and anterior cingulate cortex (Abutalebi et al., 2007; Blanco-Elorrieta et al., 2018; Wang et al., 2007). The strength of the RS effects in the bilateral hub regions did not show dominance of one over the other hemisphere. This network includes the FULG in both hemispheres, unlike the visual word form area (VWFA) which predominates in the left hemisphere (Dehaene and Cohen, 2011; Zhan et al., 2023). While the PHC in the left hemisphere may contribute to verbal memory (Alessio et al., 2006), we showed that the PHC in both hemispheres were engaged in language-based categorization of visual words. The bilateral

ATL and the LSTS are activated during comprehension of sentences requiring retrieval of information from autobiographical, emotional, and episodic memory (Ferstl et al., 2008). In our work, however, these regions were spontaneously involved during language-based word categorization in a perceptual task performed on a single word that did not demand any complicated memory processing. The insula and OFC can be activated by words of negative or positive valence (Alia-Klein et al., 2007; Gu and Han, 2007) or of object names with strong olfactory associations (Han et al., 2020).

The words used in our study, however, refer to names of tools or animals without obvious emotional contents. Our GCA results showed similar patterns of forward and backward connections between adjacent brain regions in the word-categorization network in both hemispheres, though the strength of connectivity between adjacent brain regions was slightly different between the two hemispheres.

The word-categorization network identified serves two key processes that support classification of words of two languages i.e., the computations of both intra-language similarity and inter-language difference between words. The neural processes of intra-language similarity and inter-language difference occurred spontaneously during language-based categorization of words because the one-back task employed in our study did not require explicit processing of linguistic properties of words or explicit classification of words of two different languages. However, the neural correlation distances corresponding to intra-language similarity and inter-language difference were highly correlated in the one-back and explicit classification tasks. Moreover, the neural index that combined intra-language similarity and inter-language difference (**δ**) predicted behavioral response efficiencies during an explicit word classification task. These results indicate that the two neural computations take place irrespective of task demands and provide a neural basis of behavioral performances during language-based word categorization.

Importantly, neural computations of intra-language similarity and inter-language difference occurred as early as 150 ms and completed within 300 ms after word onset, which were earlier than those involved in categorization of words in terms of semantic concepts (Giari et al., 2020) and different from the early word form processing that occurs dominantly in the left ventral occipito-temporal cortex around 170 ms after stimulus onset (Marinkovic et al., 2003; Nan et al., 2022; Pylkkänen and Marantz, 2003; Thesen et al., 2012). Other brain regions are engaged in a later time window (200–400 ms), including the left temporal cortex and ATL, and bilateral inferior prefrontal cortices and OFC (Marinkovic et al., 2003), which support semantic and phonological processing of words and sentences (Hodgson et al., 2021; Zhu et al., 2022). Our findings highlight rapid neural computations engaged in spontaneous language-based categorization of words which are independent of the processing of linguistic (e.g., semantic and phonological) information of words.

The word-categorization network serves spontaneous categorization of words of typologically diverse languages (i.e., Chinese, English, Korean, Italian) similarly in Chinese, English, and German speakers and for learned (e.g., Chinese and English) and unlearned languages (Korean and Italian). The human brain may evolve the network to enable spontaneous categorization of words of an alphabetic language and a non-alphabetic language as salient symbols of different social group identities. Indeed, some nodes of the word-categorization network have been found to respond to symbols of outgroup versus ingroup members (e.g., insula) (Merritt et al., 2021), underlie social categorization of faces and representations of social concepts (e.g., fusiform, ATLs) (Golby et al., 2001; Pobric et al., 2016; Zhou et al., 2020), and encode social information about others (e.g., OFC) (Park et al., 2020).

Chinese/Korean and English/Italian are languages that represent the two major cultural societies in the world (i.e., East Asian and Western Europe). Fast and spontaneous categorization of words of an alphabetic language and a non-alphabetic language as symbols of different cultural groups may provide a pivotal cognitive basis for classification of people for appropriate real-life social interactions.

Although the same neural network was engaged in the computations of intra-language similarity and inter-language difference during language-based word-categorization, the spatiotemporal characteristics of dynamic activities in this network underlying the two computations were not the same. The temporal procedures varied slightly across the two computations. The connectivity strength of the word-categorization network was greater during computations of intra-language similarity compared with inter-language difference. Moreover, the activations of the network started from different nodes during computations of intra-language similarity compared with inter-language difference. Therefore, although the essential function of the word-categorization network is to calculate a correlation distance between two words, this network may work in different fashions when computing the correlation distance between two words that belong to the same or different language categories.

How possible are the early neural RS effects within 200 ms after word onset observed in our study related to the processing of low-level perceptual features or high-level linguistic (e.g., orthography, semantics, phonology) properties of visual words? Our analyses of the ERPs to scrambled Chinese and English words in Experiment 2 did not show significant RS effect. Because only low-level visual features were preserved in the scrambled words, the ERP results provided no evidence that the early RS effects on the neural response to words can be attributed to habituation of perception of the low-level perceptual features. Furthermore, we found that the RS effects on the neural response to radicals and letters in Experiment 3 took place in a delayed time window and exhibited different scalp distributions (i.e., over the central region for radicals and occipital regions for letters) compared with the neural RS effects related to words. Thus the early RS effects on the neural response to words cannot be interpreted as habituation of perception of the middle-level units of Chinese and English words (i.e., radicals and letters) either. In addition, the early neural RS effects were similarly observed for both familiar (i.e., Chinese and English) and unfamiliar (i.e., Korean and Italian) languages and occurred earlier than the time window in which the processing of the linguistic properties of visual words takes place (Marinkovic et al., 2003; Hodgson et al., 2021; Zhu et al., 2022). Therefore, the early neural RS effects identified in our work were unlikely to be associated with the processing of the linguistic (e.g., orthography, semantics, phonology) properties of visual words since these properties of unfamiliar languages were unknown to the participants. Taken together, our findings of the early neural RS effects highlight an early word-level representation of alphabetic vs. non-alphabetic languages which distinguishes words from letters/radicals but is similar for familiar or unfamiliar languages. Our results, however, do not exclude the possibility that the processing of the linguistic properties of visual words may contribute to the long-latency RS effect around 300 ms after word onset. Further processing of the linguistic properties of visual words of familiar languages may follow the early language-based categorization of visual words, though this should be tested in future research.

Our findings raise other questions for future research. Our work tested only young adults. Given the finding of language-based social preferences in infants' behaviors (Howard et al., 2015; Kinzler et al., 2012; Liberman et al., 2017a), it is interesting to compare the developmental trajectories of the neural underpinnings of the word-categorization network and the core linguistic language network. It is also important to clarify how these two networks interact with each other during language processing. The current work discovered the word-categorization network using only written words. It is unclear whether the key nodes of this network beyond fusiform/lingual gyri, which support word-form processing (Zhan et al., 2023), are also engaged in spontaneous language-based categorization of spoken words. To address this issue is important for understanding whether the spoken and writing systems of language function serve as markers of social group identities vis similar neural mechanisms. Recent research using single-neuronal recording has disentangled the process of semantic information (Jamali et al., 2024) and representations of internal and vocalized speech (Wandelt et al., 2024) at the cellular scale. Similar techniques may be employed to probe the mechanisms of language-based word categorization at the level of individual neurons in different nodes of the network. At last but not least, social group classification of others may occur based on both language signals and facial

information (Rakić et al., 2011). How the word-categorization network interacts with that underlying social categorization of faces (Zhou et al., 2020) to influence real-life social behaviors deserves further investigation.

Finally, it should be noted that the current work was initiated by the previous behavioral findings which suggest that language can serve as a socially relevant category cue but focused on the neural mechanisms underlying rapid language-based categorization of visual words. Although the previous findings suggest that the language-based categorization of visual words provides a cognitive basis of social categorization of people, our work did not directly test whether and how the neural processes involved in the language-based categorization of visual words are linked to social evaluation or intergroup processes which are critical for social categorization of people. To clarify this issue should promote deep comprehension of the neural mechanisms underlying the social-categorization function of language but is beyond the scope of the current study. Future research should investigate the connection between language-based categorization of words and social categorization based on other social cues (e.g., faces), which is pivotal to understanding of social interactions in real-world situations.

In conclusion, our EEG and MEG results revealed robust RS effects in the early neural responses to visual words of the same language. The reliability of these RS effects was confirmed across words of different familiar and unfamiliar languages, in samples of speakers with different native languages, and through split-half reliability analyses (see Supplementary Materials, Fig. S19). These effects were supported by the bilateral neural networks whose activity reflected computations of correlation distances between word pairs, capturing both intra-language similarity and inter-language differences during the categorization of visual words in alphabetic and non-alphabetic languages. Together, these findings advance our understanding of spontaneous, language-based neural categorization of visual words as a key basis of the social-categorization function of language.

# Materials and Methods

## Participants

The present study recruited nine independent samples of native Chinese, English and German speakers in Experiments 1 to 9 (see Table S1 for information about participants). All participants were students recruited from universities in Beijing. All Chinese participants began to learn English in primary schools. The English speakers were from different regions (Experiment 4: 13 from North America, 10 from Southeast Asia, 7 from Europe, 1 from Australia, 1 from South Asia, 1 from the Middle East, and 1 from East Africa; Experiment 8: 23 from North America, 5 from Southeast Asia, 3 from Europe, 1 from Australia, 1 from South Asia and 1 from Hong Kong). All the German speakers in Experiment 5 were from Europe. All participants self-reported no neurological diagnoses and had normal or corrected-to-normal vision. This study was approved by the Research Ethics Committee at the School of Psychological and Cognitive Sciences, Peking University (Ethics approval number: #2022-02-07). Informed consent was obtained from all participants prior to the study. All participants were paid for their participation and were informed of their rights to quit at any time during the study. The sample size in Experiment 1 was determined in reference to a previous EEG/MEG study that investigated social categorization of faces using the same RS paradigm (Zhou et al., 2020). Sample sizes in the following studies were determined based on the results of Experiment 1.

## Stimuli

Language materials used in this study included 40 Chinese words, 40 English words and 40 German words (see Table S2 for all word stimuli). The two-character Chinese words included 20 animal names and 20 tool names. English, German, Korean, and Italian words were translated from the Chinese words to match semantic meanings. Word frequencies of these Chinese, English and German words were calculated using the SUBTLEX-CH (Cai and Brysbaert, 2010), SUBTLEX-US (Brysbaert and New, 2009), and SUBTLEX-DE (Brysbaert et al., 2011) databases.

Word frequencies (log10 frequency) were comparable for Chinese words (mean value = 0.97, from −0.62 to 2.74), English words (mean value = 1.32, from 0.09 to 2.68), and German words (mean value = 1.21, from −0.10 to 2.43). Each word subtended a visual angle of 11.53° × 4.28° at a viewing distance of 60 cm in EEG studies and of 8.67° × 3.88° at a view distance of 75 cm in MEG studies. Scrambled control stimuli were created by cutting each of the Chinese and English words into 32 squares (4 × 8) which were then shuffled and reorganized. This manipulation did not change the number of pixels and size of each stimulus. Twenty Chinese radicals and 20 English letters were selected. These stimuli were used to investigate potential contributions of shape features and middle-level units of words to language-based categorization of words.

## Procedure of the one-back task

The RS paradigm was adopted from previous studies of social categorization of faces (Zhang et al., 2023b; Zhou et al., 2020). Words of two languages (Chinese and English, English and German, or Korean and Italian), or scrambled words of two languages (Chinese and English), or word units of two languages (Chinese radicals and English letters) were used in different studies. This RS paradigm consisted of an alternating condition (Alt-Cond), in which words of two different languages (or scrambled words of two languages, or word units of two languages) were presented alternately, and a repetition condition (Rep-Cond), in which words of one language (or scrambled words of one language, or word units of one language) were presented repeatedly (Fig. 1a).

In each trial, a stimulus was displayed for 600 ms in the center of a grey background. Then, it was followed by a fixation cross which had a duration varying from 250 to 550 ms. Participants performed a one-back task (i.e., responding to a casual target stimulus that was presented in two consecutive trials) by pressing a button. In Experiments 1, 4 and 5, participants completed a Chinese-English session and a German-English session. The order of the two different sessions in each study was counter-balanced across participants. Each participant completed 3 runs in each session. Each run consisted of 8 blocks of 20 to 24 trials (including 20 non-target words and randomly 0 to 4 target words). A break of 8 s was given between two successive blocks in the way that a number count flashed with a 1-s step from 8 to 1 at the fixation position. In each run, words were presented in the Rep-Cond in 4 blocks of trials (two blocks for words of each language) and in the Alt-Cond in 4 blocks of trials. Words were displayed in a random order in each block and different blocks of trials were presented in a random order. This design gave 120 non-target trials of each language category in the Rep-Cond and Alt-Cond, respectively. In Experiments 7a and 8a, participants completed a Chinese-English session. The same design was employed in Experiment 2 using scrambled Chinese/English words and in Experiment 3 using Chinese radicals/English letters. The same design was used in Experiments 6 and 9 in which only Korean and Italian words were used.

## Procedure of the explicit word classification task

In Experiments 7b and 8b, participants were asked to perform a task to explicitly classify Chinese and English words during MEG recording. The stimuli (Chinese and English words) and procedure were the same as those in the Alt-Cond in Experiments 7a and 8a except that the inter-trial interval was 800 to 1400 ms. Participants were required to sort perceived words into Chinese or English by pressing two buttons as fast and accurately as possible. The response buttons corresponding to words of the two languages were counter-balanced across participants. Each participant completed 3 runs. Each run consisted of 40 Chinese words and 40 English words presented in a random order.

## EEG data acquisition and analyses

### EEG data acquisition and preprocessing

We conducted EEG recordings using a 10-20 system cap with 64 Ag/AgCl ring electrodes (BrainAmp DC; Brain Products GmbH, Gilching, Germany). EEG signals were digitized at a sampling rate of 500 Hz with a band-pass filter of 0.01–100 Hz and referenced online against FCz. The electrode AFz was used as the ground.

Impedances of individual electrodes were kept below 5 kΩ. We performed EEG preprocessing and data analyses using the EEGLAB toolbox (Delorme and Makeig, 2004). The EEG signals were re-referenced to the average of the left and right mastoid electrodes (TP9, TP10) and filtered with a band-pass filter at 0.5–40 Hz during offline processing. After running independent component analysis, artifacts related to eye movement or blinks and muscle activities were automatically removed using the ICLabel (Pion-Tonachini et al., 2019) plug-in for EEGLAB with the criterion that probabilities of artifact components exceeded 0.9. Only non-target trials were included in EEG data analyses.

## Univariate EEG data analyses

Event-related potentials (ERPs) in each condition were averaged separately with an epoch beginning 200 ms before stimulus onset and continuing for 1,000 ms. The baseline for measuring ERP amplitudes was the mean voltage of a 200-ms prestimulus interval. The latency of each ERP component was measured relative to stimulus onset. Trials with noise exceeding ±100 μV at any electrode were excluded from the average. This process left 118.94±3.00, 116.46±7.78, 118.59±3.57, 118.93±3.18, 117.73±5.16, and 119.65±0.83 trials on average in each condition for further EEG data analyses in Experiments 1 to 6, respectively. We used the FieldTrip toolbox (Oostenveld et al., 2011) to compare ERP amplitudes in the Rep-Cond and Alt-Cond. Cluster-based permutation *t*-tests were conducted to identify significant RS effects (i.e., decreased amplitudes in the Rep-Cond compared to Alt-Cond). A *t*-value indicating the difference in neural responses to words in the Rep-Cond and the Alt-Cond was calculated at each time point of each electrode. Adjacent points in time and space exceeding a predefined threshold ($P < 0.05$, two-tailed) were grouped into one or multiple clusters. The summed cluster *t*-values were compared against a permutation distribution to obtain the *P* values of each cluster. This distribution was generated by randomly reassigning condition markers for each participant (10,000 iterations) and the maximum summed cluster *t*-values were computed for each iteration (Maris and Oostenveld, 2007). We conducted two-tailed cluster-based permutation *t*-tests in the entire epoch (0–800 ms) at all electrodes (62 electrodes after excluding the reference channels (TP9 and TP10)) to identify significant clusters ($P < 0.05$) for all univariate EEG data analyses.

## Multivariate EEG data analyses

Based on the results of univariate analyses that identified significant neural RS effects, we further conducted multivariate representation similarity analyses (Kriegeskorte et al., 2006) to examine the neural processes of intra-language similarity and inter-language difference engaged in spontaneous language-based categorization of words. We performed cluster-based permutation *t*-tests at the EEG electrodes which showed significant main effect of RS. A neural representation dissimilarity matrix (RDM) was constructed based on the correlation distance (one minus the Pearson correlation coefficients) between neural responses (ERP amplitudes at these electrodes) to two words from the same or different languages in both the Rep-Cond and Alt-Cond. All non-target trials were used in the computation of the RDM. This resulted in a 160 × 160 RDM, which included 40 words from each language in the Alt-Cond and Rep-Cond at each time point (Fig. 2c). Because the neural processes of intra-language similarity occurred more frequently in the Rep-Cond than Alt-Cond, the RDM corresponding to intra-language similarity would be attenuated in the Rep-Cond (vs. Alt-Cond) due to habituation (or the mean correlation distances would be increased in the Rep-Cond due to habituation). By contrast, the neural processes of inter-language difference occurred more frequently in the Alt-Cond than Rep-Cond, the RDM corresponding to inter-language difference would be attenuated in the Alt-Cond (vs. the Rep-Cond) due to habituation (or the mean correlation distances would be decreased in the Alt-Cond due to habituation). The differences in RDMs corresponding to both intra-language similarity and inter-language dissimilarity would be evident between the Rep-Cond and Alt-Cond if spontaneous language-based categorization of words engages both clustered representations of words of each language (i.e., enhanced processes of intra-language similarity in the Rep-Cond (vs. Alt-Cond)) and separated representations of words of two different languages (i.e., enhanced processes of inter-language dissimilarity or difference in the Alt-Cond (vs. Rep-Cond)). To test these predictions, we calculated the time courses of the mean values in the cells of the RDMs corresponding to intra-language

dissimilarity and inter-language dissimilarity in both the Rep-Cond and Alt-Cond. After performing Fisher-Z transformation of the correlation distances, we then performed cluster-based permutation *t*-tests to examine significant difference between the time courses in 100 to 300 ms in the Rep-Cond and Alt-Cond (predefined threshold $P < 0.025$, 10,000 iterations, cluster-level $P < 0.05$, one-tailed).

After validating the time windows of significant multivariate RS effects on intra-language similarity and inter-language difference, we averaged these significant time periods to obtain three RDMs that represented the RS effects of intra-Chinese (Korean/Chinese radicals) similarity, intra-English (Italian/English letters) similarity, and Chinese vs. English (Korean vs. Italian/Chinese radicals vs. English letters) difference for each condition (the Rep-Cond or Alt-Cond). To further illustrate the RS effects of intra-language similarity and inter-language difference, with the MATLAB function *mdscale*, we performed non-metric multidimensional scaling analyses of the 160 × 160 RDM using Kruskal's normalized stress formula-1 criterion. We extracted the first two dimensions and then transformed the RDM into three 2D space scatter plots for each condition. In this way, we could see how two kinds of language words are clustered within each language and separated from each other. The variations of clustered representations of words of the same language and separated representations of words of two languages were quantified by comparing the Euclidean distances in the word space between two words of the same language and between two words of different languages in the Alt-Cond and Rep-Cond, respectively.

## MEG data acquisition and analyses

### MEG and MRI data acquisition and preprocessing

We used a whole-head MEG system with 102 magnetometers and 204 planar gradiometers (Elekta Neuromag TRIUX) to record neuromagnetic activity in a magnetically shielded room. The MEG signals were sampled at 1 kHz with an online band-pass filter of 0.1–330 Hz. We removed external interference from the raw MEG data using Maxfilter (Elekta-Neuromag) and co-registered head positions of each run with the first run using Maxmove (a subcomponent of Maxfilter). A high-resolution anatomical T1-weighted image was acquired for each participant (448× 512 mm matrix, 192 slices, 0.5 × 0.5 × 1.00 $mm^3$ spatial resolution; TR = 2530 ms, TE=2.98 ms, inversion time (TI) = 1100 ms, FOV = 25.6 × 25.6 cm, FA = 7°, scanning order: interleaved). Padded clamps were used to minimize head motion and earplugs were used to attenuate scanner noise in the MRI acquisition. Three anatomical landmarks (nasion, left and right pre-auricular points), 4 HPI coils and at least 200 points on the scalp and face were digitized using the Probe Position Identification system (Polhemus) to co-register the MEG data with MRI coordinates. We conducted offline MEG preprocessing and analyses using the Brainstorm toolbox (Tadel et al., 2011). The MEG data were low-pass filtered at 40 Hz. Eye-blink artefacts were removed by signal-space projection. Then, the MEG data were epoched in terms of the stimulus trigger codes.

### Sensor-space whole brain univariate analysis

Event-related fields (ERF) in each condition were averaged separately with an epoch beginning 200 ms before stimulus onset and continuing for 1,000 ms. The baseline for ERF measurement was the mean MEG activity of a 200-ms prestimulus interval and the latency was measured relative to stimulus onset. Trials exceeding 3500 fT at any MEG sensor were excluded from the average. This process left 118.27±3.49, 113.56±11.39, and 117.48±9.32 trials on average in each condition for further MEG data analyses in Experiments 7a, 8a, and 9, respectively. We examined the time courses of the RS effect on sensor-space MEG signals by pooling across Chinese and English words and Chinese and English speakers in Experiments 7a and 8a. Based on the EEG findings of neural RS effects, we conducted whole-brain cluster-based permutation *t*-tests to detect significant sensor-space RS effects (Alt-Cond > Rep-Cond, predefined threshold $P < 0.01$, 10,000 iterations, cluster-level $P < 0.05$, two-tailed) within 400 ms after stimulus onset.

### Source-space whole brain and ROI univariate analyses

The high-resolution anatomical T1-weighted image of each participant was used for source reconstruction of the neural RS effect shown in sensor-space MEG signals. FreeSurfer (*http://surfer.nmr.mgh.harvard.edu/*) was used for segmentation of the T1 image. After co-registration of an individual's brain anatomy and MEG sensors, the noise covariance matrix was computed from the 2-minute daily recordings of the empty room before the experiment. Brain activities were then estimated using a distributed model consisting of 15,002 current dipoles combining the time series of magnetometer and gradiometer signals using a linear inverse estimator (weighted minimum-norm current estimate, signal-to-noise ratio of 3, depth weighting of 0.5, unconstrained dipole orientations) separately for each condition and for each participant in a single-sphere head model. Individual source-space activities were then subject to baseline normalization by subtracting the mean and dividing by the standard deviation of source activations in the pre-stimulus intervals of 200 ms.

Normalized source activations were obtained by computing the norm of three dipole moments in each direction and were smoothed using an 8-mm FWHM Gaussian kernel. Individual normalized source-space data were projected to a standard brain model (ICBM152, 15,002 vertices) for the group-level analysis. Source space signals were down-sampled to 250 Hz before statistical analyses. The sources of neural RS effects were obtained by comparing the grand source signals across Chinese and English words in the Rep-Cond and Alt-Cond. Given the time window of significant RS effects on sensor-space signals, we performed whole-brain cluster-based permutation t-tests in 100–300 ms (vertex-level $P < 0.001$, 10,000 iterations, cluster-level $P < 0.05$, one-tailed) to detect significant attenuations of source-space MEG signals in the Rep-Cond than Alt-Cond. Based on the intersection between the RS effects in the source space and the Desikan-Killiany-Tourville atlas (Klein and Tourville, 2012), we created 11 regions of interest (ROI) to further assess the latency and intensity of RS effects in each brain region. Cluster-based permutation *t*-tests were performed in each brain region to testify significant RS effects (predefined $P < 0.001$, 10,000 iterations, cluster-level $P < 0.05$, one-tailed). We compared the peak intensities of the RS effects in the left and right hemisphere using 2 (left vs. right hemisphere) by 5 (OFC, insula, ATL, FULG, PHC region pairs) repeated measures analysis of variance (ANOVA) (LSTS was not included).

## Source-space multivariate analysis

We conducted multivariate analyses of source-space MEG signals to examine whether the neural network identified in the whole-brain univariate analyses underlies the neural processes of intra-language similarity and inter-language difference during language-based word categorization. Similar to the multivariate analyses of EEG signals, we computed a 160×160 RDM using MEG source space signals to Chinese and English words in the Rep-Cond and Alt-Cond. To increase the signal-to-noise ratio, we first subdivided the 11 ROIs into 313 subregions using the auto-subdivision function in Brainstorm (5 vertices in each subregion). We performed source reconstruction for each word in each condition without smoothing. The mean unsmoothed source signals of 1,000 Hz at each time point were extracted from these subregions to construct a 313-dimensional multivariate neural pattern for each word in each condition at each time point. Then, we computed a correlation distance (one minus the Pearson correlation coefficients) between patterns of MEG source signals of two words in the Rep-Cond and Alt-Cond to obtain the network 160×160 RDM at each time point. Correlation distances were calculated for two words of the same language and two words of two different languages as indices of intra-language similarity and inter-language difference, respectively. Cells in the RDM corresponding to intra-language similarity and inter-language difference were averaged to obtain the time course of the changes in correlation distances. Similarly, we conducted multidimensional scaling analyses of the global network RDMs to identify the first two components corresponding to each word. These two components were then used to construct a 2D word space in which words of the same language or of the two different languages in the Alt-Cond and Rep-Cond were plotted, respectively.

To further assess dynamic contributions of each brain region to the processing of intra-language similarity and inter-language difference during language-based word categorization, we computed local RDMs using the patterns of activity within each of the 11 ROIs. After performing Fisher-Z transformation, the mean correlation distance values specific to intra-/inter-language processing were calculated and assigned to all vertices in each ROI. This allowed us to get the whole-network correlation distance map in both the Rep-Cond and Alt-Cond for intra-language similarity and inter-language difference. Finally, we performed a whole-network cluster-based permutation *t*-test at 100–300 ms. A predefined threshold of $P < 0.025$, 10,000 iterations, and a cluster-level threshold of $P < 0.05$ (one-tailed) was used to examine the temporal and spatial characteristics of the neural RS effects related to the processing of intra-language similarity and inter-language difference.

## Granger causality analysis

To further investigate information flow in the language categorization brain network, using the MVGC toolbox (Barnett and Seth, 2014), we conducted Granger causality analyses (GCA) (Geweke, 1984; Granger, 1969) of the time courses of the RS effects (i.e., the averaged correlation distance in the Rep-Cond minus that in the Alt-Cond at each time point during100–300 ms after the word onset) in the brain regions which demonstrated significant RS effects. The nodes of this network were the same as the MEG ROIs, including the bilateral OFC, insula, ATL, PHC, FULG, and the left STS.

Time series of intra-language similarity or inter-language difference were used to estimate the pairwise-conditional Granger causality among all brain regions. To satisfy the stationarity assumption and the requirement of zero-mean time series for GCA, we first preprocessed the time series of the RS effects, including linear detrending and rescaling (the subtraction of the temporal mean and division by the temporal standard deviation), to remove drifts and slow fluctuations of the time courses and to perform normalization. We conducted the preprocessing for all 780 samples (pairwise dissimilarity between all words of one language) for intra-language similarity and all of the 1600 samples (pairwise dissimilarity between all words of the two languages) for inter-language difference. We estimated the model order for each participant using the Bayesian information criterion and selected the model order of 5 (because down-sampling was not applied to the GCA analyses, a 5-ms lag was used for prediction of the neural activity in one brain region using the neural activity in another brain region) based on the mode of estimated model orders in all participants. A vector autoregression model was fitted using the time series data and was checked for stability and symmetric positive-definite residuals covariance matrix. All participants passed the model check. Then, we calculated time-domain pairwise-conditional causalities from the parameters of the vector autoregression model using the state-space method (Barnett and Seth, 2015). The vector autoregression model was transformed into an equivalent state-space model to estimate the Granger causality values (GC values). In this way, we obtained the Granger causality matrix for each participant and the averaged Granger causality matrix of all participants.

To test the group-level significance of the averaged Granger causality, we generated 1,000 surrogate averaged Granger causality matrices by stepwise permutation of the time series of the source variable by blocks (i.e., block permutation, randomly rearranging data while preserving local temporal dependencies). The original time series was divided into consecutive and non-overlapping blocks of fixed length. We used the model order, which is 5, as the block size. In each iteration, we permuted the time series of each variable while keeping the remaining variables unchanged to calculate the GC values from the permuted variable to the remaining variables. We conducted 1,000 iterations to obtain 1,000 surrogate averaged Granger causality matrices to make the null distribution of the Granger causality. Group-level significance of the original averaged Granger causality matrix was tested by comparing the original Granger causality with the null distribution. We employed the FDR method for correction for multiple comparisons because the GCA tested the connections among multiple brain regions. We reported and illustrated significant and reliable Granger causality connections in the language categorization brain network ($P < 0.05$, FDR correction, GC values > 0.001).

# Data availability

Data and codes for data analyses in this study are available at: *https://doi.org/10.5061/dryad.34tmpg4wn*

# Acknowledgements

This work was supported by the National Natural Science Foundation of China (projects 32230043 and 32371092), Das Chinesisch-Deutsche Zentrum für Wissenschaftsförderung (M-0093), the High-performance Computing Platform of Peking University, and the National Center for Protein Sciences at Peking University. The authors thank Zhirui Zhao and Tengbin Huo for help with EEG and MEG data collection.

# Additional information

## Author contributions

SH and GZ conceived and designed the study, collected and analyzed the data, and wrote the manuscript. SH supervised all aspects of the research.

## Funding


| Funder | Grant reference number | Author |
|---|---|---|
| National Natural Science Foundation of China | 32230043 | Shihui Han |


## Author ORCID iDs

**Guo Zheng**: https://orcid.org/0000-0003-0374-5801
**Shihui Han**: https://orcid.org/0000-0003-3350-5104

# Additional files

***Supplementary Materials.***

# References

**Abutalebi J,** et al. (2007) The neural cost of the auditory perception of language switches: An event-related functional magnetic resonance imaging study in bilinguals. *J Neurosci* **27**:13762-13769 https://doi.org/10.1523/jneurosci.3294-07.2007 | PubMed

**Alessio A,** et al. (2006) Memory and language impairments and their relationships to hippocampal and perirhinal cortex damage in patients with medial temporal lobe epilepsy. *Epilepsy Behav* **8**:593-600 https://doi.org/10.1016/j.yebeh.2006.01.007 | PubMed

**Alia-Klein N,** et al. (2007) What is in a word? No versus Yes differentially engage the lateral orbitofrontal cortex. *Emotion* **7**:649-659 https://doi.org/10.1037/1528-3542.7.3.649 | PubMed

**Arioli M,** Gianelli C, Canessa N (2021) Neural representation of social concepts: a coordinate-based meta-analysis of fMRI studies. *Brain Imaging Behav* **15**:1912-1921 https://doi.org/10.1007/s11682-020-00384-6 | PubMed

**Barnett L,** Seth AK (2015) Granger causality for state-space models. *Phys Rev E* **91**:040101 https://doi.org/10.1103/PhysRevE.91.040101 | PubMed

**Barnett L,** Seth AK (2014) The MVGC multivariate Granger causality toolbox: A new approach to Granger-causal inference. *J Neurosci Methods* **223**:50-68 https://doi.org/10.1016/j.jneumeth.2013.10.018 | PubMed

**Baus C,** Ruiz-Tada E, Escera C, Costa A (2021) Early detection of language categories in face perception. *Sci Rep* **11**:9715 https://doi.org/10.1038/s41598-021-89007-8 | PubMed

**Blanco-Elorrieta E,** Emmorey K, Pylkkänen L (2018) Language switching decomposed through MEG and evidence from bimodal bilinguals. *Proc Natl Acad Sci* **115**:9708-9713 https://doi.org/10.1073/pnas.1809779115 | PubMed

**Brysbaert M,** et al. (2011) The word frequency effect: a review of recent developments and implications for the choice of frequency estimates in German. *Exp Psychol* **58**:412-424 https://doi.org/10.1027/1618-3169/a000123 | PubMed

**Brysbaert M,** New B (2009) Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behav Res Methods* **41**:977-990 https://doi.org/10.3758/BRM.41.4.977 | PubMed

**Bucholtz M,** Hall K (2010) Locating Identity in Language. In: Llamas C, Watt D (Eds). *Language and Identities* Edinburgh University Press. pp. 18-27 https://doi.org/10.1515/9780748635788-006

**Cai Q,** Brysbaert M (2010) SUBTLEX-CH: Chinese Word and Character Frequencies Based on Film Subtitles. *PLoS ONE* **5**:e10729 https://doi.org/10.1371/journal.pone.0010729 | PubMed

**Champoux-Larsson M.-F,** Ramström F, Costa A, Baus C (2022) Social Categorization Based on Language and Facial Recognition. *Journal of Language and Social Psychology* **41**:331-349 https://doi.org/10.1177/0261927x211035159

**Crinion J,** et al. (2006) Language Control in the Bilingual Brain. *Science* **312**:1537-1540 https://doi.org/10.1126/science.1127761 | PubMed

**Dehaene S,** Cohen L (2011) The unique role of the visual word form area in reading. *Trends Cogn Sci* **15**:254-262 https://doi.org/10.1016/j.tics.2011.04.003 | PubMed

**DeJesus JM,** Hwang HG, Dautel JB, Kinzler KD (2018) "American = English Speaker" Before "American = White": The Development of Children's Reasoning About Nationality. *Child Dev* **89**:1752-1767 https://doi.org/10.1111/cdev.12845 | PubMed

**Delorme A,** Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods* **134**:9-21 https://doi.org/10.1016/j.jneumeth.2003.10.009 | PubMed

**Fedorenko E,** Ivanova AA, Regev TI (2024a) The language network as a natural kind within the broader landscape of the human brain. *Nat Rev Neurosci* **25**:289-312 https://doi.org/10.1038/s41583-024-00802-4 | PubMed

**Fedorenko E,** Piantadosi ST, Gibson EAF (2024b) Language is primarily a tool for communication rather than thought. *Nature* **630**:575-586 https://doi.org/10.1038/s41586-024-07522-w | PubMed

**Ferstl EC,** Neumann J, Bogler C, von Cramon DY (2008) The extended language network: A meta-analysis of neuroimaging studies on text comprehension. *Hum Brain Mapp* **29**:581-593 https://doi.org/10.1002/hbm.20422 | PubMed

**Freedman DJ,** Riesenhuber M, Poggio T, Miller EK (2001) Categorical Representation of Visual Stimuli in the Primate Prefrontal Cortex. *Science* **291**:312-316 https://doi.org/10.1126/science.291.5502.312 | PubMed

**Friederici AD,** Gierhan SME (2013) The language network. *Curr Opin Neurobiol* **23**:250-254 https://doi.org/10.1016/j.conb.2012.10.002 | PubMed

**Geweke JF** (1984) Measures of conditional linear dependence and feedback between time series. *J Am Stat Assoc* **79**:907-915 https://doi.org/10.1080/01621459.1984.10477110

**Giari G,** Leonardelli E, Tao Y, Machado M, Fairhall SL (2020) Spatiotemporal properties of the neural representation of conceptual content for words and pictures - an MEG study. *Neuroimage* **219**:116913 https://doi.org/10.1016/j.neuroimage.2020.116913 | PubMed

**Golby AJ,** Gabrieli JD, Chiaåo JY, Eberhardt JL (2001) Differential responses in the fusiform region to same-race and other-race faces. *Nat Neurosci* **4**:845-850 https://doi.org/10.1038/90565 | PubMed

**Granger CWJ** (1969) Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica* **37**:424-438 https://doi.org/10.2307/1912791

**Greenberg RD** (2008) *Language and Identity in the Balkans: Serbo-Croatian and Its Disintegration* Oxford University Press.

**Gu X,** Han S (2007) Neural substrates underlying evaluation of pain in actions depicted in words. *Behav Brain Res* **181**:218-223 https://doi.org/10.1016/j.bbr.2007.04.008 | PubMed

**Han P,** et al. (2020) Neural processing of odor-associated words: an fMRI study in patients with acquired olfactory loss. *Brain Imaging Behav* **14**:1164-1174 https://doi.org/10.1007/s11682-019-00062-2 | PubMed

**Hodgson VJ,** Lambon Ralph MA, Jackson RL (2021) Multiple dimensions underlying the functional organization of the language network. *Neuroimage* **241**:118444 https://doi.org/10.1016/j.neuroimage.2021.118444 | PubMed

**Howard LH,** Henderson AME, Carrazza C, Woodward AL (2015) Infants' and Young Children's Imitation of Linguistic In-Group and Out-Group Informants. *Child Dev* **86**:259-275 https://doi.org/10.1111/cdev.12299 | PubMed

**Ito TA,** Bartholow BD (2009) The neural correlates of race. *Trends Cogn Sci* **13**:524-531 https://doi.org/10.1016/j.tics.2009.10.002 | PubMed

**Jamali M,** et al. (2024) Semantic encoding during language comprehension at single-cell resolution. *Nature* **631**:610-616 https://doi.org/10.1038/s41586-024-07643-2 | PubMed

**Kinzler KD** (2021) Language as a Social Cue. *Annu Rev Psychol* **72**:241-264 https://doi.org/10.1146/annurev-psych-010418-103034 | PubMed

**Kinzler KD,** Dautel JB (2012) Children's essentialist reasoning about language and race. *Dev Sci* **15**:131-138 https://doi.org/10.1111/j.1467-7687.2011.01101.x | PubMed

**Kinzler KD,** Dupoux E, Spelke ES (2007) The native language of social cognition. *Proc Natl Acad Sci* **104**:12577-12580 https://doi.org/10.1073/pnas.0705345104 | PubMed

**Kinzler KD,** Dupoux E, Spelke ES (2012) 'Native'objects and collaborators: Infants' object choices and acts of giving reflect favor for native over foreign speakers. *J Cogn Dev* **13**:67-81 https://doi.org/10.1080/15248372.2011.567200 | PubMed

**Klein A,** Tourville J (2012) 101 Labeled Brain Images and a Consistent Human Cortical Labeling Protocol. *Front Neurosci* **6**:171 https://doi.org/10.3389/fnins.2012.00171 | PubMed

**Kriegeskorte N,** Goebel R, Bandettini P (2006) Information-based functional brain mapping. *Proc Natl Acad Sci* **103**:3863-3868 https://doi.org/10.1073/pnas.0600244103 | PubMed

**Kriegeskorte N,** et al. (2008) Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey. *Neuron* **60**:1126-1141 https://doi.org/10.1016/j.neuron.2008.10.043 | PubMed

**Li H,** Zhang J, Ding G (2021) Reading across writing systems: A meta-analysis of the neural correlates for first and second language reading. *Biling Lang Cogn* **24**:537-548 https://doi.org/10.1017/S136672892000070X

**Liberman Z,** Woodward AL, Kinzler KD (2017a) The origins of social categorization. *Trends Cogn Sci* **21**:556-568 https://doi.org/10.1016/j.tics.2017.04.004 | PubMed

**Liberman Z,** Woodward AL, Kinzler KD (2017b) Preverbal infants infer third-party social relationships based on language. *Cogn Sci* **41**:622-634 https://doi.org/10.1111/cogs.12403 | PubMed

**Lorenzoni A,** Santesteban M, Peressotti F, Baus C, Navarrete E (2022) Language as a cue for social categorization in bilingual communities. *PLoS ONE* **17**:e0276334 https://doi.org/10.1371/journal.pone.0276334 | PubMed

**Malik-Moraleda S,** et al. (2022) An investigation across 45 languages and 12 language families reveals a universal language network. *Nat Neurosci* **25**:1014-1019 https://doi.org/10.1038/s41593-022-01114-5 | PubMed

**Malik-Moraleda S,** et al. (2024) Functional characterization of the language network of polyglots and hyperpolyglots with precision fMRI. *Cereb Cortex* **34** https://doi.org/10.1093/cercor/bhae049 | PubMed

**Marinkovic K,** et al. (2003) Spatiotemporal dynamics of modality-specific and supramodal word processing. *Neuron* **38**:487-497 https://doi.org/10.1016/s0896-6273(03)00197-1 | PubMed

**Maris E,** Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods* **164**:177-190 https://doi.org/10.1016/j.jneumeth.2007.03.024 | PubMed

**Merritt CC,** MacCormack JK, Stein AG, Lindquist KA, Muscatell KA (2021) The neural underpinnings of intergroup social cognition: an fMRI meta-analysis. *Soc Cogn Affect Neurosci* **16**:903-914 https://doi.org/10.1093/scan/nsab034 | PubMed

**Nan W,** et al. (2022) The spatiotemporal characteristics of N170s for faces and words: A meta-analysis study. *PsyCh Journal* **11**:5-17 https://doi.org/10.1002/pchj.511 | PubMed

**Olson IR,** McCoy D, Klobusicky E, Ross LA (2013) Social cognition and the anterior temporal lobes: a review and theoretical framework. *Soc Cogn Affect Neurosci* **8**:123-133 https://doi.org/10.1093/scan/nss119 | PubMed

**Oostenveld R,** Fries P, Maris E, Schoffelen JM (2011) FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput Intell Neurosci* **2011**:156869 https://doi.org/10.1155/2011/156869 | PubMed

**Pang C,** Zhou N, Deng Y, Pu Y, Han S (2025) Neural Tracking of Race-Related Information During Face Perception. *Neurosci Bull. Advance online publication* https://doi.org/10.1007/s12264-025-01419-y | PubMed

**Park SA,** Miller DS, Nili H, Ranganath C, Boorman ED (2020) Map Making: Constructing, Combining, and Inferring on Abstract Cognitive Maps. *Neuron* **107**:1226-1238.e1228. https://doi.org/10.1016/j.neuron.2020.06.030 | PubMed

**Pietraszewski D,** Schwartz A (2014) Evidence that accent is a dimension of social categorization, not a byproduct of perceptual salience, familiarity, or ease-of-processing. *Evol Hum Behav* **35**:43-50 https://doi.org/10.1016/j.evolhumbehav.2013.09.006

**Pion-Tonachini L,** Kreutz-Delgado K, Makeig S (2019) ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *Neuroimage* **198**:181-197 https://doi.org/10.1016/j.neuroimage.2019.05.026 | PubMed

**Pobric G,** Lambon Ralph MA, Zahn R (2016) Hemispheric Specialization within the Superior Anterior Temporal Cortex for Social and Nonsocial Concepts. *J Cogn Neurosci* **28**:351-360 https://doi.org/10.1162/jocn_a_00902 | PubMed

**Pu Y,** Han S (2025) Neural Basis of Categorical Representations of Animal Body Silhouettes. *Neurosci Bull* **41**:211-223 https://doi.org/10.1007/s12264-024-01268-1 | PubMed

**Pylkkänen L,** Marantz A (2003) Tracking the time course of word recognition with MEG. *Trends Cogn Sci* **7**:187-189 https://doi.org/10.1016/s1364-6613(03)00092-5 | PubMed

**Rakić T,** Steffens MC, Mummendey A (2011) Blinded by the accent! The minor role of looks in ethnic categorization. *J Pers Soc Psychol* **100**:16 https://doi.org/10.1037/a0021522 | PubMed

**Rhodes M,** Baron A (2019) The Development of Social Categorization. *Annu Rev Dev Psychol* **1**:359-386 https://doi.org/10.1146/annurev-devpsych-121318-084824 | PubMed

**Roberts G** (2013) Perspectives on Language as a Source of Social Markers. *Lang Linguist Compass* **7**:619-632 https://doi.org/10.1111/lnc3.12052

**Rodriguez-Fornells A,** et al. (2005) Second Language Interferes with Word Production in Fluent Bilinguals: Brain Potential and Functional Imaging Evidence. *J Cogn Neurosci* **17**:422-433 https://doi.org/10.1162/0898929053279559 | PubMed

**Sachdev I,** Bourhis RY (1990) Language and social identification. In: Abrams D, Hogg M (Eds). *Social Identity Theory: Constructive and Critical Advances* Harvester Wheatsheaf. pp. 33-51

**Scherer KR,** Giles H (1979) *Social markers in speech* Cambridge University Press.

**Shell M** (2001) Language Wars. *CR: New Centenn Rev* **1**:1-17 https://doi.org/10.1353/ncr.2003.0059

**Siok WT,** Perfetti CA, Jin Z, Tan LH (2004) Biological abnormality of impaired reading is constrained by culture. *Nature* **431**:71-76 https://doi.org/10.1038/nature02865 | PubMed

**Tadel F,** Baillet S, Mosher JC, Pantazis D, Leahy RM (2011) Brainstorm: a user-friendly application for MEG/EEG analysis. *Comput Intell Neurosci* **2011**:879716 https://doi.org/10.1155/2011/879716 | PubMed

**Tan LH,** Laird AR, Li K, Fox PT (2005) Neuroanatomical correlates of phonological processing of Chinese characters and alphabetic words: A meta-analysis. *Hum Brain Mapp* **25**:83-91 https://doi.org/10.1002/hbm.20134 | PubMed

**Thesen T,** et al. (2012) Sequential then interactive processing of letters and words in the left fusiform gyrus. *Nat Commun* **3**:1284 https://doi.org/10.1038/ncomms2220 | PubMed

**Wandelt SK,** et al. (2024) Representation of internal speech by single neurons in human supramarginal gyrus. *Nat Hum Behav* **8**:1136-1149 https://doi.org/10.1038/s41562-024-01867-y | PubMed

**Wang Y,** Xue G, Chen C, Xue F, Dong Q (2007) Neural bases of asymmetric language switching in second-language learners: An ER-fMRI study. *Neuroimage* **35**:862-870 https://doi.org/10.1016/j.neuroimage.2006.09.054 | PubMed

**Zahn R,** et al. (2007) Social concepts are represented in the superior anterior temporal cortex. *Proc Natl Acad Sci* **104**:6430-6435 https://doi.org/10.1073/pnas.0607061104 | PubMed

**Zhan M,** Pallier C, Agrawal A, Dehaene S, Cohen L (2023) Does the visual word form area split in bilingual readers? A millimeter-scale 7-T fMRI study. *Sci Adv* **9**:eadf6140 https://doi.org/10.1126/sciadv.adf6140 | PubMed

**Zhang G,** et al. (2023a) A social-semantic working-memory account for two canonical language areas. *Nat Hum Behav* **7**:1980-1997 https://doi.org/10.1038/s41562-023-01704-8 | PubMed

**Zhang T,** Han S (2021) Non-phase-locked alpha oscillations are involved in spontaneous racial categorization of faces. *Neuropsychologia* **160**:107968 https://doi.org/10.1016/j.neuropsychologia.2021.107968 | PubMed

**Zhang T,** Zhou Y, Han S (2023b) Priority of racial and gender categorization of faces: A social task demand framework. *J Pers Soc Psychol* **124**:483 https://doi.org/10.1037/pspa0000318 | PubMed

**Zhou Y,** et al. (2020) Neural dynamics of racial categorization predicts racial bias in face recognition and altruism. *Nat Hum Behav* **4**:69-87 https://doi.org/10.1038/s41562-019-0743-y | PubMed

**Zhu JD,** Seymour RA, Szakay A, Sowman PF (2020) Neuro-dynamics of executive control in bilingual language switching: An MEG study. *Cognition* **199**:104247 https://doi.org/10.1016/j.cognition.2020.104247 | PubMed

**Zhu Y,** et al. (2022) Distinct spatiotemporal patterns of syntactic and semantic processing in human inferior frontal gyrus. *Nat Hum Behav* **6**:1104-1111 https://doi.org/10.1038/s41562-022-01334-6 | PubMed

# Peer reviews

## Reviewer #1 (Public review):

Summary:

This study demonstrates, through a series of EEG and MEG experiments, that the human brain automatically categorizes words from alphabetic and non-alphabetic languages, and it unpacks the neural mechanisms of this process from multiple angles. The work examines not only univariate repetition-suppression (RS) effects, but also how repeating or alternating

languages influences the representational similarity of words within and across language categories.

Strengths:

The univariate RS effects across multiple experiments lend support to some of the main conclusions.

Comments on revised version.

The authors have made appropriate revisions and supplements in response to the issues I raised, which has largely resolved my concerns.

https://doi.org/10.7554/eLife.110320.2.sa1

## Reviewer #2 (Public review):

Summary:

This study investigates how the human brain categorizes visual words from distinct writing systems (alphabetic vs. non-alphabetic). Using a repetition suppression paradigm combined with electroencephalography and magnetoencephalography, the authors conducted nine experiments with independent participants to identify the neural network underlying language-based categorization, characterize its temporal dynamics, and test whether this process operates independently of linguistic properties such as semantic meaning and pronunciation.

Strengths:

The study employs a well-validated design with clear control conditions and systematically manipulates key variables including writing system, language familiarity, and native language background. The use of nine experiments with independent participant samples strengthens the reliability and replicability of the results. The work combines EEG and MEG, cross-validating findings across imaging modalities to support the reported neural effects. A combination of univariate, multivariate, and connectivity analyses is used to characterize neural responses and network interactions. Results are consistent across multiple language groups and for both familiar and unfamiliar languages, supporting the generalizability of the identified neural mechanism beyond specific languages or prior experience.

Comments on revised version.

Earlier versions of the manuscript framed these findings as more directly reflecting the social-categorization function of language. In the revised manuscript, the authors now more carefully distinguish language-based word categorization from broader claims regarding social categorization and explicitly acknowledge that the current experiments do not directly test social evaluation or intergroup processes. These revisions improve the conceptual precision of the work and address my major concern from the previous review.

The additional methodological clarifications and supplementary analyses also strengthen the manuscript. Overall, I believe the revised version provides solid evidence for rapid language-based categorization of visual words across different writing systems.

https://doi.org/10.7554/eLife.110320.2.sa3

## Author response:

The following is the authors' response to the original reviews.

> ***eLife Assessment***
>
> *This important study investigates how the brain categorizes written words from different writing systems (e.g., alphabetic vs. non-alphabetic), shedding potential light on the neural basis of language's social-categorization function. Overall, the evidence supporting the authors' claims is solid, though some analyses and key interpretations would benefit from fuller justification.*

Thank you for handling our manuscript! We've modified the manuscript according to the reviewers' comments and suggestions.

> ***Public Reviews:***
>
> ***Reviewer #1 (Public review):***
>
> *Summary:*
>
> *This study demonstrates, through a series of EEG and MEG experiments, that the human brain automatically categorizes words from alphabetic and non-alphabetic languages, and it unpacks the neural mechanisms of this process from multiple angles. The work examines not only univariate repetition-suppression (RS) effects, but also how repeating or alternating languages influences the representational similarity of words within and across language categories.*
>
> *Strengths:*
>
> *The univariate RS effects across multiple experiments lend support to some of the main conclusions*
>
> *Weaknesses:*
>
> *I have reservations about the logic underlying the multivariate analyses, and I believe the implications of the control experiments merit fuller discussion.*
>
> *(1) Question 1: Logic of the multivariate analyses*
>
> *The original text states:*
>
> *"The processing of intra-language similarity was quantified as correlation distances between neural responses to two words of the same language, which occurred more frequently and would be inhibited in the Rep-Cond (vs. Alt-Cond) due to habituation (Fig. 1c)...".*
>
> *I argue that this passage conflates two levels. Building a representational dissimilarity matrix (RDM) is a data-analysis step; it cannot be equated with a cognitive computation. Hence, there is no sense in which this computation occurs "more frequently" in one condition. RDM construction rests on the pairwise similarity of activity patterns, so even if a task engaged no cognitive computation of representational similarity, we could still compute an RDM. Conversely, if a task factor alters the RDM, we must explain how that factor changes the underlying neural patterns, not claim that it triggers specific cognitive processing. Therefore, I neither understand what "more frequent processing" the authors refer to, nor accept their account of the multivariate results.*
>
> *The multivariate result pattern, briefly, is that distances between words, both within and across languages, are larger under the repetition condition. One plausible interpretation is that a word representation comprises two parts: language-type (alphabetic vs. non-alphabetic) and fine-grained identity features (visual shape, orthography, semantics, phonology, etc.). Repetition of language type may, via RS, reduce the weight of the first*

> *component, thereby increasing the relative contribution of fine-grained features and amplifying inter-word differences. This could explain the multivariate findings.*

Thank you for these insightful comments regarding the logic of the multivariate analyses. In the revision, we've elaborated the rationale underlying our experimental design. Specifically, we've explained why the processing of intra-language similarity is expected to occur more frequently in the repetition condition (Rep-Cond) than in the alternation condition (Alt-Cond) whereas the reverse is true for the processing of inter-language difference. Importantly, we've clarified that the processing of intra-language similarity was assessed rather than defined by conducting the multivariate analyses. The multivariate analyses were conducted to assess correlation distances between neural responses to pairs of words, either within the same language or across different languages. We explained what smaller intra-language correlation distances and larger inter-language correlation distances mean for language-base categorization of words (see Page 7-8).

We appreciate the alternative account of the observed neural repetition suppression (RS) effects in terms of language-type versus fine-grained identity (visual shape, orthography, semantics, phonology, etc.) feature processing. We included a paragraph in the revised Discussion to discuss how possible the early neural RS effect can be attributed to the processing of the fine-grained identity features of visual words. This discussion allowed us to clarify that the early neural RS effects related to visual words of familiar and unfamiliar languages highlight the early spontaneous language-based categorization as a unique process of visual words of alphabetic and non-alphabetic languages. However, our results do not exclude the possibility that the processing of the linguistic properties of visual words may contribute to the long-latency RS effect (see Page 37-38).

Page 7-8

"The processing of intra-language similarity occurs when two words of the same language are perceived repeatedly with short interstimulus intervals. Because words of the same language were repeatedly presented in the Rep-Cond and words of two different languages were displayed in the Alt-Cond, the processing of intra-language similarity occurred more frequently and would be inhibited in the Rep-Cond (vs. Alt-Cond) due to habituation (Fig. 1c). By contrast, the processing of inter-language difference takes place when two words of different languages are perceived with short interstimulus intervals. Since words of different languages appeared more frequently in the Alt-Cond (vs. Rep-Cond), we would expect RS of the processing of inter-language difference in the Alt-Cond (vs. Rep-Cond). The neural processing of intra-language similarity was quantified as correlation distances between neural responses to two words of the same language whereas the neural processing of inter-language difference was assessed as correlation distances between neural responses to two words of two different languages. The correlation distances from the multivariate analyses were further employed to assess how words of one language are clustered and how far words of two languages are separated in a two-dimensional (2D) space during language-based word categorization. Enhanced language-based word categorization is associated with smaller intra-language correlation distances, which reflect more densely clustered words of the same language, and larger inter-language correlation distances, which manifest further separated words of two different languages."

Page 37-38

"How possible are the early neural RS effects within 200 ms after word onset observed in our study related to the processing of low-level perceptual features or high-level linguistic (e.g., orthography, semantics, phonology) properties of visual words? Our analyses of the ERPs to scrambled Chinese and English words in Experiment 2 did not show significant RS effect. Because only low-level visual features were preserved in the scrambled words, the ERP results provided no evidence that the early RS effects on the neural response to words can be

attributed to habituation of perception of the low-level perceptual features. Furthermore, we found that the RS effects on the neural response to radicals and letters in Experiment 3 took place in a delayed time window and exhibited different scalp distributions (i.e., over the central region for radicals and occipital regions for letters) compared with the neural RS effects related to words. Thus the early RS effects on the neural response to words cannot be interpreted as habituation of perception of the middle-level units of Chinese and English words (i.e., radicals and letters) either. In addition, the early neural RS effects were similarly observed for both familiar (i.e., Chinese and English) and unfamiliar (i.e., Korean and Italian) languages and occurred earlier than the time window in which the processing of the linguistic properties of visual words takes place (Marinkovic et al., 2003; Hodgson et al., 2021; Zhu et al., 2022). Therefore, the early neural RS effects identified in our work were unlikely to be associated with the processing of the linguistic (e.g., orthography, semantics, phonology) properties of visual words since these properties of unfamiliar languages were unknown to the participants. Taken together, our findings of the early neural RS effects highlight an early word-level representation of alphabetic vs. non-alphabetic languages which distinguishes words from letters/radicals but is similar for familiar or unfamiliar languages. Our results, however, do not exclude the possibility that the processing of the linguistic properties of visual words may contribute to the long-latency RS effect around 300 ms after word onset. Further processing of the linguistic properties of visual words of familiar languages may follow the early language-based categorization of visual words, though this should be tested in future research."

> *(2) Question 2:*
>
> *For unlearned languages, people cannot distinguish lexical from sub-lexical levels. What, then, determines (i) the RS-effect difference between letters and radicals in familiar languages and words in unlearned ones, and (ii) the similarity of repetition effects between words in unlearned and familiar languages? An explicit account is needed.*

Thank you for this suggestion. In the revised manuscript, we've included a dedicated paragraph addressing these two issues. Specifically, we've provided a more precise account of the differences in repetition suppression (RS) effects between words and letters/radicals in familiar languages, as well as the similar RS effects observed for unlearned and familiar languages. We believe that our findings of the early neural RS effects highlight an early word-level representation of alphabetic vs. non-alphabetic languages which distinguishes words from letters/radicals but is similar for familiar or unfamiliar languages (see Page 37-38).

Page 37-38

"How possible are the early neural RS effects within 200 ms after word onset observed in our study related to the processing of low-level perceptual features or high-level linguistic (e.g., orthography, semantics, phonology) properties of visual words? Our analyses of the ERPs to scrambled Chinese and English words in Experiment 2 did not show significant RS effect. Because only low-level visual features were preserved in the scrambled words, the ERP results provided no evidence that the early RS effects on the neural response to words can be attributed to habituation of perception of the low-level perceptual features. Furthermore, we found that the RS effects on the neural response to radicals and letters in Experiment 3 took place in a delayed time window and exhibited different scalp distributions (i.e., over the central region for radicals and occipital regions for letters) compared with the neural RS effects related to words. Thus the early RS effects on the neural response to words cannot be interpreted as habituation of perception of the middle-level units of Chinese and English words (i.e., radicals and letters) either. In addition, the early neural RS effects were similarly observed for both familiar (i.e., Chinese and English) and unfamiliar (i.e., Korean and Italian) languages and occurred earlier than the time window in which the processing of the linguistic properties of visual words takes place (Marinkovic et al., 2003; Hodgson et al., 2021; Zhu et al., 2022). Therefore, the early neural RS effects identified in our work were unlikely to

be associated with the processing of the linguistic (e.g., orthography, semantics, phonology) properties of visual words since these properties of unfamiliar languages were unknown to the participants. Taken together, our findings of the early neural RS effects highlight an early word-level representation of alphabetic vs. non-alphabetic languages which distinguishes words from letters/radicals but is similar for familiar or unfamiliar languages. Our results, however, do not exclude the possibility that the processing of the linguistic properties of visual words may contribute to the long-latency RS effect around 300 ms after word onset. Further processing of the linguistic properties of visual words of familiar languages may follow the early language-based categorization of visual words, though this should be tested in future research."

> ***Reviewer #2 (Public review):***
>
> *Summary:*
>
> *This study investigates how the human brain categorizes visual words from distinct writing systems (alphabetic vs. non-alphabetic) as a neural basis for the social-categorization function of language. Using a repetition suppression paradigm combined with electroencephalography and magnetoencephalography, the authors conducted nine experiments with independent participants to identify the neural network underlying language-based categorization, characterize its temporal dynamics, and test whether this process operates independently of linguistic properties such as semantic meaning and pronunciation.*
>
> *Strengths:*
>
> *(1) The study employs a well-validated design with clear control conditions and systematically manipulates key variables, including writing system, language familiarity, and native language background. The use of nine experiments with independent participant samples strengthens the reliability and replicability of the results.*
>
> *(2) The work combines EEG and MEG, cross-validating findings across imaging modalities to support the reported neural effects. A combination of univariate, multivariate, and connectivity analyses is used to characterize neural responses and network interactions.*
>
> *(3) Results are consistent across multiple language groups and for both familiar and unfamiliar languages, supporting the generalizability of the identified neural mechanism beyond specific languages or prior experience.*
>
> *Weaknesses:*
>
> *The authors provide compelling evidence that the identified neural network supports the categorization of words by language, including computations of intra-language similarity and inter-language difference. However, the conceptual framing of this finding as directly reflecting the social-categorization function of language may be premature. While the task captures spontaneous language categorization, it does not involve social evaluation or intergroup processes. The connection to social categorization is inferred from prior literature rather than demonstrated within the current experimental design. Clarifying this distinction would strengthen the conceptual precision of the manuscript.*

Thank you for this important comment. In the revised Introduction and Discussion, we've clarified several related issues. First, prior research suggests that language can serve as a socially relevant category cue. Second, these findings imply that rapid categorization of words by language may occur in the human brain. Third, although our results identify a neural network supporting such rapid language-based categorization of visual words, they do not directly test how this process relates to social categorization of people (see Page 3-4; Page

39). Highlighting these points help delineate the scope of our findings and point to important directions for future research.

Page 3-4

“The social-categorization function of language revealed in these behavioral studies implicates that rapid categorization of words of different languages may occur in the human brain. Furthermore, the findings of infant studies (e. g., Liberman et al., 2017b) suggest that the neural process involved in categorization of words of different languages may develop even prior to the processing of linguistic properties (e.g. semantic meanings) of words. Nevertheless, up to date, there has been little neuroimaging research examining the neural mechanisms underlying automatic and fast categorization of words of different languages.”

Page 39

“Finally, it should be noted that the current work was initiated by the previous behavioral findings which suggest that language can serve as a socially relevant category cue but focused on the neural mechanisms underlying rapid language-based categorization of visual words. Although the previous findings suggest that the language-based categorization of visual words provides a cognitive basis of social categorization of people, our work did not directly test whether and how the neural processes involved in the language-based categorization of visual words are linked to social evaluation or intergroup processes which are critical for social categorization of people. To clarify this issue should promote deep comprehension of the neural mechanisms underlying the social-categorization function of language but is beyond the scope of the current study. Future research should investigate the connection between language-based categorization of words and social categorization based on other social cues (e.g., faces), which is pivotal to understanding of social interactions in real-world situations.”

> ***Recommendations for the authors:***
>
> ***Reviewer #2 (Recommendations for the authors):***
>
> *(1) Revise the conceptual framing to clarify the relationship between the experimental results and the proposed social-categorization function of language. If the authors wish to retain the emphasis on social categorization in the title or discussion, they should explicitly explain how the observed neural mechanisms of language-based word categorization link to social evaluation, intergroup processes, or real-world social categorization. This clarification would strengthen the conceptual coherence and justify the use of social categorization within the current study's scope.*

Thank you for this and the following suggestions. In the revised Introduction and Discussion, we’ve clarified the following point: First, the findings of prior behavioral studies suggest a social-categorization function of language. Second, based on these behavioral findings, we predicted automatic and fast categorization of words by language. Our study tested this prediction using neuroimaging and investigated the neural mechanisms of language-type-based categorization of visual words. This is the main goal of our work. Third, to examine how the observed neural mechanisms of language-based word categorization link to social evaluation, intergroup processes, or real-world social categorization is important but beyond the scope of the current work. However, this is a very important question. Future research should test the connection between the neurocognitive processes involved in social categorization of people and the neural categorization of visual words by language revealed in our study. Consistently, the title of our paper “Neural categorization of visual words of alphabetic and non-alphabetic languages” and Discussion focus on contributions of our findings to understanding of the neural categorization of visual words by language rather than its connection to social categorization of people. Above all, we’ve clarified in the revision

that our study was initiated by the findings of social function of language but was limited to the neural processing of visual words (see Page 3-4; Page 39). Thanks again for this comment.

Page 3-4

"The social-categorization function of language revealed in these behavioral studies implicates that rapid categorization of words of different languages may occur in the human brain. Furthermore, the findings of infant studies (e. g., Liberman et al., 2017b) suggest that the neural process involved in categorization of words of different languages may develop even prior to the processing of linguistic properties (e.g. semantic meanings) of words. Nevertheless, up to date, there has been little neuroimaging research examining the neural mechanisms underlying automatic and fast categorization of words of different languages."

Page 39

"Finally, it should be noted that the current work was initiated by the previous behavioral findings which suggest that language can serve as a socially relevant category cue but focused on the neural mechanisms underlying rapid language-based categorization of visual words. Although the previous findings suggest that the language-based categorization of visual words provides a cognitive basis of social categorization of people, our work did not directly test whether and how the neural processes involved in the language-based categorization of visual words are linked to social evaluation or intergroup processes which are critical for social categorization of people. To clarify this issue should promote deep comprehension of the neural mechanisms underlying the social-categorization function of language but is beyond the scope of the current study. Future research should investigate the connection between language-based categorization of words and social categorization based on other social cues (e.g., faces), which is pivotal to understanding of social interactions in real-world situations."

> *(2) Clarify the consistency between the reported model order (5 ms lag) and the sampling rate after downsampling (250 Hz, corresponding to 4 ms per time point). If a discrepancy exists, clearly explain how the time-series data were processed.*

We clarified in the revision (see Page 53) that "because down-sampling was not applied to the GCA analyses, a 5-ms lag was used for prediction of the neural activity in one brain region using the neural activity in another brain region".

> *(3) For the representational similarity analysis (RSA), report reliability measures for the representational dissimilarity matrices (e.g., split-half reliability) to verify that the observed effects are stable given the number of trials per condition.*

Following this suggestion, we've conducted split-half reliability analyses and reported the results in the revised supplementary materials. The reliability analyses are also mentioned in the revised Discussion (see Page 40).

Page 40

"In conclusion, our EEG and MEG results revealed robust RS effects in the early neural responses to visual words of the same language. The reliability of these RS effects was confirmed across words of different familiar and unfamiliar languages, in samples of speakers with different native languages, and through split-half reliability analyses (see Supplementary Materials, Fig. S19). These effects were supported by the bilateral neural networks whose activity reflected computations of correlation distances between word pairs, capturing both intra-language similarity and inter-language differences during the categorization of visual words in alphabetic and non-alphabetic languages. Together, these findings advance our understanding of spontaneous, language-based neural categorization of visual words as a key basis of the social-categorization function of language."

> *(4) Provide complete statistical information for all significant results reported in the supplementary materials, including relevant test statistics (e.g., t-values, cluster p-values) in figure legends or a supplementary results table to improve transparency.*

Complete statistical information has been provided in the revised supplementary materials (see Tables S4 and S5).

> *(5) Streamline the presentation of the nine experiments in the main text to emphasize the core conceptual and methodological logic, potentially using a schematic overview or flowchart to improve readability.*

As suggested, we've included an overview of the nine experiments in the revised Introduction. This overview helps understanding of the core conceptual and methodological issues in our work (see Page 6).

Page 6

"In nine experiments we recorded EEG/MEG signals from Chinese, English, and German speakers when viewing words of an alphabetic language and a non-alphabetic language (English and Chinese words, or Italian and Korean words) or of two alphabetic languages (English and German) in the Rep-Cond and Alt-Cond. We recorded EEG signals from Chinese participants to examine temporal neural dynamics of spontaneous language-based word categorization in Experiment 1. The similar paradigm was employed in Experiments 2 and 3 to investigate whether perceptual features or radical/letters of words are sufficient to generate spontaneous language-based categorization of visual words. The results in Experiment 1 were replicated in native English and German speakers in Experiments 4 and 5, respectively. Neural dynamics of categorization of words of two unlearned languages were further investigated in Chinese participants in Experiment 6. Finally, the neural networks supporting the spontaneous categorization of words of two learned or unlearned languages were localized using MEG in Chinese and English speakers in Experiments 7-9, respectively."

> *(6) Strengthen the transition between the discussion of the social-categorization function of language and the neural mechanisms of visual word categorization in the introduction.*

Following this suggestion, we've modified the Introduction to strengthen the transition between the discussion of the social-categorization function of language and research on neural mechanisms of visual word categorization (see Page 3-4).

Page 3-4

"The social-categorization function of language revealed in these behavioral studies implicates that rapid categorization of words of different languages may occur in the human brain. Furthermore, the findings of infant studies (e. g., Liberman et al., 2017b) suggest that the neural process involved in categorization of words of different languages may develop even prior to the processing of linguistic properties (e.g. semantic meanings) of words. Nevertheless, up to date, there has been little neuroimaging research examining the neural mechanisms underlying automatic and fast categorization of words of different languages."

> *(7) Briefly define the repetition suppression (RS) paradigm when first mentioned (i.e., reduced neural response to repeated stimuli from the same category, reflecting categorical processing) to improve accessibility for non-specialist readers.*

The RS paradigm is now defined in Introduction when being mentioned for the first time in the manuscript (see Page 5-6).

Page 5-6

"The present study investigated neural dynamics of categorization of visual words of two different (an alphabetic versus a non-alphabetic, or two different alphabetic) languages by combining EEG/MEG with a repetition suppression (RS) paradigm adopted from previous studies of social categorization of faces (Zhang et al., 2023b; Zhou et al., 2020). RS refers to the attenuation in neural responses to a repeated occurrence of stimuli that engage common neuronal populations or processes due to habituation (Grill-Spector et al., 2006). The RS paradigm consisted of an alternating condition (Alt-Cond), in which visual words of two different languages were presented alternately, and a repetition condition (Rep-Cond), in which words of one language were presented repeatedly (Fig. 1a). Neural responses to stimuli of the same category were attenuated in the Rep-Cond compared to Alt-Cond due to habituation and this RS effect disentangles the neural activities underlying categorization of faces and body silhouettes of a specific social group."

> *(8) Report detailed participant demographic information, including exact age range/mean age and gender ratio for each experiment, to meet standard reporting practices in neuroscience.*

We've modified Table S1 to include the information about exact age range/mean age and gender ratio in each experiment.

> *(9) Correct minor typographical and grammatical errors, including These finding (line 59) and Chinse (line 223).*

These and other grammatical errors have been corrected in the revision.

https://doi.org/10.7554/eLife.110320.2.sa0