Toward Conceptual Networks in Brain : Decoding Imagined Words from Word Reading

Linyang He[1*], **Shujie Geng**[1*], **Jiawei Han**[1], **Miao Cao**[1†] and **Jianfeng Feng**[1,2,3,4†]
[1]Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University
[2]Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence, Ministry of Education, China
[3]MoE Frontiers Center for Brain Science, Fudan University
[4]Zhangjiang Fudan International Innovation Center
{linyang_he, sjgeng20, jwhan19, mcao, jffeng}@fudan.edu.cn

Abstract

Language is an epiphenomenon of human's subjective world which is noted as the conceptual network. Human beings realized communication of knowledge, experience, and symbolic entity of subjective psyche, across time and space by language which included but not limited to spoken or writing systems. From the perspective of computational linguistics, one concept in the conceptual network would be identically activated despite variations of modalities (i.e. comprehension, generation or production). In the current study, we conducted a semantic-access word reading task (language comprehension) and a word imagining task (language generation) in Chinese native speakers during fMRI scanning. Part-of-speech category and lexicon of stimuli in word imagining task were predicted by brain responses in the word reading task. Significantly, our learning model, which was trained from brain activation of word reading, achieved decoding both imagined words and semantically transferred imagined words. To our knowledge, this is the first report of cross-modality and semantics transferring decoding of imagined speech. Given the huge processing discrepancies between language comprehension and generation, our results demonstrated a stable conceptual network in the human brain and flexible access from linguistic ways to conceptual network, which shed light on understanding brain mechanisms of the relationship between language and thought.

1 Introduction

The relationship between language and thought is a fundamental question in philosophy, linguistics, psychology, and cognitive science and has been controversial over centuries. It is proposed that language is language itself and separates from thought, however, experimental and quantitive evidence was insufficient.

Human being's subjective psyche is internalized as neuron-based conceptual network and externalized as symbolic knowledge and experience. With the blooming development of neural imaging techniques and natural language processing, studies decoding linguistic information in the brain yielded initial results. Brain-Computer Interface (BCI) studies achieved successful decoding/prediction for vowels and consonants(Pei et al., 2011), word classification(Martin et al., 2016), and spoken phrases(Herff et al., 2015). Remarkably, Chang et al., utilized invasive high-density multielectrode EEG and recurrent neural networks to decode cortical articulatory movement representations for spoken sentences and achieved speech synthesis at sentence level(Anumanchipalli et al., 2019). Subsequently, Chang et al. developed neuroprosthesis by decoding cortical articulatory representations and applied it in a patient with anarthria and spastic quadriparesis, with a real time decoding rate of 15.2 words/min and word error rate of 25.6%(Moses et al., 2021).

However, limitations were transparent. On one hand, in previous studies, the decoding model usually worked poorly, that is, the decoder could merely distinguish among a few simple words in one single subject with invasive neuroimaging methods. Despite of individual variances, low signal-to-noise ratio and other neurophysiological noises, spatial resolution and/or local signals of EEG/MEG can only reveal partial linguistic processing information, which was the key shortcoming. What's more, target decoding linguistic features ranged from phoneme, vowels and consonants, words, phrases, sentences to imagined speech, which corresponded to various language and cognitive processing stages. Thus, it is largely unknown which linguistic feature/cognitive manipulation contributed the most to decoding. In Chang's study, articulatory movement information was considered as the main feature to achieve de-

* Equal contribution
† Corresponding authors

coding. By focusing on an exact stage (mapping from cortical articulatory movement representations to speech acoustics in vocal organs), they obtained better decoding performance and interpretability. Above all, the potential solution would be globally decoding conceptual networks in human brains.

A key issue before globally decoding a conceptual network lies in testing its stability and flexible access across linguistic ways. Researchers explored mapping between brain responses and semantic features at the word, phrase and sentence level respectively and found 'semantic systems' located 'everywhere' in the brain. Gallant et al. first depicted globally voxel-wised neural semantic maps through hours of online narrative stories listening tasks and data-driven methods(Huth et al., 2016a). Furthermore, whether semantic feature-related activation was modality-independent was explored. Gerven et al. found that prediction of semantic categories can be performed by brain activation in the left inferior temporal cortex and frontal cortex which is independent of information modalities (spoken and written names, photographs, and natural sounds)(Simanova et al., 2014). Fedorenko et al. built a universal decoder which inferred semantic information at all word, phrase and sentences levels, regardless of the information form(words v.s. images)(Pereira et al., 2018). However, semantic and visual properties of language showed complex dynamics in very early stages which were modulated by top-down information from high-order language brain regions. It is insufficient to prove the stability of conceptual network by information-form independent semantic access only in language comprehension/generation/production.

To bridge the gap, we aimed to predict semantic information in language generation from language comprehension. Specifically, in this study, Chinese native speakers were recruited to participate in fMRI-based online semantic-access word reading task (language comprehension) and word imagining task (language generation). Higher-level neural linguistic representations (conceptual network) learned from word reading were utilized to predict 1) read words, 2) imagined words and 3) semantically transferred imagined words at different linguistic levels: 1) part-of-speech(noun/verb) classification; 2) lexicon decoding across subjects. Also, two voxel selection algorithms: 1) data-driven voxel-wise linear regression (VWLR); 2) hypothesis-driven representational similarity analysis (RSA) were applied. Results showed that VWLR worked better in lexicon decoding while RSA had good performance in POS classification. Most importantly, our decoder predicted semantically transferred imagined words which were not present in word reading, indicating successful modality crossing and semantics transferring. The above results suggest the stability and effectiveness of the conceptual network and the human brain's generalization ability to transfer between thought and language.

2 Methods

2.1 Subjects

In experiment-1, semantic-access reading task, we recruited 47 native Mandarin speakers (Mean age = 23.4, 25 males) in the campus of Fudan University with vision/corrected vision over 4.8 and no history of neurological disease or psychiatric disorders. 41 subjects were evaluated as right-handed by Edinburgh Handedness Inventory and the remaining subjects were balanced. All subjects provided informed written consent before the formal experiment. Experiment-1 was approved by the Ethics Committee of the School of Life Sciences of Fudan University.

In experiment-2, imagined speech task, we recruited 24 college students (Mean age = 23.6, 10 males) with vision/corrected vision over 4.8 and no history of neurological disease or psychiatric disorders. All were identified as right-handed. Informed written consent were required before fMRI scanning. Experiment-2 was approved by the Ethics Committee of Institute of Science and Technology for Brain-inspired Intelligence of Fudan University.

2.2 Stimuli and Word Embedding

There were 40 Chinese words(see Appendix C) used in experiment-1, with 13 of them being nouns, 13 verbs, 13 adjectives and the left one's POS being ambiguous(noun/verb). In experiment-2, in addition to 40 words above, there were 20 new words, 10 of which were verbs and 10 nouns. To make sure the conceptual network played a key role in predicting new words, we first selected 160 candidate words and then chose 20 new words from these candidates. These 20 new words should fall into the semantic space range which was covered by the above 40 old words as Figure 1 shows.

The semantic space was constructed with word vectors. Previous studies showed that semantic vectors can predict human measurements(Murphy et al., 2018; Pereira et al., 2016). Here we used 300-dimensional word vectors pre-trained from Chinese Wikipedia(Li et al., 2018). In detail, the representation algorithm we set up was skip-gram with negative sampling(SGNS), while the context features including word, character and bigram were all used. All words mentioned above were commonly used two-character words selected from the *Modern Chinese Frequency Dictionary* with average frequency balanced between nouns and verbs(See Appendix C).

Figure 1: 60 words used in this study. 40 gray words were presented in both reading and imagining tasks while 20 colored ones were used only for transferred imagining. Visualization after T-SNE dimensionality reduction.

2.3 Experimental Designs

In experiment-1, stimuli from Chinese words, English words, and Chinese pinyin(the latter two were fillers) were presented on black screen with horizontal visual angle as 4.37°, with pixel, number of strokes, and word frequency matched and no semantic equivalent across conditions. There were 40 trials in each category. We applied event-related design and each stimulus was visually presented for 1000 ms. Inter stimulus interval was 4000-6000ms. Presented orders were randomized across subjects. Participants were required to press 3 different buttons corresponding to stimuli categories after semantic access. A practice experiment was performed before fMRI scanning. Only brain responses to Chinese words were analyzed in the current study.

In experiment-2, the stimuli set contained 40 Chinese words presented in experiment-1 and 20 new Chinese words and were divided into 30 pairs.

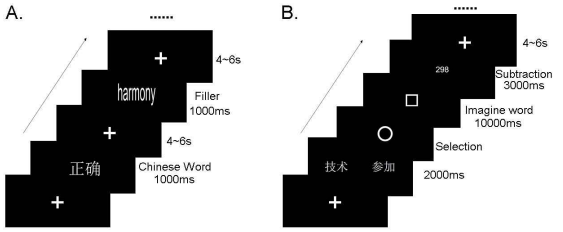

Figure 2: Experiment design

Each pair of words were visually presented on screen for 2000ms. Subjects were required to do word selection by button-press during a circle presented in screen center. After selection, a box was presented in the screen center for 10s and subjects were required to do the word imagining task:they needed to complete semantic associations of the selective word as much as possible, with subtraction manipulation followed to stop semantic associations. A practice experiment was performed before fMRI scanning.

2.4 fMRI Data Collection and Prepossessing

We conducted anatomical and functional magnetic resonance imaging (MRI) scanning in 3.0-T Siemens Prisma scanner at Zhangjiang International Brain Imaging Center (ZIC) of Fudan University, Shanghai. Experiment-1 and experiment-2 utilized the same sequences with parameters listed below. Echo planar imaging (EPI): TR = 720ms, TE = 33ms, matrix size = 110*96, number of slices = 72, slice thickness = 2mm, flip angle =52°, field of view (FOV) = 220*196 mm, and T1-weighted imaging: TR =3000ms, TE = 2.56ms, matrix size = 320*320, number of slices =208, flip angle = 8°, field of view (FOV) = 256*256 mm. Data preprocessing were carried out by Statistical Parametric Mapping-12 (SPM12, Wellcome Trust Centre for Neuroimaging, London, UK. http://www.fil.ion.ucl.ac.uk/spm). Slice timing, realignment, co-registration, segmentation, normalization and smooth were performed.

2.5 Two Voxel Selection Algorithms

The fMRI data is usually highly dimensional and has a lot of noise. To facilitate effective decoding, the brain image data will be reduced to a low dimension before subsequent statistical analysis(Huth et al., 2016b; Pereira et al., 2018). We investigated two voxel selection algorithms. One is data-driven voxel-wise linear regression, the other is representation similarity analysis, which is hypothesis-driven.

Before these two algorithms were implemented, a simple non-zero mask was applied to the whole brain to reduce the computation complexity.

2.5.1 Voxel-wise Linear Regression(VWLR)

Voxel-wise encoding and decoding models have been used widely in computational cognitive neuroscience to relate brain measurements to computational models(Wu et al., 2006; Naselaris et al., 2011, 2015; Huth et al., 2016a). In the current study, for each voxel, we performed L1-regularization linear regression(LASSO) between its 1-level neural signal values of all subjects and 300-dimensional embedding vectors of 40 overt reading words using 10-fold cross-validation. Then we ranked all voxels according to their average training R^2 scores. Top 1000/1500/2000/2500/5000 informative voxels would be used for follow-up training and testing.

2.5.2 Representational Similarity Analysis

Representational similarity analysis (RSA) was initially proposed by Kriegeskorte et al. (2008). RSA is a computational model that describes the relevance between stimuli and corresponding brain responses. In addition to analyzing the differences in the representation of different types of stimuli in neural signals, the correlation between behavior and neural measurements can also be constructed. In general, we want to use the RSA model to localize the most relevant voxels correlating to semantic word vectors. These voxels would later be used in decoder training and testing.

Representational Dissimilarity Matrix(RDM) Before performing RSA, we constructed RDMs of both neural data and behavioral data. RDM extracts the information carried by a given representation, whether in the brain or stimuli. It describes the geometric distances in multi-dimensional response space of different experimental conditions. The distance or dissimilarity is often computed as $1 - similarity$. In fact, there is no much different statistical impact whether using measure of similarity or measure of dissimilarity but the latter seems more popular considering its wide usage in other technique(latent semantic analysis, etc)(Kriegeskorte et al., 2008).

Neural RDM and Searchlight Given a certain stimulus, for one single voxel of a participant, there's a corresponding 1-level neural signal. To enhance the generalization of RDM representations, however, here we used the searchlight algorithm which was firstly proposed in Kriegeskorte et al. (2006). Briefly, the searchlight algorithm scans a spherical region of interest(ROI), rather than a single voxel, of a given radius surrounding a given voxel. The statistics of the entire sphere will be treated as the statistics of the voxel in the middle and will be applied to further statistical analysis including regression and classification. It can also be used in RSA to generate neural RDM(Clarke and Tyler, 2014). In this study, for each subject, given a word stimulus, there's an extended 1-level vector for each voxel. We set the radius of the sphere to 3 voxels, so the vector is of $((3 \times 2) + 1)^3 = 343$ dimension. In the POS classification task, there are 13 nouns and 13 verbs. For either POS, the neural representation of a voxel in one subject is of $[13, 343]$ shape. After computing the dissimilarity(1 - Spearman's ρ) pairwise, we can get the neural RDMs($[13, 13]$) for both noun and verb conditions. In the lexicon decoding task, the neural RDMs were constructed similarly while we treated all the words stimuli as one condition. Therefore, the neural representation of a voxel is a $[40, 343]$ matrix, generating $[40, 40]$ neural RDM.

Behavioral RDM In this study, behavioral RDMs were generated through computing $1 - cosine\ similarity$ between 300-dimensional word vectors. Just like neural RDMs, in POS classification tasks, after performing pairwise comparison between the 13 nouns and calculating the dissimilarity, we got the noun RDM($[13, 13]$). Verb RDM ($[13, 13]$) could be obtained by the same way. In the lexicon decoding task, a $[40, 40]$ word RDM was constructed.

Representational Similarity Analysis After we got the neural RDMs and behavioral RDMs, we performed RSA. We took down the triangular matrix of all RDMs and converted them into one dimension. In the POS task, for each voxel, we computed the Spearman rank-order correlation coefficient ρ between the flattened noun/verb RDM and their corresponding neural RDMs of each subject. Fisher r-z transformation was applied to the coefficient. We then made a one-sided T-test ($H_0 : \mu > 0$) across subjects. Top $N/2$ ($N \in \{1000, 1500, 2000, 2500, 5000\}$) voxels with highest z-scores and $p < 0.05$ were selected in both noun and verb conditions and were then merged into top N informative voxels for further analysis. In lexicon decoding task, similar steps were adopted except for only one condition, and top N

voxels were directly selected.

2.6 Two Decoding Tasks

In order to explore how different voxel selection methods perform at different linguistic levels, we designed two decoding tasks.

2.6.1 Part-of-Speech classification

The first one was a coarse-grained part-of-speech classification task. As previous studies suggested, the neural substrates of Chinese noun and verb processing vary a lot from each other(Yu et al., 2011, 2012, 2013; Yang et al., 2017), while the distinction between many adjectives and nouns is very vague and ambiguous. In addition, the size balance between different labels should also be considered. Thus, we let the decoder implement a binary classification just between nouns and verbs. In this task, a support vector machine(SVM) was selected as the learning algorithm. The kernel function of the SVM classification model was radial basis function(RBF).

2.6.2 Lexicon decoding

Apart from POS classification, we designed a more fine-granted lexicon decoding task. In this task, the learning algorithm needed to predict the corresponding 300-dimensional semantic vector(rather than a label) given a brain image. The predicted vector would be compared with all other word vectors later. We set the decoder as a lasso linear regression, which was L1-regularization.

2.7 Decoder Training

In the POS classification task, 1-level signals of 13 nouns and 13 verbs from all subjects in overt reading(experiment-1) were the training set. Neural data of 23 nouns(13 of which were the same as the training set while 10 of which were new imagined words) and 23 verbs from subjects in imagined speech task(experiment-2) were the test set. In the lexicon decoding task, fMRI data of all overt reading 40 words were trained while neural signals from 60 words in imagined speech were tested. In both decoding tasks, decoders were trained using a leave-one-subject-out(LOSO) cross-validation procedure. There were 47 subjects in experiment-1 for decoder training. Hence the overt reading data set was divided into 47 blocks. Data in each block came from the same subject. 46 blocks were used as the training set and the remaining one used for evaluation in overt reading task.

2.8 Decoding and Evaluation

POS Classification Accuracy For the POS classification task, we used the accuracy score in the classification confusion matrix as the test metric.

Lexicon Decoding Rank Accuracy For the lexicon decoding task, following the setting in Pereira et al. (2018), a rank accuracy was defined. More formally, for a given brain image, the decoder would predict a 300-dimensional word vector. We computed the cosine similarity scores between the predicted vector and all word vectors in the test set. We then ranked all the words in the test set according to the cosine similarity scores decreasingly. Suppose the corresponding real word for this certain image ranks i th, and there are n words in the test set in total, then the lexicon decoding evaluation score can be given as rank accuracy: $r = \frac{n-i}{n-1}$

Considering we let the learning algorithm decode both word reading and imagined speech, there would be evaluation results for both types of speech, too.

Reading Words Decoding Decoders' overt reading performance was quantified as the mean value of remaining subjects' test scores in LOSO cross-validation, in both POS classification and lexicon decoding tasks.

Imagined Words Decoding Decoders trained with overt reading data were then used to predict the same 40 words in the training set, but in imagining condition. Predicted labels or semantic vectors were compared with the real ones to compute the evaluation score using above metrics.

Transferred Imagined Words Decoding Since there were 20 additional words in imagined speech not shown in the overt reading presentation, the training set did not cover these 20 words. Here we defined the decoding of new imagined words as transferred imagined speech decoding, which implied the semantics transferring based on the conceptual network. We utilized the same decoder in normal imagined speech decoding, while predicting 20 different imagined words.

3 Results and Analysis

3.1 POS Classification

As Figure 3 shows, both RSA and voxel-wise LR feature selection could achieve classifying different POS tags in all kinds of speeches (cross-validation T-test, $p < 0.05, H_0 : \mu > 0.5$). Significantly, transferred imagined speech decoding was realized.

The semantics transferring indicates the conceptual network working.

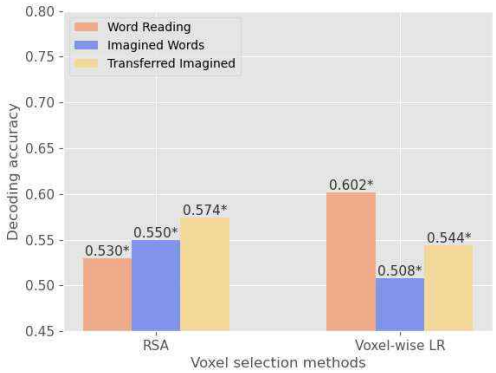

Figure 3: Decoding accuracy of POS classification. Red bar is for overt reading decoding, blue one for imagined speech while yellow one for transferred imagined speech.

In detail, voxel-wise LR worked better than RSA in decoding read words, while in normal imagined speech and transferred imagined speech, RSA outperformed. What's interesting is that RSA's decoding performance on both imagined words($57.4\%, 55.0\%$) was even better than that of read words(53%). This may suggest that the neural processing difference between the concepts represented by 'verbs' and 'nouns' was greater in the imaginary modality than in the reading modality, and this difference was captured by RSA.

Another point worth paying attention to is that in RSA, transferred imagined words' decoding accuracy(0.574) was higher than normal imagined words(0.55). Considering the decoder was trained with words which didn't appear in transferred speech, this result might indicate that these 10 nouns and 10 verbs in transferred speech differ more from each other than those 26 words in normal imagined speech, from the perspective of 'concept'. Once the decoder extracted the difference information of POS successfully, it performed better in transferred speech.

Figure 4 and Figure 5 show the effect of the number of selected voxels on the decoding accuracy. With the increase in the number of voxels, the accuracy of decoders using the RSA voxel selection method doesn't improve very much. The best number of voxels for RSA is around 1500 to 2000. This means that in the RSA method, more voxel might not bring extra information gain. For a relatively coarse-grained level such as part of speech, a proper amount of voxel is enough to dig out the differences in the neural representations of different POS tags in the conceptual network.

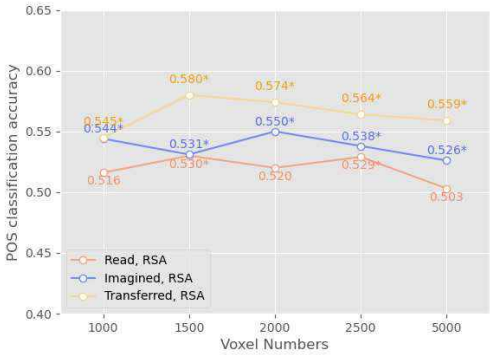

Figure 4: POS performance with different voxel numbers. RSA voxel selection algorithm focused.

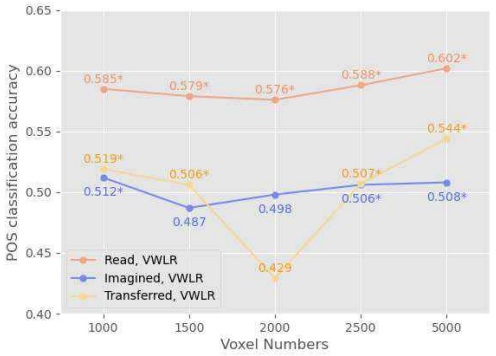

Figure 5: POS performance with different voxel numbers. Voxel-wise LR voxel selection algorithm focused.

As for the voxel-wise LR, with the number of voxels growing, the accuracy decreases first, and then gradually increases, achieving the best performance at 5000 voxels. Voxel-wise LR is a data-driven feature selection algorithm, within the same modality, more voxels mean more information gain to some extent. However, when cross-modality is considered(overt read $\rightarrow$ imagine), the situation might be different.

3.2 Lexicon Decoding

Performance of lexicon decoding can be found as Figure 6 presents. As we can see, the RSA method didn't work at all in lexicon decoding while voxel-wise LR worked quite well. For lexicon decoding task, all words were considered to be under the same condition('lexicon') in RSA. RSA localizes voxels that activate similarly for a certain class of stimuli, which means it can be used for classifi-

cation between different conditions. So, it's not surprised to see that within one condition, RSA can't help decode.

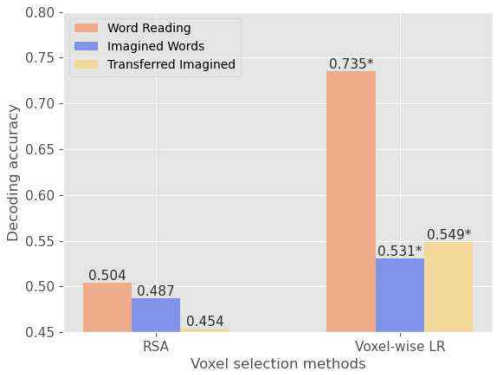

Figure 6: Accuracy of lexicon decoding

As for voxel-wise LR, the decoding performance for overt reading reached 73.5%. Our work achieved similar accuracy compared to previous studies(Pereira et al., 2018). However, different from the previous work which trained and test on the same participant, our experiment paradigm was cross-subject. This result provides reliable evidence for the stability and similarity of the conceptual networks across subjects, which has also been suggested in other studies(Lu et al., 2021).

On the other hand, voxel-wise LR significantly (T-test across all samples for rank scores, $p < 0.05, H_0 : \mu > 0.5$) decoded imagined words(53.1%). To our knowledge, this is the first report of achieving decoding imagined speech at a fine-granted lexicon level. Some previous works have also explored the decoding of imagined speech, but they were limited to whether the coarse-grained distinction between concrete words and abstract words(T et al., 2021); simple yes, no and a third word(Sereshkeh et al., 2019); or several simple phrases(Dash et al., 2020). In addition, the decoders in these studies were all trained on imagined speech, rather than overt speech like in our work, which was cross-modality.

Most significantly, transferred imagined word decoding was achieved successfully at $54.9\%(p < 0.05)$. Predicting imagined new words has never appeared in previous work to our known. This result suggests that when the conceptual network was captured by learning algorithm, it could help predict new imagined words.

Figure 7 and Figure 8 demonstrate the trend of lexicon decoding performance with the number of

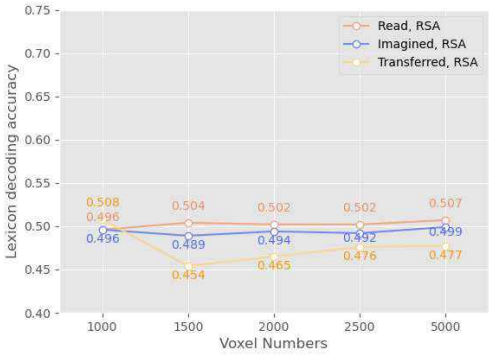

Figure 7: Lexicon decoding accuracy with various reduced dimensionality. The voxel selection method is RSA.

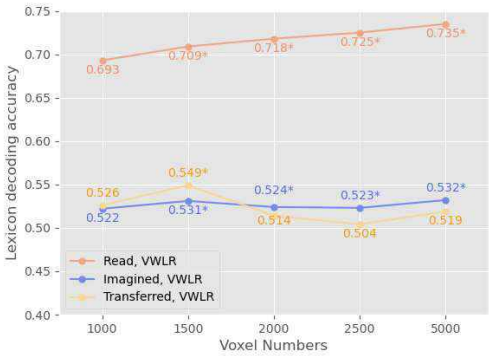

Figure 8: Lexicon decoding accuracy with various reduced dimensionality. The voxel selection method is voxel-wise LR.

voxels increasing. RSA didn't work any the time. In VWLR, when the voxel number grew up, the learning model performed better and better in decoding reading words. What should be noticed is that in imagined words decoding, the best performance was achieved at 1500 voxels, rather than 5000 voxels in word reading. This indicates that a larger voxel number might cause the decoder overfit to the reading modality. Especially, when transferred word decoding was considered, only the learning model trained with 1500 voxels performed statistically significant results($p = 0.006$).

3.3 Voxel Selection Results

Figure 9 and Figure 10 show the top 5000 selected voxel distribution over the whole brain. Informative voxels in RSA were ranked by $z-$score, while in VWLR, R^2 metric was used to sort the voxels. Figure 9 shows that Chinese nouns activate in more general area while in Chinese verbs condition, left inferior frontal gyrus(LIFG) and posterior superior and middle temporal gyri (LpSTG&MTG)

were activated, which is consistent with Yu et al. (2012). As shown in Figure 10, in lexicon decoding, VWLR localized relevant voxels in more widely-distributed areas than RSA. This result can be supported by Huth et al. (2016a).

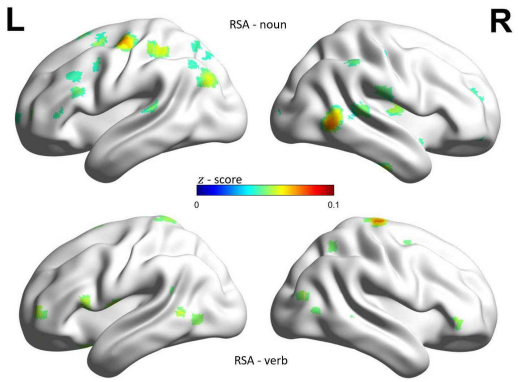

Figure 9: Selected voxels of RSA for nouns and verbs.

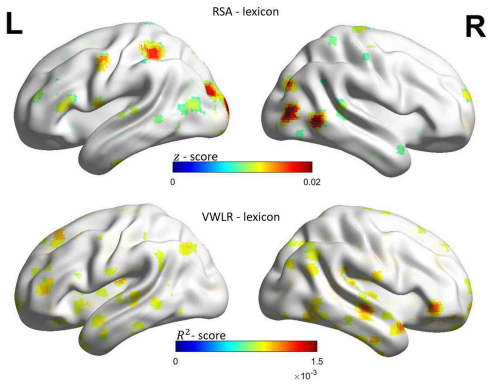

Figure 10: Selected voxels for lexicon decoding.

4 Discussion

To explore the neural conceptual network, we conducted prediction of imagined words and transferred imagined words using brain activation corresponding to word reading. Our decoder worked at both part-of-speech level and lexicon level and results are robust throughout different number of voxels selected to prediction model. It indicated that internal neural conceptual network was the terminal serving for semantic 'departure' and 'arrival' and language processing (comprehension, generation and production) played the role of flight carrying semantic information. To our best knowledge, it is the first time to provide neurobiological evidence for the ancient philosophy issue 'relationship between thought and language'.

Another interesting results were prediction of lexicon only worked based on data-driven VWLR voxel selection algorithm but predictions of part-of-speech were successful based on both data-driven VWLR voxel selection algorithm and hypothesis-driven RSA voxel selection algorithm. Tracing back to the nature of these two voxel selection algorithms, RSA aimed to establish links between neural responses and part-of-speech processing while VWLR directly reached valid voxels corresponding to lexical semantic vectors. It is reasonable that fine-grained lexical level brain activation can predict coarse-grained semantic category level neural activities. Also, it suggested that semantic information in neural conceptual network might be hierarchical organized and clustered into different semantic dimensions like the real world conceptual network depict by natural language processing. Furthermore, brain regions derived from VWLR were distributed onto more broad cortical areas, which is consistent with previous studies describing semantic dimensions and semantic maps(Wang et al., 2018; Xu et al., 2018). Brain regions derived from RSA-noun processing and RSA-verb processing converged in bilateral inferior frontal gyrus one of which was considered as the hub of general language generation and diverged in other valid voxels, indicating discrepancy anatomical basis underlying part-of-speech processing. To surmise, the neural conceptual network characterized by hierarchical functional organization and its coupling anatomical basis.

Two questions need to be paid close attention. First, based on trade-off between research goal and experimental cost, word sets used in current were scale-limited. Future research should develop real world conceptual network based on large corpora, figure out mapping rules between real world conceptual network and neural conceptual network and then construct robust and individual neural conceptual network. Second, we selected conservative voxel selection algorithms (data-driven VWLR and hypothesis-driven RSA) and decoding algorithms (SVM and LASSO) in our study and state-of-the-art neural network algorithm was discarded. Neural network algorithms do achieve higher accuracy of prediction but aim of current study was to test stability of conceptual network. Better performance of neural network algorithm can hardly be interpreted as mediations of conceptual network. In future, neural network algorithms should be utilized to construct conceptual network in brain by learning brain response for real world conceptual

network and parameters in hidden layers may help understand linguistic processing.

5 Conclusion

Our study explored plausible conceptual network by establishing links between neural responses in language comprehension and generation at part-of-speech level and lexicon level. Significant predictions of imagined words from word reading suggest feasibility to construct stable internalized conceptual network. More broadly, successful predictions of transferred imagined speech indicate potential hierarchical and/or clustered structure of conceptual network. Taken together, our study provides novel evidence to expand understandings of the nature of conceptual network and its relationship with linguistic processing. Future studies should focus on construction of the neural conceptual network which adapt to the real-world conceptual network and multilevel linguistic processing.

Acknowledgements

We would like to thank supports from the Natural Science Foundation of China [grant numbers 81901826, 61932008], the Natural Science Foundation of Shanghai [grant number 19ZR1405600, 20ZR1404900], the Shanghai Municipal Science and Technology Major Project [No.2018SHZDZX01], ZJLab and Shanghai Center for Brain Science and Brain-inspired Technology.

Contributions

L.H. designed the study; built the stimuli; implemented voxel selection, decoder training and testing; analyzed and visualized the results. S.G. and J.H. collected the fMRI data. S.G designed psychological experiment, prepossessed the fMRI data. L.H. and S.G. performed the theoretical analysis and wrote the paper. M.C. and J.F. participated in the discussion. We would also like to thank Prof. Xuexin Zhang at the department of psychology at Fudan for his advice.

References

Gopala K Anumanchipalli, Josh Chartier, and Edward F Chang. 2019. Speech synthesis from neural decoding of spoken sentences. *Nature*, 568(7753):493–498.

Alex Clarke and Lorraine K. Tyler. 2014. Object-specific semantic coding in human perirhinal cortex. *The Journal of Neuroscience*, 34(14):4766–4775.

Debadatta Dash, Paul Ferrari, and Jun Wang. 2020. Decoding imagined and spoken phrases from non-invasive neural (meg) signals. *Frontiers in Neuroscience*, 14:290–290.

Christian Herff, Dominic Heger, Adriana De Pesters, Dominic Telaar, Peter Brunner, Gerwin Schalk, and Tanja Schultz. 2015. Brain-to-text: decoding spoken phrases from phone representations in the brain. *Frontiers in neuroscience*, 9:217.

Alexander G Huth, Wendy A De Heer, Thomas L Griffiths, Frédéric E Theunissen, and Jack L Gallant. 2016a. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, 532(7600):453–458.

Alexander G Huth, Tyler Lee, Shinji Nishimoto, Natalia Y Bilenko, An T Vu, and Jack L Gallant. 2016b. Decoding the semantic content of natural movies from human brain activity. *Frontiers in systems neuroscience*, 10:81.

Nikolaus Kriegeskorte, Rainer Goebel, and Peter Bandettini. 2006. Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, 103(10):3863–3868.

Nikolaus Kriegeskorte, Marieke Mur, and Peter A Bandettini. 2008. Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 2:4–4.

Shen Li, Zhe Zhao, Renfen Hu, Wensi Li, Tao Liu, and Xiaoyong Du. 2018. Analogical reasoning on chinese morphological and semantic relations. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 138–143. Association for Computational Linguistics.

Junfeng Lu, Zehao Zhao, Jie Zhang, Bin Wu, Yanming Zhu, Edward F Chang, Jinsong Wu, Hugues Duffau, and Mitchel S Berger. 2021. Functional maps of direct electrical stimulation-induced speech arrest and anomia: a multicentre retrospective study. *Brain*, 144(8):2541–2553.

Stephanie Martin, Peter Brunner, Iñaki Iturrate, José del R Millán, Gerwin Schalk, Robert T Knight, and Brian N Pasley. 2016. Word pair classification during imagined speech using direct brain recordings. *Scientific reports*, 6(1):1–12.

David A Moses, Sean L Metzger, Jessie R Liu, Gopala K Anumanchipalli, Joseph G Makin, Pengfei F Sun, Josh Chartier, Maximilian E Dougherty, Patricia M Liu, Gary M Abrams, et al. 2021. Neuroprosthesis for decoding speech in a paralyzed person with anarthria. *New England Journal of Medicine*, 385(3):217–227.

Brian Murphy, Leila Wehbe, and Alona Fyshe. 2018. Decoding language from the brain.

Thomas Naselaris, Kendrick N Kay, Shinji Nishimoto, and Jack L Gallant. 2011. Encoding and decoding in fmri. *Neuroimage*, 56(2):400–410.

Thomas Naselaris, Cheryl A Olman, Dustin E Stansbury, Kamil Ugurbil, and Jack L Gallant. 2015. A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes. *Neuroimage*, 105:215–228.

Xiaomei Pei, Dennis L Barbour, Eric C Leuthardt, and Gerwin Schalk. 2011. Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans. *Journal of neural engineering*, 8(4):046028.

Francisco Pereira, Samuel Gershman, Samuel Ritter, and Matthew Botvinick. 2016. A comparative evaluation of off-the-shelf distributed semantic representations for modelling behavioural data. *Cognitive neuropsychology*, 33(3-4):175–190.

Francisco Pereira, Bin Lou, Brianna Pritchett, Samuel Ritter, Samuel J Gershman, Nancy Kanwisher, Matthew Botvinick, and Evelina Fedorenko. 2018. Toward a universal decoder of linguistic meaning from brain activation. *Nature communications*, 9(1):1–13.

Alborz Rezazadeh Sereshkeh, Rozhin Yousefi, Andrew T Wong, and Tom Chau. 2019. Online classification of imagined speech using functional near-infrared spectroscopy signals. *Journal of Neural Engineering*, 16(1):16005–16005.

Irina Simanova, Peter Hagoort, Robert Oostenveld, and Marcel AJ Van Gerven. 2014. Modality-independent decoding of semantic information from the human brain. *Cerebral cortex*, 24(2):426–434.

Proix T, Saa Jd, Christen A, Martin S, Pasley Bn, Knight Rt, Tian X, Poeppel D, Doyle Wk, Devinsky O, Arnal Lh, Mégevand P, and Giraud A. 2021. Imagined speech can be decoded from low- and cross-frequency features in perceptual space. *bioRxiv*.

Xiaosha Wang, Wei Wu, Zhenhua Ling, Yangwen Xu, Yuxing Fang, Xiaoying Wang, Jeffrey R Binder, Weiwei Men, Jia-Hong Gao, and Yanchao Bi. 2018. Organizational principles of abstract words in the human brain. *Cerebral Cortex*, 28(12):4305–4318.

Michael C-K Wu, Stephen V David, and Jack L Gallant. 2006. Complete functional characterization of sensory neurons by system identification. *Annu. Rev. Neurosci.*, 29:477–505.

Yangwen Xu, Xiaosha Wang, Xiaoying Wang, Weiwei Men, Jia-Hong Gao, and Yanchao Bi. 2018. Doctor, teacher, and stethoscope: neural representation of different types of semantic relations. *Journal of Neuroscience*, 38(13):3303–3317.

Huichao Yang, Qixiang Lin, Zaizhu Han, Hongyu Li, Luping Song, Lingjuan Chen, Yong He, and Yanchao Bi. 2017. Dissociable intrinsic functional networks support noun-object and verb-action processing. *Brain and language*, 175:29–41.

Xi Yu, Yanchao Bi, Zaizhu Han, and Sam-Po Law. 2013. An fmri study of grammatical morpheme processing associated with nouns and verbs in chinese. *PloS one*, 8(10):e74952.

Xi Yu, Yanchao Bi, Zaizhu Han, Chaozhe Zhu, and Sam-Po Law. 2012. Neural correlates of comprehension and production of nouns and verbs in chinese. *Brain and language*, 122(2):126–131.

Xi Yu, Sam Po Law, Zaizhu Han, Caozhe Zhu, and Yanchao Bi. 2011. Dissociative neural correlates of semantic processing of nouns and verbs in chinese—a language with minimal inflectional morphology. *NeuroImage*, 58(3):912–922.

A Stimuli

Table 1 and 2 show the details of our stimuli.

No.	Word	Translation	POS	Frequency
1	生产	Produce	verb	0.096
2	告诉	Tell	verb	0.042
3	参加	Participate	verb	0.029
4	准备	Prepare	verb	0.029
5	战斗	Battle	verb	0.024
6	创造	Create	verb	0.022
7	形成	Form	verb	0.020
8	前进	Advance	verb	0.020
9	发动	Start	verb	0.017
10	相信	Believe	verb	0.017
11	改变	Change	verb	0.014
12	成立	Found	verb	0.014
13	举行	Hold	verb	0.013
			Mean	0.027
14	经济	Economy	noun	0.075
15	国家	Nation	noun	0.074
16	眼睛	Eye	noun	0.054
17	技术	Techonology	noun	0.046
18	情况	Situation	noun	0.046
19	资产	Property	noun	0.031
20	速度	Speed	noun	0.022
21	现象	Phenomenon	noun	0.021
22	方向	Direction	noun	0.021
23	工程	Engineering	noun	0.018
24	利益	Profit	noun	0.016
25	妇女	Woman	noun	0.015
26	幸福	Happiness	noun	0.012
			Mean	0.035
27	伟大	Great	adj.	0.039
28	正确	Correct	adj.	0.029
29	清楚	Clear	adj.	0.026
30	容易	Easy	adj.	0.026
31	立刻	Instant	adj.	0.020
32	认真	Conscientious	adj.	0.017
33	巨大	Tremendous	adj.	0.016
34	积极	Positive	adj.	0.016
35	迅速	Rapid	adj.	0.015
36	热情	Enthusiastic	adj.	0.014
37	奇怪	Strange	adj.	0.012
38	危险	Dangerous	adj.	0.011
39	普通	Normal	adj.	0.009
			Mean	0.019
40	服务	Service	n./v.*	0.015

Table 1: 40 words used both in overt reading and imagined speech. *ambiguous POS, dropped in the noun/verb classification task

No.	Word	Translation	POS	Frequency
1	方法	Method	noun	0.038
2	温度	Temperature	noun	0.013
3	工人	Worker	noun	0.060
4	矛盾	Contradiction	noun	0.021
5	农业	Agriculture	noun	0.024
6	感情	Sentiment	noun	0.012
7	现代	Modern times	noun	0.031
8	世界	World	noun	0.076
9	身体	Body	noun	0.022
10	队伍	Team	noun	0.016
			Mean	0.031
11	建设	Construct	verb	0.046
12	产生	Generate	verb	0.022
13	要求	Demand	verb	0.035
14	进攻	Attack	verb	0.012
15	解放	Liberate	verb	0.049
16	发表	Publish	verb	0.013
17	进行	Conduct	verb	0.066
18	认为	Consider	verb	0.025
19	记得	Remember	verb	0.011
20	扩大	Enlarge	verb	0.011
			Mean	0.029

Table 2: 20 new words used only in imagined speech

B Read Words Decoding Visualization

As Figure 11 shows, here we demonstrate the word reading decoding results of one subject in LOSO cross-validation. The left plot was generated by comparing the original 40 word vectors with themselves. The matrix was colored with the rank score of cosine similarity. The more similar the two word vectors are, the higher the rank score is. The right plot was constructed similarly with true word vectors compared with predicted word vectors. Decoding accuracy of this subject was 79.8%. As we can see, the values on the diagonal of the decoded matrix are relatively high, and some pattern in the decoded matrix was similar to the original one.

C Imagined Words Decoding Group-Level Analysis

In order to explore how our decoder performed under the overall word imagining condition, we also designed a group-level study. For each imagined word, we averaged the brain activation of all subjects in the test set. Inspired by the ERP method in EEG research, this helps reduce individual differences and random factors, and capture the cognitive response of the word imagination event itself as much as possible. For imagined words decoding, the group-level decoding performance achieved at 53.3%, while the transferred one achieved at 65.8%, which is much higher than that of individual-level(54.9%). Like in reading words decoding, similar comparing matrices were built for imagined words decoding as Figure 12 and Figure 13 show but at a group level. As we can see, for transferred imagined words, the decoded matrix's pattern(left one in Figure 13) was somewhat similar to the self-compared matrix on the left.

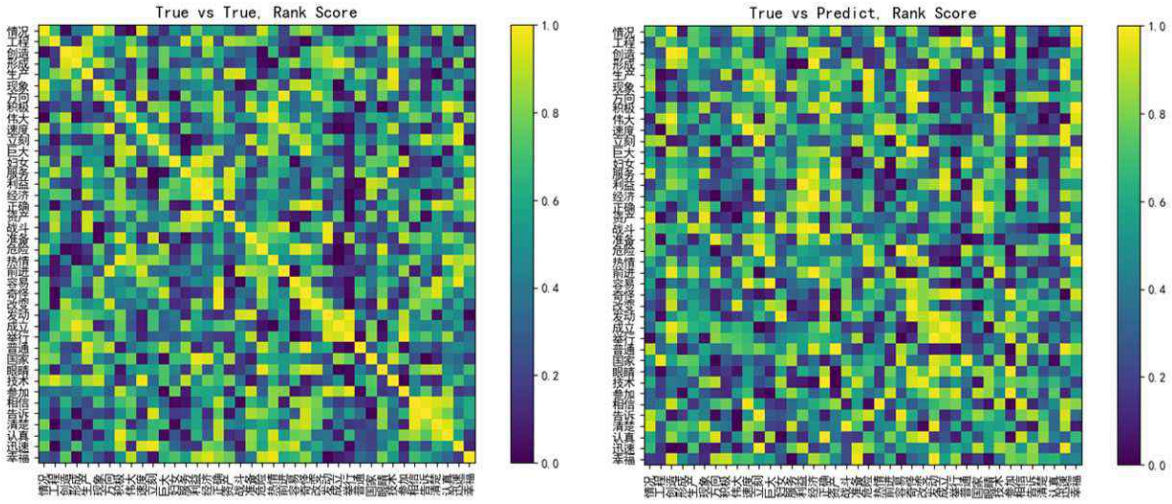

Figure 11: Ranked cosine similarity correlation for reading words decoding of one subject in the LOSO cross-validation.

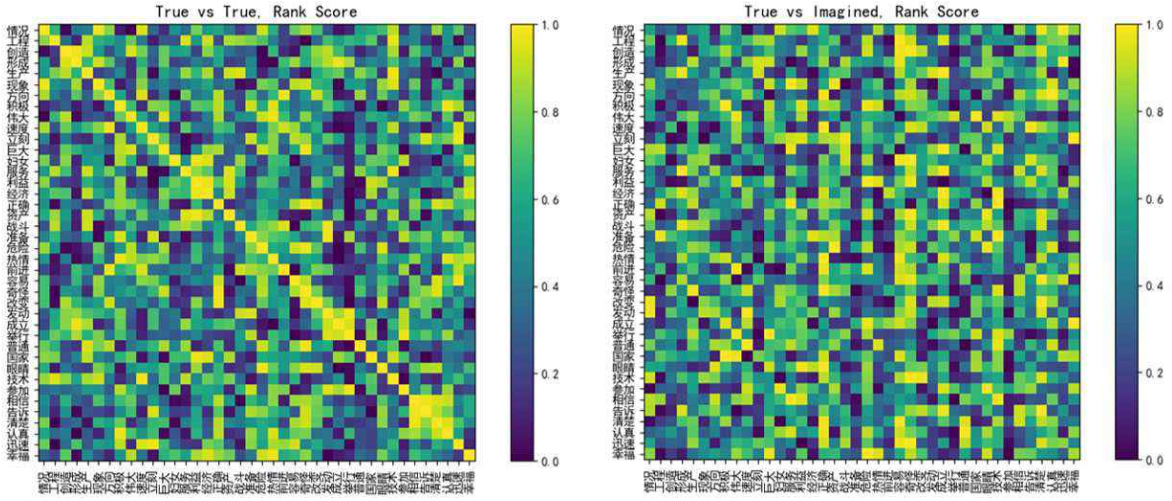

Figure 12: Ranked cosine similarity correlation for imagined words decoding of the test group.

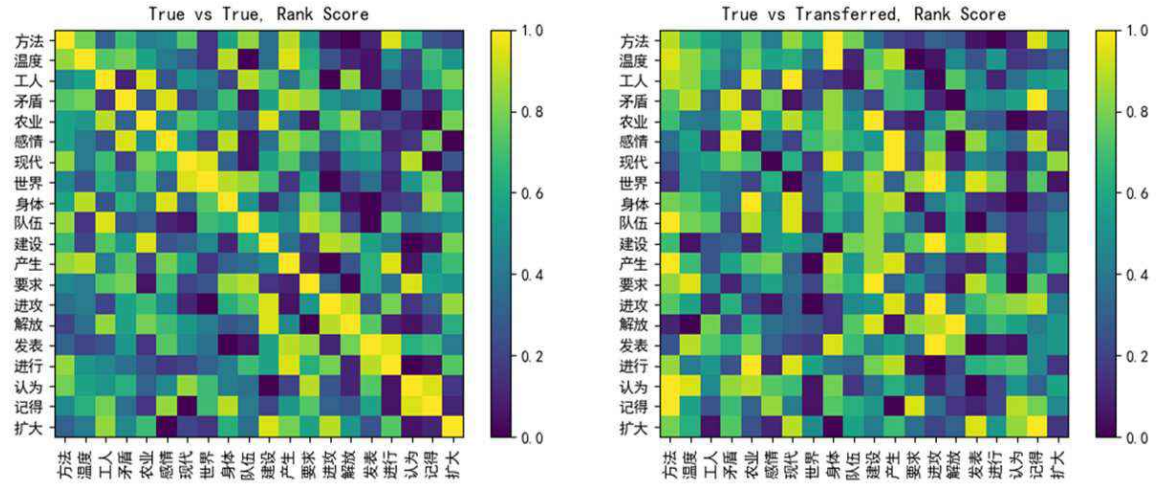

Figure 13: Ranked cosine similarity correlation for transferred imagined words decoding of the test group.