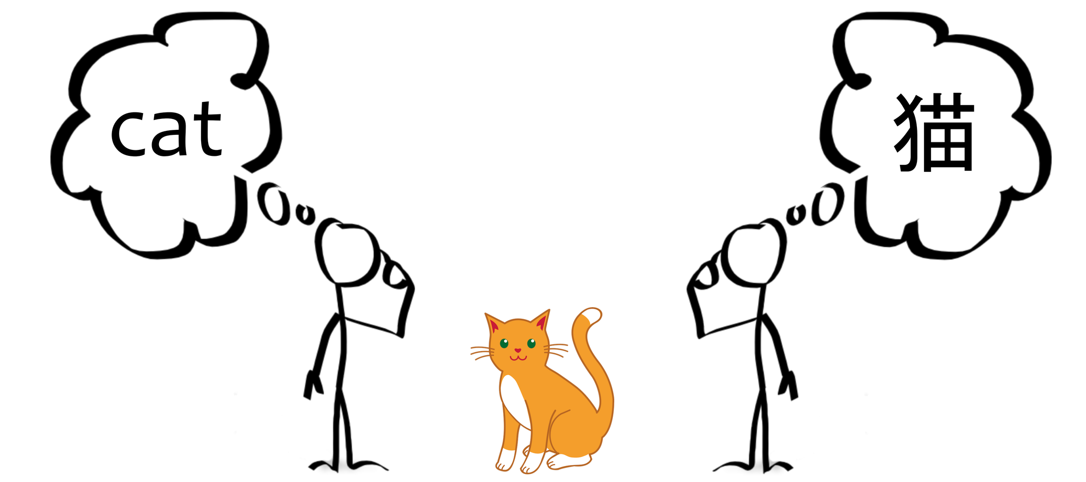
Chinese and English speakers share representations for word-elicited concepts, but semantic models struggle to capture this similarity

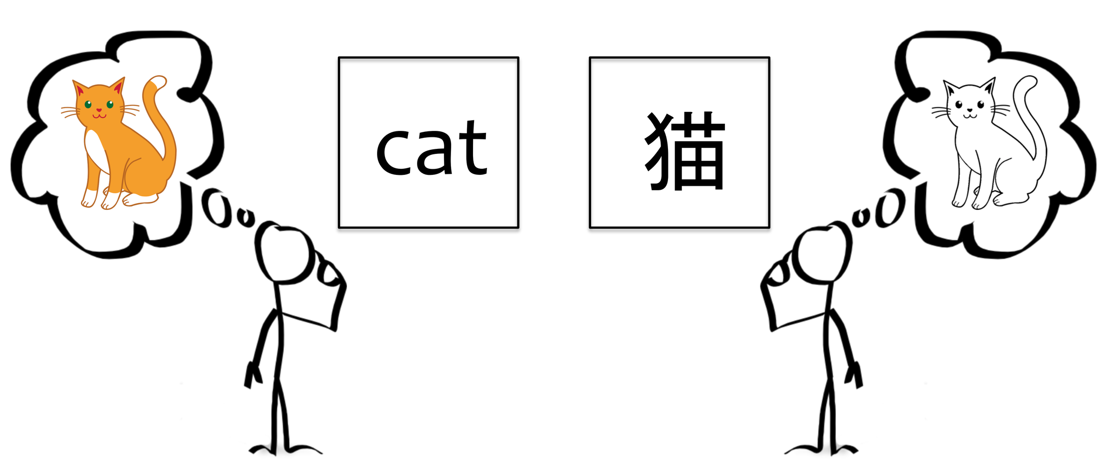
Benjamin D. Zinszer[1], Andrew J. Anderson[1], Olivia Kang[2,3], Thalia Wheatley[2], Rajeev D. S. Raizada[1]
1. University of Rochester; 2. Dartmouth College, 3. Harvard University

Background

One world, two languages:

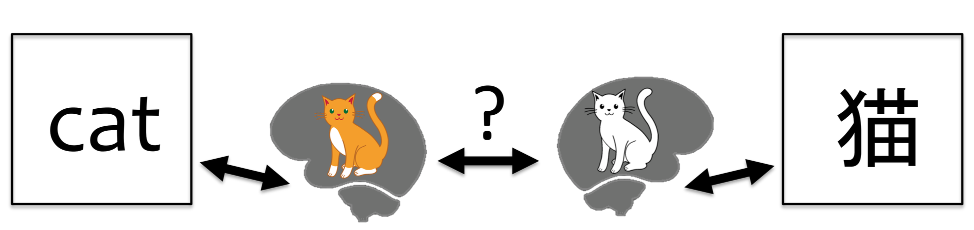


Although two languages may completely differ in surface appearance (phonology, orthography) they occur in the same world and therefore represent common / shared conceptual information.

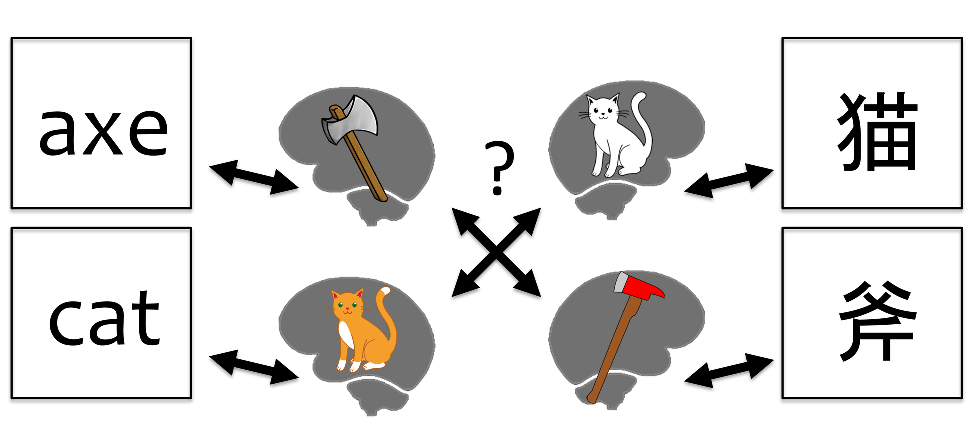


Conversely, speakers of different languages should share highly similar mental representations for translation-equivalent words in their respective languages (allowing for some individual or regional variations).

Neural decoding of fMRI allows us to measure neural-activity as a proxy of mental representation:



fMRI activity patterns elicited by words read by speakers of different languages might serve as a bridge between word labels in one language and word-labels in another. **Do they?**



Can we learn guess translations of words by comparing the brains of speakers in each language when they read the words?

Acknowledgments

We are grateful to Donias Doko and Carol Jew for data collection, Xixi Wang for language consulting, and Peiyao Chen, Anqi Li, Yinghui Qiu, and Tianyang Zhang for translation.

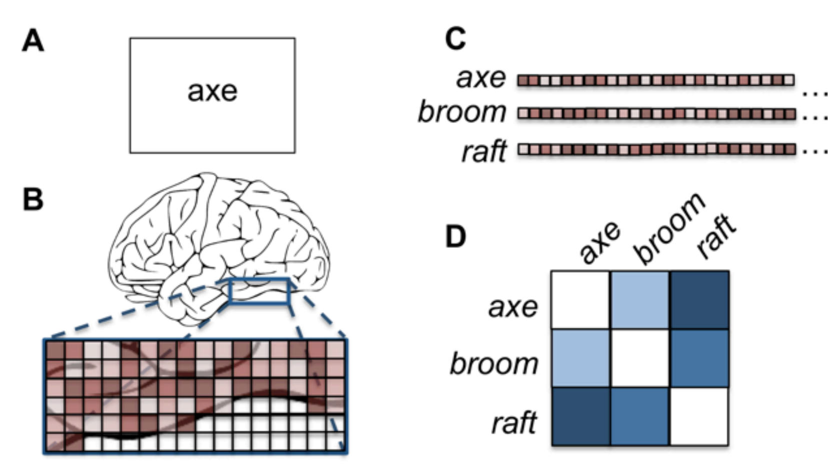
This work was supported in part by NSF Award 1228261 (P.I. Raizada).

* Contact: Benjamin Zinszer – bzinszer@gmail.com

Imaging Data

In a previous study (Zinszer, Anderson, Kang, Wheatley, & Raizada, 2015, *Cognitive Science Society*), we used representational similarity analysis to match fMRI responses to translation-equivalent words across native speakers of Chinese and English.

Representational Similarity Analysis

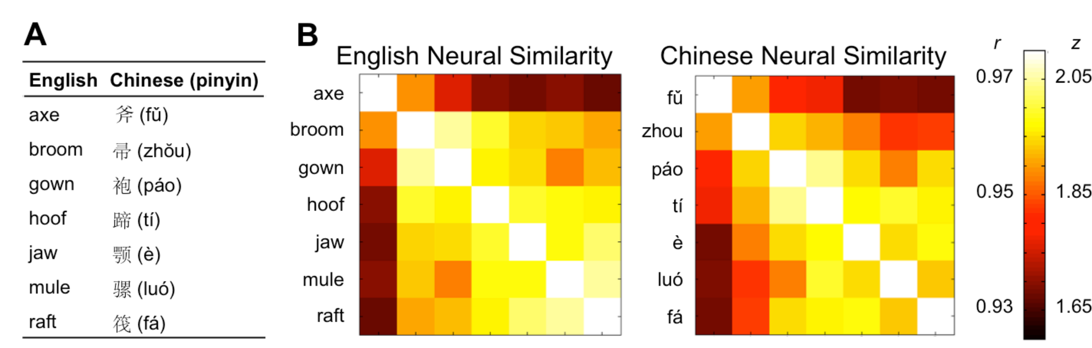


(A) Each word is presented to a participant in her native language

(B) BOLD responses are measured in each voxel of the brain.

(C) Vector of length *v* for BOLD response in *v* voxels of an ROI.

(D) Vectors are pairwise-correlated for a participant's similarity structure.

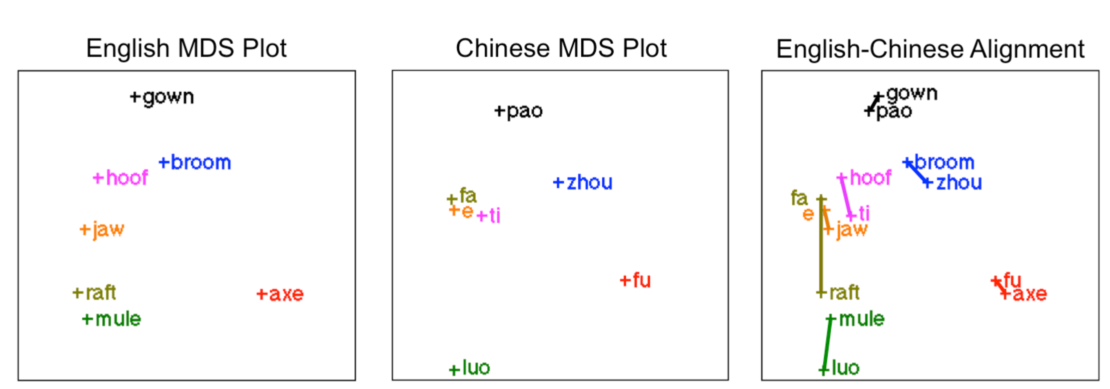


(A) Stimuli: 7 English and 7 Chinese (Simplified script) concrete nouns

(B) Similarity structures based on whole-brain data (all cortical voxels). The representational similarity structures of 11 English and 11 Chinese participants were averaged to produce respective group models.

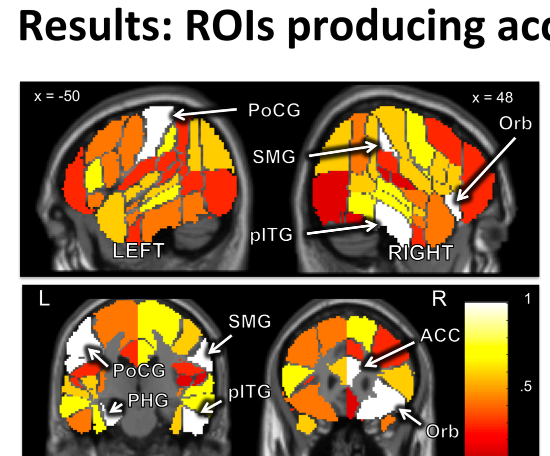
Results: Alignment of Chinese and English structures

By permuting the labels (rows and columns) in each language, we can look for alignments of the two similarity structures that produce the best overall fit (as in Zinszer *et al.*, 2015).



MDS plots (first 2 dimensions of 7) illustrates alignment between Chinese and English in one of the regions of interest (post-central gryrus).

Results: ROIs producing accurate translations



- We tested alignment in each of the 96 Harvard-Oxford ROIs across the cortex. Translation accuracy is depicted.
- Six regions produced 100% decoding accuracy (white), and several more above chance (>71%; see Zinszer *et al.*, 2015).

Semantic Models

Decoding neural data with a semantic model

In the same way that similarity structures can be used to compare brain responses across languages, similarity structures can also be used to compare brain responses to other models.

Behavior-based models

mule - axe

Noun-Noun (NN) ratings: 11 native speakers of each language judge the similarity of each pair of nouns (order randomized) in their respective language.

mule - lift

Noun-Verb (NV) ratings: 11 native speakers of each language judge the relatedness of each nouns to 25 verbs used in the corpus model below. 25-dimension vectors are pairwise-correlated for similarity structure.

Corpus-based models

Two parallel corpora of Internet text in Chinese and English were used to build semantic models for each of the seven nouns: http://corpus.leeds.ac.uk/internet.html (Sharoff, 2006)

Leeds corpus model:

- Nouns are represented by a vector of 25 noun-verb cooccurrences (Mitchell *et al.*, 2008).
- Verbs translated to Chinese to build parallel representations.

English	Chinese	English	Chinese
see	看	enter	进入
say	说	move	移动
taste	尝	listen	倾听
wear	穿	approach	接近
open	开	fill	填
run	跑	clean	清理
near	靠近	lift	举
eat	吃	rub	擦
hear	听	smell	闻
drive	驾驶	fear	怕
ride	骑	push	推
touch	碰	manipulate	操纵
break	打破		

Broadened Leeds corpus model:

- Representations for nouns based on a small group of similar nouns to increase frequency in the corpus and yield more representative co-occurrence values.
- Nouns in English and Chinese may not be direct translations, but groups are relatively equivalent.

English	Chinese
axe, knife, sword	斧，锯，剑
broom, mop, vacuum	帚，拖把，簸箕
gown, dress, robe	袍，裙，礼服
hoof, paw, foot	蹄，爪，脚
jaw, mouth, bone	颚，口，骨
mule, horse, donkey	骡，马，驴
raft, boat, barge	筏，船，艇

Comparison

How strongly do models correlate across languages?

Cross-Language Correlation

Measure	r	p
MRI - Whole brain	0.89	< 0.001
NN rating	0.85	< 0.001
NV rating	0.61	0.003
Broadened Leeds	0.44	0.045
Leeds	-0.08	0.726

- Neuroimaging data produce highly correlated representations.
- Behavioral ratings are also highly correlated across languages.
- Only the broadened corpus model was correlated across languages.

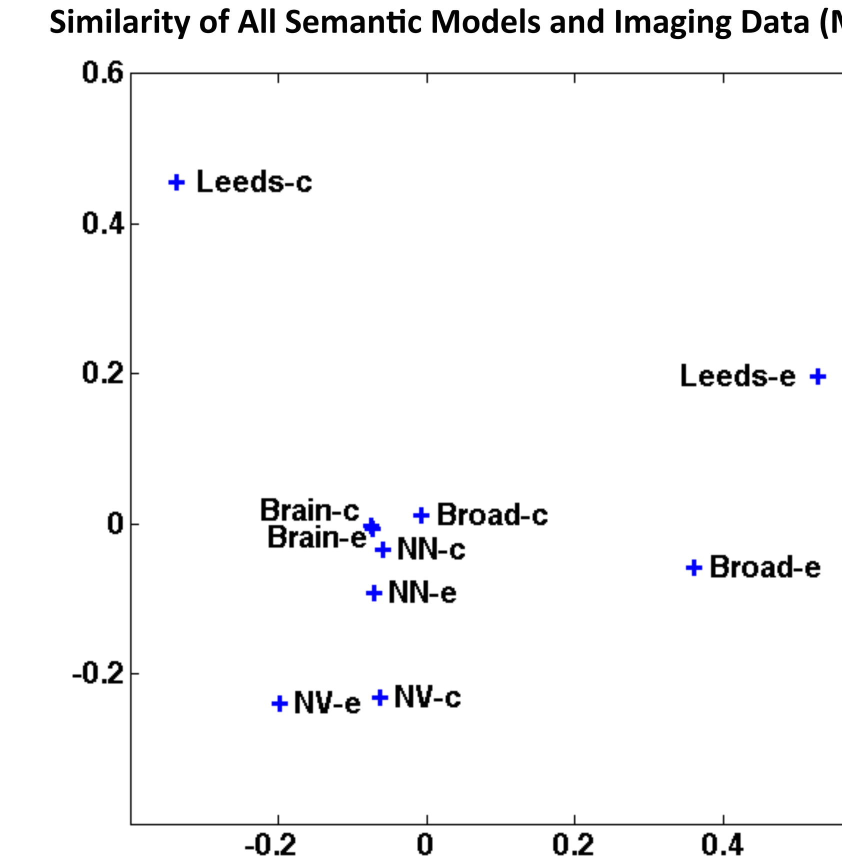
How strongly do models correlate with Imaging Data?

Correlation of Semantic Models to Imaging Data

Model	Whole brain r	Whole brain p	Harv.-Oxf. ROIs mean r	Harv.-Oxf. ROIs s.d.	Harv.-Oxf. ROIs max |r|
English					
NN	0.15	0.52	0.16	0.13	0.47
NV	0.11	0.64	0.05	0.13	0.49
Broad	-0.24	0.30	-0.14	0.17	0.60
Leeds	-0.24	0.30	-0.13	0.15	0.41
Chinese					
NN	-0.11	0.62	-0.09	0.13	-0.36
NV	-0.08	0.72	-0.10	0.14	-0.43
Broad	-0.47	0.03	-0.45	0.11	-0.69
Leeds	-0.17	0.46	-0.15	0.10	-0.38

- None of the semantic models are similar to the whole-brain data.
- **English:** Behavioral and corpus models were moderately correlated to imaging data in a few top-performing ROIs.
- **Chinese**: All models were uncorrelated, or even negatively correlated with the imaging data in certain ROIs.

Similarity of All Semantic Models and Imaging Data (MDS)



MDS plot of similarity between semantic models and whole brain representations in Chinese (c) and English (e). Brain-e and Brain-c representations are overlapping.

Conclusions

Imaging Data: Correspondence between neural representations of word-elicited concepts are very high across languages.

Behavioral Models: Behavioral ratings of noun similarity across languages are also highly consistent, but these intuitions about similarity do not account for the similarity in neural representations.

Corpus Models: The present corpus model does not capture similarity across languages, nor does it strongly relate to neural data at all.

References: Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K. M., Malave, V. L., Mason, R. A., & Just, M. A. (2008). Predicting human brain activity associated with the meanings of nouns. *Science, 320*(5880), 1191-1195.; Sharoff, S. (2006) Creating general-purpose corpora using automated search engine queries. In Marco Baroni and Silvia Bernardini, (eds), *WaCky! Working papers on the Web as Corpus*. Gedit, Bologna.; Zinszer, B. D., Anderson, A. J., Kang, O., Wheatley, T., & Raizada, R.D.S. (2015). You say potato, I say tudou: How speakers of different languages can share the same concept. *Proceedings of the 37th Annual Conference of the Cognitive Science Society*.; Harvard-Oxford brain atlas: http://www.fmrib.ox.ac.uk/fsl/