Bootstrapping into Filler-Gap: An Acquisition Story

Marten van Schijndel and Micha Elsner
Department of Linguistics
The Ohio State University

June 24, 2014

Background

Filler-Gap

A non-local dependency that potentially spans an unbounded # of lexemes.

e.g. That's {the ball} John kicked ___.
e.g. That's {the ball} Mary said John kicked ___.

This is hard because:

- Filler must be remembered
- Where is the gap?

Motivation

How could children learn this?

Goal

- Simple model of filler-gap

Types of Filler-Gap (for us)

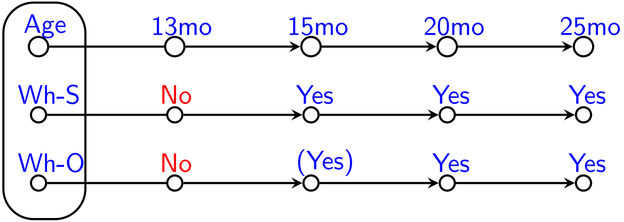
Questions

Wh-S: {What} ___ ate the apple?
Wh-O: {What} did the monkey eat ___?

Relatives

Wh-rS: Find {the boy} who ___ bumped the girl.
Wh-rO: Find {the boy} who the girl bumped ___.
That-rS: Find {the boy} that ___ bumped the girl.
That-rO: Find {the boy} that the girl bumped ___.

Acquisition Pattern



Developmental timeline of wh- question comprehension

Parentheses = marginal comprehension

That-relatives acquired slower than wh-relatives

[Seidl et al., 2003, Gagliardi et al., 2014]

Acquisition Pattern

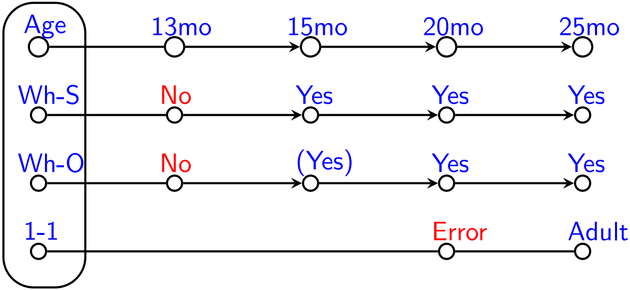
1-1 Role Bias

Subject Object

- John gorped
- Mary gorped John
- John and Mary gorped

Interpreted by Gertner and Fisher (2012) as 'Agent-first bias'
But we will show: can be modeled as 1-1 role bias

Acquisition Pattern



Developmental timeline of 1-1 role bias errors (21, 25)
Children stop this error by 25 months
[Naigles, 1990, Gertner and Fisher, 2012]

Model Motivation

What are children learning?

Complex Grammatical Constraints

Under certain conditions:

Arguments may occur in non-canonical syntactic positions.

e.g., questions introduce an expected future gap (SLASH, A-bar).

Problem:

Syntax isn't great yet

- Role conjunction not comprehended

  [Gertner and Fisher, 2012]

- Ditransitives not generalized until later

  [Goldberg et al., 2004, Bello, 2012]

Model Motivation

What are children learning?

Different Possible Orderings

The flower **hit** the apple.
What **hit** the apple.
What did the flower **hit**?

Plausible:
Word ordering patterns are fairly widespread (e.g. SOV, SVO, etc)

Previously used in BabySRL [Connor et al., 2008, 2009, 2010]

Model

- Inspired by Gradual Learning Algorithm [Boersma, 1997]
- Structure mapping: nouns used to learn verbs [Yuan et al., 2012]
- Roles assigned via ordered, latent distributions

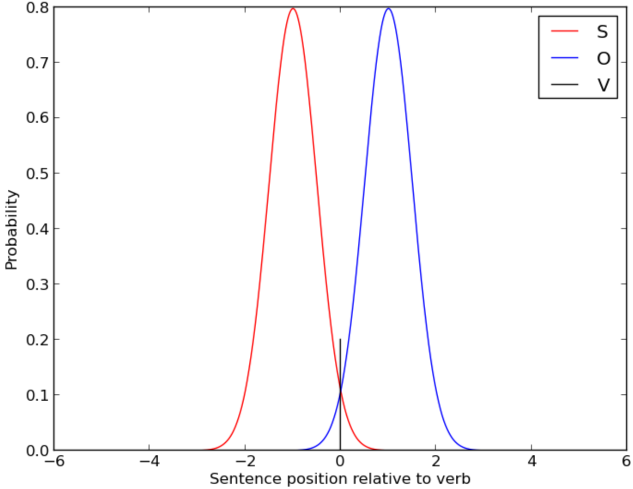
Assumptions

- (14m) Children can chunk nouns [Waxman and Booth, 2001]
- (pre-25m) Ns and roles are 1-to-1 [Gertner and Fisher, 2012]
- (9m) Abstract factors (#N) are used by learners [Xu, 2002]
- (4-5y) Children are bad at recursion [Diessel and Tomasello, 2001]

Implementation Assumptions

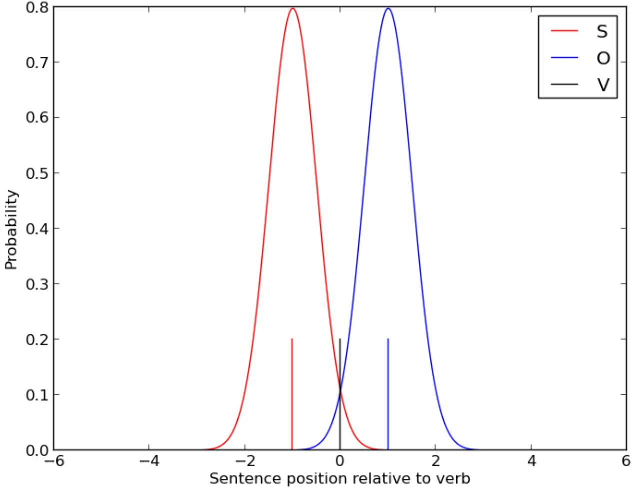
- Generate position of arguments relative to verb
- Sampled from Gaussian distributions
- Samples assumed to be independent

Model



0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Probability
−6
−4
−2
0
2
4
6
Sentence position relative to verb
S
O
V

Model



$P(-1 \mid S) \cdot P(1 \mid O)$

The cat bumped the dog.

Model

Possible parses...

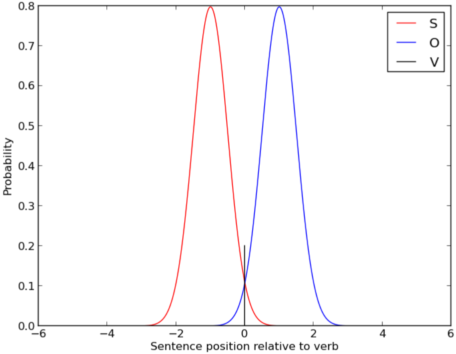
P(SVO) = P(−1 | S) · P(1 | O)
The cat bumped the dog.

P(OVS) = P(−1 | O) · P(1 | S)
The cat bumped the dog.

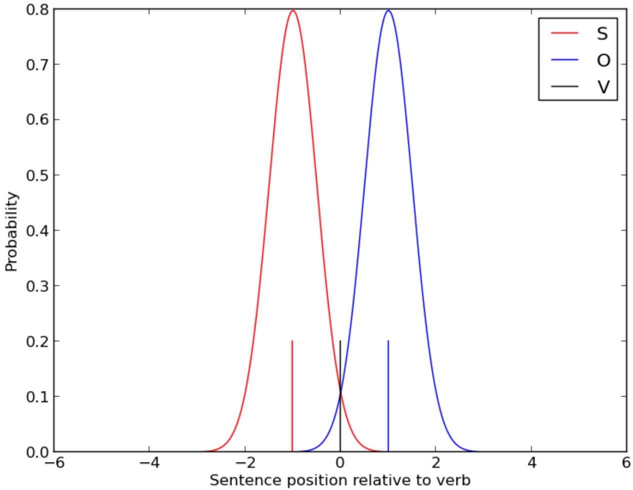
P(VO) = P(−1 | skip) · P(1 | O)
The cat bumped the dog.

P(SV) = P(−1 | S) · P(1 | skip)
The cat bumped the dog.

...



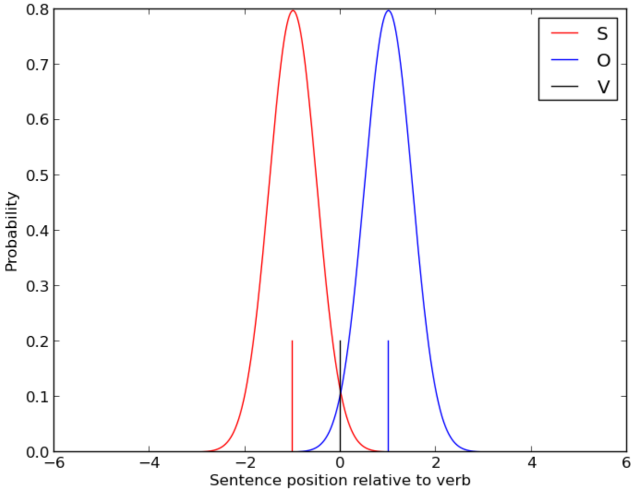
Model



$P(-1 \mid S) \cdot P(1 \mid O)$

The cat bumped the dog.

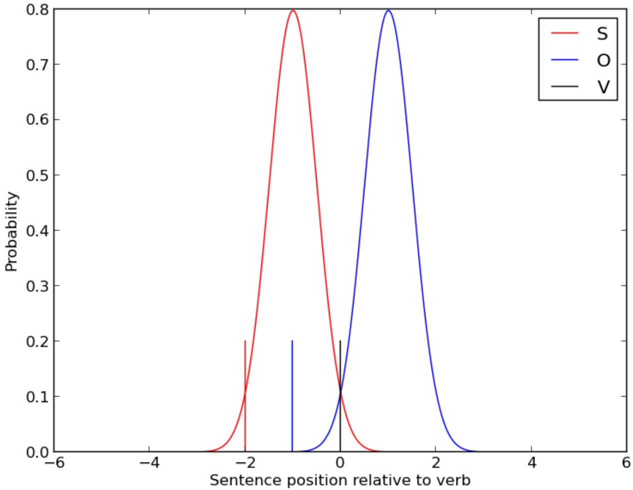
Model



$$P(-1 \mid S) \cdot P(1 \mid O)$$

Wh-S: Which cat bumped the dog?

Model



$P(-3 \mid S) \cdot P(-1 \mid O)$

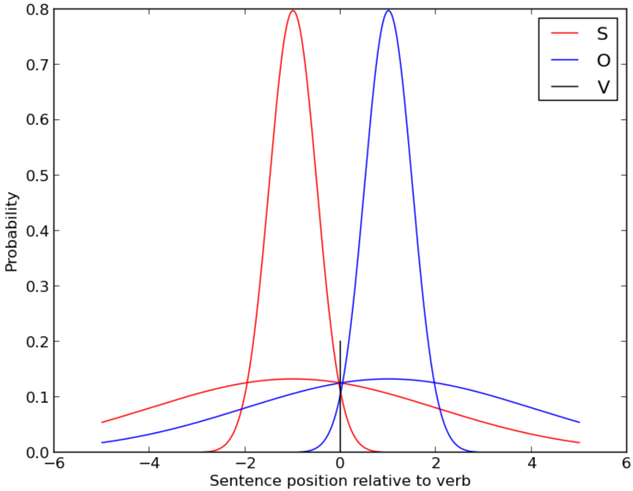
Wh-O: Which cat did the dog bump?*

Model

Initialization 2.0

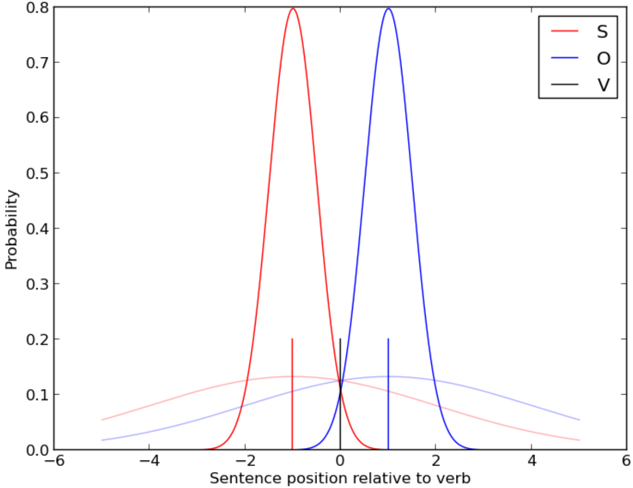
- Split distributions into mixtures of distributions
  - 1) strong due to canonical evidence
  - 2) weak, but finds arguments from anywhere

Model



0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Probability
−6
−4
−2
0
2
4
6
Sentence position relative to verb
S
O
V

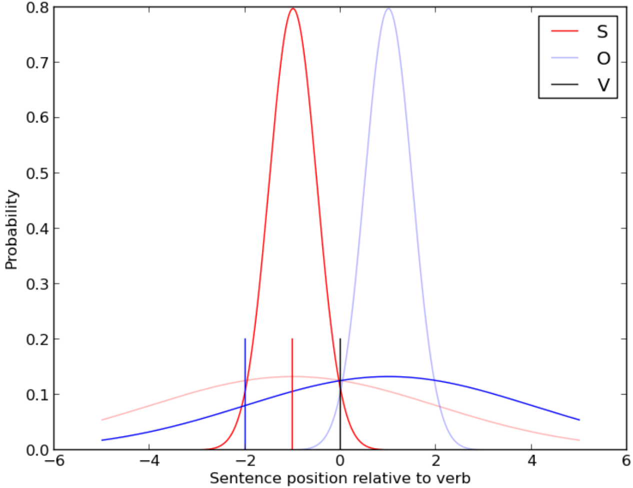
Model



$$P(-1 \mid S_C) \cdot P(1 \mid O_C)$$

Wh-S: Which cat bumped the dog?

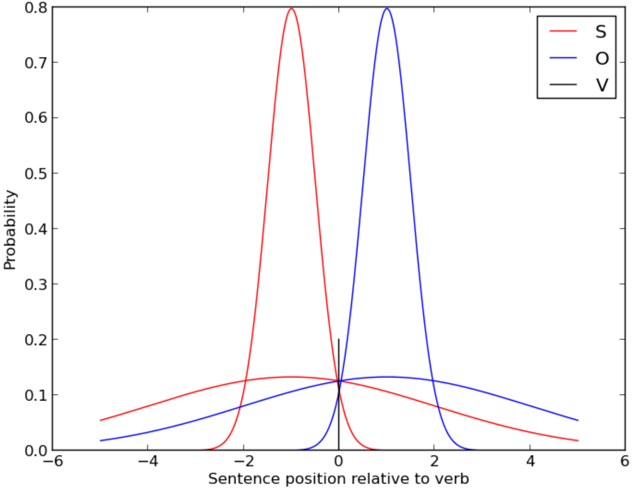
Model



$$P(-3 \mid O_N) \cdot P(-1 \mid S_C)$$

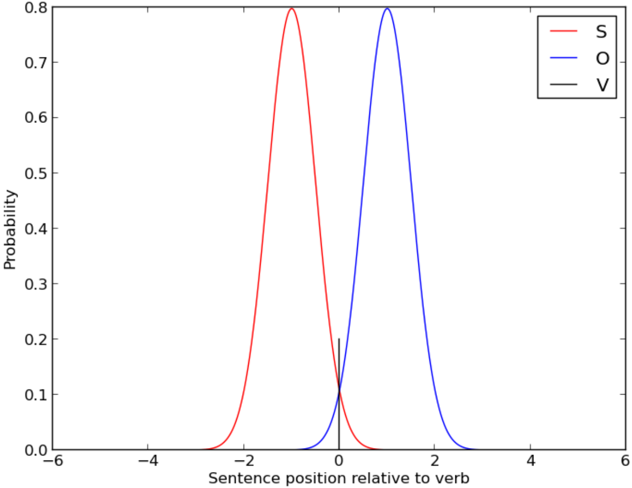
Wh-O: Which cat did the dog bump?

Model



0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Probability
−6
−4
−2
0
2
4
6
Sentence position relative to verb
S
O
V

Model

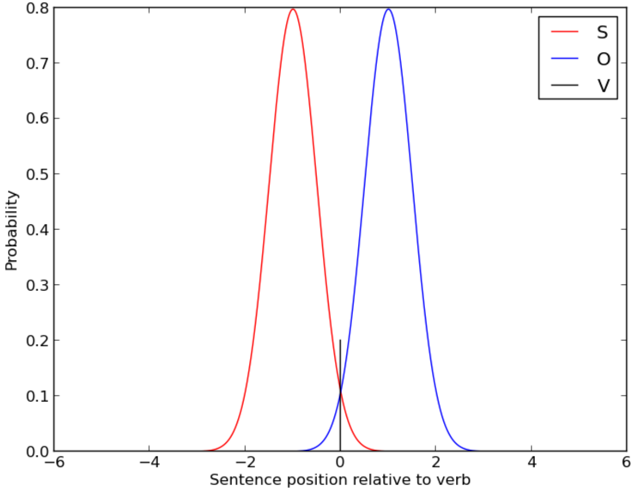


With priors, our initial model looks like this.

Evaluation

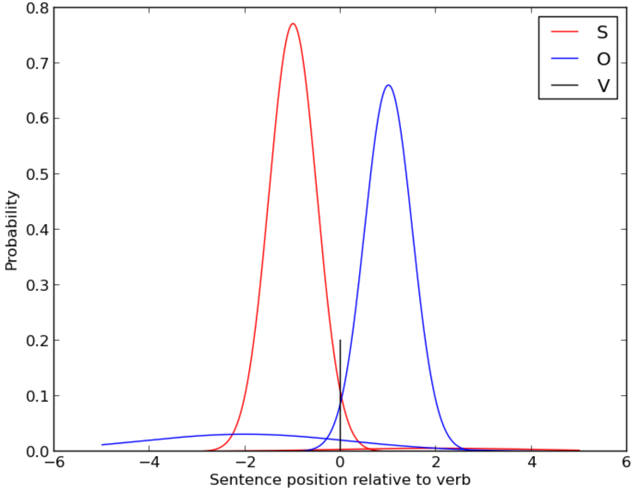
1. Extract CDS from Eve corpus
   (‘you’, ‘S’) (‘get’, ‘V’) (‘one’, ‘O’) .
   (‘what’, ‘O’) are (‘you’, ‘S’) (‘doing’, ‘V’) ?
   (‘you’, ‘S’) (‘have’, ‘V’) another cookie right on the table .
2. Chunk nouns (NLTK)
   (N;you)(V;get)(N;one) .
   (N;what)(X;are)(N;you)(V;doing) ?
   (N;you)(V;have)(N;cookie)(X;right)(X;on)(N;table) .
3. Run inference (EM)
   - Estimate labels using distributions over previous observations
   - Estimate new distributions using labelled data

Results



S
O
V
Probability
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
−6
−4
−2
0
2
4
6
Sentence position relative to verb

Results



S
O
V
Probability
Sentence position relative to verb
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
−6
−4
−2
0
2
4
6

Results: Quantitative

Overall Accuracy

Arguments correctly labelled

	P	R	F
Initial	**.56**	.66	.60
Trained	.54	**.71**	**.61***

Eve (n = 4820)

	P	R	F
Initial	**.55**	.62	.58
Trained	.53	**.67**	**.59***

Adam (n = 4461)

* ($p < .01$)

Results: Quantitative

But those numbers reflect overall performance...

We can try a coarse filler-gap filter.

Extract sentences where either:

- O precedes V
- S not immediately followed by V

Filler-gap Corpora

	P	R	F
Initial	.53	.57	.55
Trained	**.55**	**.67**	**.61***

Eve FG (n = 1345)

	P	R	F
Initial	.53	.52	.52
Trained	**.54**	**.63**	**.58***

Adam FG (n = 1287)

* ($p < .01$)

Results: Quantitative

Eve FG Corpus

Subject/Object

	P	R	F
Initial	**.66**	**.83**	**.74**
Trained	.64	.84	.72†

Subject (n = 691)

	P	R	F
Initial	.35	.31	.33
Trained	**.45**	**.52**	**.48***

Object (n = 654)

That/Wh-

	P	R	F
Initial	.63	.45	.53
Trained	**.73**	**.75**	**.74***

Wh- (n = 689)

	P	R	F
Initial	.43	.48	.45
Trained	**.44**	**.57**	**.50†**

That (n = 125)

* ($p < .01$) † ($p < .05$)

1-1 Role Bias

How often is NNV labelled as SOV? (1-1 role bias error)

- Connor et al (2008, 2009): 63-82% error (agent-first bias)
- Our initial model: 66% error (1-1 bias)

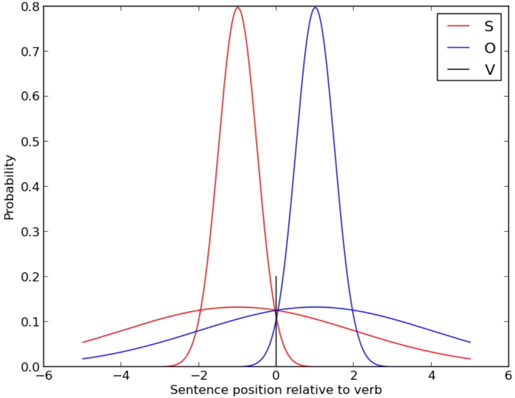
Current model is comparable to Baby SRL

Initialization Analysis

Very Robust

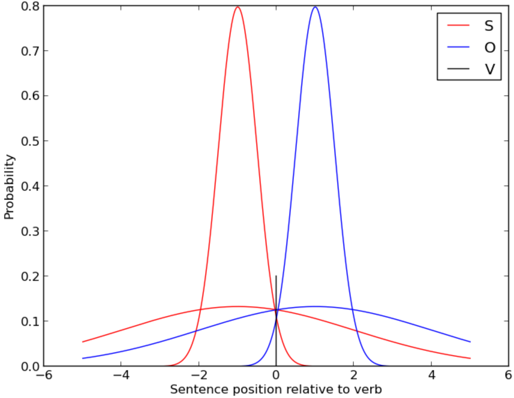
- positions: -3,3 ; -1,1 ; -0.1,0.1
- variance: 0.5 – 4
- caveat: filler preverbal prob must outweigh skip-penalty

Model Selection



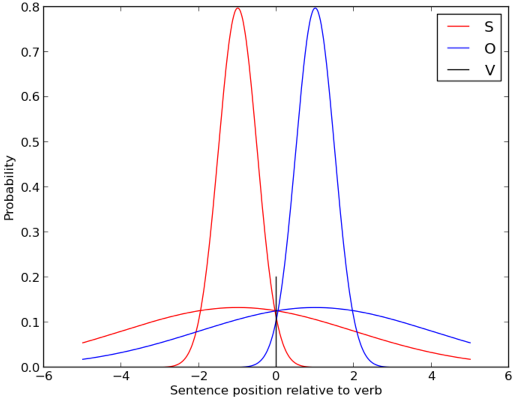
Do we really want this setup?

Model Selection



Is the non-canonical subject useful? (According to BIC)

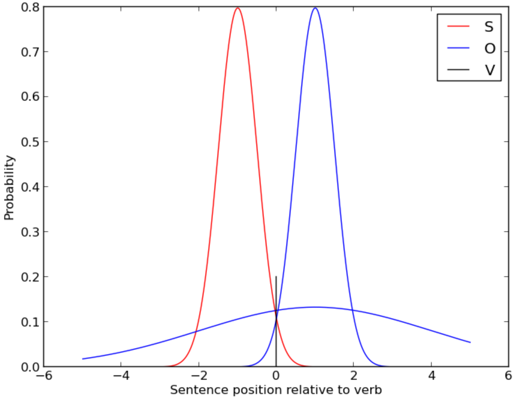
Model Selection



"Helps" capture imperatives...
But kids know imperatives...

'Put the cookie on the table!'
'[You] put the cookie on the table!'

Model Selection



Then non-canonical subject isn't useful (according to BIC)

Suggests dynamic Gaussian generation is possible

Future Work

- Add lexicalization
- Dynamically generate Gaussians
- Model non-English (verb-medial) languages
- Bootstrap linear grammar into hierarchic grammar

Conclusion

It is possible to acquire filler-gap without (complex) syntax.
The current model offers additional benefits:

- Reflects developmental S-O asymmetry
- Reflects developmental That-Wh asymmetry
- Robust to varied initializations

Questions?

Thanks to:

- Peter Culicover
- William Schuler
- Laura Wagner
- Attendees of the OSU 2013 Fall Ling. Colloquium Fest

This work was partially funded by an OSU Dept. of Linguistics Targeted Investment for Excellence (TIE) grant for collaborative interdisciplinary projects conducted during the academic year 2012-13.

Results: 1-1 bias

How often NNV is labelled SOV

Current Model

	Error Rate
Initial	.66
Trained	.13

(n = 1000)

Trained Baby SRL

	Error Rate
Arg-Arg	.65
Arg-Verb	0

[Connor et al., 2008]

	Error Rate
Arg-Arg	.82
Arg-Verb	.63

[Connor et al., 2009]

Results: 1-1 bias

Agent Prediction

	Recall
Initial	.67
Trained	.65

Transitive (n = 1000)

	Recall
Initial	1
Trained	.96

Intransitive (n = 1000)

[Connor et al., 2010]

	Recall
Weak (10) lexical	.71
Strong (365) lexical	.74
Gold Args	.77

Transitive

	Recall
Weak (10) lexical	.59
Strong (365) lexical	.41
Gold Args	.58

Intransitive

1-1 Role Bias Summary

How often is the agent correctly labelled?

Transitives (1173 sents)

- Connor et al. (2010): 71-77%
  - Lexicalization helps
- Initial current model: 67%
  Trained current model: 65%
  - Completely unlexicalized

Intransitives (1513 sents)

- Connor et al. (2010): 41-59%
- Initial current model: 100%
  Trained current model: 96%

Current model is comparable to Baby SRL for transitives
Current model does much better on intransitives

Gertner and Fisher (2012)

[Gertner and Fisher, 2012]

The boy/girl is gorping.

Gertner and Fisher (2012)

[Gertner and Fisher, 2012]

The girl is gorping the boy.

Gertner and Fisher (2012)

[Gertner and Fisher, 2012]

The girl and the boy are gorping.

Bibliography I

Bello, S. (2012).
Identifying indirect objects in French: An elicitation task.
In *Proceedings of the 2012 annual conference of the Canadian Linguistic Association*.

Boersma, P. (1997).
How we learn variation, optionality, and probability.
Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam, 21:43–58.

Connor, M., Gertner, Y., Fisher, C., and Roth, D. (2008).
Baby srl: Modeling early language acquisition.
In *Proceedings of the Twelfth Conference on Computational Natural Language Learning*.

Bibliography II

Connor, M., Gertner, Y., Fisher, C., and Roth, D. (2009).
Minimally supervised model of early language acquisition.
In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning*.

Connor, M., Gertner, Y., Fisher, C., and Roth, D. (2010).
Starting from scratch in semantic role labelling.
In *Proceedings of ACL 2010*.

Diessel, H. and Tomasello, M. (2001).
The acquisition of finite complement clauses in english: A corpus-based analysis.
Cognitive Linguistics, 12:1–45.

Bibliography III

Gagliardi, A., Mease, T. M., and Lidz, J. (2014).
Discontinuous development in the acquisition of filler-gap dependencies: Evidence from 15- and 20-month-olds.
Harvard unpublished manuscript: http://www.people.fas.harvard.edu/~gagliardi.

Gertner, Y. and Fisher, C. (2012).
Predicted errors in children's early sentence comprehension.
Cognition, 124:85–94.

Goldberg, A. E., Casenhiser, D., and Sethuraman, N. (2004).
Learning argument structure generalizations.
Cognitive Linguistics, 14(3):289–316.

Naigles, L. R. (1990).
Children use syntax to learn verb meanings.
The Journal of Child Language, 17:357–374.

Bibliography IV

Seidl, A., Hollich, G., and Jusczyk, P. W. (2003).
Early understanding of subject and object wh-questions.
Infancy, 4(3):423–436.

Waxman, S. R. and Booth, A. E. (2001).
Seeing pink elephants: Fourteen-month-olds' interpretations of novel nouns and adjectives.
Cognitive Psychology, 43:217–242.

Xu, F. (2002).
The role of language in acquiring object kind concepts in infancy.
Cognition, 85:223–250.

Yuan, S., Fisher, C., and Snedeker, J. (2012).
Counting the nouns: Simple structural cues to verb meaning.
Child Development, 83(4):1382–1399.