Statistical Learning and Language

(in spite of arbitrariness)

Davide Crepaldi, Simona Amenta and Marco Marelli
[davide.crepaldi@sissa.it]
[https://lrlac.sissa.it]

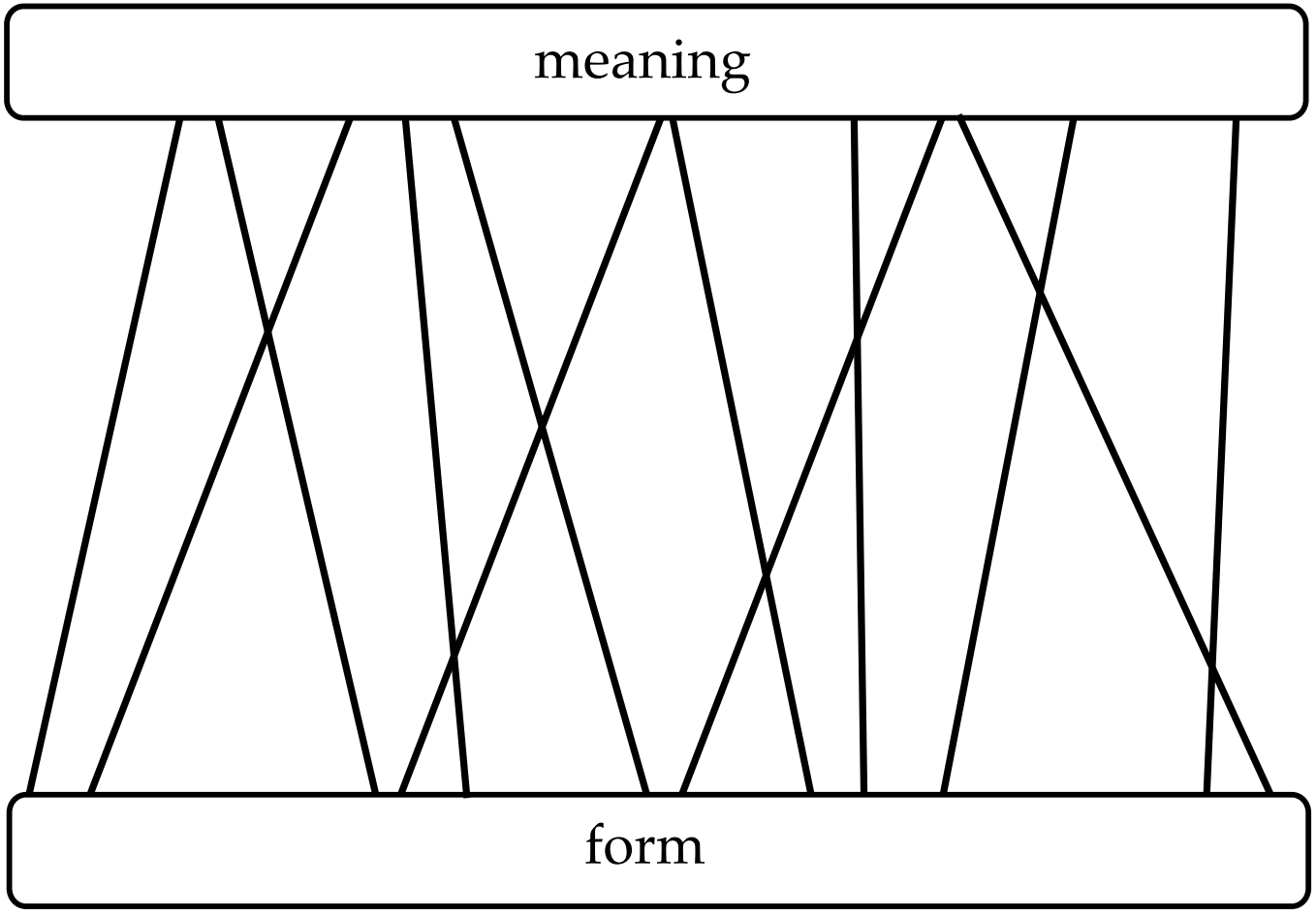
SISSA, Trieste, Italy
Universiteit Gent, Belgium
Università di Milano Bicocca, Italy

EvoLang XII
Toruń, Poland, 17 April 2018

Language as a symbolic system

table

Language as a symbolic system



Non-random co-occurrence

- Mary has baked biscuits and muffins for the family on Christmas day
- Ball the idea after flew nice a chemist

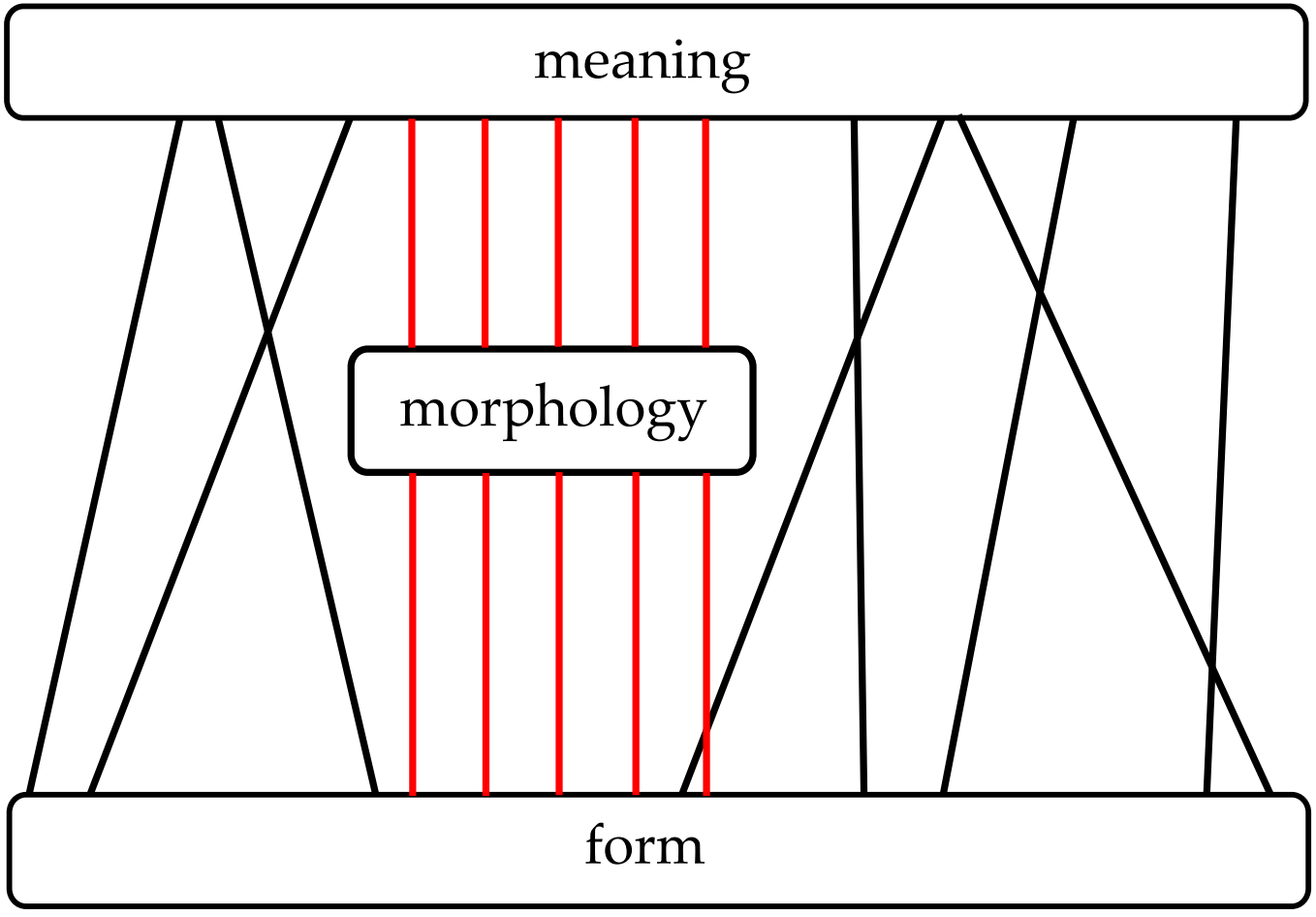
Non-random co-occurrence

		Number of co-occurrences
dog	cat	7927
	walk	2356
	collar	612
	fox	252
	bear	73
	pizza	10

Islands of regularity **within** words

- Dealer
- Payer
- Seeker
- Learner
- Killer
- Holder

A breach into arbitrariness



Far away from perfect predictability

Vessel
Jewel
Tunnel

Whisper
Never
Corner
Mother

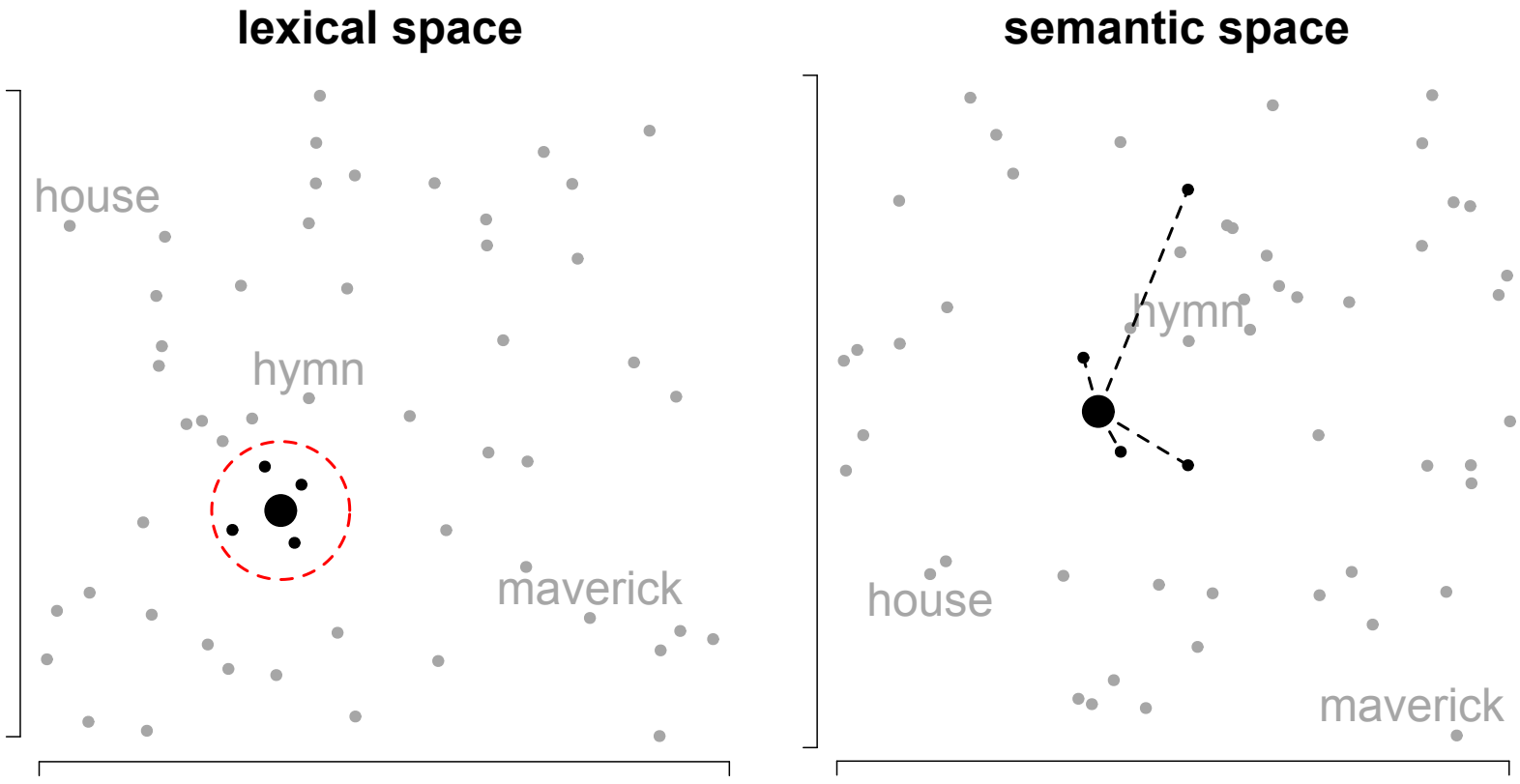
Darker
Larger

The core question

Despite its **fundamental arbitrariness**, language does feature **regularities** that are **probabilistic** in nature.

Does the brain pick up on these regularities?

Orthography–Semantic Consistency (OSC)



OSC, mathematical formulation

$$OSC(t) = \frac{\sum_{x=1}^{k} f_{r_x} \cos(\vec{t}, \vec{r_x})}{\sum_{x=1}^{k} f_{r_x}}$$

(Marelli et al., 2015)

A morphological take

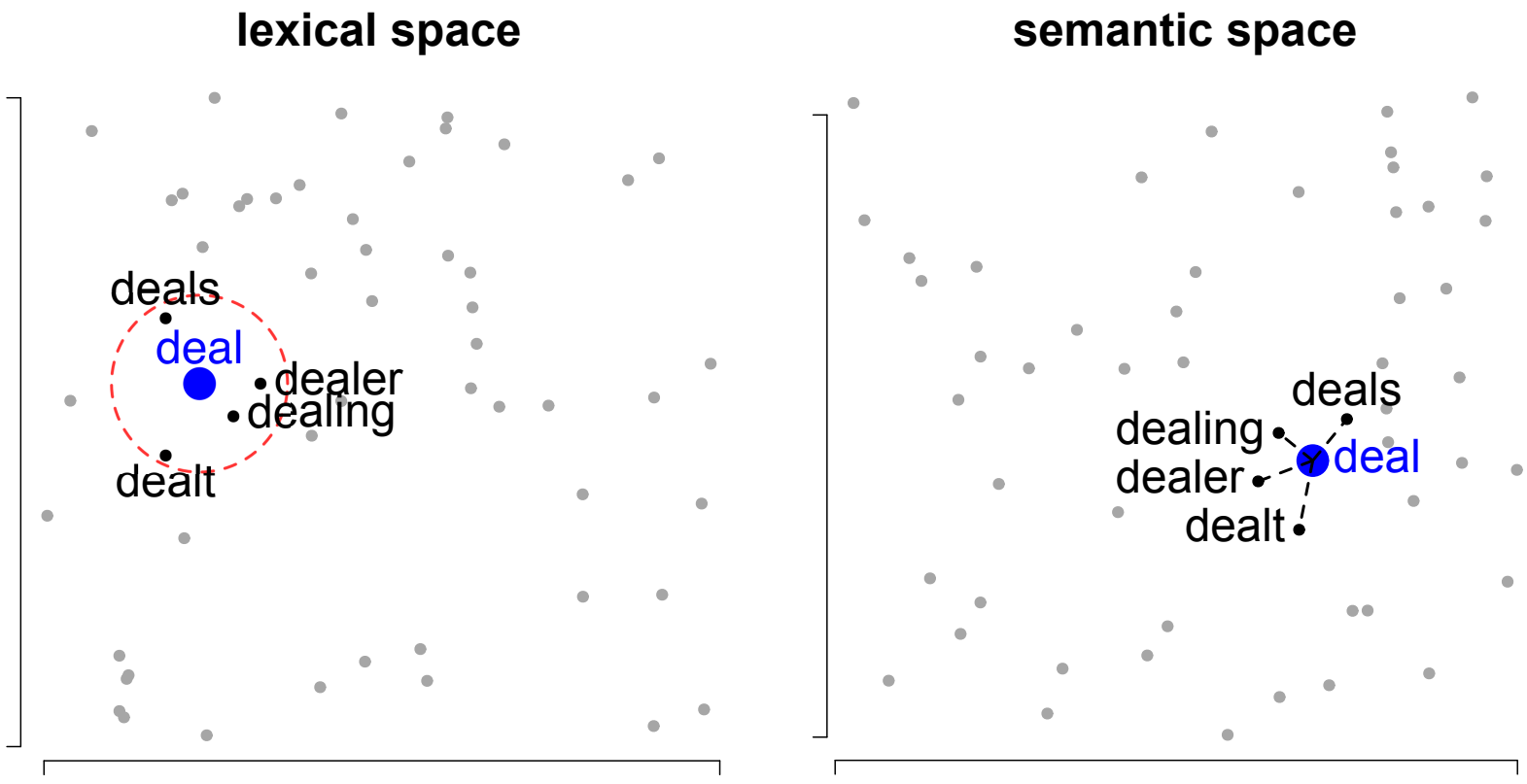
CORN

- corns
- corner
- cornwall
- cornish
- ...

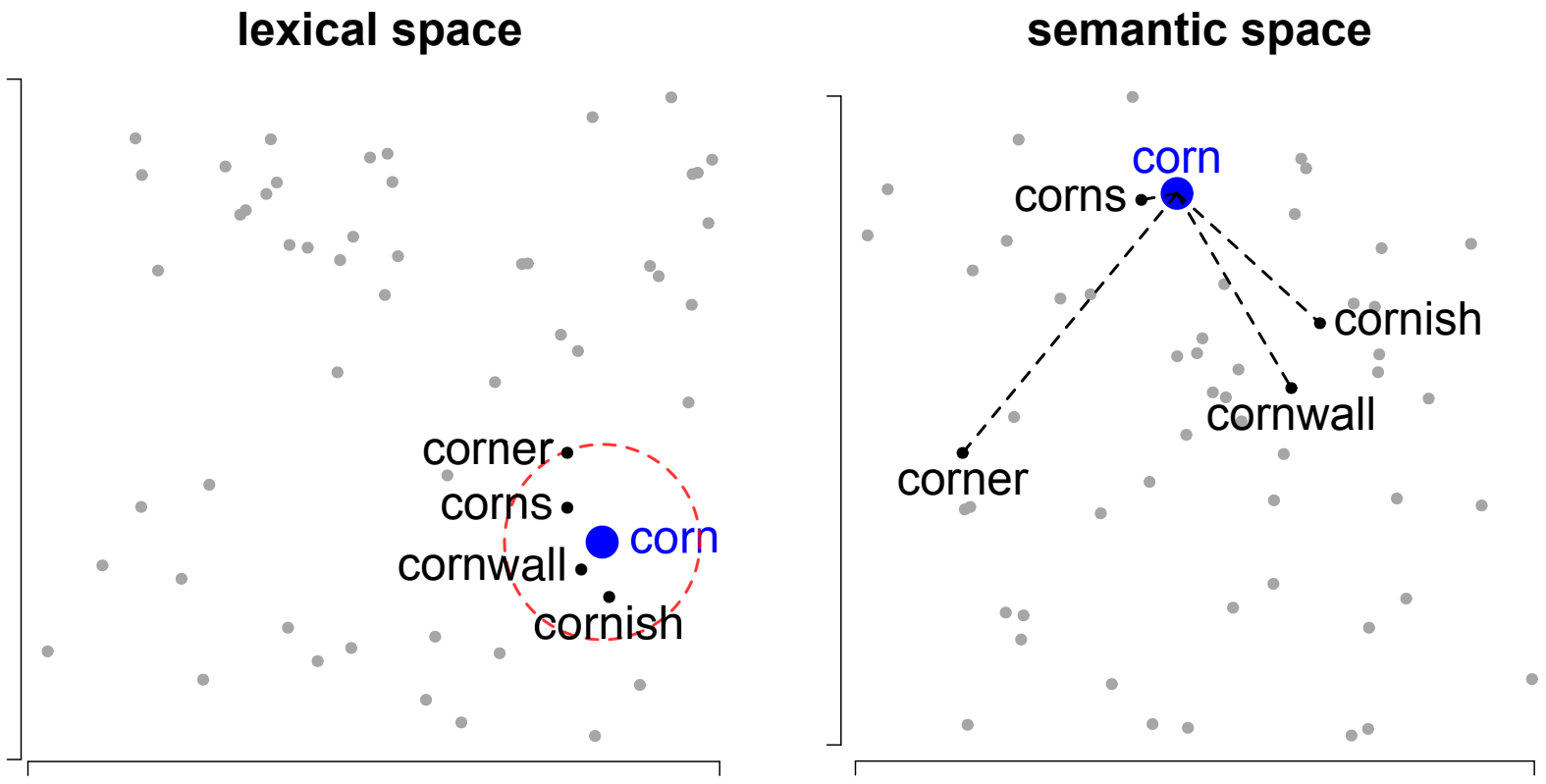
but not

- torn
- cork
- ...

A consistent space



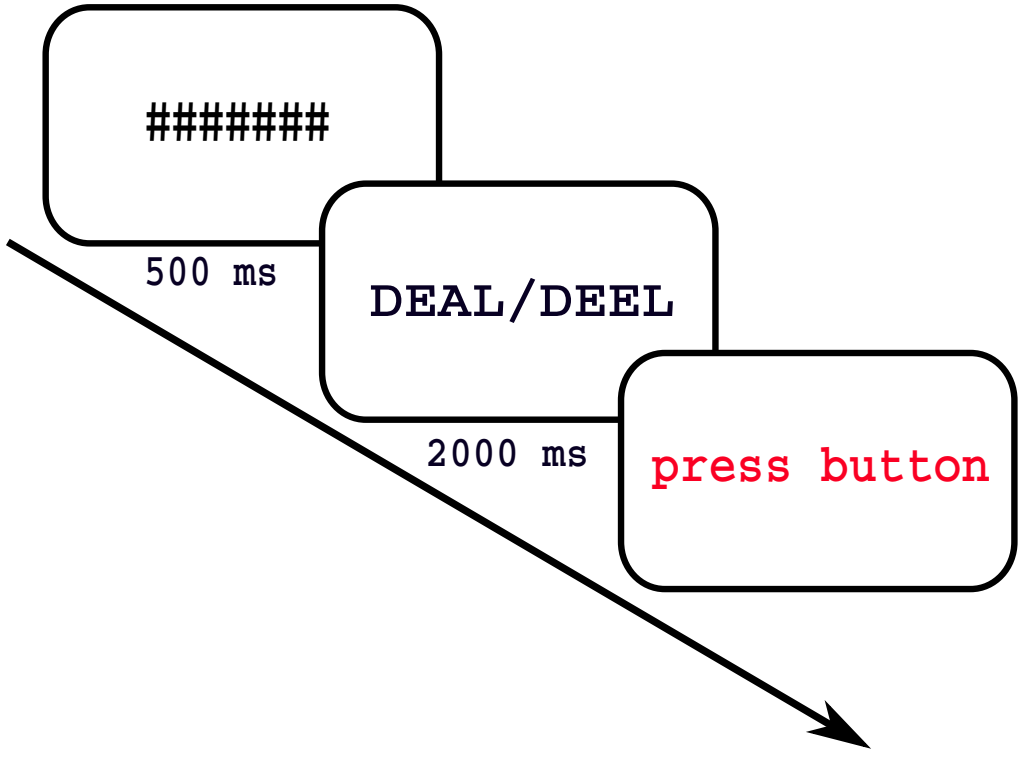
An inconsistent space



Three experiments

- Identification times on isolated words
- Proficiency in L2
- Brain electrophysiology during sentence reading

Lexical decision



(Balota et al., 2007)

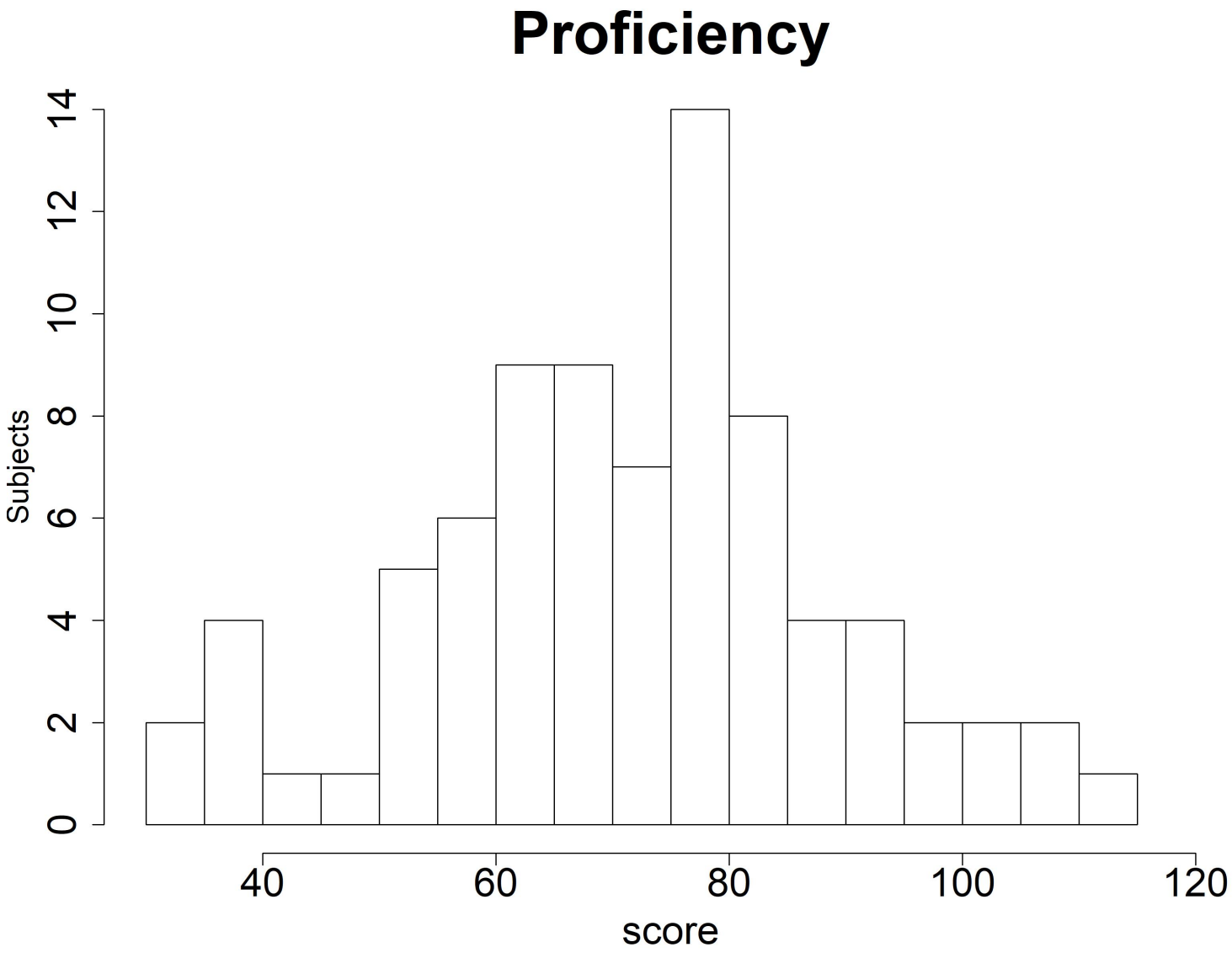
OSC gets unique variance

Table 6. *Results of the regression analysis on the lexical decision latencies extracted from the BLP for a large set of random words*

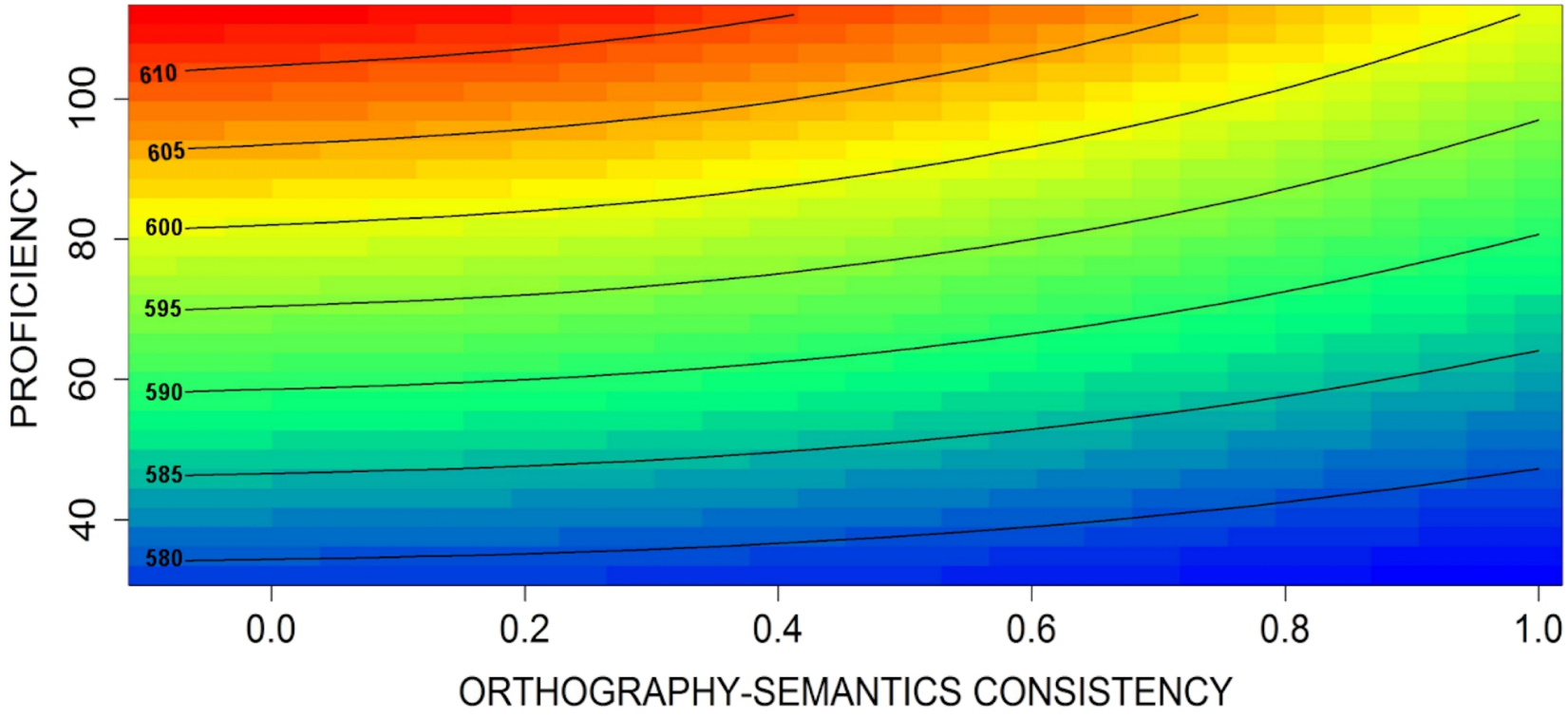
	Estimate	*Std error*	t *value*	p *value*
Intercept	6.5922	.0109	602.89	.0001
Word frequency	−0.0308	.0009	33.41	.0001
Word FS	−0.0041	.0021	1.97	.0495
Word length	0.0035	.0013	2.74	.0061
OSC	−0.0254	.0066	3.84	.0002

(Marelli et al., 2015)

OSC tracks language learning

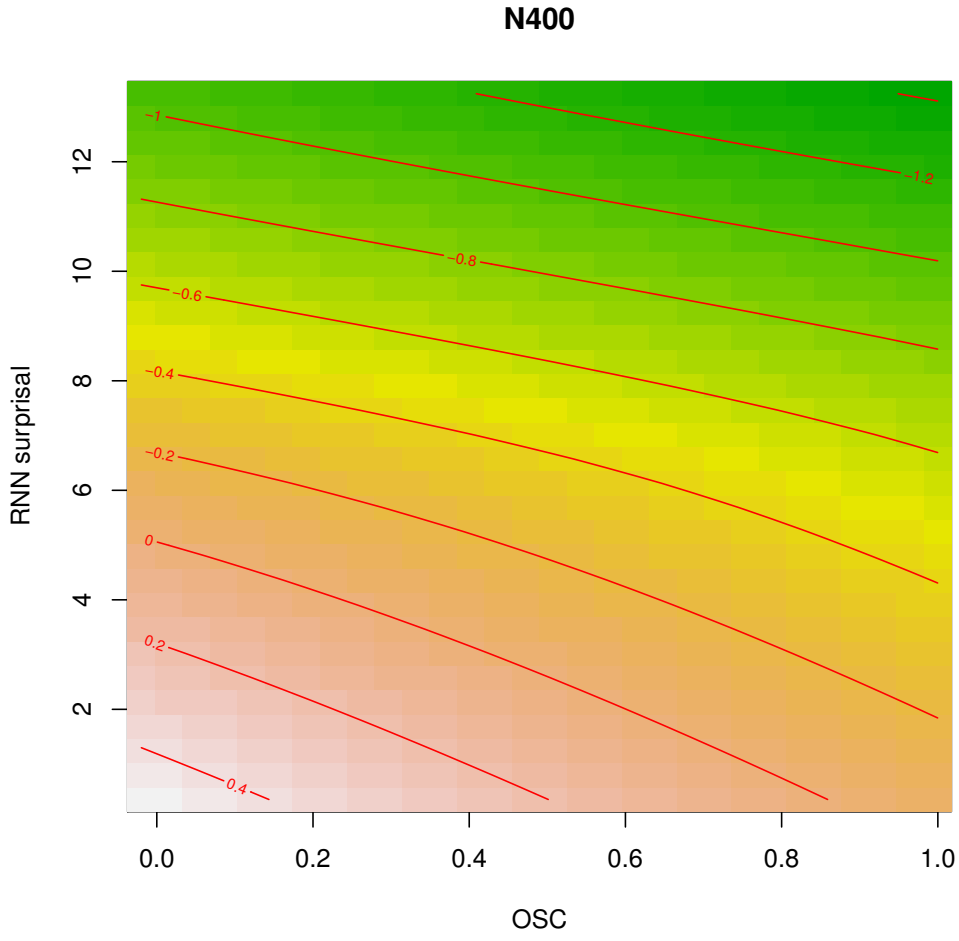


OSC tracks language learning



(Viviani and Crepaldi, 2018)

OSC explains brain electrophysiology



(Crepaldi et al., 2018; Frank et al., 2015)

Sum up of the data

- Words are recognized more quickly if they live in a consistent part of the lexical-semantic network
- Peole become more sensitive to consistency as their proficiency in a language grows
- Consistency affects word identification on-line during the processing of connected speech

The broader message

The brain does code for consistency in the lexical system

Predictive coding?

Implications

Morphology

From a finite set of well-defined objects and operations, to part of a more general and probabilistic form-to-meaning mapping effort carried out by the brain

Language Evolution

Brain counterpart of the emergence of structure in the cultural evolution of language (Kirby et al., 2008)

Language

General-purpose learning mechanisms subserve linguistic processing (Ellison, 2013)

Acknowledgments

Statistical Learning and Language

(in spite of arbitrariness)

Davide Crepaldi, Simona Amenta and Marco Marelli
[davide.crepaldi@sissa.it]
[https://lrlac.sissa.it]

SISSA, Trieste, Italy
Universiteit Gent, Belgium
Università di Milano Bicocca, Italy

EvoLang XII
Toruń, Poland, 17 April 2018

Scan this to get this talk on your phone

References

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. I., Kessler, B., Loftis, B., Neely, J., Nelson, D. L., Simpson, G. B., and Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, 39:445–459.

Crepaldi, D., Budinich, L., Amenta, S., and Marelli, M. (2018). Orthography??to–semantics consistency in sentence reading. *In preparation*.

Ellison, T. M. (2013). *Categorisation as topographic mapping between uncorrelated spaces*, pages 131–141. Springer.

Frank, S. L., Otten, L. J., Galli, G., and Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and Language*, 140:1–11.

Kirby, S., Cornish, H., and Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences USA*, 105:10681–10686.

Marelli, M., Amenta, S., and Crepaldi, D. (2015). Semantic transparency in free stems: The effect of Orthography–Semantics Consistency on word recognition. *Quarterly Journal of Experimental Psychology*, 68:1571–1583.

Viviani, E. and Crepaldi, D. (2018). How do we decompose foreign words? *In preparation*.