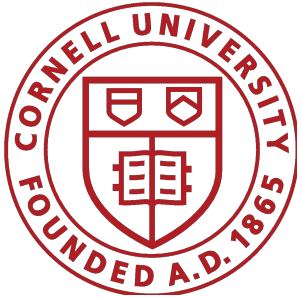


Processing MWEs: Neurocognitive Bases of Verbal MWEs and Lexical Cohesiveness within MWEs

Shohini Bhattasali, Murielle Fabre & John Hale
✉ sb2295@cornell.edu

COLING 2018: LAW-MWE-CxG ● August 26, 2018

Introduction

- Human study of natural language comprehension
- Brain basis or areas that correspond to different aspects of MWE comprehension
- Present neuroimaging study using fMRI

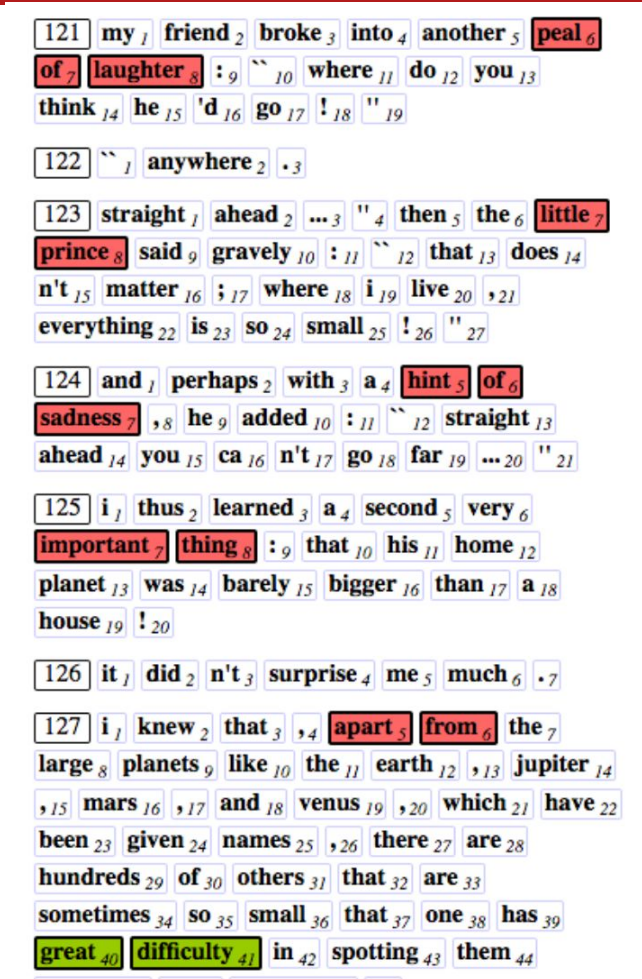


Fig. 1: Text annotated with MWEs (Bhattasali et al., 2017)

Introduction

- Use text attributes, correlate to real-time speech events and map them to observable brain processes

"Once when I was six years old, I saw a magnificent picture…"

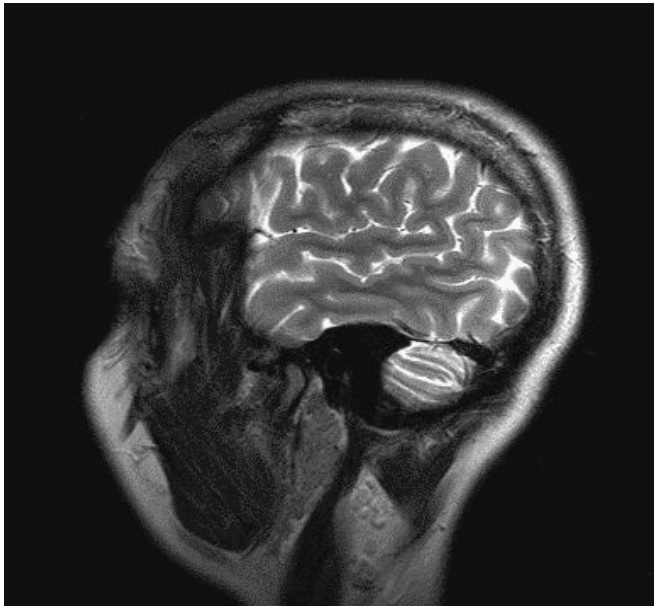


Fig. 2: Sample fMRI scan

Introduction

- Natural language comprehension relies on at least two cognitive processes:
  - Retrieval of memorized elements
  - Structural composition

- MWEs like *break the ice*, *boa constrictor, safe and sound*, *see to it, in spite of* can help us address the neural correlates of these processes.

- However, MWEs are a heterogeneous family of word clusters.

Roadmap

- Introduction
- ➢ **Research Questions**
- Background
- fMRI experiment
- Results
- Conclusion

Research Questions

Research Question 1:

Do MWEs with different levels of cohesiveness correspond to different brain areas?

Research Questions

Research Question 2:

Does comprehension of verbal MWEs implicate separate brain areas from non-verbal MWEs?

Roadmap

- Introduction
- Research Questions
- ➢ **Background**
- fMRI experiment
- Results
- Conclusion

Background

- Psycholinguistic studies have shown that MWEs are produced and understood faster than matched control phrases due to their frequency, familiarity, and predictability (Siyanova-Chanturia and Martinez, 2014)

- Eg. *bride and groom* vs. *groom and bride*
  salt and pepper vs. *pepper and salt*
  (Siyanova-Chanturia et al., 2011)

Background

- MWEs span across different grammatical categories
- Our dataset: 56% verbal MWEs; verbal idioms, verb participle constructions, light verb constructions, verb nominal constructions etc

(1) You must **see to it** that you regularly **pull out** the baobabs as soon as they can be told **apart from** the rose bushes to which they look very similar to when they are young.

(2) "**Good morning**", said the **little prince** politely, who then turned around, but saw nothing.

Background

- AM such as Pointwise Mutual Information (PMI; Church & Hanks 1990) can be used to capture the varying degrees of compositionality within MWEs:

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

- MWEs that receive a higher PMI score are seen as lexically more cohesive, which is interpreted as more noncompositional (less compositional)

Background

PMI	multiword expression receiving this score
26.59474426	heart skipped a beat
23.79983038	have nothing to do with
21.25998782	forehead with a handkerchief
21.17721316	burst into tear
20.17480668	once upon a time
20.15121667	boa constrictor
18.85209561	peal of laughter
-2.336733827	be order
-2.493268369	do calculation
-2.721901963	be object
-2.982215241	be hundred
-3.152845604	a well
-3.501675488	drink anything
-3.635409951	have plan

Table 1: MWEs with high & low PMI scores

Background

In summary:

- MWEs are processed differently from other phrases
- Can be distinguished based on grammatical category
- Their compositionality can be quantified with a metric like PMI

Roadmap

- Introduction
- Research Questions
- Background
- ➢ **fMRI experiment**
- Results
- Conclusion

fMRI Experiment

Dataset:

- The audio stimulus was Antoine de Saint-Exupery's *The Little Prince*, translated by David Wilkinson and read by Nadine Eckert-Boulet.

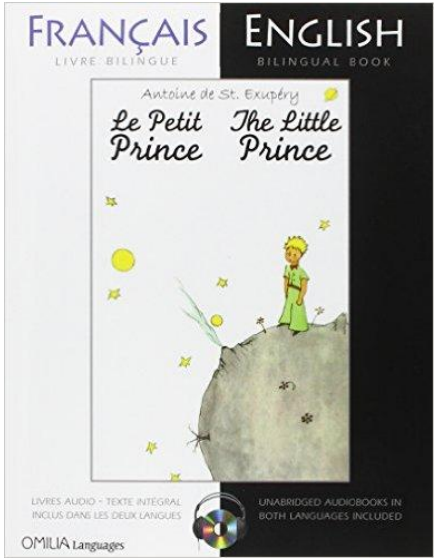


Fig. 3: Cover of *The Little Prince*

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

fMRI Experiment

Dataset:

- PMI scores based on corpus frequency counts from the Corpus of Contemporary English (Davies, 2008), and were calculated using `mwetoolkit` (Ramisch et al., 2010; Ramisch, 2012).

- The Stanford POS tagger and the NLTK POS tagger were used to annotate the words within the MWEs with their grammatical categories (Bird and Loper, 2004; Manning et al., 2014)

fMRI Experiment

Experimental Design:

- Participants (n=51, 32 female) were college-aged, right-handed, native English speakers

- Listened to *The Little Prince*'s audiobook for 1 hour 38 minutes across nine sections. (15,388 words total)
- Comprehension was confirmed through multiple-choice questions (90% accuracy, SD 3.7%).

fMRI Experiment

Statistical Analysis:

- The General Linear Model (GLM) typically used in fMRI data analysis is a time series linear regression (Poldrack et al., 2011).

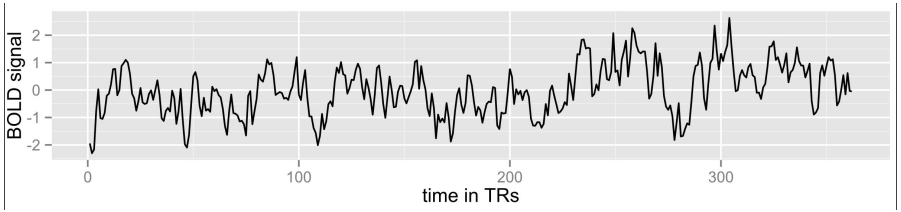


Fig. 4: Sample BOLD signal

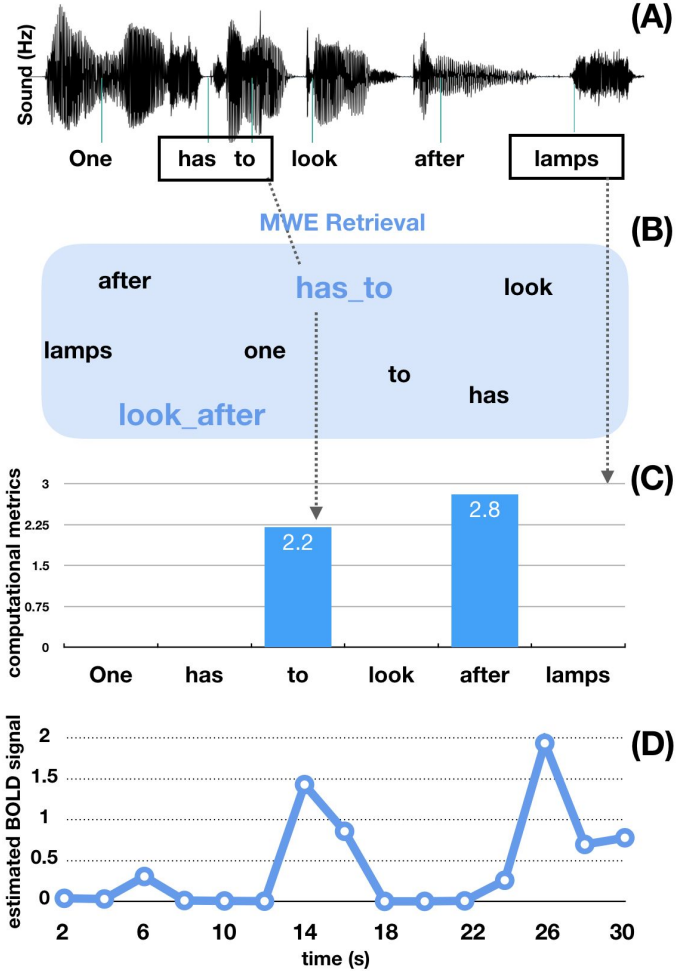


Fig 5: Pipeline of analysis adapted from Bhattasali et al., (in

fMRI Experiment

Analysis 1: Cohesiveness within MWEs

- 742 MWEs annotated with PMI scores
- Word rate
- Bottom-up parser action count
- Word frequency
- Intonation (Pitch)
- Acoustic Intensity (Volume)

Analysis 2: Verbal vs Non-verbal MWEs

- Verbal MWEs: 416/742 MWEs (56%)
- Non-verbal MWEs: 326/742 (44%)
- Word rate
- Bottom-up parser action count
- Word frequency
- Intonation (Pitch)
- Acoustic Intensity (Volume)

Roadmap

- Introduction
- Research Questions
- Background
- fMRI experiment
- ➢ **Results**
- Conclusion

Results

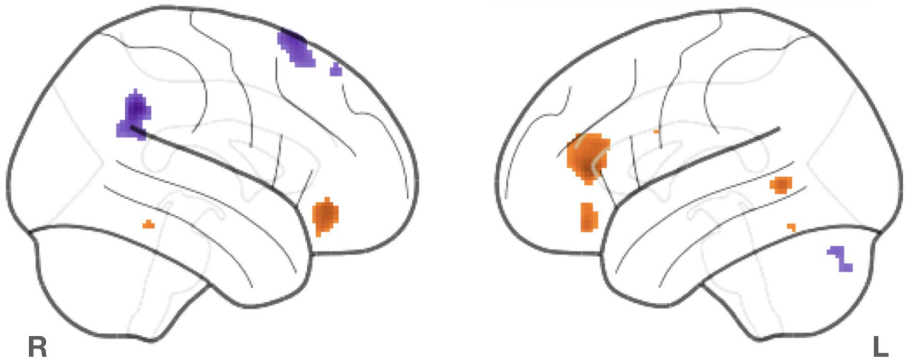


Fig. 6: Significant cluster for cohesion within MWEs after FWE correction for multiple comparisons with $p < 0.05$

Group-level results for Lexical Cohesion with MWEs:

- Increasing cohesiveness, as seen through positive activation with PMI (in purple), elicits the Precuneus and Supplementary Motor Area
- Decreasing cohesiveness, as seen through negative activation with PMI (in orange), correlates with activity in well-known nodes of the language network, such as Broca's area and the posterior Temporal Gyrus.

Results

Group-level results for Verbal MWEs vs Non-verbal MWEs:

- Verbal MWEs appear right-lateralized compared to non-verbal ones in IPL and in IFG triangularis.
- Non-verbal MWEs yielded a wider pattern of activation, including bilateral Supramarginal Gyrus extending to STG and right SMA.

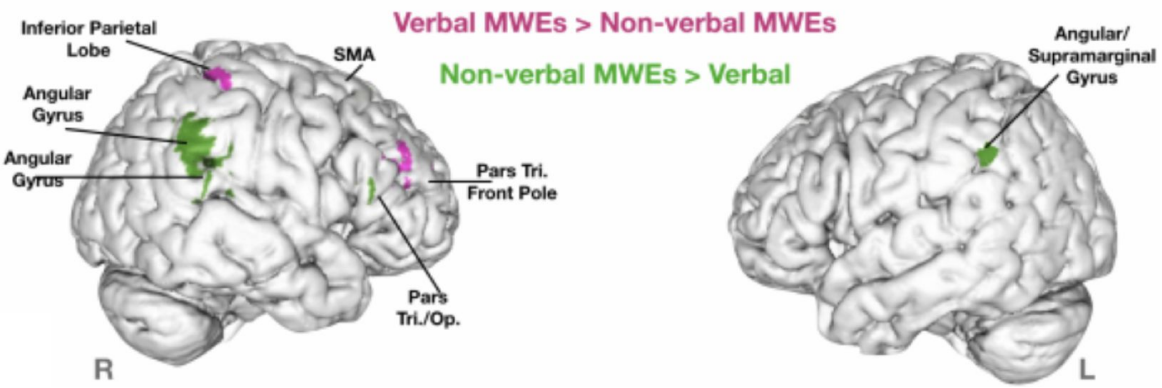


Fig. 6: Significant cluster for verbal & non-verbal MWEs after FWE correction for multiple comparisons with p <0.05

Roadmap

- Introduction
- Research Questions
- Background
- fMRI experiment
- Results

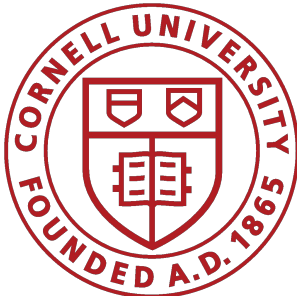
➢ **Conclusion**

Conclusion

- Provide neuroimaging evidence to illustrate that MWEs can be distinguished based on two different aspects:
  - Cohesiveness
  - Grammatical category: Verbal vs Non-verbal

Conclusion

- Repurpose PMI as an association measure to describe MWEs in terms of cohesion
  - shows that it is a cognitively informative metric to model cohesiveness and compositionality within word clusters in natural language.



Thank you!