A Concurrent NLP-fMRI Approach to the Brain's Mathematical Network

M2 CogMaster Internship Defense

Samuel Debray

September 16, 2022

École Normale Supérieure

The Brain Networks for Advanced Mathematics

Amalric, M., & Dehaene, S. (2016). Origins of the brain networks for advanced mathematics in expert mathematicians. Proceedings of the National Academy of Sciences, 113(18), 4909–4917.

fMRI analysis of subjects listening to advanced mathematical statements.

Subjects mathematicians or controls (matched in qualification).

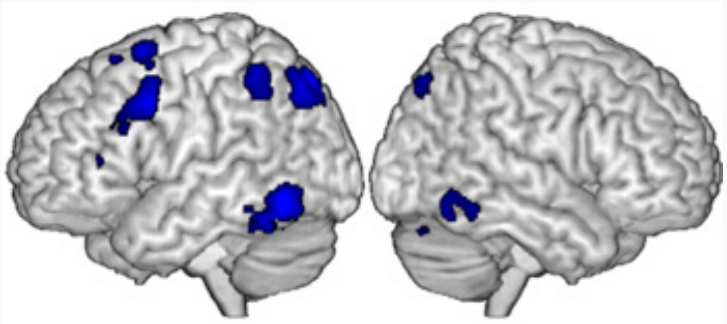

Figure 1: Interaction (maths – non-maths in mathematicians) > (maths – non-maths in controls).

Evidence for a mathematical network, elicited by advanced and elementary mathematics.

Network disjoint with language areas.

Aims:

(i) use Machine Learning models to capture mathematical semantics;

(ii) see if Natural Language Processing representations correlate with behavioural and fMRI data.

Natural Language Models

Word embedding model

GloVe – GLObal VEctors

Algorithm that parses a corpus and returns a dictionary

word $\mapsto$ vector encoding context.

Very common in the litterature to analyse language processing in the brain.

Text embedding model

The Transformer

State-of-the-art deep learning model.

Computes semantic representation of texts.

Endowed with great mathematical abilities.

Semantic Analysis of Mathematics

Method

Creation of a vocabulary using `GloVe`.

1. Parse all French Wikipedia pages to find mathematical pages.
2. Lemmatise the pages (e.g. ”computed” $\rightarrow$ ”compute”, ”Theorems” $\rightarrow$ ”theorem”).
3. Train a `GloVe` model on the pages, with output vectors in 50 dimensions.
4. Retain (manually) the 1,000 most frequent words deemed desirable.
5. Analyse them...

PCA

Idea: (orthogonal) directions which capture the maximal amount of variance.

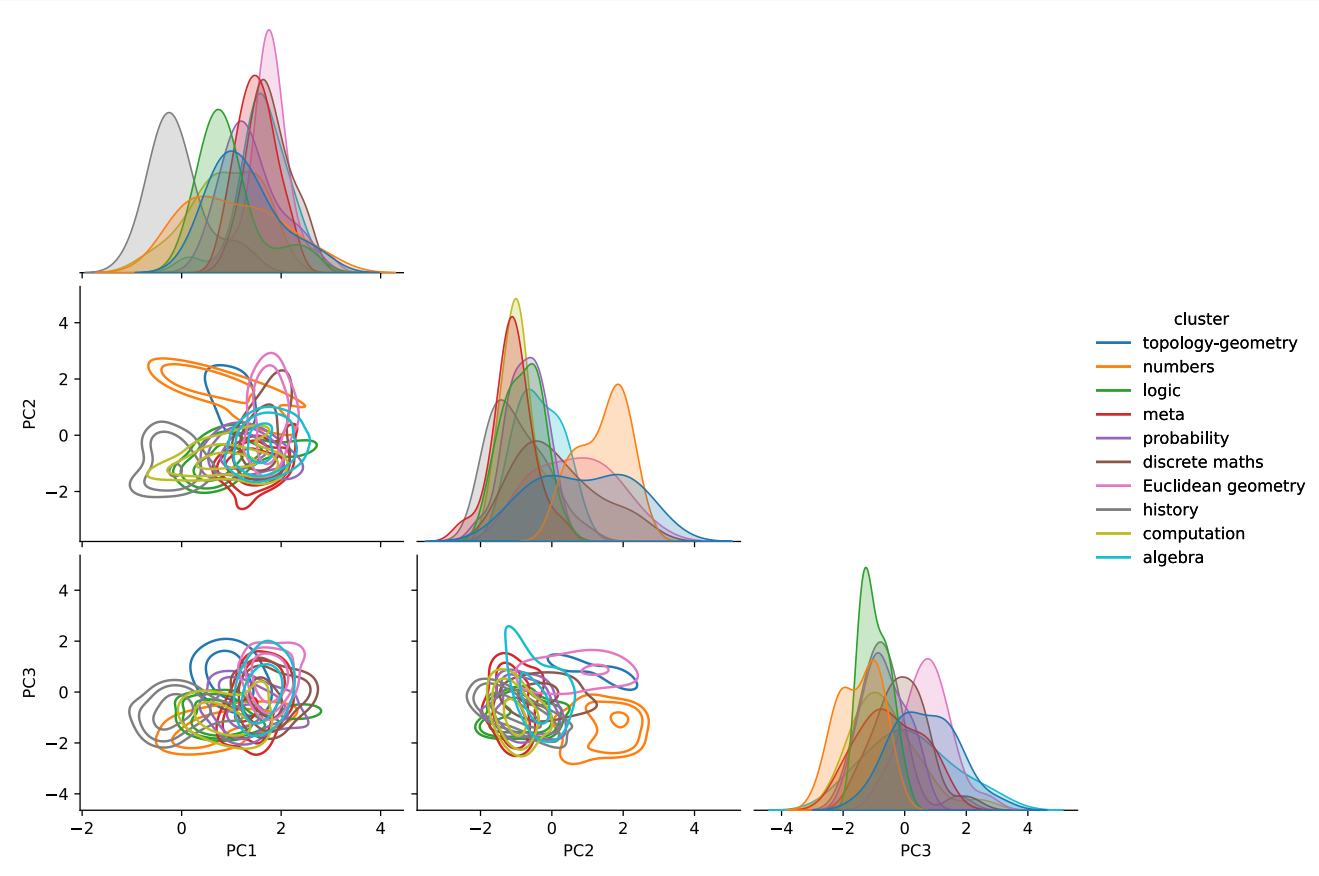

Figure 2: Kernel density plot of the 20 most frequent words of each cluster.

Analysis: PCA + reduction to 34 dim. + spectral clustering (10 clust.)

Spectral Clustering – Semantic Map of Mathematics

Idea: almost like `k-means`, i.e. find $S := \amalg_{i=1}^{k} S$ s.t.

$$S = \arg \min_{\substack{X = \amalg_{i=1}^{k} X_i \\ X \text{ partition of data}}} \sum_{i=1}^{k} \sum_{x \in X_i} \|x - \mu(X_i)\|_2^2.$$

Analysis: spectral clustering (10 clust.) + `tSNE` + Voronoi tessellation

Mathematics and the `Transformer`

How we used the `Transformer`

The model used was `GPT-fr`.

$$\begin{array}{lcccl} \text{Input} & \xrightarrow{\text{Model}} & \text{Logits} & \xrightarrow{\mathcal{L}:\, x \mapsto -\log(x)} & \text{NLL} \\ x_1 & & \Pr_\theta(x_1) & & -\log(\Pr_\theta(x_1)) \quad := \text{NLL}(x_1) \\ x_2 & & \Pr_\theta(x_2|x_1) & & -\log(\Pr_\theta(x_2|x_1)) \quad := \text{NLL}(x_2) \\ \vdots & & \vdots & & \vdots \\ x_t & & \Pr_\theta(x_t|x_{<t}) & & -\log(\Pr_\theta(x_1|x_{<t})) \quad := \text{NLL}(x_t) \end{array}$$

The output score is

$$\text{output}(X) := \max_{1 \le i \le t} \text{NLL}(x_i), \qquad X = x_1 \cdots x_t$$

it captures the model's surprisal on the input statement.

Model evaluated on same statements as subjects of Amalric and Dehaene's experiments.

The `Transformer` as a Classifier

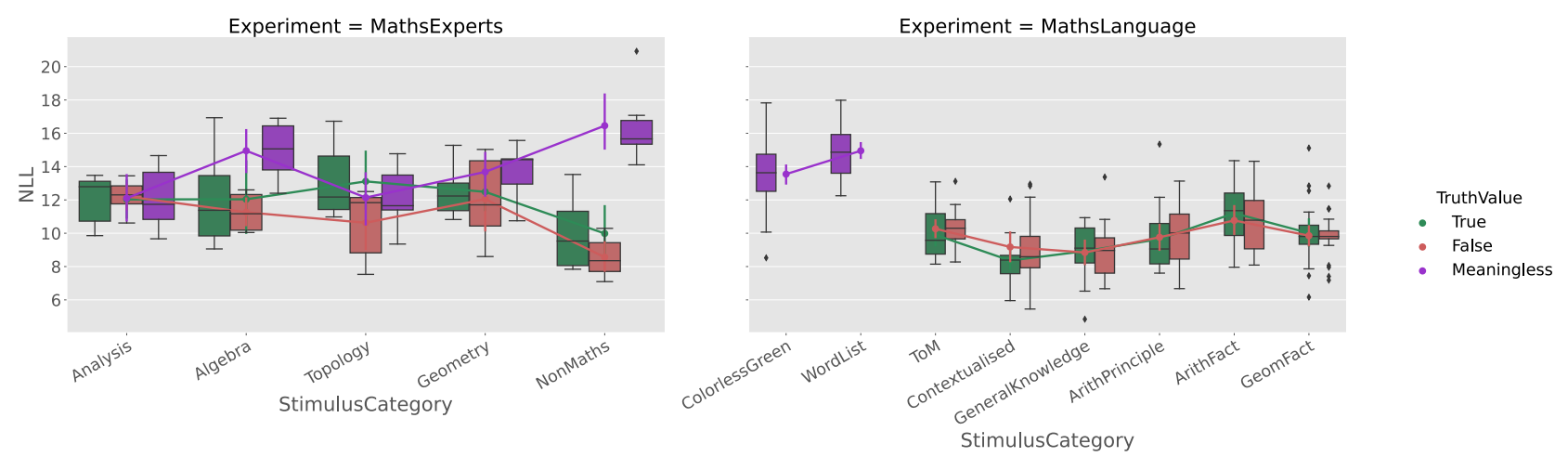

Figure 3: Surprisal computed by `GPT-fr` as a function of stimuli's category and truth value.

Found effects of:

- truth value — *test meaning's effect* — *restrict to meaningful*
- category — *restrict to mathematical*

Strong effect of meaning (meaningful vs. meaningless).

No effect of truth value when restricted to meaningful mathematical stimuli.

The Transformer and Human Subjects

Figure 4: Percentage of subjects evaluating the stimulus is not true against GPT-fr's output.

Analysis of fMRI Data

Question

Do the first principal components of the global `GloVe` embedding enable to predict the distinction between mathematical and non-mathematical stimuli reported by Amalric and Dehaene?

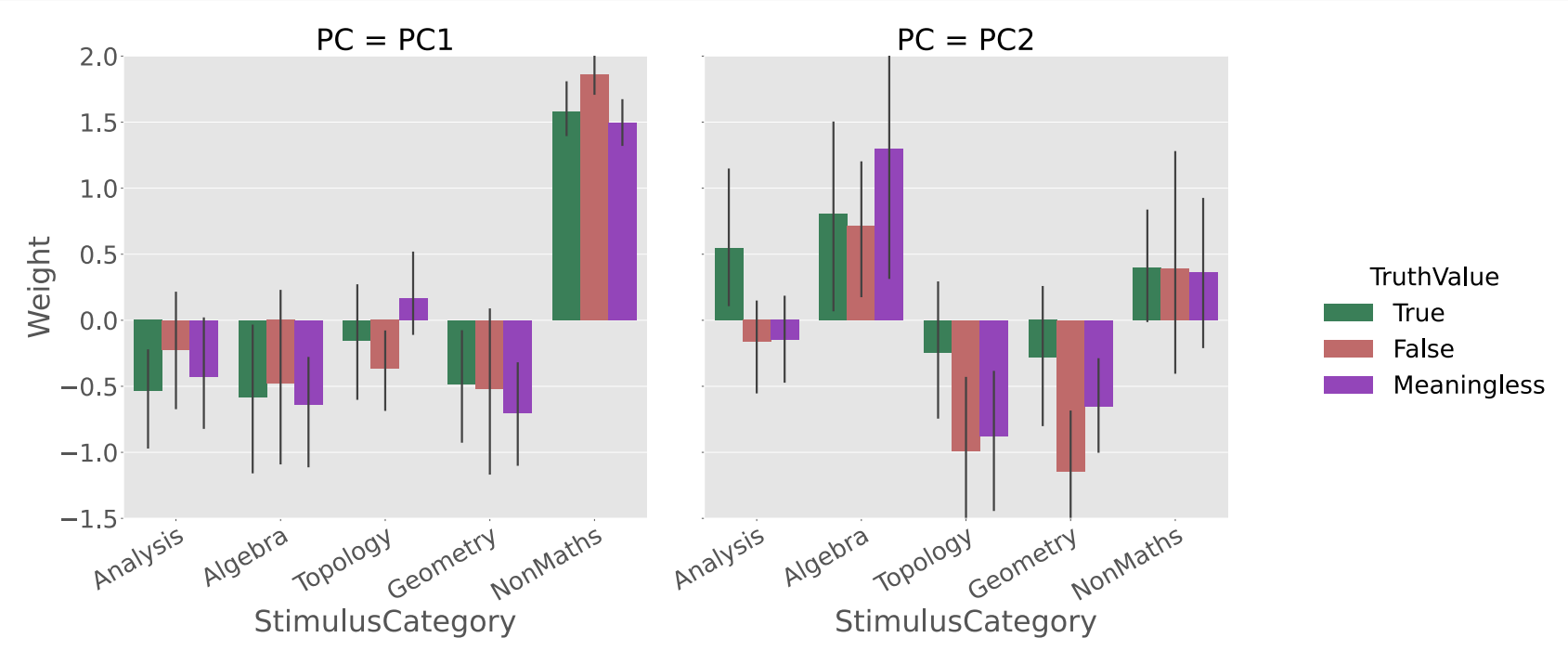

Figure 5: Projection of the stimuli's embedding onto PC1 and PC2 of the global `GloVe` model. PC1 explains 18.5% of the observed variance and PC2 explains 6.4%.

GLM Description

Only analysed Amalric and Dehaene's MathsExperts.

`GloVe` model trained on global corpus (maths + non-maths).

Regressors of interest:

- Categoric: Meaningful/Meaningless
- Parametric: PC1–3 for meaningful stimuli

No categorical regressor to tell whether a stim. is math. or not.

Regions Showing a Significant Effect of PC1 in Mathematicians

Effect of `MeaningfulPC1` but no effect from the others.

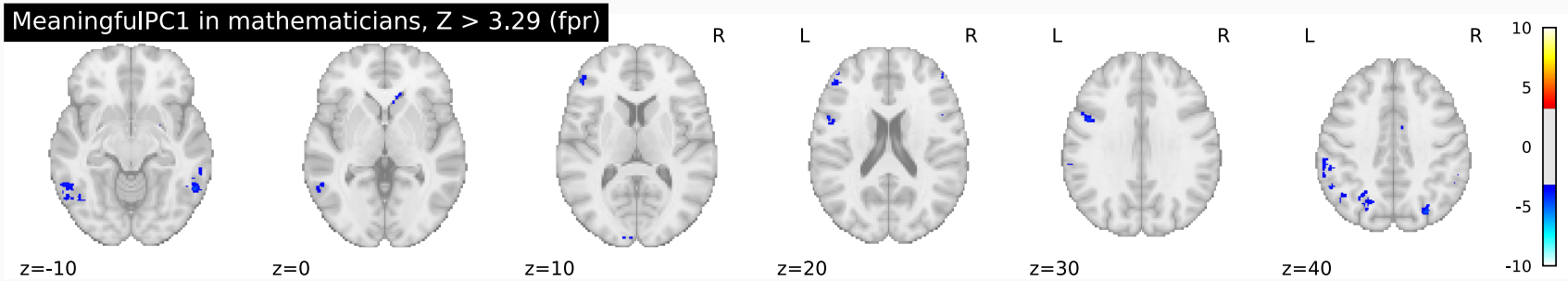

Figure 6: Group analysis, Z-values, $n = 15$.

ROI analysis: `MeaningfulPC1` has an effect in mathematicians but not in controls. No effect of the other `MeaningfulPC`s.

Limitations & Possible Continuations

Take Home Messages

- `GloVe` does capture a fair amount of mathematical semantics and spectral clustering bring out a classification of mathematics.

- `GPT-fr` is able to make the distinction between meaningful and meaningless statements.

- The first principal component of the `GloVe` embeddings of the global vocabulary makes a clear distinction between mathematical and non-mathematical stimuli, and enables to retrieve Amalric and Dehaene's mathematical network.

Limitations & Possible Continuations

Limitations

- Limited ressources in French.
- Sentence judgement not optimal for the `Transformer`.
- For fMRI: looking at group analyses + not much data.

Possible continuations

- Redo in English, and use for instance `GPT-3`.
- Train a model of the `Transformer`.
- PCA too brutal? Find another way to reduce dimensions...
- Huth et al.'s approach: run PCA across voxels and do within-sub. analyses.

Questions?

Links

- Report: https://perso.crans.org/sdebray/files/M2InternshipReport.pdf
- Semantic map: https://perso.crans.org/sdebray/projects/MathsNLP/ClusteredMapMathematics.svg
- Dendrogram: https://perso.crans.org/sdebray/projects/MathsNLP/DendrogramMathematics.svg

Cortical map of language

Work of Huth et al. 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.

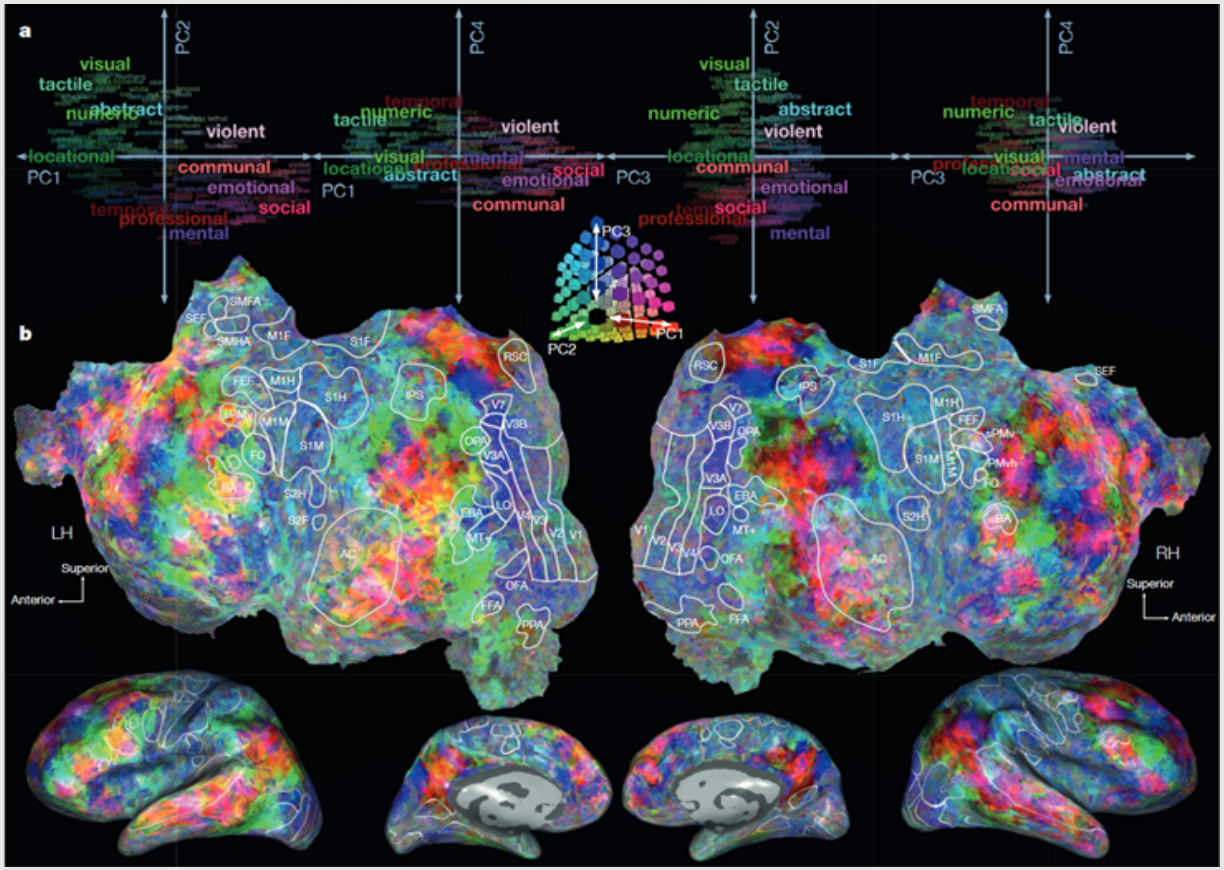

Figure 7: Cortical map of language from fMRI activation.

AI and Natural Language Processing (cont.)

Brain decoder

Work of Pereira et al. Pereira, F., Lou, B., Pritchett, B., Ritter, S., Gershman, S. J., Kanwisher, N., Botvinick, M., & Fedorenko, E. (2018). Toward a universal decoder of linguistic meaning from brain activation. Nature Communications, 9(1), 963.

Figure 8: Semantic space from a 30, 000-words vocabulary.

Creation of a decoder of linguistic meaning from brain activation.

Full Semantic Map of Mathematics