Songsheng YING
Supervisor: Sabine Ploux
Laboratory: Centre d'analyse et de mathématique sociales
Intended Memory Defense Session: July 2019
Language: English
Suggested Reviewer: [TODO]

Similarity and Association: Principles of Distributed Organisation of Semantics in the Human Brain

Introduction

Background. General semantic knowledge associates verbal and non-verbal stimuli to concepts internal to human cognitive system. In human language processing, lexicons with or without contexts are linked to their meanings by the lexicon semantic system. How brain processes semantics remains an open question. Tentatives to localize a stable semantic memory lead neuro- and computational linguists to a hub-and-spoke model (see Ralph, Jefferies, Patterson, & Rogers, 2017 for a review). A neural architecture of transmodal semantic memory across concepts with similar semantic significance, with the locus centered on bilateral ventrolateral anterior temporal lobe (vlATL) is suggested by pathological studies on *semantic dementia* (SD), *herpes simplex virus encephalitis* (HSVE) and other semantic disorders (Patterson, Nestor, & Rogers, 2007). While semantics' relevance to perception and action suggest a widely distributed, modality-specific neural network such as visual cortices (Borghesani et al., 2016). Pereira et al. (2018) built a BOLD-to-word decoder with GloVe (Pennington, Socher, & Manning, 2014), Huth, Nishimoto, Vu, & Gallant (2012) and Huth, De Heer, Griffiths, Theunissen, & Gallant (2016) used a 985-dimensional word-level co-occurrence based embedding space and narrative-story listening functional magnetic resonance imaging (fMRI) to build association maps. These results found an extensively distributed informative voxels in language, task, visual and other networks.

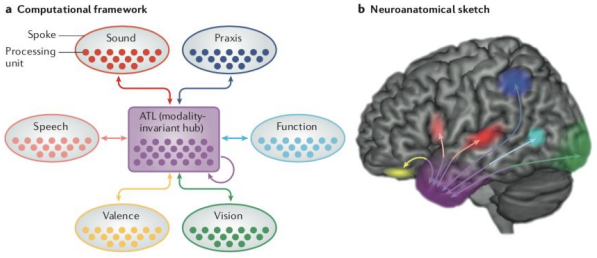

Figure 1. Adopted from (Ralph, Jefferies, Patterson, & Rogers, 2017), Hub-and-spoke model. **a**. Modality-specific informations are encoded separately in different processing layers. Such information is fed to a transmodal semantic hub, which contains conceptual knowledge, which reactivates complementary information in other spoke layers. b. A neuroanatomical representation of the

hub-and-spoke model, where the hub is located near ATL, spoke components distributed across different cortices.

(Peelen & Caramazza, 2012) used conceptual/perceptual contrast in a fMRI study to confirm that the conceptual information is most likely to be store in BA20, BA38 (ATL), while perceptual information in posterior occipitotemporal cortex. Studies on HSVE compared with SD shows an intact performance at basic semantic levels (such as dog, knife) but not at subordinate level (for example poodle or bread knife), suggesting the ATL hub might be organizing concepts hierarchically.
Rationale. This project tends to build a semantic mapping based on hub-and-spoke model, on explicitly defining the feature space of hub and spoke-components.
Key research question. What are the semantic information encoded by the semantic hub and other components?
General hypotheses. Semantic hub encodes conceptual-similarity by conceptual super-subordinate hierarchies, while other components encode perceptual-specificities depending on their modality and all other associative relations. The semantic hub is localized near left and/or right ATL, while other components correspond to other cortical areas.

Methods

Model
Consistent with (Lofi, 2015), we define conceptual-similarity in terms of taxonomical properties (such as *cat/tiger*, *museum/theater*, but not *computer/software*, as computer is a type of hardware), and association in terms of relevance of two concepts (such as *computer/software*). In order to dissociate semantic hub activation patterns from other components, we propose a novel word-embedding scheme that rejets conceptual-similarity but keep association to build a semantic encoding model for fMRI data. This new embedding space, together with conceptual-similarity embedding space and classic statistical word-embeddings will be separately tested on similarity and association benchmarks if the dissociation of two aspects is indeed implemented.
Due to the lack of availability of benchmarks in French, we first build the embedding models in English to test the dissociation algorithm, then replicate the method with French data for fMRI encoding.
Todorovic & Pallier (2018) built a word-fMRI encoding model with GloVe and English-stimuli fMRI data. Similarly, we will construct regression-based machine learning models to encode word-embedding vectors into individual voxel BOLD signals of each participant with French fMRI. We will then identify and interpret the systematic differences of voxel activation profile.
Key features. We will be comparing the encoding performance for each voxel of three different embedding spaces. All three embeddings contain non-semantic dimensions including auditory signal existence, word-speed, acoustic signal energy and bottom-up syntax parser. The conceptual-similarity is word-embedding space constructed from WordNet, which is a tree-structure ontology organized by synonym sets and super-subordinate relations. The classic statistical word-embedding is adapted from GloVe. The pure association embedding is based on GloVe, but decorrelated with the conceptual-similarity space, which is presumed to encode only semantic associations.

Rationale of feature selection. The semantic hub is hypothesized to encode conceptual-similarity. While traditional word-embedding encoding and decoding studies found a distributed mapping between brain regions and semantic vectors, including the loci of semantic hub. We want to test if conceptual-information-based embeddings match better with semantic hub activations, and if traditional embeddings' encoding performance near the semantic hub region, would decrease significantly after decorrelation with conceptual-similarity embeddings.
Programming language. We will use Python 3 to build the decorrelation algorithm.
External scripts. As one candidate for semantic conceptual-similarity embeddings, WordNet embeddings will be constructed using algorithm provided by (Saedi, Branco, António Rodrigues, & Silva, 2018). We will build such space considering only synonymy, hypernymy, hyponymy, verb participle, adjective/adverb derivation and pertainym relationships in WordNet. Meronyms, holonyms and other relationships are rejected as they are more associative. The selected vocabulary will overlap at maximum with audio stimuli provided to fMRI recording participants.
The embedding to fMRI encoding regression algorithm is implemented by Verdier, Lakretz, & Pallier (2018).
Assumptions. We assume that conceptual-similarity information encoded by classic word-embeddings is contained by conceptual-similarity embeddings, such that after decorrelation process, the residual embedding space would comprise only non-conceptual (thus purely association) data.

Input data
Embedding construction. For conceptual-similarity embeddings, we will use English and French WordNet as source data to build WordNet embeddings (Miller, 1995), (Pradet, De Chalendar, & Baguenier-Desormeaux, 2014). They are thesaurus-like database organised hierarchically based on super-subordinate relations. In addition we will also test the performance of (Saedi, Branco, António Rodrigues, & Silva, 2018)'s algorithm with synonym databases, which are available in English and French, created by thesauri fusion and symmetrisation (Ploux & Ji, 2003).
For classic word-embeddings, we use GloVe embeddings that are provided with open access[1]. They are co-occurrence frequency based statistical measures derived dense semantic vectorial representations.
The non-semantic data will be provided by Todorovic & Pallier (2018).
Embedding validation. With built and decorrelated word-embedding models, we will use vectorial distance to evaluate word-pair similarity and association with multiple benchmarking datasets (Lofi, 2015). For conceptual-similarity benchmarks, we use datasets provided by (Rubenstein & Goodenough, 1965), (Agirre et al., 2009) and (Hill, Reichart, & Korhonen, 2015). For association benchmarks, there is few available datasets due to the less clear definition of association (or relatedness in other terms), we adapt datasets from (Agirre et al., 2009) and (Halawi, Dror, Gabrilovich, & Koren, n.d.). The benchmarks are word pairs associated with a similarity or association score. Vectorial distance scores will be matched against benchmarks with pearson and spearman correlation.

---

[1] English GloVe: https://nlp.stanford.edu/projects/glove/, French DepGloVe with lemma: http://alpage.inria.fr/frmgwiki/content/word-embeddings-avec-depglove

fMRI data. We will be using fMRI data acquired in (Todorovic & Pallier, 2018), in which 20 native French speakers listen to «the Little Prince» during the whole-brain fMRI recording. The data is preprocessed by Christophe Pallier with ME-ICA pipeline (Kundu, Inati, Evans, Luh, & Bandettini, 2012).

Measures

Embedding validation. For conceptual-similarity, classic and decorrelated word-embedding models, the pearson and spearman correlation tests will give scores of semantic similarity and association.

Embedding-to-BOLD regression. Regression from each embedding scheme to individual fMRI data will give a correlation of determination (R^2-value) for each voxel. We will compare the R^2-value of each embedding model and build a voxel-wise activation profile map as similar in (Jain & Huth, 2018). This would allow us to discover if there is a conceptual-similarity based semantic representation in the previously discovered semantic hub.

Predictions

If the conceptual-similarity space is well built, we expect it to give significantly above null results over similarity benchmarks, and near null results over association benchmarks.
If the dissociation algorithm works as expected, the dissociated association embedding space, would have significantly lower performance on similarity task when compared with conceptual-similarity embedding space, and have comparably similar performance on association task when compared with the undissociated original mixed embedding space.
If transmodal hubs store pure conceptual, transmodal information hierarchically, and other functional neural networks encode other information, then ontologies such as WordNet (Miller, 1995) is analogical to hubs organizational structure. ROIs, which have a preference for conceptual-similarity based embedding models such as WordNet embeddings, should compose an transmodal semantic hub near ATL. Other regions significantly encoded by classic word-embeddings models should have a preference for decorrelated association embedding space.

Analyses

Each built embedding space is tested on semantic similarity and association benchmarks with pearson and spearman correlation. Inter-embedding-space benchmark result comparisons would be tested for significance.
Embeddings with adequate performance in either similarity or association domain would then be used to encode fMRI BOLD signals. The R^2-values will be tested for significance. We subtract obtained R^2-values of conceptual-similarity model from association model to build a contrast map with a comparison significance mask. We will run an ANOVA on all voxel R^2-value between-model differences with subject, embedding type as factors. The voxels with significant main effect of embedding type would draw an additional contour on the contrast map to help determine the localization of a semantic hub graphically.

Interpretation

If the hypothesis is correct and our assumptions on data manipulations are exact, we should see activation preference for conceptual-similarity embeddings in brain regions near ATL,

centered on vlATL. Other significantly correlated voxels found by classic word-embedding models should have a preference for association embeddings. This would further suggest the hierarchical concept organization in the transmodal semantic hub.

Expected contributions

Songsheng Ying. Word-Embedding preparation, embedding space decorrelation, fMRI data analysis and interpretation, master thesis.
Sabine Ploux. Result analysis and linguistic interpretation.
Christophe Pallier. fMRI data acquisition and preprocessing, neuro-linguistic interpretation, fMRI encoding regression scripts.
Laurent Bonnasse-Gahot. Embedding space decorrelation.

Bibliography

Agirre, E., Alfonseca, E., Hall, K., Kravalova, J., Paşca, M., & Soroa, A. (2009). A Study on Similarity and Relatedness Using Distributional and WordNet-based Approaches. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 19–27). Stroudsburg, PA, USA: Association for Computational Linguistics. Retrieved from http://dl.acm.org/citation.cfm?id=1620754.1620758

Borghesani, V., Pedregosa, F., Buiatti, M., Amadon, A., Eger, E., & Piazza, M. (2016). Word meaning in the ventral visual path: a perceptual to conceptual gradient of semantic coding. *NeuroImage*, *143*, 128–140. https://doi.org/10.1016/j.neuroimage.2016.08.068

Halawi, G., Dror, G., Gabrilovich, E., & Koren, Y. (n.d.). *Large-Scale Learning of Word Relatedness with Constraints*.

Hill, F., Reichart, R., & Korhonen, A. (2015). SimLex-999: Evaluating Semantic Models With (Genuine) Similarity Estimation. *Computational Linguistics*, *41*(4), 665–695. https://doi.org/10.1162/COLI_a_00237

Huth, A. G., De Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, *532*(7600), 453–458. https://doi.org/10.1038/nature17637

Huth, A. G., Nishimoto, S., Vu, A. T., & Gallant, J. L. (2012). A Continuous Semantic Space Describes the Representation of Thousands of Object and Action Categories across the Human Brain. *Neuron*, *76*(6), 1210–1224. https://doi.org/10.1016/j.neuron.2012.10.014

Jain, S., & Huth, A. (2018). Incorporating Context into Language Encoding Models for fMRI. *BioRxiv*, 327601. https://doi.org/10.1101/327601

Kundu, P., Inati, S. J., Evans, J. W., Luh, W.-M., & Bandettini, P. A. (2012). Differentiating BOLD and non-BOLD signals in fMRI time series using multi-echo EPI. *Neuroimage*, *60*(3), 1759–1770.

Lofi, C. (2015). Measuring Semantic Similarity and Relatedness with Distributional and Knowledge-based Approaches. *Information and Media Technologies*, *10*(3), 493–501. https://doi.org/10.11185/imt.10.493

Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM*,

38(11), 39–41.
Peelen, M. V., & Caramazza, A. (2012). Conceptual object representations in human anterior temporal cortex. *Journal of Neuroscience*, *32*(45), 15728–15736.
Pereira, F., Lou, B., Pritchett, B., Ritter, S., Gershman, S. J., Kanwisher, N., … Fedorenko, E. (2018). Toward a universal decoder of linguistic meaning from brain activation. *Nature Communications*, *9*(1), 963. https://doi.org/10.1038/s41467-018-03068-4
Ploux, S., & Ji, H. (2003). A Model for Matching Semantic Maps between Languages (French/English, English/French). *Computational Linguistics*, *29*(2), 155–178. https://doi.org/10.1162/089120103322145298
Pradet, Q., De Chalendar, G., & Baguenier-Desormeaux, J. (2014). Wonef, an improved, expanded and evaluated automatic french translation of wordnet. In *Proceedings of the Seventh Global Wordnet Conference* (pp. 32–39).
Ralph, M. A. L., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, *18*(1), 42–55. https://doi.org/10.1038/nrn.2016.150
Rubenstein, H., & Goodenough, J. B. (1965). Contextual Correlates of Synonymy. *Commun. ACM*, *8*(10), 627–633. https://doi.org/10.1145/365628.365657
Saedi, C., Branco, A., António Rodrigues, J., & Silva, J. (2018). WordNet Embeddings. In *Proceedings of The Third Workshop on Representation Learning for NLP* (pp. 122–131). Melbourne, Australia: Association for Computational Linguistics. Retrieved from http://www.aclweb.org/anthology/W18-3016
Todorovic, S., & Pallier, C. (2018). *Analyses IRMf lors de l'écoute de texte naturel*.
Verdier, A., Lakretz, Y., & Pallier, C. (2018). *Encodage d ' activité neuronale à partir de réseaux LSTM* (pp. 1–21). Saclay: Neurospin.