Learning Language Statistics when World Statistics are Equal
Michael S. Amato, Jon A. Willits, Maryellen C. MacDonald, Rachel S. Sussman
Language and Cognitive Neuroscience Laboratory, University of Wisconsin - Madison

Motivation
Many current theories claim that language behavior is shaped by the
distributional structure of our experiences. But...

Example sentence: Roger swept the floor with a broom.

Q: How do these 2 sources of knowledge inform language comprehension?
a) only world knowledge
b) only language statistics knowledge
c) representations formed by combining the two
d) knowledge of both separately accessible

Goal of study:
Independently manipulate language and world statistics
Understand how these information sources interact for language behavior

Prior Studies
Andrews, Vigliocco, & Vinson (2009)
Computational model trained on both types of information better
matched behavioral similarity measures
Argued semantic representations derived from optimal combination
Amato, Willits, & MacDonald (2009)
Different priming effects from identical verb in different aspectual forms
Type of priming task moderated effect
Both studies highlight importance of both sources of information
Different views of how that information is organized as long term knowledge

Current Studies
Adopt position d above: knowledge of both separately accessible
Hypotheses
1) manipulating language statistics will affect behavior
2) manipulating event statistics will affect behavior
3) effects of each will interact with task

Goal:
Independently manipulate language and world statistics
Understand how these information sources interact for language behavior

Method - Training
4 novel verbs
15 training trials each (60 total)
On each trial, simultaneous presentation of:
visual sentence
audio sentence
video of event being performed
Manipulated instrument bias in LANGUAGE or EVENT

Method - Test
Self-paced reading on 16 sentences
Contained temporarily ambiguous "with" phrases
Randomly ordered, interleaved with 82 fillers
Followed by offline surveys

Independently Manipulated Instrument Bias of verbs & events

Experiment 1 - Language Manipulation

| trial type A | Mary fleeked the horse with tongs because her sister thought it was funny. | 100% Instrument Mention Verbs | 20% Instrument Mention Verbs |
| trial type B | Mary fleeked the horse because her sister thought it was funny. | 0% Instrument Mention Verbs | 80% Instrument Mention Verbs |

Prediction: Survey asking about verbs will reflect training biases; survey asking about events will not.

Experiment 2 - Event Manipulation

| trial type A | Mary fleeked the horse because her sister thought it was funny. | 100% Instrument Mention Events | 20% Instrument Mention Events |
| trial type B | Mary fleeked the horse because her sister thought it was funny. | 0% Instrument Mention Events | 80% Instrument Mention Events |

Prediction: Survey asking about events will reflect training biases.

Overall Prediction: Graded effects during self-paced reading for both manipulations.

Experiment 1 Survey1:
knowledge of verbs
If you were to talk about [VERB]ing some object, how likely would you be to mention the thing used to [VERB] it with?

Experiment 2 Survey:
knowledge of events
How necessary is it for a person to use some sort of tool to [VERB] with, as opposed to their bare hands?

Experiment 1 Survey Results
Ratings in language-focusing survey correlated with distributions in training F(1,204)=3.72, p=.06.
Ratings in event-focusing survey did not.

Experiment 2 Survey Results
Robust learning of event instrument bias, t(53)=7.80, p<.001.

Self Paced Reading Times
Both experiments

Self-Paced Reading Results
Clear effect of language manipulation in ambiguous "with" phrase F(1,210) = 17.33, p<.001
All or nothing- verbs with 20% bias patterned with 100%
Participants with 0% language bias faster in ambiguous region
No clear effect of event manipulation detected in this online task

Discussion
Participants readily learned language statistics.
Used them during online comprehension.

Information about a verb’s tendency to mention an instrument in language was not encapsulated with the corresponding event’s tendency to be performed with an instrument.
Evidenced by differences in the survey results.

Supports a model of human knowledge structured by how it is learned. Argues against a single optimal representation of concepts.

Demonstrates that language users have access to separate statistics for verbs and the events they label, and can flexibly draw on those different sources of information.

References

Funding
Thank you NIH - T32-HD049899, 5R01-HD47425
Thank you NRSA - F31DC009936-01

Thank you NRSA - F31DC009936-01