Subcategorization frame entropy in online verb learning

Aaron Steven White — Johns Hopkins University
Valentine Hacquard **Philip Resnik** **Jeffrey Lidz** — University of Maryland, College Park

Findings

Low subcategorization frame entropy in child-directed speech is a feature not a bug

less data + low entropy > less data + higher entropy
more data + low entropy < more data + higher entropy

Background

Syntactic Bootstrapping

Knowledge: BELIEF ⟶ S(+TENSE)
Data: {think, believe, know, hope} S(+TENSE)
Inference: BELIEF ⟵ think

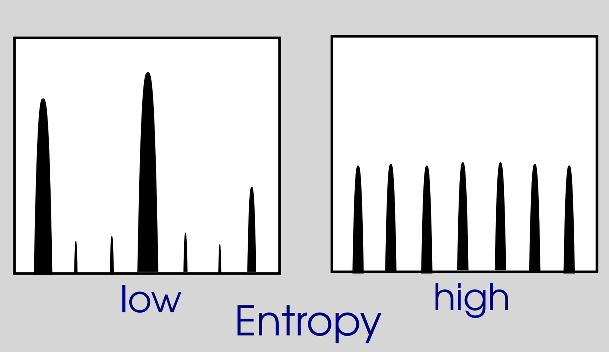
Landau & Gleitman 1985; Gleitman 1990

Prediction

More data on verb's subcategorization frame distribution
leads to
better approximation of verb's meaning

Problem

Less opportunity in child-directed speech to observe a verb's full subcategorization frame distribution

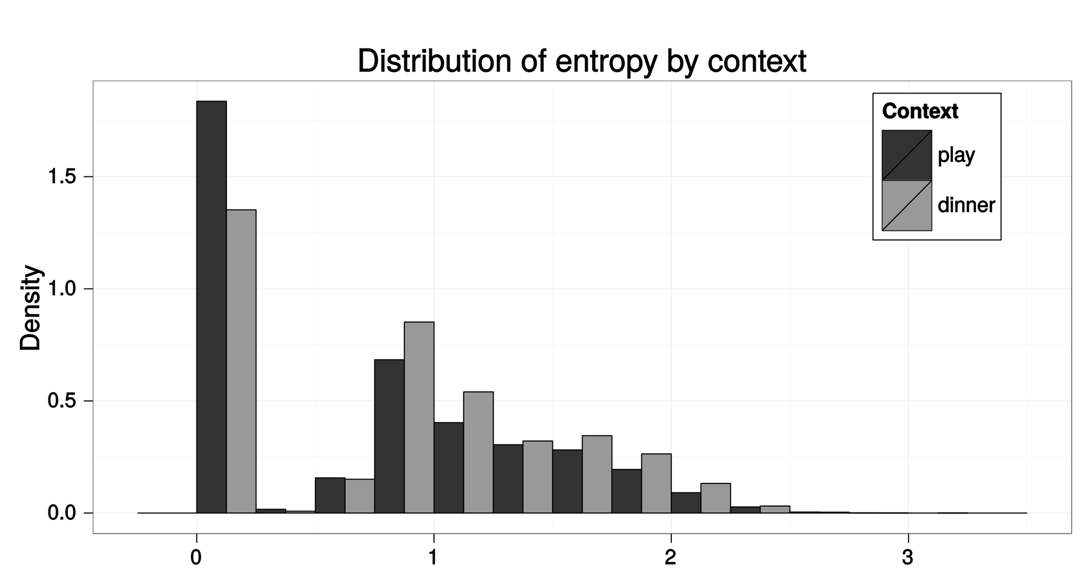


Subcategorization frame distributions in child-directed speech are **less entropic** than those in adult-directed

Buttery & Korhonen 2005; Buttery 2006

Entropy and context

Different contexts of utterance contain different amounts of by-verb subcategorization frame entropy on average



Model

mixed effects zero-inflated gamma model of entropy

random intercepts for child and verb

Logistic model coefficients

Term	Estimate	95% CI
Intercept (play)	-3.764	[-3.988, -3.436]
dinner	0.793	[0.522, 0.995]
log(freq)	3.364	[3.188, 3.619]

Inverse gamma model coefficients

Term	Estimate	95% CI
Intercept (play)	1.205	[1.185, 1.227]
dinner	-0.026	[-0.038, -0.014]
log(freq)	-0.139	[-0.151, -0.127]

Verbs' subcat frame entropy is higher on average in meal contexts than in play contexts

Idea

Use difference in average subcat frame entropy to test effect of entropy on verb learning

Focus on verbs with high subcategorization frame entropy: clause-embedding verbs (e.g. *think*, *know*, *say*, *tell*, *want*)

Norming study #1

One-shot Human Simulation Paradigm (HSP)

Gillette et al. 1999, Snedeker & Gleitman 2004, Papafragou et al. 2007, Medina et al. 2011, Trueswell et al. 2013

florped
I ____ you I'm not having a new baby now.
told

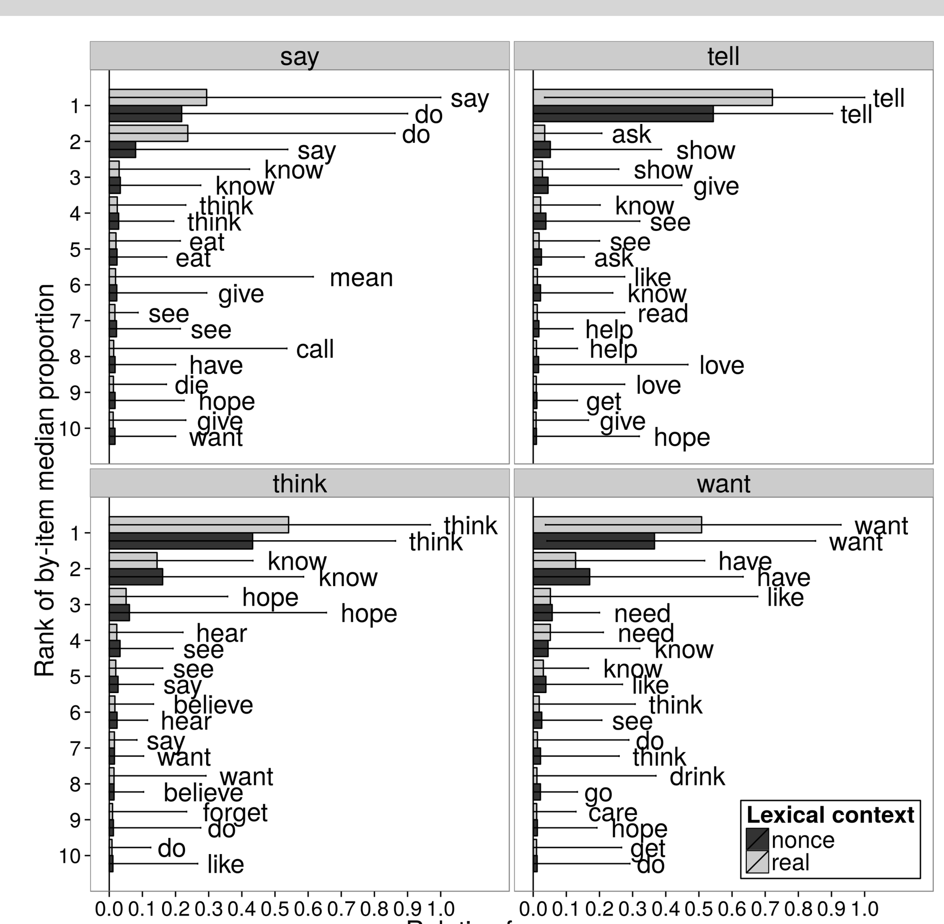
What word is most like florp in meaning?

Sample 20 sentences for each of 10 most frequent clause-embedding verbs from each utterance context (*play* v. *meal*)

Create nonce versions of each sentence by replacing words from lexical category with nonce words

I ____ you I'm not bloaving a jope fimpy now.

677 participants from Amazon Mechanical Turk provided 30 responses per sentence



Model

mixed effects logistic model of accuracy

random intercepts for participant and verb

Logistic model coefficients

Term	Estimate	Std. Error	pval
Intercept	**-2.401874**	**0.581084**	**<0.001**
real	**1.064432**	**0.187932**	**<0.001**
play	-0.004282	0.191773	0.982
real:play	0.096546	0.266819	0.717

Higher accuracy with real compared to nonce but no difference based on context of utterance

Norming study #2

Aim: measure informativity of each sentence for use in constructing HSP training sets

cf. Medina et al. 2011, Trueswell et al. 2013

155 participants from Amazon Mechanical Turk provided 5 similarity judgments per pair

Extract inaccurate responses from norming study #1 and collect semantic similarity to real verb

Use accuracy + this similarity as measure of item informativity

Experiment

Spatial Human Simulation Paradigm (HSP)

White 2015

4800 participants from Amazon Mechanical Turk each saw one of 160 training sets then rated the similarity between the learned verb and real verbs tested in another task

Create training sets by crossing

10 verbs
- 2 contexts of utterance (*play*, *dinner*)
- 2 lexical contexts (*nonce*, *real*)
- 2 training set sizes (*small*, *large*)
- 2 informativity levels (*high*, *low*)

I florped you I'm not having a new baby now.
I florped her that we wouldn't be there.
Did you florp the teacher?
....

How similar are...
florp and **think**
florp and **want**
...

How similar are...
tell and **think**
tell and **want**
...

Find correlation between normalized similarity judgments and normalized real-real similarity judgments from White et al. 2015

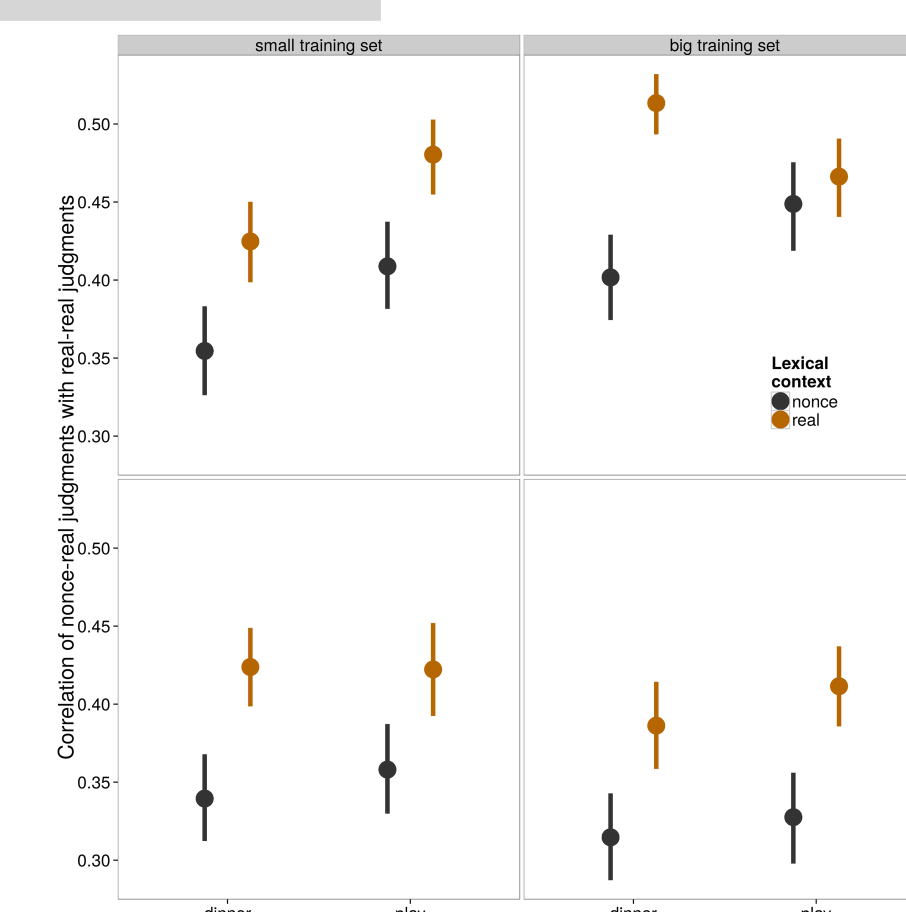
Model

mixed effects linear model of correlation

random intercepts for participant and verb

Linear model coefficients

	Estimate	Std. Error	t value
Intercept	**0.340497**	**0.027188**	**12.524**
play	0.013672	0.019568	0.699
real	**0.081704**	**0.018699**	**4.369**
high	-0.009126	0.018662	-0.489
big	-0.032717	0.018588	-1.760
play:real	-0.023468	0.026659	-0.880
play:high	**0.056136**	**0.026892**	**2.087**
real:high	0.007215	0.026247	0.275
play:big	0.022864	0.027024	0.846
real:big	-0.012370	0.026169	-0.473
high:big	**0.091242**	**0.026036**	**3.504**
play:real:high	0.004018	0.037495	0.107
play:real:big	0.015536	0.037133	0.418
play:high:big	-0.029892	0.037021	-0.807
real:high:big	0.031643	0.037142	0.852
play:real:high:big	**-0.085784**	**0.052785**	**-1.625**



With less data, low entropy contexts are most useful for inferring verb meaning. With more data, higher entropy contexts are most useful for inferring verb meaning, but only if participants also get lexical information.

Conclusions

Rather than being a hindrance, low entropy may be necessary for getting syntactic bootstrapping off the ground, while high entropy feeds later fine-tuning of the learned meanings using lexical information.

References

Buttery, P. and A. Korhonen. 2005. Large-scale analysis of verb subcategorization differences between child directed speech and adult speech. In *Proceedings of IWIRVFVC.*

Buttery, P. 2006. *Computational models for first language acquisition.* Doctoral Dissertation, University of Cambridge.

Gillette, J., H. Gleitman, L. Gleitman, and A. Lederer. 1999. Human simulations of vocabulary learning. *Cognition* 73:135–176.

Gleitman, L. 1990. The structural sources of verb meanings. *Language Acquisition*1:3–55.

Landau, B., and Lila R. Gleitman. 1985. *Language and Experience: Evidence from the Blind Child*. Harvard University Press.

Medina, T.N., J. Snedeker, J.C. Trueswell, and L.R. Gleitman. 2011. How words can and cannot be learned by observation. *PNAS*. 108:9014-9019.

Papafragou, A., K. Cassidy, and L. Gleitman. 2007. When we think about thinking: The acquisition of belief verbs. *Cognition* 105:125–165.

Snedeker, Jesse, and Lila Gleitman. 2004. Why it is hard to label our concepts. *Weaving a lexicon* 257–294.

Trueswell, J.C., T.N. Medina, A. Hafri, and L.R. Gleitman. 2013. Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive Psychology* 66:126-156.

White, A.S. 2015. *Information and incrementality in syntactic bootstrapping.* Doctoral Dissertation, University of Maryland.

White, A.S., V. Hacquard, and J. Lidz. 2015. Projecting attitudes. submitted to *Cognitive Science*

This work was supported by an NSF DDRIG grant (1456013)

29th CUNY Conference on Human Sentence Processing at the University of Florida, 5 March 2016