Bootstrapping into Filler-Gap: An Acquisition Story

Marten van Schijndel
Department of Linguistics
The Ohio State University

November 22, 2013

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

- Simplest model of filler-gap?

Background

Psychology

Children can't use filler-gap until 5 years
[de Villiers and Roeper, 1995]

Computational Linguistics

An uncommon phenomenon that doesn't boost performance much
[Rimell et al., 2009, Nivre et al., 2010, Nguyen et al., 2012]

Experimental Results

[Seidl et al., 2003]
Preferential looking paradigm

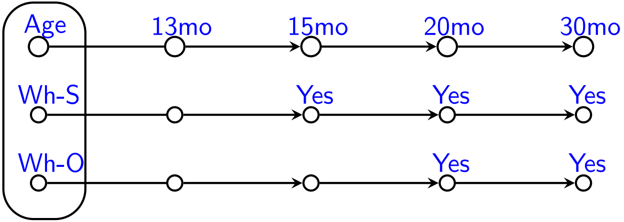
Wh-

Wh-S: What hit the apple?
Wh-O: What did the flower hit?

Control

Where is the flower?

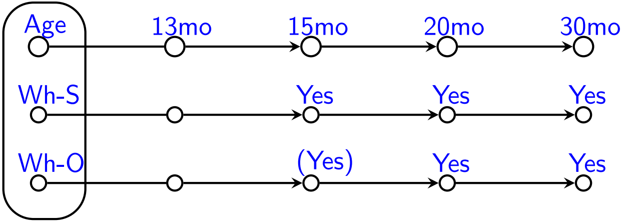
Acquisition Pattern?



Developmental timeline of wh- question comprehension (13, 15, 20)

[Seidl et al., 2003]

Acquisition Pattern



Developmental timeline of wh- question comprehension (15, 20)
Parentheses = marginal comprehension
[Gagliardi et al., 2011]

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).

Different Possible Orderings

The flower **hit** the apple.

What **hit** the apple.

What did the flower **hit**?

Model Motivation

Different Word Orderings

- SOV: Japanese
  Hindi
  German
- SVO: English
  Mandarin
  Spanish
- VSO: Zapotec
  Irish
- VOS: Malagasy
  Baure

Model Motivation

OT: Different Constraint Orderings

Yield different phonological realizations [Boersma, 1997]
e.g. nasal place assimilation

an+pa	*Gesture(tip)	*Replace(cor)
[anpa]	*!	
[ampa]		*

an+pa	*Replace(cor)	*Gesture(tip)
[anpa]		*
[ampa]	*!	

Model

- Gradual Learning Algorithm [Boersma, 1997]
- Structure mapping: nouns used to learn verbs [Yuan et al., 2012]

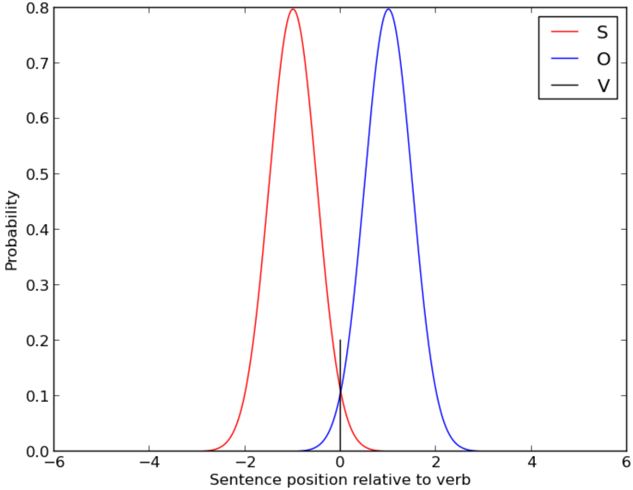
Assumptions

- Children can identify nouns [Shi et al., 1998]
- Ns and roles are 1-to-1 [Gertner and Fisher, 2012]
- Abstract factors (#N) are used by learners [Xu, 2002]
- Children are bad at recursion [Diessel and Tomasello, 2001]

Implementation Assumptions

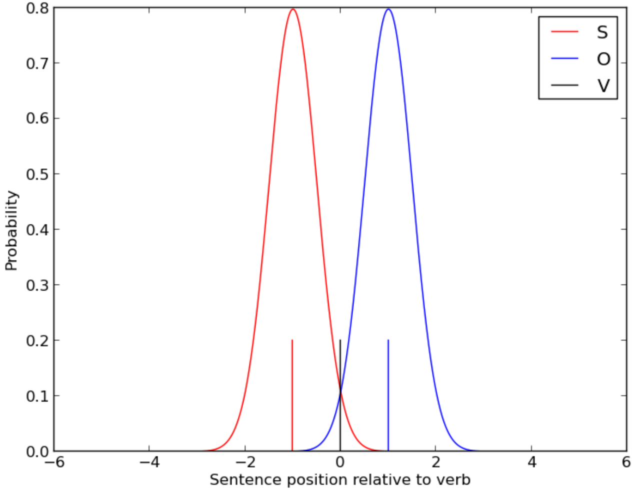
- Distributions are Gaussian

Model



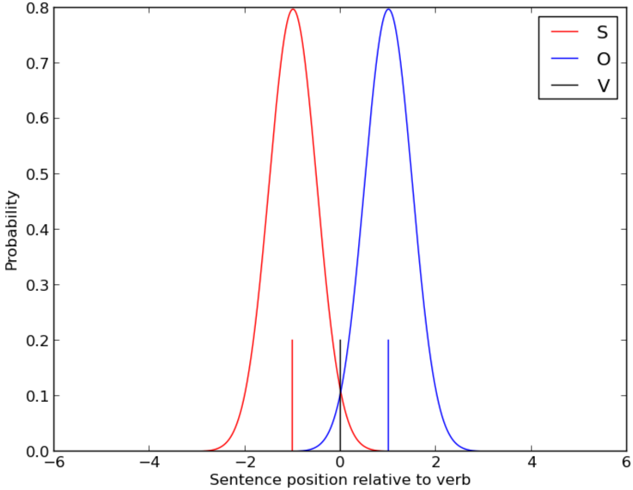
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Probability
−6
−4
−2
0
2
4
6
Sentence position relative to verb
S
O
V

Model



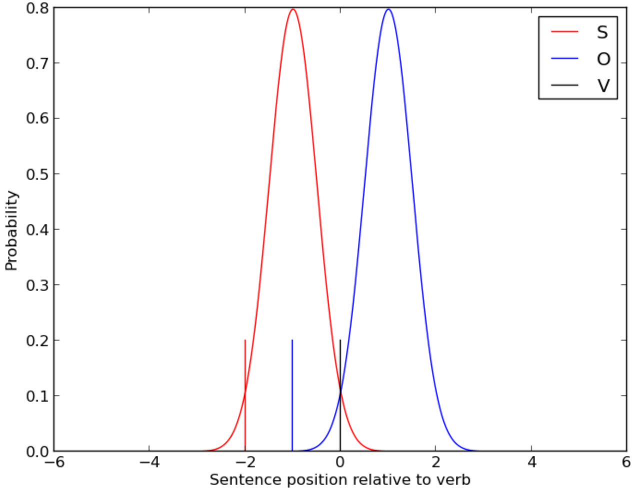
The cat bumped the dog.

Model



Wh-S: Which cat bumped the dog?

Model



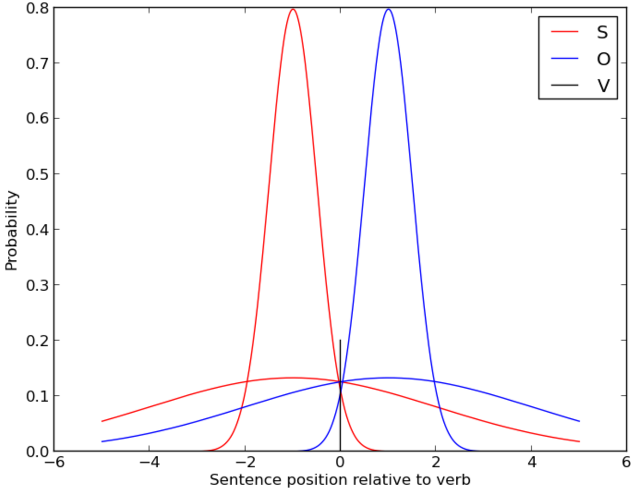
Wh-O: Which cat did the dog bump?*

Model

Initialization 2.0

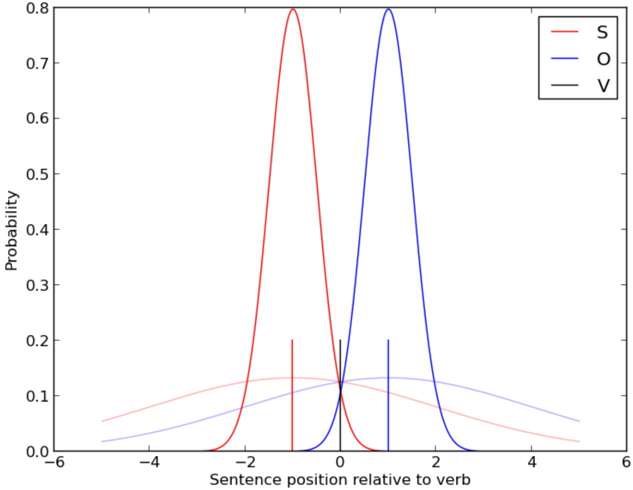
- Split distributions into mixtures of distributions
  - 1) strong due to canonical evidence
  - 2) weak, but finds arguments from anywhere

Model



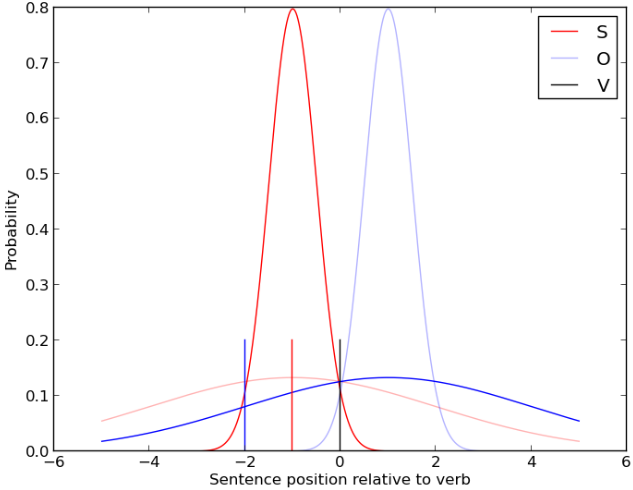
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Probability
−6
−4
−2
0
2
4
6
Sentence position relative to verb
S
O
V

Model



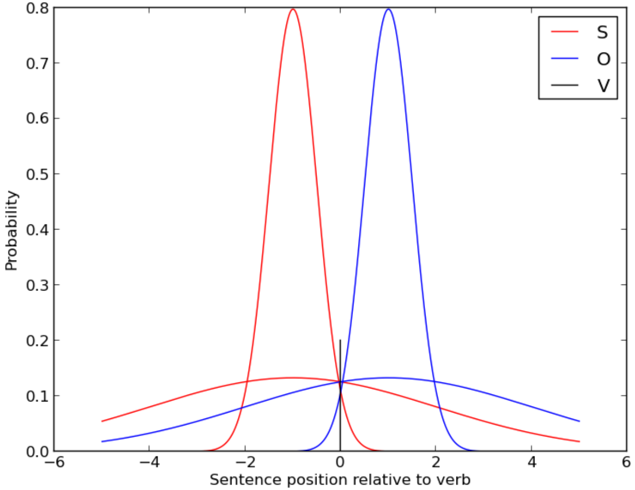
Wh-S: Which cat bumped the dog?

Model



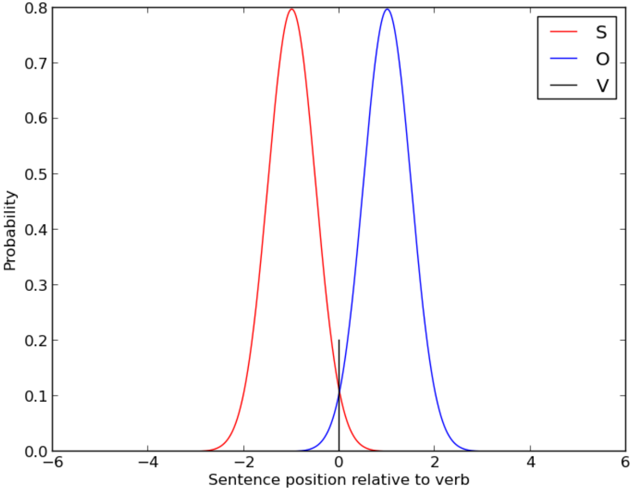
Wh-O: Which cat did the dog bump?

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

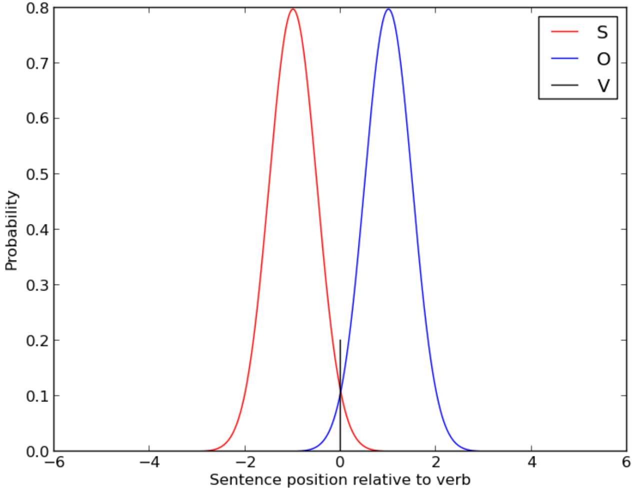
1. Extract CDS from Eve corpus
   ('you', 'S') ('get', 'V') ('one', 'O') .
   ('what', 'O') are ('you', 'S') ('doing', 'V') ?
   ('you', 'S') ('have', 'V') another cookie right on the table .
2. Chunk nouns (NLTK)
   (N;you)(V;get)(N;one) .
   (N;what)(X;are)(N;you)(V;doing) ?
   (N;you)(V;have)(N;cookie)(X;right)(X;on)(N;table) .
3. Run inference

Inference

Expectation-Maximization

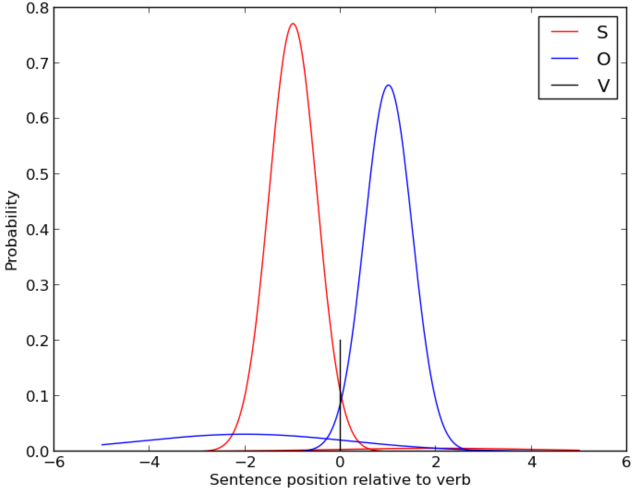
- Estimate labels using distributions over previous observations
- Estimate new distributions using labelled data
- Iterate until converged ($\sim$4 iterations)

Results



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

Results


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

Relative Development

[Gagliardi and Lidz, 2010, Gagliardi et al., 2011]

T-Rel

T-S: Show me the dog that bumped the cat.
T-O: Show me the cat that the dog bumped.

W-Rel

Wh-S: Show me the dog who bumped the cat.
Wh-O: Show me the cat who the dog bumped.

Results

- 'Wh-' and 'that' relative comprehension ∼15 months
- 'Wh-' easier than 'that'

Relative Differences

That: Confusion with dem/det?

- That is a book.
- Gimme that!
- Gimme that book!
- Find the cookie that the mouse ate.

Wh-: Helped by questions?

- Who kicked the bucket?
- Who did the burglar assault?
- Find the mouse who the cat ate.

Results: Quantitative

Overall Accuracy

Arguments correctly labelled

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

Eve (n = 3944)

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

Adam (n = 3622)

* ($p < .01$)

Results: Quantitative

Agent Prediction

	Recall
Initial	.67
Trained	.65

Transitive (n = 1000)

	Recall
Initial	1
Trained	.96

Intransitive (n = 1000)

[Connor et al., 2010] (pseudo-comparable)

	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

Results: Quantitative

But those numbers reflect overall performance...

We can try a coarse filler-gap filter.

Extract sentences where:

- 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

	n	P	R	F
Subject	691	.66	.83	.74
Object	654	.35	.31	.33

Initial Model

	P	R	F
Subject	.64	.84	.72†
Object	.45	.52	.48*

Trained Model

That/Wh-

	n	P	R	F
Wh-	363	.63	.45	.52
That	68	.43	.48	.45

Initial Model

	P	R	F
Wh-	.73	.75	.74*
That	.44	.57	.50†

Trained Model

* ($p < .01$) † ($p < .05$)

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
  - positions: -3,3 ; -1,1 ; -0.1,0.1
  - sd: filler preverbal prob must outweigh skip-penalty

Questions?

Thanks to everyone who gave feedback on this project:
Lacqueys, Clippers, Dave Howcroft, Evan Jaffe, William Schuler, and Peter Culicover, but especially Micha Elsner

Connor et al '10

How does this model compare to Connor et al '10?

Connor et al are interested in modeling SRL acquisition and in replicating 1-1 role bias error (21 months).

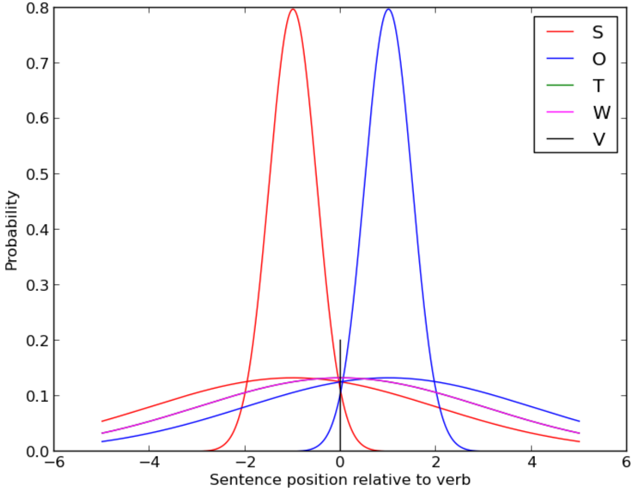
Plausibility

- Connor et al '10 productively learn 5 roles
  - This increases their specificity
  - Children do not generalize above 2 roles until after 31 months (earliest) [Goldberg et al., 2004, Bello, 2012]
- Connor et al's results raise questions about structure mapping
  Single N is patient 40% of the time?

1-1 Role Bias

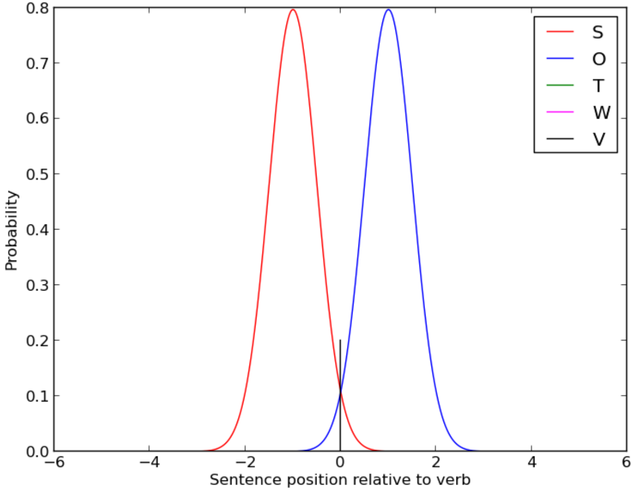
- Connor et al (gold training): 63-82% 1-1 bias error
- Our initial model: 77% 1-1 bias error

Model: Relativizers



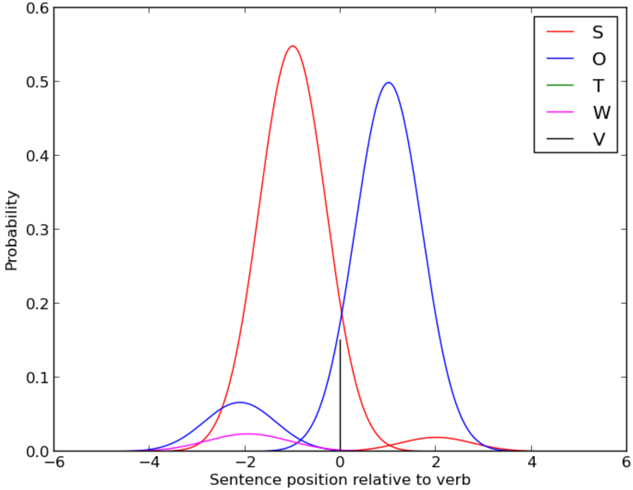
Initial model with function Gaussians

Model: Relativizers



Initial relative model with priors

Results: Relativizers



Trained model with function Gaussians

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. (2010).
Starting from scratch in semantic role labelling.
In *Proceedings of ACL 2010*.

de Villiers, J. and Roeper, T. (1995).
Barriers, binding, and acquisition of the dp-np distinction.
Language Acquisition, 4(1):73–104.

Bibliography II

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

Gagliardi, A. and Lidz, J. (2010).
Morphosyntactic cues impact filler-gap dependency resolution in 20- and 30-month-olds.
In *Poster session of BUCLD35*.

Gagliardi, A., Mease, T. M., and Lidz, J. (submitted 2011).
U-shaped development in the acquisition of filler-gap dependencies: Evidence from 15- and 20-month-olds.
Language Learning and Development.

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

Bibliography III

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

Nguyen, L., van Schijndel, M., and Schuler, W. (2012).
Accurate unbounded dependency recovery using generalized categorial grammars.
In *Proceedings of the 24th International Conference on Computational Linguistics (COLING '12)*, Mumbai, India.

Nivre, J., Rimell, L., McDonald, R., and Gómez-Rodríguez, C. (2010).
Evaluation of dependency parsers on unbounded dependencies.
In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING'10)*, pages 833–841.

Bibliography IV

Rimell, L., Clark, S., and Steedman, M. (2009).
Unbounded dependency recovery for parser evaluation.
In *Proceedings of EMNLP 2009*, volume 2, pages 813–821.

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

Shi, R., Werker, J. F., and Morgan, J. L. (1998).
Newborn infants sensitivity to perceptual cues to lexical and grammatical words.
Cognition, 72(2):B11–B21.

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

Bibliography V

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