Statistical Learning and Language (in spite of arbitrariness?)

Davide Crepaldi

[davide.crepaldi@sissa.it]
[http://lrlac.sissa.it]

International School for Advanced Studies (SISSA), Trieste Language, Learning and Reading lab

Beyond Reading, Venezia, 22–23 September 2017





A little game to start off

catampo



Arbitrariness

- elephant
- table
- heat
- drum

Arbitrariness. Really?

- elephant
- table
- heat
- drum
- preheat
- juicer
- (noncatampo)
- bioweapon
- guesstimate
- (quirinarie)

The core idea

Morphology* has created probabilistic regularities in language form, and form-to meaning mapping. The brain codes for these regularities, and uses them during processing.

Statistical Learning in Language

Eye Tracking in Children Learning to Read

The experiment, but not so much of

Natural reading

- Stories (=connected text)
- Just read and understand (=no strange task to carry out)

Eye tracking

- Many children, create a database to share
- Across a natural spectrum of age
- Across a natural spectrum of reading proficiency
- Check sensitivity to statistical regularities

For today

- Data from 22 kids (out of the 80 tested so far)
- nGrams
 - ► ALBERO:
 - 2grams: AL, LB, BE, ER, RO
 - 3grams: ALB, LBE, BER, ERO
 - Agrams: ALBE, LBER, BERO
 - Average nGram frequency across whole words



- School trip
- The scientist gathers data, the kids gather experience
- SISSA Medialab
- 7 sessions, 140 kids in total







 1745 tokens, from 728 different words, across 12 short stories

Participant sample



Raven Score

Frequency and length



Age effects



Early processing?



nGrams effects



Grain size



To wrap up

- Frequency effects (which is statistical learning) in very young kids, and in early measures of processing
- nGram frequency seems to affect eye movements in children
- Children seem to track better the stats of larger chunks (jumping to lexicality?).

A new approach to reading

- Scripts can be seen as fully–fledged visual systems
- They can be studied as such (without language)
- The way we learn to deal with them can be captured through statistical learning
- The way we learn to map them onto language can be captured through statistical learning

A new approach to reading

- Scripts can be seen as fully–fledged visual systems
- They can be studied as such (without) language
- The way we learn to deal with them can be captured through statistical learning
- The way we learn to map them onto language can be captured through statistical learning

Masked Priming and Proficiency in L2

Masked priming



Experimental design

- piattino-PIATTO
- fontina–FONTE
- samba-SAMBUCA

Age/method of acquisition

- What age were you first exposed to English?
- Were you primarily exposed to ENG at home or in school?
- Please rate the relative dominance of ITA vs. ENG in your current everyday life experience
- Do you speak any other language other than IT and ENG?

Profiency

- Phonemic discrimination
- Phonemic fluency
- Spelling to dictation
- Vocabulary
- Morphological awareness
- Oral comprehension
- Reading comprehension

Masked priming overall, L1



Masked priming overall, L2



L2 priming and proficiency



And what about Age of Acquisition?

It just doesn't work

- L2 masked priming is (very) different from L1 masked priming
- L2 masked priming is modulated by proficiency, not much by AoA
- ► For low-proficiency L2, all about letter orthography
- As proficiency increases, form priming vanishes and morphological priming emerges...
- ... but no morpho-orthography for L2

Acknowledgments

- Valentina Pescuma, Eva Viviani, Maria Ktori, Marijana Sjekloća, Francesca Franzon (SISSA), Benedetta Cevoli (now at RHUL), Eleonora Lomi (now at UCL).
- Valentina Parma (SISSA) and Simona Cerrato (SISSA Medialab).





Statistical Learning and Language (in spite of arbitrariness?)

Davide Crepaldi

[davide.crepaldi@sissa.it]
[http://lrlac.sissa.it]

International School for Advanced Studies (SISSA), Trieste Language, Learning and Reading lab

Beyond Reading, Venezia, 22–23 September 2017



