Holistic and Compositional Representations in Multiword Expressions

Wednesday Bushong[1,2], Emily Morgan[1], Melodie Yen[3], Roger Levy[1]

[1]Department of Linguistics, University of California San Diego [2]Department of Brain and Cognitive Sciences, University of Rochester [3]Department of Linguistics, University of Arizona

Introduction

Usage-based theories of language claim that phrases longer than one word might be represented explicitly in long-term memory (e.g., [1]).

Like many other structures, multiword expressions show frequency effects in processing and production [2,3,4]. But this isn't necessarily evidence that phrases longer than one word are explicitly represented. For example, it's possible that some other feature of the phrase (e.g., semantic or event representation) is easier to process, with frequency a natural consequence of this.

How can we test whether multiword expressions are explicitly stored as holistic units?

Binomial Expressions

Phrases of the form "X and Y" / "Y and X"

Benefits:

- Same lexical items
- Identical formal semantic & syntactic structure
- Only difference between orders is *exact string of words*

If frequent multiword expressions can be stored as holistic units, we should see a word-order specific priming effect in frequently attested multiword expressions but *not* novel expressions.

Materials

48 attested binomial expressions (lifetime exposures: min 31, mean 2,000)
48 novel binomial expressions (lifetime exposures: 0)

	"X and Y" order	"Y and X" order
Attested	mix and match	match and mix
	family and friends	friends and family
	vitamins and minerals	minerals and vitamins
Novel	hesitate and readjust	readjust and hesitate
	bishops and seamstresses	seamstresses and bishops
	vegetables and kale	kale and vegetables

Design & Procedure

Participants: 207 native English speakers from Amazon Mechanical Turk were paid $4.00 each for their participation.

Training Phase: Participants read items in one item or the other in three different sentence contexts. Each item was presented in the same order in all training trials.

Testing Phase: Participants read (in self-paced reading) each trained item once in either the same (*match*) or different (*mismatch*) order as in training. They also read items that were not seen in training to establish a baseline.

	Match	Mismatch	Untrained
Training	"... mix and match ..."	"... match and mix ..."	*(No sentences with*
	"... mix and match ..."	"... match and mix ..."	*{mix and match}*
	"... mix and match ..."	"... match and mix ..."	*in either order)*
Testing	"... mix and match ..."	"... mix and match ..."	"... mix and match ..."

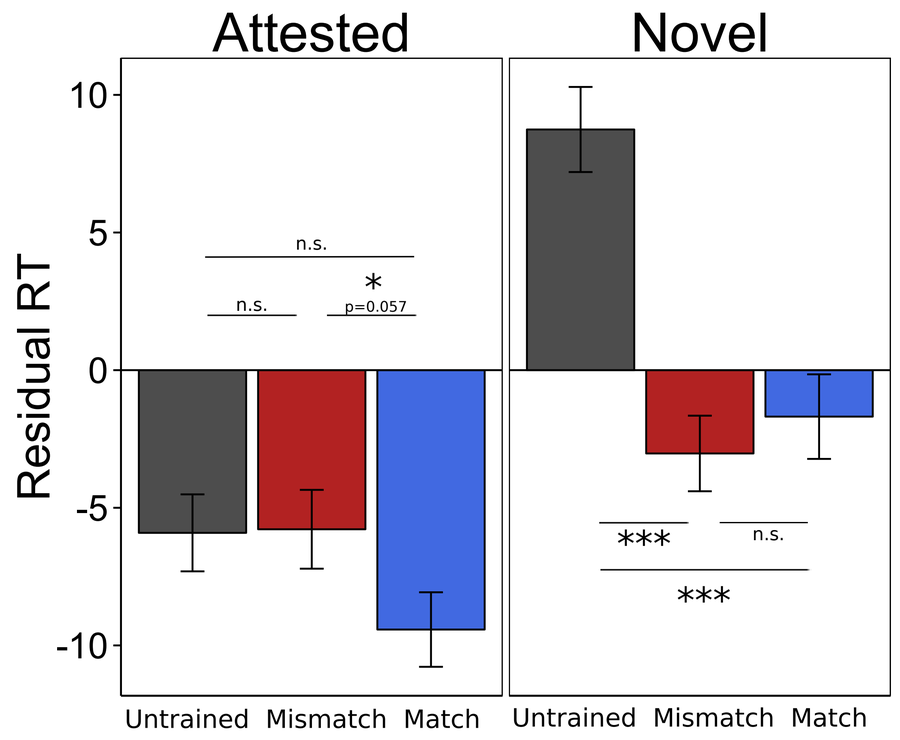
These conditions are repeated for the other testing order, and for novel items.

References

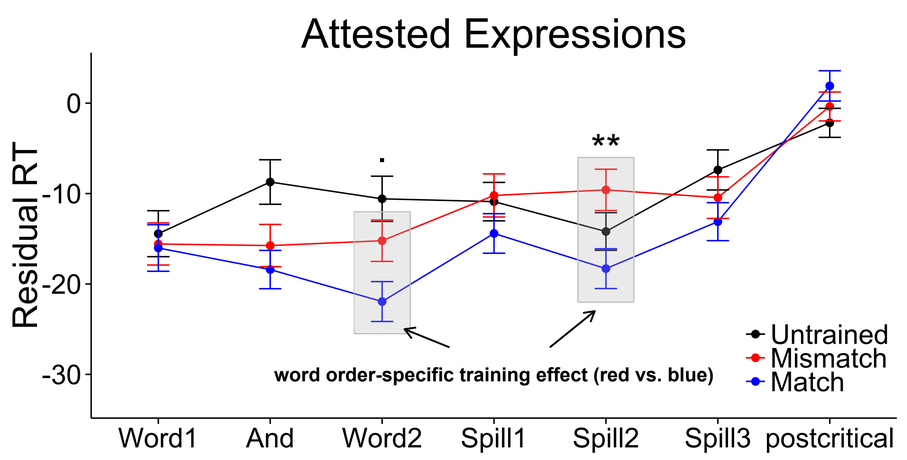
[1] Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, 711-733.
[2] Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, 62(1), 67-82.
[3] Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. J. (2011). Seeing a phrase time and again matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(3), 776-784.
[4] Arnon & Cohen Priva, U. (2013). More than words: The effect of multi-word frequency and constituency on phonetic duration. *Language and Speech*, 56(3), 349-371.

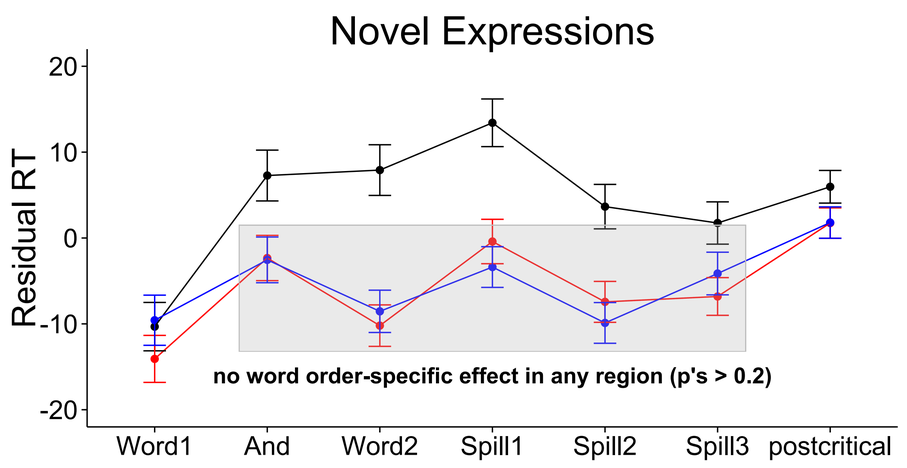
Results

Per-Word Average Across Critical Regions



Word-by-Word





– Attested expressions show a word order-specific training benefit, with no overall gain from training as compared to no training.

– Novel expressions show only an overall gain from training in either order, with no word order-specific training benefit in any region.

Conclusions

1. Attested expressions show a word order-specific training benefit, but no overall training benefit independent of word order. This suggests attested multiword expressions can be stored and accessed as **holistic units**.

2. Novel expressions did not show a word-order specific training benefit, suggesting they are not stored and accessed as holistic units.

Future Work

Why do novel expressions show a larger overall training benefit?

- Larger increase in experience with novel items? ($0 \rightarrow 4$ exposures more important than $2{,}000 \rightarrow 2{,}004$)
- Lexical frequency effect? Lower frequency words show larger repetition priming effects, and on average novel expressions contained lower frequency words than attested.
- Strength of lexical priming might be modulated by expression frequency?