Check for updates

RESEARCH ARTICLE

Individual typological differences in a neurally distributed semantic processing system: Revisiting the *Science* article by Mitchell *et al.* on computational neurolinguistics [version 1; peer review: 1 approved with reservations]

Hiroyuki Akama

Tokyo Institute of Technology, Tokyo, 152-8552, Japan

V1 First published: 24 Apr 2018, 7:487
<https://doi.org/10.12688/f1000research.14584.1>
Latest published: 24 Apr 2018, 7:487
<https://doi.org/10.12688/f1000research.14584.1>

Abstract

Background: Revisiting the 2008 *Science* article by Mitchell *et al.* on computational neurolinguistics, individual typological differences were found as striking characteristics in the patterns of informative voxels crucial for the distributed semantic processing system.

Methods: The results of different feature selection methods (ANOVA and Stability) were compared based on the open datasets of each subject for evaluating how these features were decisive in predicting human brain activity associated with language meaning.

Results: In general, the two selection results were similar and the voxel-wise ranks were correlated but they became extremely dispersive for a subgroup of subjects exhibiting mediocre precision when examined without regularization. Quite interestingly, looking at the anatomical location of these voxels, it appears that the modality-specific areas were likely to be monitored by the Stability score (indexing “identity”), and that the ANOVA (emphasizing “difference”) tended to detect supramodal semantic areas.

Conclusions: This minor finding indicates that in some cases, seemingly poor data may deeply and systematically conceal information that is significant and worthwhile. It may have potential for shedding new light on in the controversy pertaining to cognitive semantics, which is divided into modality-biased (embodied) and amodal symbol theories.

Keywords

fMRI, MVPA, feature selection, individual variability, semantics

Peer review discontinued

Peer review at F1000Research is author-driven. Currently no reviewers are being invited. [What does this mean?](#)

Approval Status ?

1

version 1

24 Apr 2018

?
[view](#)

1. Chiu-Hsieh Hsu, Academia Sinica, Taipei, Taiwan

Any reports and responses or comments on the article can be found at the end of the article.

This article is included in the **INCF** gateway.

This article is included in the **Preclinical Reproducibility and Robustness** gateway.

Corresponding author: Hiroyuki Akama (akama.h.aa@m.titech.ac.jp)

Author roles: Akama H: Conceptualization, Formal Analysis, Investigation, Methodology, Project Administration, Software, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2018 Akama H. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

How to cite this article: Akama H. **Individual typological differences in a neurally distributed semantic processing system: Revisiting the *Science* article by Mitchell *et al.* on computational neurolinguistics [version 1; peer review: 1 approved with reservations]** F1000Research 2018, 7:487 <https://doi.org/10.12688/f1000research.14584.1>

First published: 24 Apr 2018, 7:487 <https://doi.org/10.12688/f1000research.14584.1>

Introduction

It is widely acknowledged that despite some challenges in multi-voxel pattern analysis (MVPA), the issue of individual variability, raised as a penalty to classification accuracy in cross-subject modelling, is challenging to overcome, particularly when targeting concepts and meanings conveyed by language¹. Admittedly, the precision rates in MVPA could be mostly uniform for experiments successfully performed at the individual first level, as in the case of the classically recognized *Science* article authored by T. Mitchell and his group (*Predicting Human Brain Activity Associated with the Meanings of Nouns*)². In their study, distributed brain activation patterns were observed in nine subjects conceptualizing 60 concrete nouns in an fMRI scanner, and these neural patterns were regressed on a co-occurrence probability in a text corpus between each of these fMRI nouns and semantic features (25 basic verbs). The highly significant prediction accuracy that they obtained at the subject level was associated with a pattern separability built on a set of informative voxels as features, which are distributed across numerous brain areas and differ across subjects. The *Science* study and other computational neurolinguistic reports^{3–7} could generate classifiers with high precision (owing to the L2-regularization technique that obliterates individual variability) and draw plausible semantic maps in the individual brain.

However, functionality at work in the individual brain remains to be specified further beyond unanimously acceptable modelling results. In this study, an alternative view is put forward to uncover systematicity and provide typology in individual variability through a reanalysis of the open data used in the Mitchell *et al.*'s *Science* study. Several previous studies have reported that despite the almost invariably accurate performance of the subjects in conceptualization tasks, a fundamental difference is seen in the magnitude of how the information on the feature voxels is dispersive based on their anatomical locations^{1,5}. Unquestionably, the experimental paradigm developed by Mitchell *et al.* hinged on modality specific factors, which promoted distributiveness of semantic processing systems. Their stimulus set, which used captioned drawings with considerable visual effects and concrete nouns with implications of some motion/perception, rendered the experiment sensitive to embodied cognition⁸. The attribute generation task, which consisted of thinking about the properties of the object, could allow free association and perceptual simulation⁹. Yet, even without accounting for the modality specific factors, there remains some debate pertaining to the supramodal semantic centers in general as to whether the left temporal pole and anterior temporal gyrus (hub and spoke model)^{10–11} or the left middle temporal gyrus and the left angular gyrus (high level convergence model)^{12–13} are nodal for the loci of a genuinely semantic process. The primary goal of the present study was to determine whether such topology of semantic processing can be elucidated through a typology of fMRI subjects and to examine how that typology is determined by the subjects' hidden neural responses characteristic to the selection of the most informative voxels.

Methods

The datasets were nine .mat files corresponding to the nine subjects (P1–P9), downloaded from the website of Carnegie Mellon University. A '.mat' file was created (by using **MATLAB** R2015a) from each subject's data, called 'runByVoxByNouns-P<number>.mat,' as a three-dimensional array corresponding to x: runs (six repeated presentations) by y: voxels by z: words (60 fMRI nouns). This procedure facilitated the computation of informative voxel locations identified by different feature-selection methods, which were the F values of ANOVA to measure and evaluate the mean value across the items repeatedly presented in a learning set and the Stability scores to distinguish voxels that exhibited consistently similar activation patterns to the items for machine learning¹⁴. Mitchell *et al.* adopted the latter method to select the top 500 voxels, which involves computing the Pearson's correlation coefficient of the activation vectors for the stimulus nouns over 15 ($=_6P_2$) pairs of fMRI presentation runs repeated 6 times. They computed the average pairwise correlation for each voxel over all pairs of rows in the matrix composed of six presentation runs by 58 nouns (reserving two nouns for testing). In this study, cross-validation was not performed except for computing the modeling accuracy for each subject using the ordinary least square method (OLS) without L2-regularization. The top 500 voxels were selected from the overall runs by the ANOVA and the Stability scores, and the number of the voxels not shared by the two selection results (subtraction operation for the two sets, i.e., type 1: ANOVA set – Stability set and type 2: Stability set – ANOVA set) was counted as "divergence" for each subject, as shown in **Table 1**. The voxel-wise ranks in the two selection results were compared with each other with Spearman's rank correlation

Table 1. Details of the data on the per-subject top 500 feature voxels.

Subjects	RHO	P value	divergence	Mean ANOVA	Mean Stability	OLS
P1	0.396	0	13	3.83	4.72	0.82
P2	0.460	0	36	2.54	2.98	0.78
P3	0.243	4.1E-08	21	2.32	2.67	0.78
P4	0.242	4.5E-08	27	2.45	2.90	0.78
P5	0.141	0.002	82	1.79	1.81	0.75
P6	-0.022	0.628	183	1.75	1.69	0.64
P7	0.392	0	59	1.79	1.76	0.71
P8	0.023	0.607	146	1.62	1.50	0.6
P9	-0.028	0.533	124	1.66	1.59	0.65

The top 500 voxels were selected by the ANOVA and the Stability and the modelling accuracy based on the ordinary least square method (OLS) without adjustment of L2-regularization. The subject-wise 'rho' and the corresponding *p* values represent Spearman's rank correlation coefficient between the voxel-wise ranks in the two feature selection results. 'Divergence' implies the number of the selected voxels extracted by the subtraction operation of ANOVA set – Stability set or Stability set – ANOVA set.

coefficient and the subject-wise ρ and the corresponding p values were computed. The ANOVA and Stability feature-ranking data were all mapped to anatomical regions according to the automated anatomical labeling (AAL) atlas¹⁵.

Results

The raw modelling accuracy, based on the ordinary least square methods (OLS) without L2-regularization, decayed with the magnitude of “divergence” between the ANOVA and the Stability score and with the decrease in rank similarity between the voxels selected by each method (Table 1). Admittedly, good modelling accuracy is associated with high F values and Stability

scores for the selected voxels, as is seen in P1 eliciting the best precision by a wide margin and the next best group of P2–P4. However, there was no mean difference in feature scores among the subjects of the middle group, P5, P6, and P7, although P5 and P7 exhibited significant Spearman’s rank correlation coefficients, divergence less than 100, and OLS precision higher than 70%, while none of these conditions was true for P6. When visualizing the score distributions of the top 500 voxels with a notched box plot, it is apparent that the poor-performance group (P6, P8, and P9) was characterized by narrowness of the boxes and the low upper whiskers (Figure 1) and dispersiveness of the voxel-wise ranks in the two selection results (Figure 2).

Figure 1. Notched box plot for the score distributions of the top 500 voxels. The voxels were selected by the ANOVA (above) and the Stability score (below). The horizontal axis corresponds to the numbered participants. This figure was depicted using the MATLAB function of 'boxplot.' Data points exceeding the whiskers are displayed using +. The singularity of Participant 1 is conspicuous with the best scores by far, while the typology of the other participants was easily recognized classifying them into two groups (fair: P2, P3, and P4; poor: P5–P9). However, a finer-grained observation allows us to stress the narrowness of the boxes and the low upper whiskers for P6, P7, and P9 and bundle them as a small subgroup of interest.

Figure 2. Per-subject scatter plots of the ranks of the top 500 voxels. The residuals to the regression line of $y=x$ represent how closer or farther the two kinds of ranks of the selected voxels. The dots right to the red vertical line ($y = 500$) represent the voxels (type 1) selected by the ANOVA but not by the Stability score as the top 500. The dots over the red horizontal line ($x = 500$) correspond to the voxels (type 2) selected by the Stability score but not by the ANOVA under the same criteria. Irregularity in distribution is found in P6, P8, and P9.

Dataset 1. Reanalysis of Mitchell *et al.* data

<http://dx.doi.org/10.5256/f1000research.14584.d201767>

Discussion

Our results support the notion that particular types of individuals differ markedly in their way of recruiting voxels with respect to different feature selection methods, i.e., Stability scores and F values of ANOVA. The Stability scores examine the extent to which each voxel reacted to the same stimulus across runs in a constant manner; therefore, it is the “identity” of an object that is emphasized by this index as invariable through repetition. Conversely, the F values of ANOVA pertain to the magnitude of between-group variance across the responses to the 60 nouns with respect to the within-group variance across the 6 presentations of each individual noun; therefore, the “difference” is likely to be captured by that index although it should be inextricably linked with the “identity” side. In consequence, regardless of the feature selection method, mostly the same voxels with a similar top 500 ranking order could be selected from the brains of P1–P5 and P7, but the remaining subjects (P6, P8, and P9) showed important divergence from the list of the 500 important voxels selected by the two methods. The mean index values for the top voxels were significantly larger in the former subjects than the latter ones for each method; the difference in raw classification accuracy without the regularization effect was conspicuous between these subject groups. However, an in-depth analysis revealed some questions to be delineated, since P5 and P7 may be treated equally as members of an interesting subgroup in that, despite the highly significant rank correlation for the selected voxels, the mean values were relatively low and not significantly different from those of P6 in the poor-performance group.

When tapping into the anatomical regions from which feature voxels were selected, the most intriguing property was that high precision in modelling was guaranteed rather by extra-linguistic regions. It is noteworthy that in P1 (recording 82% as classification accuracy by the OLS with no regularization), the majority of the top 487 informative voxels shared by the Stability score and the ANOVA were found in the visual areas of the temporal and occipital lobes and several in the frontal and parietal lobes. Differently from all the other subjects, the left inferior frontal gyrus, pars triangularis (“Frontal_Inf_Tri_L”), and the left precuneus (“Precuneus_L”), frequently considered as involved in executive functions of language activity, were not recorded in the overlapping selected voxel areas of P1. When focusing on the poor-performance subject group (P6, P8, and P9), which exhibited a large divergence (larger than 1 standard

deviation from the mean) between the voxel selections by the two methods, it appeared that the modality-specific areas were likely to be monitored by the Stability score (indexing “identity”), and that the ANOVA (emphasizing “difference”) tended to detect supramodal semantic areas. The voxels type 1 (selected by the ANOVA but not by the Stability score as the top 500) and type 2 (selected by the Stability score but not by the ANOVA under the same criteria) voxels were mostly extracted from different anatomical regions. The frequency distribution tables (Figure 3) represent the number of type 1 and type 2 voxels selected from P6, P8, and P9 and attributed to each anatomical area in the AAL brain atlas. The ANOVA highlighted as locations of type 1 voxels, some areas for amodal or supramodal semantic processing, especially the left middle temporal gyrus (“Temporal_Mid_L”) which was the most populous by far in this category. In contrast, the Stability score tended to introduce bias to the vision-related areas, notably the left middle occipital gyrus (“Occipital_Mid_L”), which may reflect stimulus modality or perceptual symbols in embodied cognition.

This divergence allows us to shed a new light on a traditionally controversial subject in neural semantics; where is the border that separates the brain regions selective to purely conceptual functions and sensory-driven, modality-dominant, so extrinsic to meaning processing? The compatibility of the Stability score with the perceptual modalities may suggest with the embodiment view (to which Mitchell *et al.* also referred for the neural signature of the verb “eat”) that the “identification” of a concept is materially founded upon a sensory-perceptual system and real-life experience with its instances of referent (or stimuli) to shape its cognitively grounded symbols. However, the voxel information brought by the ANOVA enables us to propose an alternative view for the discrimination power in language, having an affinity for the amodal (not to say disembodied) symbol theory. Descended from the school of Saussure, this theory (often relying on lexical co-occurrence information from language corpora as in the case of the *Science* study) postulates that the value of a symbol (or a linguistic sign) is not derived from its intrinsic sense but from language itself as a computability system of “difference.” It is quite intriguing that the reanalysis of the *Science* data assessed, through the variability of subjects performing a language task, a salient discrepancy between the brain regions informative of manifold essence in semantic processing. Indeed, we are not yet in a position to argue these philosophically opposite views only through a succinct review such as this report. However, at least we may conclude here that such a fundamental issue was, quite interestingly, readdressed by reanalyzing the data from a subject group that elicited inconsistency to the feature selection methods

Figure 3. Frequency tables of the informative voxels extracted from the poor-performance subjects. The voxels were classified into type 1 (above, ANOVA subtracted from the Stability score) and type 2 (below, Stability score subtracted from the ANOVA) and attributed to each anatomical area in the AAL brain atlas (the voxels labelled 'Not_Found' were removed from the lists). The type 1 voxels tend to belong to the supramodal semantic regions (such as "Temporal_Mid_L"), whereas the type 2 ones are characterized by the dominance of the visual area (such as "Occipital_Mid_L").

and relatively low precision rates to fMRI machine learning classifiers.

Data availability

The dataset of Mitchell *et al.* was downloaded from

<http://www.cs.cmu.edu/afs/cs/project/theo-73/www/science2008/data.html>

Dataset 1: Reanalysis of Mitchell *et al.* data. [10.5256/f1000research.14584.d201767](https://doi.org/10.5256/f1000research.14584.d201767)¹⁶

Competing interests

No competing interests were disclosed.

Grant information

The author(s) declared that no grants were involved in supporting this work.

Acknowledgement

The author would like to thank Editage (www.editage.jp) for English language editing.

References

1. Akama H, Murphy B: **Emerging Methods for Conceptual Modelling in Neuroimaging.** *Behaviormetrika*. Springer. 2016; **44**(1): 117–133.
[PubMed Abstract](#) | [Publisher Full Text](#)
2. Mitchell TM, Shinkareva SV, Carlson A, *et al.*: **Predicting human brain activity associated with the meanings of nouns.** *Science*. 2008; **320**(5880): 1191–5.
[PubMed Abstract](#) | [Publisher Full Text](#)
3. Bullinaria JA, Levy JP: **Limiting factors for mapping corpus-based semantic representations to brain activity.** *PLoS One*. 2013; **8**(3): e57191.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
4. Anderson AJ, Binder JR, Fernandez L, *et al.*: **Predicting Neural Activity Patterns Associated with Sentences Using a Neurobiologically Motivated Model of Semantic Representation.** *Cereb Cortex*. 2017; **27**(9): 4379–4395.
[PubMed Abstract](#) | [Publisher Full Text](#)
5. Pereira F, Detre G, Botvinick M: **Generating text from functional brain images.** *Front Hum Neurosci*. 2011; **5**: 72.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
6. Akama H, Miyake M, Jung J, *et al.*: **Using Graph Components Derived from an Associative Concept Dictionary to Predict fMRI Neural Activation Patterns that Represent the Meaning of Nouns.** *PLoS One*. 2015; **10**(4): e0125725.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
7. Huth AG, de Heer WA, Griffiths TL, *et al.*: **Natural speech reveals the semantic maps that tile human cerebral cortex.** *Nature*. 2017; **532**(7600): 453–8.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
8. Willems RM, Casasanto D: **Flexibility in embodied language understanding.** *Front Psychol*. 2011; **2**: 116.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
9. Engelen JA, Bouwmeester S, de Bruin AB, *et al.*: **Perceptual simulation in developing language comprehension.** *J Exp Child Psychol*. 2011; **110**(4): 659–75.
[PubMed Abstract](#) | [Publisher Full Text](#)
10. Patterson K, Nestor PJ, Rogers TT: **Where do you know what you know? The representation of semantic knowledge in the human brain.** *Nat Rev Neurosci*. 2007; **8**(12): 976–87.
[PubMed Abstract](#) | [Publisher Full Text](#)
11. Rice GE, Lambon Ralph MA, Hoffman P: **The Roles of Left Versus Right Anterior Temporal Lobes in Conceptual Knowledge: An ALE Meta-analysis of 97 Functional Neuroimaging Studies.** *Cereb Cortex*. 2015; **25**(11): 4374–91.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
12. Binder JR, Desai RH, Graves WW, *et al.*: **Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies.** *Cereb Cortex*. 2009; **19**(12): 2767–96.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
13. Binder JR, Desai RH: **The neurobiology of semantic memory.** *Trends Cogn Sci*. 2011; **15**(11): 527–36.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
14. Pereira F, Mitchell T, Botvinick M: **Machine learning classifiers and fMRI: a tutorial overview.** *NeuroImage*. 2009; **45**(1 Suppl): S199–209.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
15. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, *et al.*: **Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain.** *NeuroImage*. 2002; **15**(1): 273–89.
[PubMed Abstract](#) | [Publisher Full Text](#)
16. Akama H: **Dataset 1 in: Individual typological differences in a neurally distributed semantic processing system: Revisiting the Science article by Mitchell *et al.* on computational neurolinguistics.** *F1000Research*. 2018.
[Data Source](#)

Peer review discontinued Peer review at F1000Research is author-driven. Currently no reviewers are being invited. What does this mean?

Version 1

Reviewer Report 18 July 2018

<https://doi.org/10.5256/f1000research.15871.r33424>

© 2018 Hsu C. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Chiu-Hsieh Hsu

Institute of Linguistics, Academia Sinica, Taipei, Taiwan

The manuscript is a reanalysis of the study of Mitchell et al. (2008). The author claims that using different feature selection methods can help to identify areas that are functional distinctive in individual level. The finding of the presented study provides substantial implication to studies that focus on how neural activation encoding the semantic content. I only have few minor comments.

1. (on page 3) please provide details of the cross-validation method of the OLS analysis.
2. In Discussion, the author describe about the distribution of "type 1 voxels" and "type 2 voxels" in some participants. I wonder if the results in the rest of the participants show a similar pattern?
3. I would like to suggest that the author can provide figures or tables that show the distribution of three types of voxels (type 1, type 2, and convergent voxels) of each participant. This information would help the reader to see the consistency/divergence of encoding models in each participant.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research