# Modeling Conventionalization and Predictability in Multiword Expressions at the brain level

Murielle Popa-Fabre [1] Shohini Bhattasali [2] John Hale [3] Christophe Pallier [4]

[1]CNRS Université Paris Diderot–Paris 7 [2]University of Maryland [3]University of Georgia [4]CEA Inserm Cognitive Neuroimaging Unit

CRCNS 2019

## Introduction

Multiword Expressions (MWEs) demonstrate finer-grained distinctions such as degrees of conventionalization and predictability.

Examples: *break the ice, boa constrictor, safe and sound, see to it, in spite of*

These gradiences could be reflected during cognitive processing and are captured using computational metrics, Association Measures.

### Question

Are these computational measures and their cognitive instantiations are discernible at the cerebral level during sentence processing?

## Data Collection

Participants (n=51) were college-aged, right-handed, native English speakers.

Listened to a spoken recitation of *The Little Prince* for 1 hour and 38 minutes across nine separate sections; 15,388 words in total.

Comprehension was confirmed through multiple-choice questions at the end of each section (90% accuracy, SD = 3.7%).

## Multiword Expressions

742 MWEs were identified in the dataset through a transition-based MWE analyzer (Al Saied et al., 2017) trained on Children's Book Test dataset (Hill et al., 2015).

Association Measures calculated using corpus frequency counts from COCA:

1. Pointwise Mutual Information (PMI):

$$PMI = log_2 \frac{c(w_n^1)}{E(w_n^1)} \quad (1)$$

2. Dice's Coefficient:

$$Dice = \frac{n \times c(w_n^1)}{\Sigma_{i=1}^n c(w_i)} \quad (2)$$

## Group-level Results

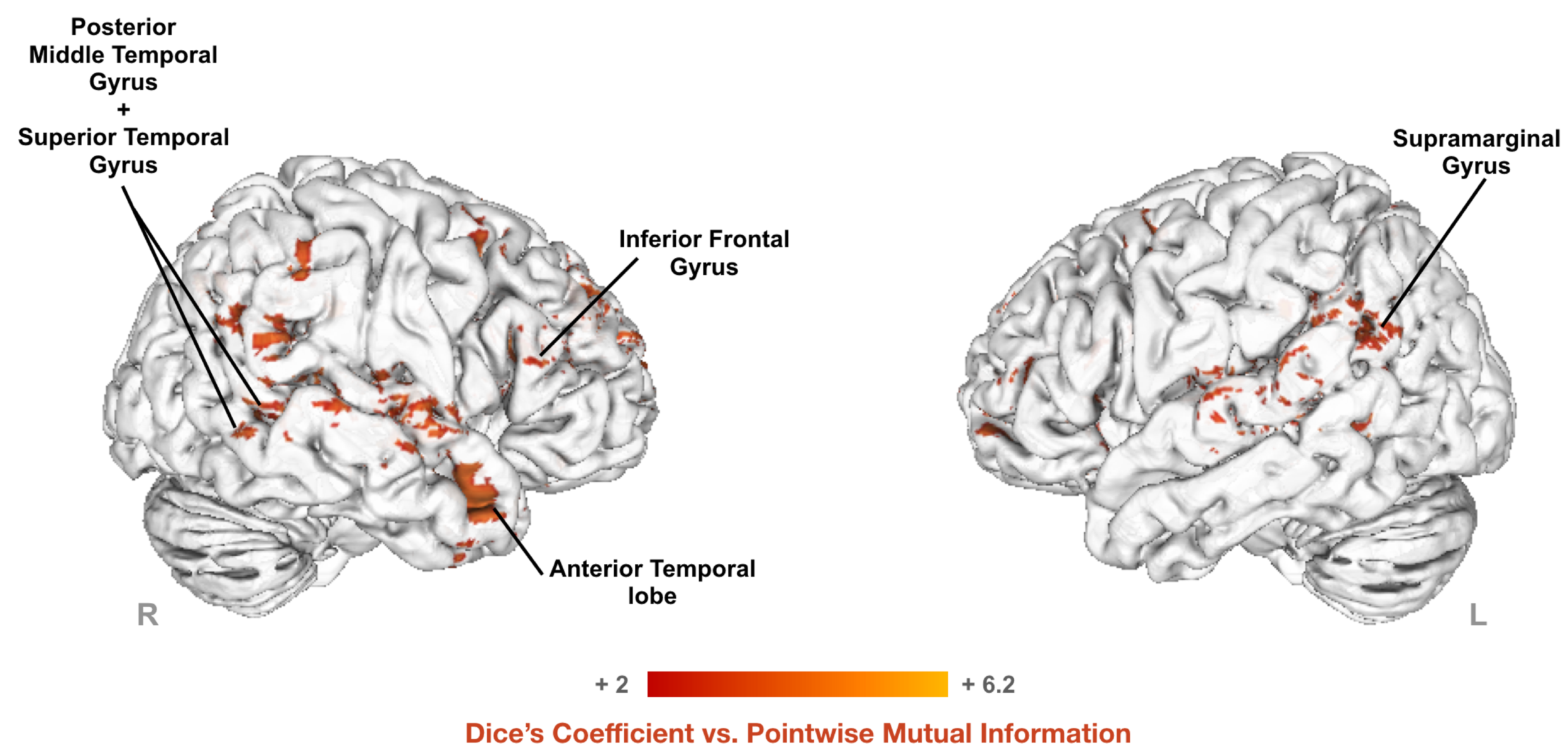


Figure 1: Z-map showing regions having a significant effect for Dice's coefficient versus Pointwise Mutual Information after Bonferroni correction with $p < 0.05$

| Regions for Dice >PMI | Cluster size (in voxels) | MNI Coordinates x | y | z | z-scores |
|---|---|---|---|---|---|
| R Superior Temporal Gyrus (BA 38) | 47 | 48 | 10 | -26 | 5.80 |
| R Middle Temporal Gyrus | 84 | 54 | -18 | -10 | 6.09 |
| R Middle Temporal Gyrus (BA 22) | 98 | 48 | -36 | 2 | 5.85 |
| R Superior Temporal Gyrus (BA 22) | 70 | 48 | -12 | 2 | 5.83 |
| R Middle Temporal Gyrus (BA 22) | 16 | 58 | -46 | 2 | 5.14 |
| L Superior Temporal Gyrus | 13 | -62 | -18 | 6 | 5.64 |
| R Superior Frontal Gyrus | 10 | 20 | 56 | 12 | 5.53 |
| R Inferior Frontal Gyrus (BA 45) | 10 | 48 | 20 | 14 | 5.64 |
| L Supramarginal Gyrus | 22 | -56 | -56 | 22 | 5.37 |
| R Inferior Parietal Lobule/ Superior Temporal Gyrus (BA 40) | 10 | 62 | -46 | 22 | 5.44 |
| R Inferior Parietal Lobule/ Superior Temporal Gyrus (BA 40) | 16 | 54 | -46 | 22 | 5.45 |
| R Superior Frontal Gyrus | 35 | 20 | 42 | 34 | 5.69 |
| R Cingulate Gyrus | 17 | 2 | -34 | 34 | 5.85 |
| R Precenus | 22 | 32 | -72 | 36 | 5.76 |
| L Inferior Parietal Lobule | 12 | -34 | -58 | 46 | 5.17 |

Table 1: Significant clusters for Dice's Coefficient versus Pointwise Mutual Information after Bonferroni correction with $p < 0.05$

## fMRI Analysis

Preprocessing was carried out with AFNI version 16 and ME-ICA v3.2 (Kundu et al., 2011).

GLM Analysis:

- MWE predictors convolved with SPM12's canonical HRF, regressed against observed BOLD signal during passive story listening.
- To account for sentence-level compositional processes, a regressor formalizing syntactic structure building included (Hale, 2014).
- Includes four regressors of non-interest: word offset, frequency, pitch, intensity

$r^2$ Model Comparison: For every subject, we compute how much the inclusion of each variable of interest (i.e. Dice and PMI) increases the cross-validated $r^2$.

### Conclusion

- Dice's Coefficient, formalizing the degree of predictability, is a better predictor of cerebral activation for processing MWEs.
- Indicates that Dice is a more cognitively plausible computational metric.
- Suggests that a bimodal distribution of gradience is more suitable for processing MWEs

## Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 1607441.