Are Frequent Phrases Directly Retrieved like Idioms?
An Investigation with Self-paced Reading and Language Models

Giulia Rambelli, Emmanuele Chersoni, Marco S.G. Senaldi, Philippe Blache, Alessandro Lenci
Processing Expressions with Different Degrees of Compositionality

**Idioms**

Andy *stole the thunder.*

- Ease in processing
  - facilitation effects in reading (Conklin & Schmitt, 2008; Titone et al., 2019)
  - more positive electric signal in brain activity (Vespignani et al., 2010)

- How are idioms represented in the lexicon?

**Frequent expressions**

Andy *stole the wallet.*

- Ease in processing
  - Lexical boundles (Tremblay et al., 2011)
  - 4-word expressions (Bannard and Matthews, 2008; Arnon and E. V. Clark, 2011)

- How frequent a sequence should be to be stored in the lexicon?

**non-compositional view**

(Swinney and Cutler, 1979; Cacciari and Tabossi, 1988, i.a.)

**hybrid models**

(Liben & Titone, 2008; Titone et al., 2019)
Research Question

**Question** Do IDIOMS and FREQUENT expression have the same facilitation effect in processing?

<table>
<thead>
<tr>
<th>Experimental Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. idiomatic expressions (ID)</td>
</tr>
<tr>
<td>2. compositional and highly frequent expressions (HF)</td>
</tr>
<tr>
<td>3. compositional and low frequent expressions (LF)</td>
</tr>
</tbody>
</table>

**Experiments** For direct objects in the 3 conditions, we compare

1. **Reading times** (RTs) collected by Self-Paced Reading (SPR) experiment
2. **Surprisal values** of Neural Language Models (NLMs)
Exp 1: Self-paced Reading (SPR)

Material 48 VERB+det+NOUN idioms and corresponding HF and LF bigrams -> 144 stimuli

Method Moving-window SPR paradigm

Participants 90 L1 English speakers from North America (M=29.6 ± 7.55). Delivered remotely.

Hypothesis RT(ID) < RT(HF) < RT(LF)
Exp 1: SPR Results

Result Participants responded similarly to idioms and frequent phrases but more slowly to the unfrequent expressions. There are facilitation effects in the comprehension of both figurative meaning of idioms and the compositional one of HF.

Explanations
1. same mechanism
2. facilitation effects are similar but depend on different mechanisms
Exp 2: Modeling RTs with NLMs

Material The same 144 stimuli sentences

Architectures

- autoregressive models -\(\rightarrow\) GPT2 (small, medium, large, xl)
- bidirectional models -\(\rightarrow\) BERT-base-case and T5-base
- recurrent neural networks (RNN) -\(\rightarrow\) tinyLSTM (Stephen et al., 2017); GRNN (Gulordava et al., 2018)

Method Measure the Surprisal (Hale, 2001; Levy, 2008) of a word

\[
\text{Surprisal}(w_i) = -\log P(w_i|\text{context})
\]

Hypothesis The Surprisal values are distributed in the same way of huma reading times (RTs)
Exp 2: GPT2 Surprisals

Results
1. All the GPT2 models produce Surprisal(ID) < Surprisal(HF)…
   … with the exception of GPT2-small

Wilcoxon Signed-Rank Test, *: p < .05
Exp 2: BERT, T5 & RNNs Surprisals

Results (continue)

2. BERT and T5 show a Surprisal(HF) < Surprisal(ID)
3. GRNN is similar to T5
4. Only tinyLSTM is comparable to human RTs
Exp 2: The Role of Context

**Question:** Are NLMs sensible to context?

**Method:** fed NLMs only with the target sentence

**Results:**
- RNN and bidirectional models produce the same Surprisal with or without the context sentence.
- GPT2 models have lower Surprisal scores giving a context sentence.
Contributions

• People read idioms and frequent compositional units at comparable speed
  • How are represented in the mental lexicon?
• Both idiomatic and frequent expressions are highly expected by GPT2 models, not by bidirectional models
  • GPT2-small has comparable to RTs -> inverse scaling effect (Oh and Schuler, 2022)
• Context seems to affect little or not at all the Surprisal scores
Find more info in our paper or get in touch at the MWE 2023 on May 6th!

Email: giulia.rambelli4@unibo.it
Twitter: @g_rambelli