Neural basis of social-categorization function of language

Guo Zheng, Shihui Han

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

Address correspondence to:
Shihui Han Ph.D. Prof.
School of Psychological and Cognitive Sciences
Peking University
52 Haidian Road, Beijing 100080, China
Email: shan@pku.edu.cn
Phone: (86) 10-6275-9138
Fax: (86) 10-6276-1081

Word number of the main text: ????
Number of display items: 8

Abstract

Languages provide social-category markers that tag people as one or another social group. We investigate how the brain sorts words into one or another language category as a neural basis of this social-categorization function of language. We show that a neural network, including the anterior temporal, insular, orbital frontal, and ventral occipito-temporal cortices in both hemispheres, computes 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 occur as early as 150 ms post-stimulus, recruit within-hemisphere functional connections, operate independently of words' semantic meanings and pronunciations, and exhibit consistently across individuals with diverse language backgrounds. These findings advance our understanding of the neural mechanisms of the social-categorization function of language.

Keywords: Visual words; Social categorization; EEG; MEG; Anterior temporal lobe

Main text

Language is a sophisticated system for transmission of information (e.g., thoughts, knowledge, and feelings) among individuals (1). 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 (2-8). 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 (9, 10). In extreme cases, individuals who spoke a different language were classified into an out-group and persecuted (11, 12). The social-categorization function of language emerges early during human development (4, 13). Infants expect two individuals who speak the same language to be affiliated (14) and favor native over non-native speakers during social learning (15) and acts of giving (16). Despite the behavioral evidence for the connection between language and social categorization, neural mechanisms underlying the social-categorization function of language remain unknown.

A key component of the social-categorization function of language is to sort written or spoken words with different pronunciations or semantic meanings into one or another language category. Language-based categorization of words 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 on the linguistic communication function of language. Studies in this field 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 (17, 18). This network responds to diverse languages (19, 20), activates similarly to native and non-native languages (21, 22), and exhibits similar patterns of responses to spoken and written words/sentences (17). 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 (23-25). 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 (23). The left temporal and frontal activities are particularly important for semantic and phonological processing of words and sentences (26) around 400 ms after word presentation (27) as well as social-semantic working-memory (28). 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 (29-34) may play important roles in language-based categorization of words.

The present study investigated neural dynamics of categorization of visual words of two different languages by combining EEG/MEG with a repetition suppression (RS) paradigm adopted from previous studies of social categorization of faces (34, 35). This 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 (34-37). 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 performed univariate analyses of the RS effects on neural responses 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 intra-group (or within-category) similarity and inter-group (or between-category) difference between perceived stimuli (34, 38-40), we further conducted multivariate representational similarity analyses (41) 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 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). By contrast, the processing of inter-language difference indexed by the correlation distances between neural responses to words of two different languages would take place more frequently when the brain encounters words of two languages in the Alt-Cond (vs. Rep-Cond) and thus would be inhibited. The results of the multivariate analyses allowed us 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.

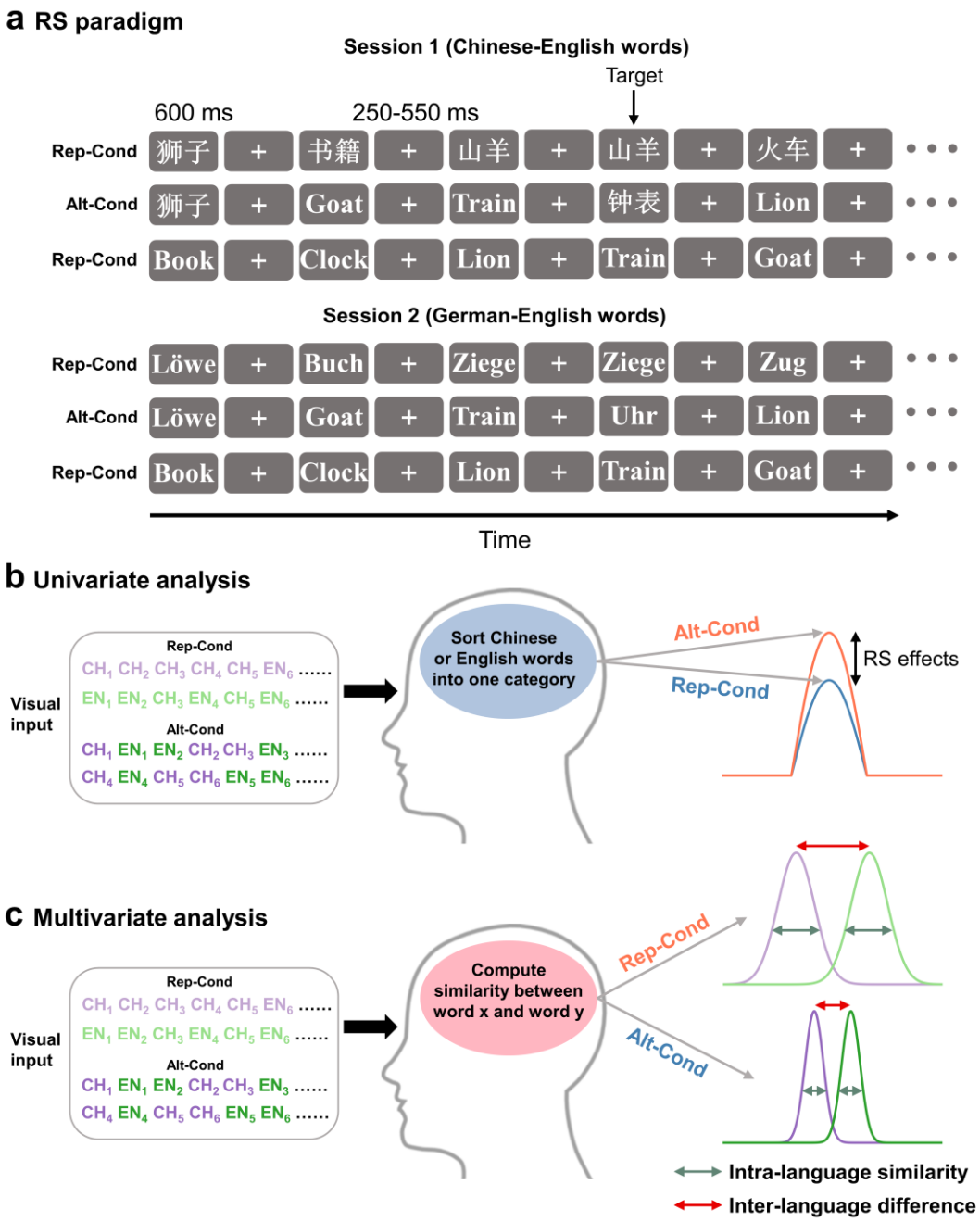

Figure 1. Illustrations of the RS paradigm and univariate/multivariate analyses of brain activities in response to words. (a) The RS paradigm. Words of two different languages (e.g., Chinese vs. English, or English vs. German) were presented alternately in the Alt-Cond, and words of one language were 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. 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.

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).

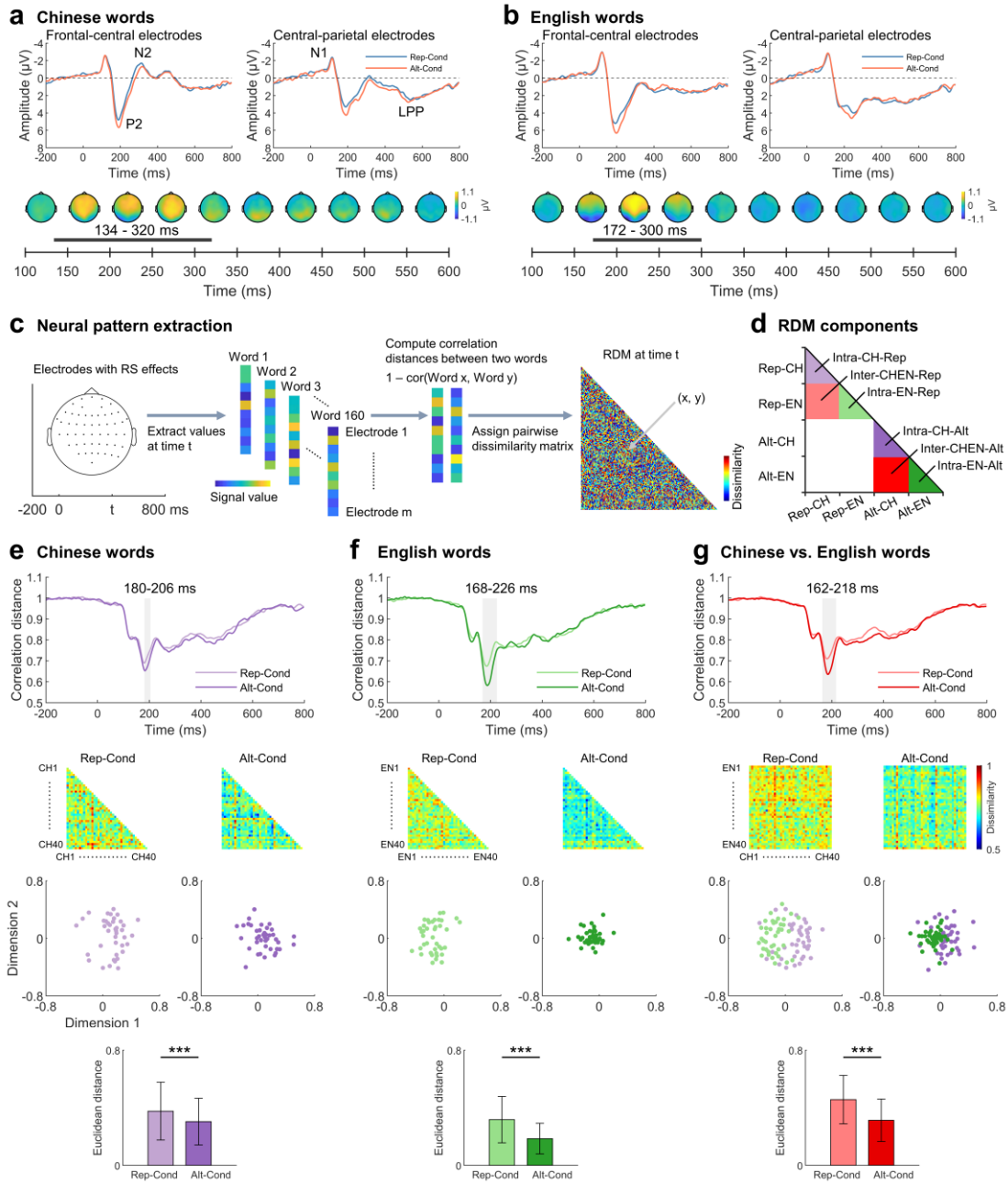

Figure 2. EEG results of Chinese speakers in Experiment 1. (a) and (b) Results of univariate analyses. Top panels illustrate electrophysiological responses to Chinese and English words in the Alt-Cond and Rep-Cond, respectively. Bottom panels show scalp distributions of significant RS effects on neural 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. *** $P < 0.001$.

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 Chinse (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 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, in Experiment 4 and 5, we recorded EEG signals to words from native English speakers (N=34) who learned Chinese (but not German) and native German speakers (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. S6–9 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.

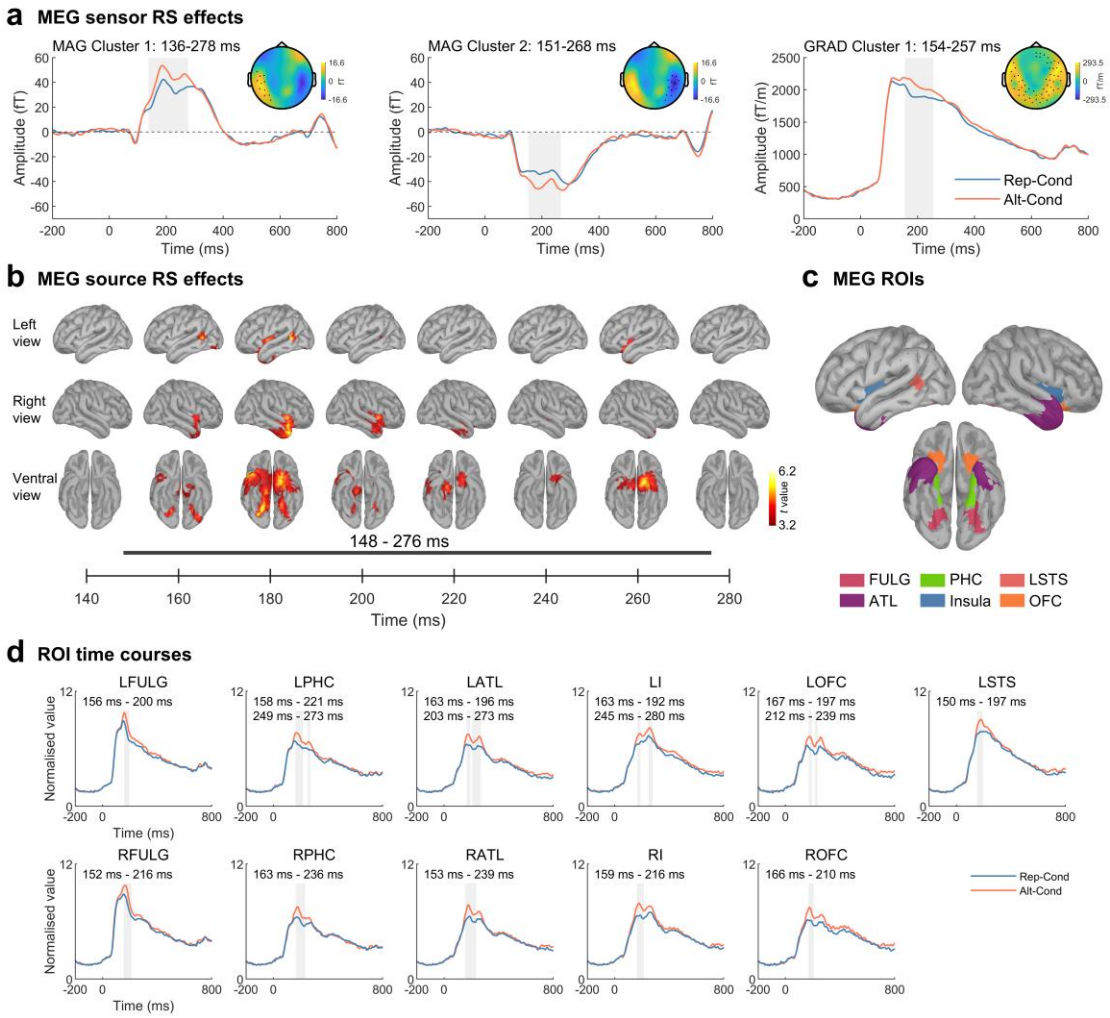

Figure 3. MEG results combined Chinese and English words across Chinese and English speakers in Experiments 7a and 8a. (a) Significant clusters of RS effects on magnetometer signals (the left two panels) and gradiometer signals (the right panel). (b) Results of the whole-brain source analyses. Shown 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;

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 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 (42), 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.

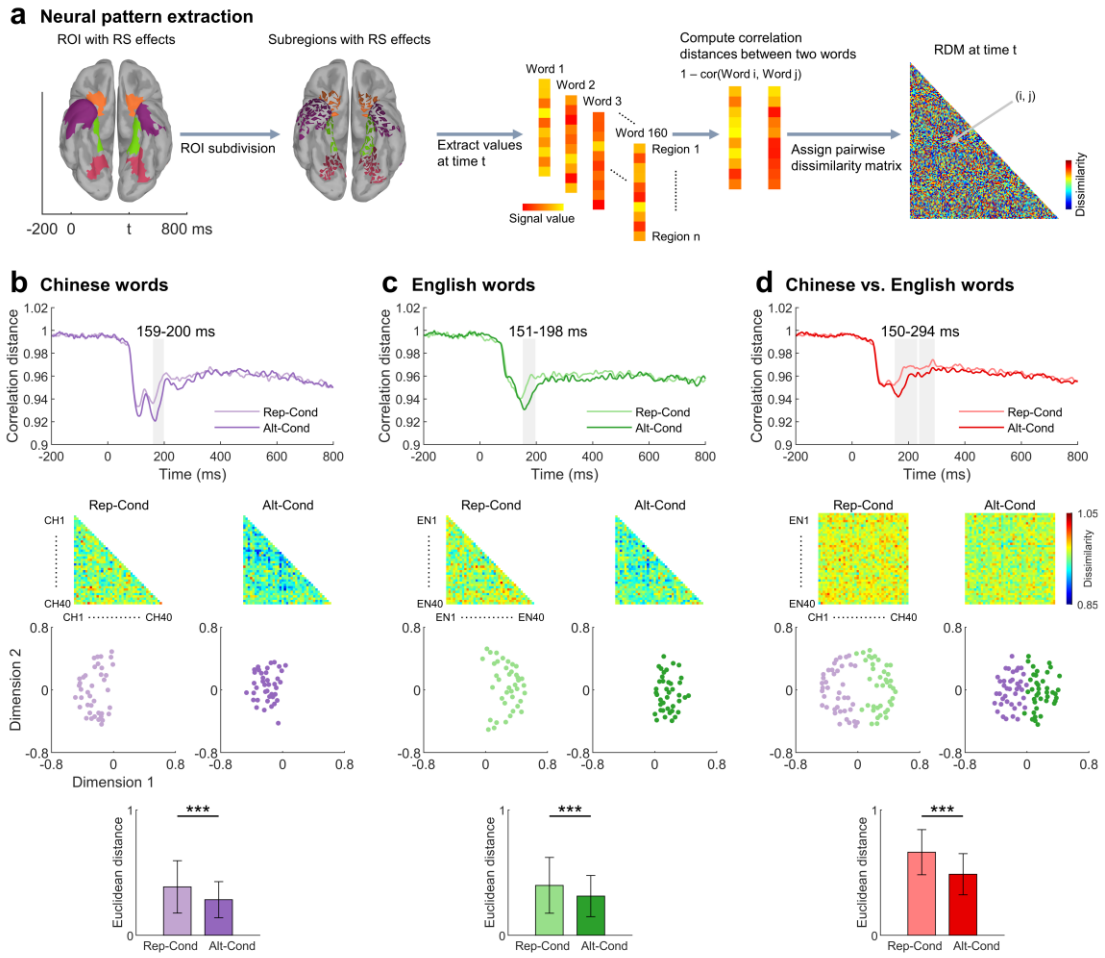

Figure 4. Results of the analyses of the global network RDM across Chinese and English speakers in Experiments 7a and 8a. (a) Illustration of the procedure to calculate network RDMs based on source-space MEG signals. (b), (c), and (d) Results of multivariate analyses. The top two panels show the time 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. ***$P < 0.001$.

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.

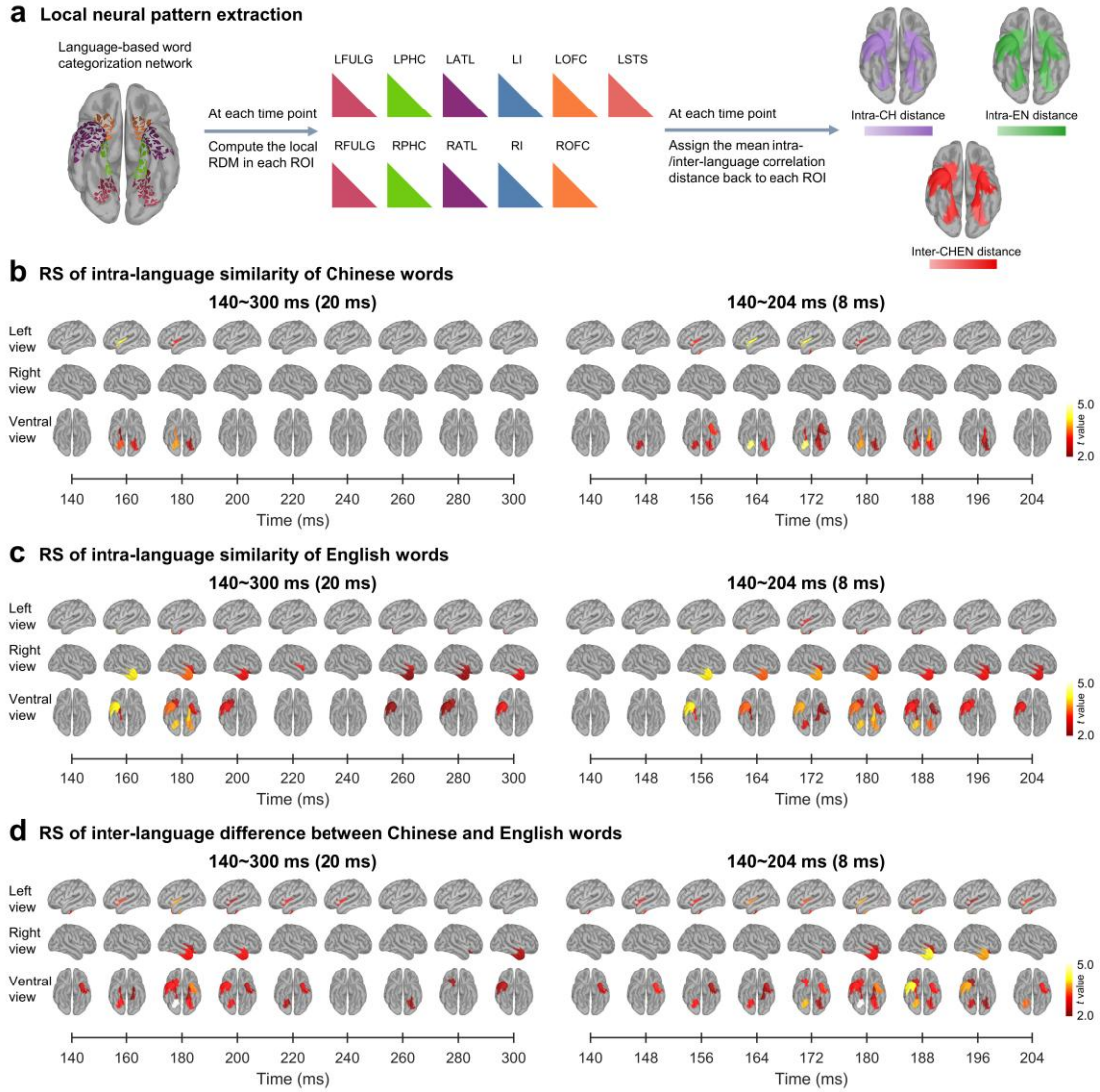

Figure 5. Results of the analyses of local RDMs across Chinese and English speakers in Experiments 7a and 8a. (a) Illustration of the procedure to calculate local RDMs based on source-space MEG signals. (b) and (c) Dynamic RS effects related to intra-language similarity of Chinese and English words in different ROIs. (d) Dynamic RS effects related to inter-language difference in different ROIs.

We further 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.

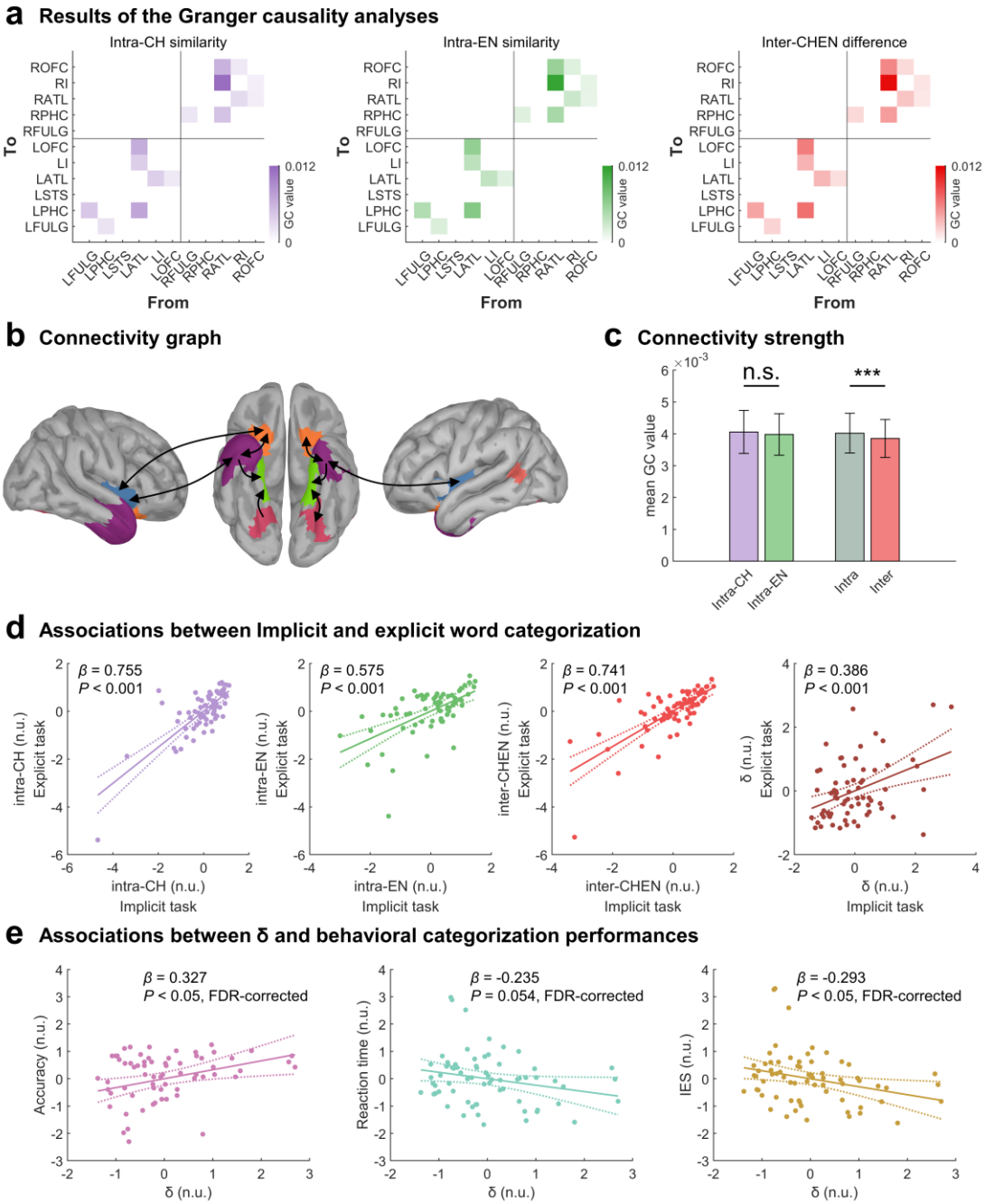

Figure 6. Results of GCA and linear regression analyses across Chinese and English speakers in Experiments 7a and 8a. (a) Connectivity values above a threshold (GC values > 0.001) corresponding to intra-language similarity and inter-language difference. Each square represents connectivity from one brain 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.

We further performed a Granger causality analysis (GCA) (43) 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:

$$\boldsymbol{\delta} = \text{Inter_LD} - (\text{Intra_LS}_{CH} + \text{Intra_LS}_{EN})/2$$

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 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. S15–17). 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 as a critical basis of the social-categorization function of language. 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 (17, 26, 44) 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 (45), lateral frontal cortices (46-48), and anterior cingulate cortex (47, 49, 50). 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 (51, 52). While the PHC in the left hemisphere may contribute to verbal memory (53), 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 (54). 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 (55, 56) or of object names with strong olfactory associations (57). 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 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 (58) and different from the early word form processing that occurs dominantly in the left ventral occipito-temporal cortex around 170 ms after stimulus onset (23-25, 59). 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 (23), which support semantic and phonological processing of words and sentences (26, 27). 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) (60), underlie social categorization of faces and representations of social concepts (e.g., fusiform, ATLs) (32, 34, 61), and encode social information about others (e.g., OFC) (62). 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.

Our findings raise new questions for future research. Our work tested only young adults. Given the finding of language-based social preferences in infants' behaviors (15, 16, 63), 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 (52), 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 (64) and representations of internal and vocalized speech (65) 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 (66). How the word-categorization network interacts with that underlying social categorization of faces (34) to influence real-life social behaviors deserves further investigation.

References

(1) E. Fedorenko, S. T. Piantadosi, E. A. F. Gibson, Language is primarily a tool for communication rather than thought. *Nature* **630**, 575-586 (2024). doi:10.1038/s41586-024-07522-w

(2) M. Bucholtz, K. Hall, "Locating identity in language" In *Language and identities*. C. Llamas & D. Watt Eds. (Edinburgh University Press, 2010), pp. 18-27.
(3) J. M. DeJesus, H. G. Hwang, J. B. Dautel, K. D. Kinzler, “American = English Speaker” before “American = White”: The development of children's reasoning about nationality. *Child Development* **89**, 1752-1767 (2018). doi:10.1111/cdev.12845
(4) K. D. Kinzler, Language as a social cue. *Annual Review of Psychology* **72**, 241-264 (2021). doi:10.1146/annurev-psych-010418-103034
(5) K. D. Kinzler, J. B. Dautel, Children’s essentialist reasoning about language and race. *Developmental science* **15**, 131-138 (2012). doi:10.1111/j.1467-7687.2011.01101.x
(6) K. D. Kinzler, E. Dupoux, E. S. Spelke, The native language of social cognition. *Proceedings of the National Academy of Sciences* **104**, 12577-12580 (2007). doi:10.1073/pnas.0705345104
(7) D. Pietraszewski, A. Schwartz, Evidence that accent is a dimension of social categorization, not a byproduct of perceptual salience, familiarity, or ease-of-processing. *Evolution and Human Behavior* **35**, 43-50 (2014). doi:10.1371/journal.pone.0254513
(8) G. Roberts, Perspectives on language as a source of social markers. *Language and Linguistics Compass* **7**, 619-632 (2013). doi:10.1111/lnc3.12052
(9) I. Sachdev, R. Y. Bourhis, "Language and social identification" In *Social identity theory: Constructive and critical advances*. D. Abrams & M. Hogg Eds. (Harvester Wheatsheaf, 1990), pp. 33-51.
(10) K. R. Scherer, H. Giles, *Social markers in speech* (Cambridge University Press, 1979).
(11) R. D. Greenberg, *Language and identity in the balkans: Serbo-croatian and its disintegration* (Oxford University Press, 2008).
(12) M. Shell, Language wars. *CR: The New Centennial Review* **1**, 1-17 (2001). doi:10.1353/ncr.2003.0059
(13) M. Rhodes, A. Baron, The development of social categorization. *Annual Review of Developmental Psychology* **1**, 359-386 (2019). doi:10.1146/annurev-devpsych-121318-084824
(14) Z. Liberman, A. L. Woodward, K. D. Kinzler, Preverbal infants infer third-party social relationships based on language. *Cognitive Science* **41**, 622-634 (2017). doi:10.1111/cogs.12403
(15) L. H. Howard, A. M. E. Henderson, C. Carrazza, A. L. Woodward, Infants’ and young children's imitation of linguistic in-group and out-group informants. *Child Development* **86**, 259-275 (2015). doi:10.1111/cdev.12299
(16) K. D. Kinzler, E. Dupoux, E. S. Spelke, ‘Native’objects and collaborators: Infants' object choices and acts of giving reflect favor for native over foreign speakers. *Journal of Cognition and Development* **13**, 67-81 (2012). doi:10.1080/15248372.2011.567200
(17) E. Fedorenko, A. A. Ivanova, T. I. Regev, The language network as a natural kind within the broader landscape of the human brain. *Nature Reviews*

Neuroscience **25**, 289-312 (2024). doi:10.1038/s41583-024-00802-4
(18) A. D. Friederici, S. M. E. Gierhan, The language network. *Current Opinion in Neurobiology* **23**, 250-254 (2013). doi:10.1016/j.conb.2012.10.002
(19) S. Malik-Moraleda, D. Ayyash, J. Gallée, J. Affourtit, M. Hoffmann, Z. Mineroff, O. Jouravlev, E. Fedorenko, An investigation across 45 languages and 12 language families reveals a universal language network. *Nature Neuroscience* **25**, 1014-1019 (2022). doi:10.1038/s41593-022-01114-5
(20) W. T. Siok, C. A. Perfetti, Z. Jin, L. H. Tan, Biological abnormality of impaired reading is constrained by culture. *Nature* **431**, 71-76 (2004). doi:10.1038/nature02865
(21) H. Li, J. Zhang, G. Ding, Reading across writing systems: A meta-analysis of the neural correlates for first and second language reading. *Bilingualism: Language and Cognition* **24**, 537-548 (2021). doi:10.1017/S136672892000070X
(22) S. Malik-Moraleda, O. Jouravlev, M. Taliaferro, Z. Mineroff, T. Cucu, K. Mahowald, I. A. Blank, E. Fedorenko, Functional characterization of the language network of polyglots and hyperpolyglots with precision fMRI. *Cerebral Cortex* **34** (2024). doi:10.1093/cercor/bhae049
(23) K. Marinkovic, R. P. Dhond, A. M. Dale, M. Glessner, V. Carr, E. Halgren, Spatiotemporal dynamics of modality-specific and supramodal word processing. *Neuron* **38**, 487-497 (2003). doi:10.1016/s0896-6273(03)00197-1
(24) W. Nan, Y. Liu, X. Zeng, W. Yang, J. Liang, Y. Lan, S. Fu, The spatiotemporal characteristics of N170s for faces and words: A meta-analysis study. *PsyCh Journal* **11**, 5-17 (2022). doi:10.1002/pchj.511
(25) L. Pylkkänen, A. Marantz, Tracking the time course of word recognition with MEG. *Trends in Cognitive Sciences* **7**, 187-189 (2003). doi:10.1016/s1364-6613(03)00092-5
(26) V. J. Hodgson, M. A. Lambon Ralph, R. L. Jackson, Multiple dimensions underlying the functional organization of the language network. *NeuroImage* **241**, 118444 (2021). doi:10.1016/j.neuroimage.2021.118444
(27) Y. Zhu, M. Xu, J. Lu, J. Hu, V. P. Y. Kwok, Y. Zhou, D. Yuan, B. Wu, J. Zhang, J. Wu, L. H. Tan, Distinct spatiotemporal patterns of syntactic and semantic processing in human inferior frontal gyrus. *Nature Human Behaviour* **6**, 1104-1111 (2022). doi:10.1038/s41562-022-01334-6
(28) G. Zhang, Y. Xu, X. Wang, J. Li, W. Shi, Y. Bi, N. Lin, A social-semantic working-memory account for two canonical language areas. *Nature Human Behaviour* **7**, 1980-1997 (2023). doi:10.1038/s41562-023-01704-8
(29) M. Arioli, C. Gianelli, N. Canessa, Neural representation of social concepts: A coordinate-based meta-analysis of fMRI studies. *Brain Imaging and Behavior* **15**, 1912-1921 (2021). doi:10.1007/s11682-020-00384-6
(30) I. R. Olson, D. McCoy, E. Klobusicky, L. A. Ross, Social cognition and the anterior temporal lobes: A review and theoretical framework. *Social Cognitive and Affective Neuroscience* **8**, 123-133 (2013). doi:10.1093/scan/nss119
(31) C. Pang, N. Zhou, Y. Deng, Y. Pu, S. Han, Neural tracking of race-related

information during face perception. *Neuroscience Bulletin* **Advance online publication** (2025). doi:10.1007/s12264-025-01419-y
(32) G. Pobric, M. A. Lambon Ralph, R. Zahn, Hemispheric specialization within the superior anterior temporal cortex for social and nonsocial concepts. *Journal of Cognitive Neuroscience* **28**, 351-360 (2016). doi:10.1162/jocn_a_00902
(33) R. Zahn, J. Moll, F. Krueger, E. D. Huey, G. Garrido, J. Grafman, Social concepts are represented in the superior anterior temporal cortex. *Proceedings of the National Academy of Sciences* **104**, 6430-6435 (2007). doi:10.1073/pnas.0607061104
(34) Y. Zhou, T. Gao, T. Zhang, W. Li, T. Wu, X. Han, S. Han, Neural dynamics of racial categorization predicts racial bias in face recognition and altruism. *Nature Human Behaviour* **4**, 69-87 (2020). doi:10.1038/s41562-019-0743-y
(35) T. Zhang, Y. Zhou, S. Han, Priority of racial and gender categorization of faces: A social task demand framework. *Journal of Personality and Social Psychology* **124**, 483 (2023). doi:10.1037/pspa0000318
(36) Y. Pu, S. Han, Neural basis of categorical representations of animal body silhouettes. *Neuroscience Bulletin* **41**, 211-223 (2025). doi:10.1007/s12264-024-01268-1
(37) T. Zhang, S. Han, Non-phase-locked alpha oscillations are involved in spontaneous racial categorization of faces. *Neuropsychologia* **160**, 107968 (2021). doi:10.1016/j.neuropsychologia.2021.107968
(38) D. J. Freedman, M. Riesenhuber, T. Poggio, E. K. Miller, Categorical representation of visual stimuli in the primate prefrontal cortex. *Science* **291**, 312-316 (2001). doi:10.1126/science.291.5502.312
(39) T. A. Ito, B. D. Bartholow, The neural correlates of race. *Trends in Cognitive Sciences* **13**, 524-531 (2009). doi:10.1016/j.tics.2009.10.002
(40) N. Kriegeskorte, M. Mur, D. A. Ruff, R. Kiani, J. Bodurka, H. Esteky, K. Tanaka, P. A. Bandettini, Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* **60**, 1126-1141 (2008). doi:10.1016/j.neuron.2008.10.043
(41) N. Kriegeskorte, R. Goebel, P. Bandettini, Information-based functional brain mapping. *Proceedings of the National Academy of Sciences* **103**, 3863-3868 (2006). doi:10.1073/pnas.0600244103
(42) A. Klein, J. Tourville, 101 labeled brain images and a consistent human cortical labeling protocol. *Frontiers in Neuroscience* **6**, 171 (2012). doi:10.3389/fnins.2012.00171
(43) L. Barnett, A. K. Seth, The MVGC multivariate Granger causality toolbox: A new approach to Granger-causal inference. *Journal of Neuroscience Methods* **223**, 50-68 (2014). doi:10.1016/j.jneumeth.2013.10.018
(44) L. H. Tan, A. R. Laird, K. Li, P. T. Fox, Neuroanatomical correlates of phonological processing of Chinese characters and alphabetic words: A meta-analysis. *Human Brain Mapping* **25**, 83-91 (2005). doi:10.1002/hbm.20134
(45) J. Crinion, R. Turner, A. Grogan, T. Hanakawa, U. Noppeney, J. T. Devlin, T. Aso, S. Urayama, H. Fukuyama, K. Stockton, K. Usui, D. W. Green, C. J.

Price, Language control in the bilingual brain. *Science* **312**, 1537-1540 (2006). doi:10.1126/science.1127761
(46) A. Rodriguez-Fornells, A. v. d. Lugt, M. Rotte, B. Britti, H.-J. Heinze, T. F. Münte, Second language interferes with word production in fluent bilinguals: Brain potential and functional imaging evidence. *Journal of Cognitive Neuroscience* **17**, 422-433 (2005). doi:10.1162/0898929053279559
(47) Y. Wang, G. Xue, C. Chen, F. Xue, Q. Dong, Neural bases of asymmetric language switching in second-language learners: An ER-fMRI study. *NeuroImage* **35**, 862-870 (2007). doi:10.1016/j.neuroimage.2006.09.054
(48) J. D. Zhu, R. A. Seymour, A. Szakay, P. F. Sowman, Neuro-dynamics of executive control in bilingual language switching: An MEG study. *Cognition* **199**, 104247 (2020). doi:10.1016/j.cognition.2020.104247
(49) J. Abutalebi, S. M. Brambati, J.-M. Annoni, A. Moro, S. F. Cappa, D. Perani, The neural cost of the auditory perception of language switches: An event-related functional magnetic resonance imaging study in bilinguals. *The Journal of Neuroscience* **27**, 13762-13769 (2007). doi:10.1523/jneurosci.3294-07.2007
(50) E. Blanco-Elorrieta, K. Emmorey, L. Pylkkänen, Language switching decomposed through MEG and evidence from bimodal bilinguals. *Proceedings of the National Academy of Sciences* **115**, 9708-9713 (2018). doi:doi:10.1073/pnas.1809779115
(51) S. Dehaene, L. Cohen, The unique role of the visual word form area in reading. *Trends in Cognitive Sciences* **15**, 254-262 (2011). doi:10.1016/j.tics.2011.04.003
(52) M. Zhan, C. Pallier, A. Agrawal, S. Dehaene, L. Cohen, Does the visual word form area split in bilingual readers? A millimeter-scale 7-T fMRI study. *Science Advances* **9**, eadf6140 (2023). doi:10.1126/sciadv.adf6140
(53) A. Alessio, L. Bonilha, C. Rorden, E. Kobayashi, L. L. Min, B. P. Damasceno, F. Cendes, Memory and language impairments and their relationships to hippocampal and perirhinal cortex damage in patients with medial temporal lobe epilepsy. *Epilepsy Behavior* **8**, 593-600 (2006). doi:10.1016/j.yebeh.2006.01.007
(54) E. C. Ferstl, J. Neumann, C. Bogler, D. Y. von Cramon, The extended language network: A meta-analysis of neuroimaging studies on text comprehension. *Human Brain Mapping* **29**, 581-593 (2008). doi:10.1002/hbm.20422
(55) N. Alia-Klein, R. Z. Goldstein, D. Tomasi, L. Zhang, S. Fagin-Jones, F. Telang, G. J. Wang, J. S. Fowler, N. D. Volkow, What is in a word? No versus yes differentially engage the lateral orbitofrontal cortex. *Emotion* **7**, 649-659 (2007). doi:10.1037/1528-3542.7.3.649
(56) X. Gu, S. Han, Neural substrates underlying evaluation of pain in actions depicted in words. *Behavioural Brain Research* **181**, 218-223 (2007). doi:10.1016/j.bbr.2007.04.008
(57) P. Han, I. Croy, C. Raue, M. Bensafi, M. Larsson, A. Cavazzana, T. Hummel, Neural processing of odor-associated words: An fMRI study in patients with acquired olfactory loss. *Brain Imaging and Behavior* **14**, 1164-1174 (2020).

doi:10.1007/s11682-019-00062-2
(58) G. Giari, E. Leonardelli, Y. Tao, M. Machado, S. L. Fairhall, Spatiotemporal properties of the neural representation of conceptual content for words and pictures - an MEG study. *Neuroimage* **219**, 116913 (2020). doi:10.1016/j.neuroimage.2020.116913
(59) T. Thesen, C. R. McDonald, C. Carlson, W. Doyle, S. Cash, J. Sherfey, O. Felsovalyi, H. Girard, W. Barr, O. Devinsky, R. Kuzniecky, E. Halgren, Sequential then interactive processing of letters and words in the left fusiform gyrus. *Nature Communications* **3**, 1284 (2012). doi:10.1038/ncomms2220
(60) C. C. Merritt, J. K. MacCormack, A. G. Stein, K. A. Lindquist, K. A. Muscatell, The neural underpinnings of intergroup social cognition: An fMRI meta-analysis. *Social Cognitive and Affective Neuroscience* **16**, 903-914 (2021). doi:10.1093/scan/nsab034
(61) A. J. Golby, J. D. Gabrieli, J. Y. Chiao, J. L. Eberhardt, Differential responses in the fusiform region to same-race and other-race faces. *Nature Neuroscience* **4**, 845-850 (2001). doi:10.1038/90565
(62) S. A. Park, D. S. Miller, H. Nili, C. Ranganath, E. D. Boorman, Map making: Constructing, combining, and inferring on abstract cognitive maps. *Neuron* **107**, 1226-1238.e1228 (2020). doi:10.1016/j.neuron.2020.06.030
(63) Z. Liberman, A. L. Woodward, K. D. Kinzler, The origins of social categorization. *Trends in Cognitive Sciences* **21**, 556-568 (2017). doi:10.1016/j.tics.2017.04.004
(64) M. Jamali, B. Grannan, J. Cai, A. R. Khanna, W. Muñoz, I. Caprara, A. C. Paulk, S. S. Cash, E. Fedorenko, Z. M. Williams, Semantic encoding during language comprehension at single-cell resolution. *Nature* **631**, 610-616 (2024). doi:10.1038/s41586-024-07643-2
(65) S. K. Wandelt, D. A. Bjånes, K. Pejsa, B. Lee, C. Liu, R. A. Andersen, Representation of internal speech by single neurons in human supramarginal gyrus. *Nature Human Behaviour* **8**, 1136-1149 (2024). doi:10.1038/s41562-024-01867-y
(66) T. Rakić, M. C. Steffens, A. Mummendey, Blinded by the accent! The minor role of looks in ethnic categorization. *Journal of Personality and Social Psychology* **100**, 16 (2011). doi:10.1037/a0021522
(67) Q. Cai, M. Brysbaert, SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLOS ONE* **5**, e10729 (2010). doi:10.1371/journal.pone.0010729
(68) M. Brysbaert, B. New, 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. *Behavior Research Methods* **41**, 977-990 (2009). doi:10.3758/BRM.41.4.977
(69) M. Brysbaert, M. Buchmeier, M. Conrad, A. M. Jacobs, J. Bölte, A. Böhl, The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology* **58**, 412-424 (2011). doi:10.1027/1618-3169/a000123

(70) A. Delorme, S. Makeig, EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods* **134**, 9-21 (2004). doi:10.1016/j.jneumeth.2003.10.009
(71) L. Pion-Tonachini, K. Kreutz-Delgado, S. Makeig, ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *Neuroimage* **198**, 181-197 (2019). doi:10.1016/j.neuroimage.2019.05.026
(72) R. Oostenveld, P. Fries, E. Maris, J. M. Schoffelen, Fieldtrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence Neuroscience* **2011**, 156869 (2011). doi:10.1155/2011/156869
(73) E. Maris, R. Oostenveld, Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods* **164**, 177-190 (2007). doi:10.1016/j.jneumeth.2007.03.024
(74) F. Tadel, S. Baillet, J. C. Mosher, D. Pantazis, R. M. Leahy, Brainstorm: A user-friendly application for MEG/EEG analysis. *Computational Intelligence Neuroscience* **2011**, 879716 (2011). doi:10.1155/2011/879716
(75) J. F. Geweke, Measures of conditional linear dependence and feedback between time series. *Journal of the American Statistical Association* **79**, 907-915 (1984). doi:10.1080/01621459.1984.10477110
(76) C. W. J. Granger, Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**, 424-438 (1969). doi:10.2307/1912791
(77) L. Barnett, A. K. Seth, Granger causality for state-space models. *Physical Review E* **91**, 040101 (2015). doi:10.1103/PhysRevE.91.040101
(78) D. C. Van Essen, A population-average, landmark- and surface-based (PALS) atlas of human cerebral cortex. *NeuroImage* **28**, 635-662 (2005). doi:10.1016/j.neuroimage.2005.06.058
(79) E. Fedorenko, I. A. Blank, Broca's area is not a natural kind. *Trends in Cognitive Sciences* **24**, 270-284 (2020). doi:10.1016/j.tics.2020.01.001
(80) A. D. Friederici, The brain basis of language processing: From structure to function. *Physiological Reviews* **91**, 1357-1392 (2011). doi:10.1152/physrev.00006.2011
(81) A. Ardila, B. Bernal, M. Rosselli, Should Broca's area include Brodmann area 47? *Psicothema* **29**, 73-77 (2017). doi:10.7334/psicothema2016.11
(82) H. Duffau, L. Capelle, D. Denvil, P. Gatignol, N. Sichez, M. Lopes, J. P. Sichez, R. Van Effenterre, The role of dominant premotor cortex in language: A study using intraoperative functional mapping in awake patients. *Neuroimage* **20**, 1903-1914 (2003). doi:10.1016/s1053-8119(03)00203-9
(83) F. Pulvermüller, Brain mechanisms linking language and action. *Nature Reviews Neuroscience* **6**, 576-582 (2005). doi:10.1038/nrn1706

Acknowledgments: The authors thank Zhirui Zhao and Tengbin Huo for help with

EEG and MEG data collection.

Funding: This work was supported by the National Natural Science Foundation of China (projects32230043 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.

Author contributions: S.H. and G.Z. conceived and designed the study, collected and analyzed the data, and wrote the manuscript. S.H. supervised all aspects of the research.

Competing interests: The authors declare no conflicts of interest.

Data and materials availability: Data and codes for data analyses in this study are available at:
http://datadryad.org/share/xB81CGC1LlwHF-D_G1A_fJJhvRSu5wnXs8jjG7lRjIo