# A Self-Paced Reading Study on Processing Constructions with Different Degrees of Compositionality

Giulia Rambelli, Alessandro Lenci, Philippe Blache

giulia.rambelli@phd.unipi.it

Università di Pisa

Aix*Marseille université Socialement engagée

## Introduction

- Many theoretical questions are still debated in the study of language comprehension: *How do speakers process compositional sequences? Which factors can facilitate language* comprehension?
- While the principle of compositionality is traditionally considered as the primary mean of explaining language processing, behavioral data support the idea that it is just a *default* option within a more complex scenario, where a series of noncompositional mechanisms can be used in processing.
  - analogy with stored exemplars, shallow processing, activation of a network of mutual expectations, etc.
- Studied *facilitation effects*:

1. idioms are read faster [Conklin and Schmitt, 2008] than transparent phrases and elicit a more positive electric signal in brain activity [Vespignani et al., 2010]

⇓

direct access to their holistic structure once recognized as an idiom

2. multi-word expressions are usually read faster than comparable sequences of lesser frequency [Arnon and Snider, 2010, Tremblay et al., 2011]

⇓

the more often a word is encountered, the more *entrenched* that representation is and the more easily it is retrieved [Bannard and Matthews, 2008]

- We assume facilitation effects are not limited to formulaic expressions but also occur when processing highly prototypical and yet compositional phrases
  - e.g., 'kick the ball' or 'chew the gum'
- Only [Jolsvai et al., 2020] compared both idioms ('on my mind') and frequent constructions ('is really nice') with respect to fragments ('know it gets').

## Experimental Hypothesis

The experiment is designed to examine reading times (RTs) of verb-noun constructions with a different degree of compositionality.

We compare 3 conditions:

1. **idiomatic expressions** (ID) 'bury the hatchet'
2. **compositional and highly frequent expressions** (HF) 'bury the treasure'
3. **compositional and low frequent expressions** (LF) 'bury the machete'

The novelty of this work is twofold:

1. we compare in the same experiment both idiomatic and high frequent expressions,
2. we do not deal with fixed multi-word sequences as lexical bundles but specifically with verb-argument constructions.

**Goal** Investigate whether idiomatic and frequent compositional expressions are processed in the same way: Are these processes founded on distinct mechanisms, or is there common access to both?
**Hypothesis** RTs are longer for compositional expressions than for idiomatic ones, and RTs are longer for infrequent expressions than for frequent ones.

## Acknowledgements

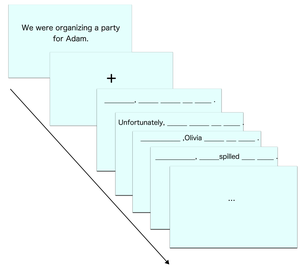
Research supported by grants ANR-16-CONV-0002 (ILCB) and the Excellence Initiative of Aix-Marseille University (A*MIDEX).

## Method

**Participants** 90 L1 English speakers from the US and Canada, aged 18 to 50 (M=29.6 ± 7.55). The experiment was delivered remotely, and participants were recruited using Prolific.

**Material** 48 VERB+determinant+NOUN idioms and corresponding high and low-frequency bigrams, objects matched by frequency and character length. Constraint: V-N association score for HF≥V-N association score for ID.

**Method** Each stimulus consisted of a context sentence presented for the participant to read in one instance and a sentence with the target phrase embedded, displayed word-by-word using the moving-window SPR paradigm.




| | Context | Precritical region -Critical region - Postcritical region |
|---|---|---|
| ID | Finn changed his life after his father's death. | All of a sudden he **kicked the habit** and stopped smoking cigarettes. |
| HF | It was the first match for Finn. | All of a sudden he **kicked the ball** into the net and won the match. |
| LF | That day Finn had completely lost his temper. | All of a sudden he **kicked the sister** of his best friend in the head. |

Table 1. Example of stimuli.

## Data Analysis

We analyzed the cleaned RTs of the phrase-final words using a linear mixed-effects model:

$$RT_{log} \sim Condition + Age + WordLength + VerbFreq_{log} + PosInList + (1|Sbj) + (0 + BigrFreq_{log}|Subj) + (1|Item)$$

| | Estimate | SE | df | t-value | p-value |
|---|---|---|---|---|---|
| Intercept(ID) | **5.569*** ** | 0.03747 | 96.63 | 148.632 | <0.001 |
| HF | 0.002 | 0.01364 | 138.6 | 0.17 | 0.865 |
| LF | **0.031*** | 0.01525 | 122.4 | 2.019 | 0.046 |
| age | **0.013**** | 0.004732 | 86.98 | 2.685 | 0.009 |
| PosInList | **-0.006*** ** | 0.0003511 | 3891 | -17.899 | <0.001 |
| WordLength | **0.017*** ** | 0.004516 | 136.9 | 3.863 | <0.001 |
| VerbFreqlog | **-0.009 *** | 0.003879 | 137.9 | -2.398 | 0.016 |

Table 2. Fixed effects for final model. *$p$<.05, **$p$<.01, and ***$p$<.001.

- Participants responded similarly to idioms (M: 262.17 ms) and frequent phrases (M: 262.69 ms) but more slowly to the unfrequent expressions (M: 270.42 ms).
- Changing the reference level with HF condition, there is still a statistical difference between condition HF and LF ($\beta$=0.028, p<0.1), even if it is smaller than the one observed above.
- The advantage for infrequent phrases was relatively small: maybe context sentences reduce the effort to interpret unpredictable expressions.

## Results

### Example of distribution of RTs (in ms)

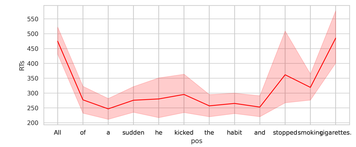



Figure 1. ID condition.

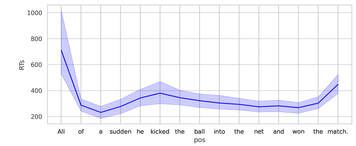



Figure 2. HF condition.

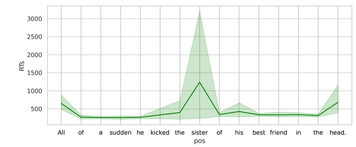



Figure 3. LF condition.

### Fixed effects

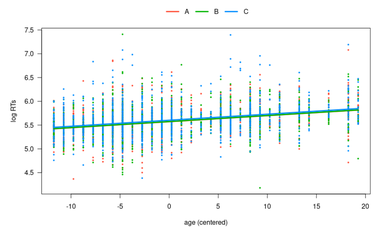



Figure 4. Older adults are slower than younger speakers.

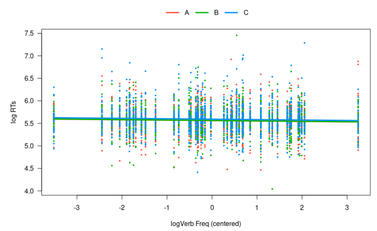



Figure 5. Objects preceded by a frequent verb are read faster.

## Discussion

Analysis reveals no difference between processing the figurative meaning of idioms and the compositional one of HF; there are facilitation effects in the comprehension of both expressions.
Two plausible explanations:

- both idioms and HF expressions are stored as unanalyzed wholes and directly retrieved once recognized, following usage-based perspective;
- processing HF relies on a co-activated network of representations operating with analogy-based mechanisms leading to sentence meaning construction; facilitation effects for ID and HF are similar but depend on different mechanisms.

The present experiment offers many exciting avenues for conducting additional exploratory and targeted analyses. Although the results of this study require further interrogation, we wish these findings contributed to the existing research in compositionality.

## References

[Arnon and Snider, 2010] Arnon, I. and Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of memory and language*, 62(1):67–82.

[Bannard and Matthews, 2008] Bannard, C. and Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological science*, 19(3):241–248.

[Conklin and Schmitt, 2008] Conklin, K. and Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied linguistics*, 29(1):72–89.

[Jolsvai et al., 2020] Jolsvai, H., McCauley, S. M., and Christiansen, M. H. (2020). Meaningfulness beats frequency in multiword chunk processing. *Cognitive Science*, 44(10).

[Tremblay et al., 2011] Tremblay, A., Derwing, B., Libben, G., and Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language learning*, 61(2):569–613.

[Vespignani et al., 2010] Vespignani, F., Canal, P., Molinaro, N., Fonda, S., and Cacciari, C. (2010). Predictive mechanisms in idiom comprehension. *Journal of Cognitive Neuroscience*, 22(8):1682–1700.