A Self-Paced Reading study on Processing Constructions with different degrees of Compositionality

Giulia Rambelli[1,2], Alessandro Lenci[1], Philippe Blache[2]
[1]University of Pisa, [2]Aix-Marseille university

Introduction. Our research aims at challenging the classical principle of compositionality in sentence processing. While traditional linguistics considered the principle of compositionality as the primary mean of explaining linguistic productivity, we argue this is just a default option into a more complex scenario, where a series of noncompositional mechanisms (analogy with stored exemplar, shallow processing, activation of a network of mutual expectations..) can take place in processing not only formulaic language expressions, but a larger set of expressions with a completely transparent meaning. Psycholinguistic literature has studied two noncompositional cases: idiomatic processing, associated with faster reading time [1] and a more positive electric signal in brain activity [2] to transparent phrases, and frequency effects, i.e., multiword sequences are usually read faster than comparable sequences of lesser frequency [3,4]. We assume that facilitation effects are not limited to formulaic expressions but are present in processing compositional phrases. Consequently, we implemented a Self-Paced Reading experiment to compare the Reading Times (RTs) of constructions with a different degree in compositionality: idiomatic expressions (ID), compositional highly frequent expressions (HF), and compositional low-predictable expressions (LF). To the best of our knowledge, no previous work had compared both idioms and frequent constructions, except for [5], but our element of novelty is the focus on predicate-argument construction. Given the previous literature, we hypothesized that *RTs are longer for compositional sentences than for idiomatic sentences, and RTs are longer for infrequent phrases than for frequent ones.*

Design. We selected 48 idiomatic VERB+determinant+NOUN phrases and corresponding high-frequency and low-frequency bigrams with the same verb. A stimulus consisted of: a context sentence presented for the participant to read in one instance and a sentence with the target phrase embedded (Table1), displayed word-by-word using the moving-window SPR paradigm [6]. The stimuli were split into three counterbalanced lists randomly initialized at each time. The experiment was delivered remotely, and participants were recruited using Prolific. We collected responses for 90 subjects from the United States and Canada, all self-reported L1 speakers of English aged between 18 and 50.

Data Analysis. We removed the outliers and examined the RTs of the last word of phrases using linear mixed models. *Condition*, *Age*, *WordLength*, *VerbFrequency,* and *PositionInList* were entered in the models as fixed effects; *Subject* and *Item* were treated as random effects with a by-subject random slope for *BigramFrequency*. RTs' difference between ID and HF turned out to be not statistically significant (Table2), while it is statistically different between ID and LF. Changing the reference level with HF condition, there is still a smaller statistical difference between HF and LF. Moreover, we observed that 1) older adults are slower than younger speakers, and 2) RTs at the end of the experiment are faster than at the beginning.

Discussion. Analysis reveals no difference between processing the figurative meaning of idioms and the compositional one of HF; there are *facilitation effects* in comprehension of both expressions. Even if this observed measure cannot say what is happening at the brain level, it opens to a wide discussion about underlying mechanisms in language processing. It may suggest that HF expressions are stored as unanalyzed wholes and directly retrieved once recognized as idioms, following usage-based models [7,8], or there could be a co-activated network of representations, and other cognitive mechanisms could act (e.g., analogy) instead of relying on the principle of compositionality. RTs for infrequent phrases were significantly slower, even if the advantage was relatively small: we presume that information introduced in context sentences reduces the effort to interpret unpredictable expressions.

Table 1 Example of stimuli. Context sentence is useful to bias the figurative meaning of each idiom for cases in which the literal interpretation is possible.

Condition	Context sentence	Target sentence (word-by-word)
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, h*e **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 2 Results of mixed-effects model for the RTs of the phrases last word.

Fixed Effects					
	Estimate	**SE**	**df**	***t*-value**	***p*-value**
Intercept(ID)	5.569	0.03747	96.63	148.632	0.000000***
HF	0.00232	0.01364	138.6	0.17	0.865247
LF	0.0308	0.01525	122.4	2.019	0.045659*
age	0.0127	0.004732	86.98	2.685	0.008691**
PosInList	-0.006285	0.0003511	3891	-17.899	0.000000***
WordLength	0.01744	0.004516	136.9	3.863	0.000172***
$VerbFreq_{log}$	-0.009301	0.003879	137.9	-2.398	0.017840*

Acknowledgements Research supported by grants ANR-16-CONV-0002 (ILCB) and the Excellence Initiative of Aix-Marseille University (A*MIDEX). We would like to thank Marco Senaldi and Christelle Zielinski for their helpful comments and support.

References

[1] Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers?. *Applied linguistics*, *29*(1), 72-89.
[2] Vespignani, F., Canal, P., Molinaro, N., Fonda, S., & Cacciari, C. (2010). Predictive mechanisms in idiom comprehension. *Journal of Cognitive Neuroscience*, 22(8), 1682-1700.
[3] Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of memory and language*, 62(1), 67-82.
[4] Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language learning*, *61*(2), 569-613.
[5] Jolsvai, H., McCauley, S. M., & Christiansen, M. H. (2020). Meaningfulness beats frequency in multiword chunk processing. *Cognitive Science*, 44(10).
[6] Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of experimental psychology: General*, 111(2), 228–238.
[7] Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford University Press on Demand.
[8] Bybee, J. (2010). *Language, usage and cognition*. Cambridge University Press.
[9] Bannard, C. and D. Matthews (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.