Cognitive Compositional Semantics using Continuation Dependencies

William Schuler, Adam Wheeler
Dept Linguistics, The Ohio State University

August 25, 2014

Introduction

Goal: model how brains represent complex scoped quantified propositions

- Use only cued associations (dependencies from cue to target state)
  [Marr, 1971, Anderson et al., 1977, Murdock, 1982, McClelland et al., 1995, Howard and Kahana, 2002]
  (no direct implementation of unconstrained beta reduction)
- Interpret by traversing cued associations in sentence, match to memory (assume learned traversal process, sensitive to up/down entailment)
- Despite austerity, can model scope using 'continuation' dependencies
- Seems to make reassuring predictions:
  - conjunct matching is easy, even in presence of quantifiers
  - quantifier upward/downward entailment (monotone incr/decr) is hard
  - disjunction is as hard as quantifier upward/downward entailment
- Empirical evaluation shows no coverage or learnability gaps
  - cognitively motivated model is about as accurate as state of art

Background: why dependencies?

Model connections in associative memory w. matrix [Anderson et al., 1977]:

$$v = M\,u \tag{1}$$

$$(M\,u)_{[i]} \stackrel{\text{def}}{=} \textstyle\sum_{j=1}^{J} M_{[i,j]} \cdot u_{[j]} \tag{1$'$}$$

Build cued associations using outer product [Marr, 1971]:

$$M_t = M_{t\text{–}1} + v \otimes u \tag{2}$$

$$(v \otimes u)_{[i,j]} \stackrel{\text{def}}{=} v_{[i]} \cdot u_{[j]} \tag{2$'$}$$

Merge results of cued associations using pointwise / diagonal product:

$$w = \text{diag}(u)\,v \tag{3}$$

$$(\text{diag}(v)\,u)_{[i]} \stackrel{\text{def}}{=} v_{[i]} \cdot u_{[i]} \tag{3$'$}$$

Background: why dependencies?

Dependency relations with label ℓ_i from u_i to v_i can be stored as vectors r_i:

$$R \stackrel{\text{def}}{=} \sum_i v_i \otimes r_i \quad (4a)$$

$$R' \stackrel{\text{def}}{=} \sum_i r_i \otimes \ell_i \quad (4b)$$

$$R'' \stackrel{\text{def}}{=} \sum_i r_i \otimes u_i \quad (4c)$$

And retrieved/traversed using accessor matrices R, R', R'' [Schuler, 2014]:

$$v_i \approx R \ \text{diag}(R' \ \ell_i) \ R'' \ u_i \quad (5)$$

This cue sequence can be simplified as dependency function:

$$v_i = (\mathbf{f}_{\ell_i} \ u_i) \quad (6)$$

Background: predications and graph matching

Dependencies can combine into *predications* [Copestake et al., 2005]:

$$(f\ u\ v_1\ v_2\ v_3\ \ldots) \;\Leftrightarrow\; (\mathbf{f_0}\ u) = \mathbf{v}_f \wedge (\mathbf{f_1}\ u) = v_1 \wedge (\mathbf{f_2}\ u) = v_2 \wedge (\mathbf{f_3}\ u) = v_3 \wedge \ldots \quad (7)$$

For example:

$$(\textsc{Contain}\ u\ v_1\ v_2) \;\Leftrightarrow\; (\mathbf{f_0}\ u) = \mathbf{v}_{\textsc{Contain}} \wedge (\mathbf{f_1}\ u) = v_1 \wedge (\mathbf{f_2}\ u) = v_2 \quad (8)$$

Dependencies incrementally matched to memory during comprehension:

$$v_t = R\ R''\ v_{t\text{–}1} \quad (9a)$$

$$A_t = A_{t\text{–}1} + R\ \mathrm{diag}(R'\ R'^{\top}\ R''\ v_{t\text{–}1})\ R''\ A_{t\text{–}1}\ v_{t\text{–}1} \otimes v_t \quad (9b)$$

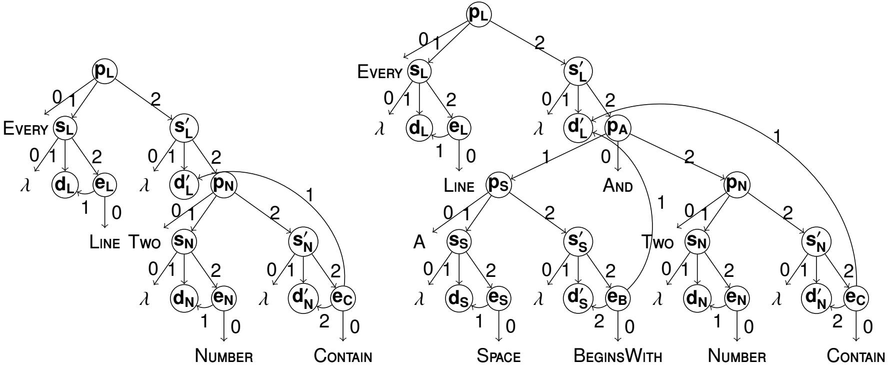
(or reverse, during production).

Need conditional traversal for entailment [MacCartney and Manning, 2009].

Scoped quantified predications: 'direct' style

Can implement a 'direct' semantics based on lambda calculus [Koller, 2004]:

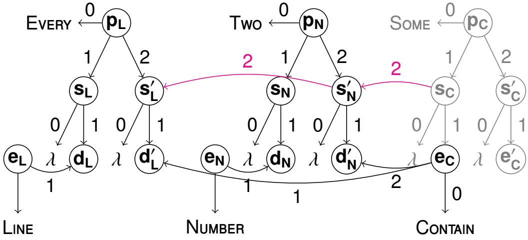
$(\textsc{Every}\ \mathbf{p}_\mathsf{L}\ \mathbf{s}_\mathsf{L}\ \mathbf{s}'_\mathsf{L}) \wedge (\textsc{Set}\ \mathbf{s}_\mathsf{L}\ \mathbf{d}_\mathsf{L}\ \mathbf{e}_\mathsf{L}) \wedge (\textsc{Line}\ \mathbf{e}_\mathsf{L}\ \mathbf{d}_\mathsf{L}) \wedge (\textsc{Set}\ \mathbf{s}'_\mathsf{L}\ \mathbf{d}'_\mathsf{L}\ \mathbf{p}_\mathsf{N}) \wedge$
$(\textsc{Two}\ \mathbf{p}_\mathsf{N}\ \mathbf{s}_\mathsf{N}\ \mathbf{s}'_\mathsf{N}) \wedge (\textsc{Set}\ \mathbf{s}_\mathsf{N}\ \mathbf{d}_\mathsf{N}\ \mathbf{e}_\mathsf{N}) \wedge (\textsc{Number}\ \mathbf{e}_\mathsf{N}\ \mathbf{d}_\mathsf{N}) \wedge (\textsc{Set}\ \mathbf{s}'_\mathsf{N}\ \mathbf{d}'_\mathsf{N}\ \mathbf{e}_\mathsf{C}) \wedge (\textsc{Contain}\ \mathbf{e}_\mathsf{C}\ \mathbf{d}'_\mathsf{L}\ \mathbf{d}'_\mathsf{N})$

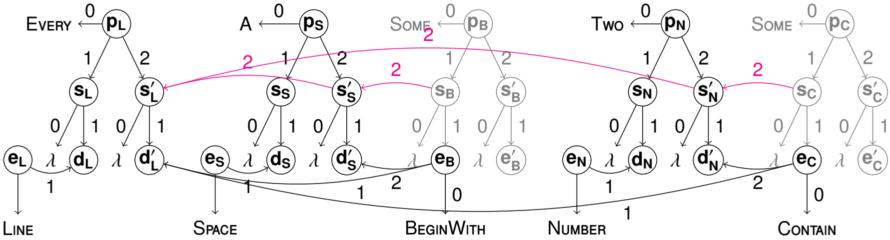


Hard to learn to match conjunct (left) in conjoined representation (right).

Scoped quantified predications: 'continuation' style

Change redundant dependency '2' at lambdas to instead point up to context:





Upward dependencies look like 'continuation-passing' style [Barker, 2002].

Bestiary of referential states

Set referents are now context-sensitive. . .

- **ordinary discourse referents** $d \in D$ [Karttunen, 1976]:
  - referents with no arguments
- **eventualities** $e \in E$ [Davidson, 1967, Parsons, 1990]:
  - referents with beginning, end, duration
  - one argument for each participant, ordered arbitrarily
- **reified sets or groups** $s \in S$ [Hobbs, 1985]:
  - referents with cardinalities, can be co-referred by plural anaphora
  - has ***iterator*** argument d_1
  - has ***scope*** argument s_2, sim. to continuation parameters [Barker, 2002]
  - has ***superset*** argument s_3 specifying superset
- **propositions** $p \in P$ [Thomason, 1980]:
  - referents that can be believed or doubted
  - form of generalized quantifier [Barwise and Cooper, 1981]
  - has ***restrictor*** argument s_1
  - has ***nuclear scope*** argument s_2

Translation to lambda calculus

Lambda calculus terms Δ can be derived from predications Γ:

- Initialize Δ with lambda terms (sets) that have no outscoped sets in Γ:

$$\frac{\Gamma, (\textsc{Set}\ s\ i\ \_\ \_)\ ;\ \Delta}{\Gamma, (\textsc{Set}\ s\ i\ \_\ \_)\ ;\ (\lambda_i\ \textsc{True}), \Delta}\ (\textsc{Set}\ \_\ \_\ s\ \_) \notin \Gamma$$

- Add constraints to appropriate sets in Δ:

$$\frac{\Gamma, (f\ i_0\ ..\ i\ ..\ i_N)\ ;\ (\lambda_i\ o), \Delta}{\Gamma\ ;\ (\lambda_i\ o \wedge (\mathbf{h}_f\ i_0\ ..\ i\ ..\ i_N)), \Delta}\ i_0 \in E$$

- Add constraints of supersets as constraints on subsets in Δ:

$$\frac{\Gamma, (\textsc{Set}\ s\ i\ \_\ \_), (\textsc{Set}\ s'\ i'\ s''s)\ ;\ (\lambda_i\ o \wedge (\mathbf{h}_f\ i_0\ ..\ i\ ..\ i_N)), (\lambda_{i'}\ o'), \Delta}{\Gamma, (\textsc{Set}\ s\ i\ \_\ \_), (\textsc{Set}\ s'\ i'\ s''s)\ ;\ (\lambda_i\ o \wedge (\mathbf{h}_f\ i_0\ ..\ i\ ..\ i_N)), (\lambda_{i'}\ o' \wedge (\mathbf{h}_f\ i_0\ ..\ i'\ ..\ i_N)), \Delta}$$

- Add quantifiers over completely constrained sets in Δ:

$$\frac{\begin{array}{c}\Gamma, (\textsc{Set}\ s\ i\ \_\ \_), (f\ p\ s'\ s''), (\textsc{Set}\ s'\ i'\ s\ \_), (\textsc{Set}\ s''\ i''\ s'\ s')\ ;\\ (\lambda_i\ o), (\lambda_{i'}\ o'), (\lambda_{i''}\ o''), \Delta\end{array}}{\Gamma, (\textsc{Set}\ s\ i\ \_\ \_)\ ;\ (\lambda_i\ o \wedge (\mathbf{h}_f\ (\lambda_{i'}\ o')\ (\lambda_{i''}\ o''))), \Delta}\ \begin{array}{l} p \in P,\ (f'..\ i'..) \notin \Gamma,\\ (f''..\ i''..) \notin \Gamma.\end{array}$$

For example: $(\textsc{Every}\ (\lambda_{d_L}\ \textsc{Some}\ (\lambda_{e_L}\ \textsc{BeingALine}\ e_L\ d_L))$
$(\lambda_{d'_L}\ \textsc{Two}\ (\lambda_{d_N}\ \textsc{Some}\ (\lambda_{e_N}\ \textsc{BeingANum}\ e_N\ d_N))$
$(\lambda_{d'_N}\ \textsc{Some}\ (\lambda_{e_H}\ \textsc{Having}\ e_H\ d'_L\ d'_N))))$

Predictions

This model makes reassuring predictions (to be evaluated in future work)...

- Conjunct matching is easy, automatic, learned early.
  Evidence: errors until about 21 months [Gertner and Fisher, 2012].
- Upward/downward entailment on 1st/2nd argument is much harder:
  More than two perl scripts work. $\vdash$ *More than two scripts work.*
  Fewer than two scripts work. $\vdash$ *Fewer than two perl scripts work.*
  Not simple matching; speaker must learn conditional matching rules.
  Evidence: 'quantifier spreading' [Inhelder and Piaget, 1958, Philip, 1995]
  (children until ~10yrs don't reliably constrain restrictor with noun, etc.).
- Disjunction is similarly difficult:
  Every line begins with at least 1 space or contains at least 2 dashes.
  Can be translated to conjunction using de Morgan's law:
  No line begins with less than 1 space and contains less than 2 dashes.
  Yields downward-entailing quantifiers, requiring conditional matching.
- Other phenomena? Evaluation shows no coverage/learnability gaps.

Dependency graph composition: lexical items

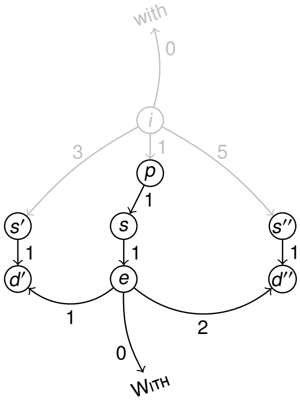
Semantics here extends categorial grammar of [Nguyen et al., 2012]. . .

Lexical items associate syntactic arguments with semantic arguments:

$$x \;\Rightarrow\; u\varphi_1...\varphi_n : \lambda_i\, (\mathbf{f_0}\ i){=}x \\ \wedge\, (\mathbf{f_0}\ (\mathbf{f_1}\ (\mathbf{f_1}\ (\mathbf{f_1}\ i)))){=}x \\ \wedge\, (\mathbf{f_1}\ (\mathbf{f_1}\ (\mathbf{f_1}\ (\mathbf{f_1}\ i)))){=}(\mathbf{f_1}\ (\mathbf{f_3}\ i)) \wedge \ldots \\ \wedge\, (\mathbf{f}_n\ (\mathbf{f_1}\ (\mathbf{f_1}\ (\mathbf{f_1}\ i)))){=}(\mathbf{f_1}\ (\mathbf{f}_{2n+1}\ i))$$

For example:

$$\text{with} \;\Rightarrow\; \textbf{A-aN-bN} : \lambda_i\, (\mathbf{f_0}\ i){=}\text{with} \\ \wedge\, (\mathbf{f_0}\ (\mathbf{f_1}\ (\mathbf{f_1}\ (\mathbf{f_1}\ i)))){=}\textsc{With} \\ \wedge\, (\mathbf{f_1}\ (\mathbf{f_1}\ (\mathbf{f_1}\ (\mathbf{f_1}\ i)))){=}(\mathbf{f_1}\ (\mathbf{f_3}\ i)) \\ \wedge\, (\mathbf{f_2}\ (\mathbf{f_1}\ (\mathbf{f_1}\ (\mathbf{f_1}\ i)))){=}(\mathbf{f_1}\ (\mathbf{f_5}\ i)).$$



Argument composition: constrain nuclear scope

Arguments apply constraints of predicates to nuclear scope of arguments:

$$d : g \quad c\text{-}\mathbf{a}d : h \Rightarrow c : \lambda_i \, (g \, (\mathbf{f}_{2n} \; i)) \wedge (h \; i) \wedge (\mathbf{f}_{2n+1} \; i){=}(\mathbf{f_2} \, (\mathbf{f_1}, (\mathbf{f}_{2n} \; i))) \qquad \text{(Aa)}$$

$$c\text{-}\mathbf{b}d : g \quad d : h \Rightarrow c : \lambda_i \, (g \, i) \wedge (h \; (\mathbf{f}_{2n} \; i)) \wedge (\mathbf{f}_{2n+1} \; i){=}(\mathbf{f_2} \, (\mathbf{f_1} \, (\mathbf{f}_{2n} \; i))) \qquad \text{(Ab)}$$

For example:

$$\dfrac{\dfrac{\text{with}}{\textbf{A-aN-bN} : \lambda_i \, (\mathbf{f_0} \; i){=}\text{with}} \quad \dfrac{\text{a number}}{\textbf{N} : \lambda_i \, (\mathbf{f_0} \; i){=}\text{num}}}{\textbf{A-aN} : \lambda_i \, (\mathbf{f_0} \; i){=}\text{with} \wedge (\mathbf{f_0} \; (\mathbf{f_4} \; i)){=}\text{num} \wedge (\mathbf{f_5} \; i){=}(\mathbf{f_2} \; (\mathbf{f_1} \; (\mathbf{f_4} \; i)))} \; \text{Ab}$$

Modifier composition: constrain restrictor

Modifiers apply constraints of modifier to restrictor of modificand:

$$u\text{-}\mathbf{a}d : g \quad c : h \Rightarrow c : \lambda_j\ \exists_i\ (\mathbf{f_2}\ i){=}j \wedge (g\ i) \wedge (h\ j) \wedge (\mathbf{f_3}\ i){=}(\mathbf{f_1}\ (\mathbf{f_1}\ (\mathbf{f_2}\ i))) \quad \text{(Ma)}$$

$$c : g \quad u\text{-}\mathbf{a}d : h \Rightarrow c : \lambda_j\ \exists_i\ (\mathbf{f_2}\ i){=}j \wedge (g\ j) \wedge (h\ i) \wedge (\mathbf{f_3}\ i){=}(\mathbf{f_1}\ (\mathbf{f_1}\ (\mathbf{f_2}\ i))) \quad \text{(Mb)}$$

For example:

$$\dfrac{\dfrac{\text{lines}}{\mathbf{N} : \lambda_i\ (\mathbf{f_0}\ i){=}\text{lines}} \quad \dfrac{\text{with a number}}{\mathbf{A\text{-}aN} : \lambda_i\ (\mathbf{f_0}\ i){=}\text{with ...}}}{\mathbf{N} : \lambda_i\ (\mathbf{f_0}\ i){=}\text{lines} \wedge \exists_j\ (\mathbf{f_0}\ j){=}\text{with ...} \wedge\ (\mathbf{f_2}\ j){=}i \wedge (\mathbf{f_3}\ j){=}(\mathbf{f_1}\ (\mathbf{f_1}\ (\mathbf{f_2}\ j)))}\ \text{Mb}$$

Scope dependencies calculated (non-incrementally)

First define a partition of the set of group referents in a sentence into sets $\{s, s', s''\}$ of referents s whose iterators $(\mathbf{f_1}\ s)$ are connected by semantic dependencies.

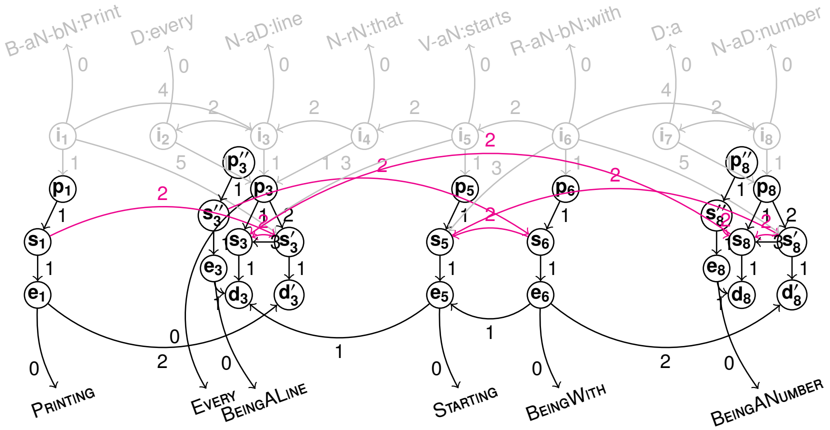
Construct scope dependencies from these partitions using a greedy algorithm:

1. start with an arbitrary referent from this partition
2. select the highest-ranked referent of that partition that is not yet attached
3. designate it as the new highest-scoping referent in that partition
4. attach it as outscoping the previous highest-scoping referent (if exists)
5. if referent has superset/subset that was not yet a highest-scoping referent:
   - switch to the partition of superset/subset referent and carry on
6. if referent has superset/subset referent that *is* the highest-scoping referent:
   - connect it to its subset/superset with a scope dependency
   - merge the two referents' partitions

Eventually you'll have one partition of connected scope dependencies.

Complete representation

Automatically generated from categorial grammar [Nguyen et al., 2012]:



Evaluation

Any coverage or learnability gaps?

Compare model predictions to [Manshadi and Allen, 2011] scripting corpus:

Print [$_1$ *every line*] *that starts with* [$_2$ *a number*] .

scoping relations: $1 > 2$

Nice domain b/c quantifiers are frequent and natural!

350 training sentences, 94 non-duplicate test sentences.

Then introduce lexicalization into preference rankings using training data:

- bilexical weights based on frequency $\tilde{\mathsf{F}}(h, h')$ head h' outscoped by h (e.g. *lines* often outscoped by *files*, b/c files contain multiple lines)

Results

Per-sentence scope accuracy (perfect recall), given gold-standard parse:

System	AR
This system, w/o lexicalization	60*
[Manshadi and Allen, 2011] baseline	63
[Manshadi et al., 2013]	72
This system, w. lexicalization	**72***

* statistically significant difference ($p = 0.001$ by two-tailed McNemar's test)

Lexicalized system gets about state of the art accuracy!

Conclusion

Cognitive compositional semantics using continuation dependencies

- seems to make reassuring predictions:
  - conjunct matching is easy, even in presence of quantifiers
  - quantifier upward/downward entailment is hard
  - disjunction is as hard as quantifier upward/downward entailment
- empirical evaluation shows no coverage or learnability gaps

Future work:

- incremental interpreter, similar to [van Schijndel and Schuler, 2013]
- this will essentially treat quantifier scope as coreference
- experiments: look for coreference-like behavior in quantifier scope

Thanks!

Bibliography I

Anderson, J. A., Silverstein, J. W., Ritz, S. A., and Jones, R. S. (1977).
Distinctive features, categorical perception and probability learning: Some applications of a neural model.
Psychological Review, 84:413–451.

Barker, C. (2002).
Continuations and the nature of quantification.
Natural Language Semantics, 10:211–242.

Barwise, J. and Cooper, R. (1981).
Generalized quantifiers and natural language.
Linguistics and Philosophy, 4.

Copestake, A., Flickinger, D., Pollard, C., and Sag, I. (2005).
Minimal recursion semantics: An introduction.
Research on Language and Computation, pages 281–332.

Bibliography II

Davidson, D. (1967).
The logical form of action sentences.
In Rescher, N., editor, *The logic of decision and action*, pages 81–94.
University of Pittsburgh Press, Pittsburgh.

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

Hobbs, J. R. (1985).
Ontological promiscuity.
In *Proc. ACL*, pages 61–69.

Howard, M. W. and Kahana, M. J. (2002).
A distributed representation of temporal context.
Journal of Mathematical Psychology, 45:269–299.

Bibliography III

Inhelder, B. and Piaget, J. (1958).
The growth of logical thinking from childhood to adolescence.
Basic Books.

Karttunen, L. (1976).
Discourse referents.
In McCawley, J. D., editor, *Notes from the Linguistic Underground (Syntax and Semantics, vol. 7)*. Academic Press, New York.

Koller, A. (2004).
Constraint-based and graph-based resolution of ambiguities in natural language.
PhD thesis, Universität des Saarlandes.

Bibliography IV

MacCartney, B. and Manning, C. D. (2009).
An Extended Model of Natural Logic.
In *Proceedings of the Eighth International Conference on Computational Semantics*, IWCS-8 '09, pages 140–156. Association for Computational Linguistics.

Manshadi, M. and Allen, J. F. (2011).
Unrestricted quantifier scope disambiguation.
In *Graph-based Methods for Natural Language Processing*, pages 51–59.

Manshadi, M., Gildea, D., and Allen, J. F. (2013).
Plurality, negation, and quantification: Towards comprehensive quantifier scope disambiguation.
In *Proceedings of ACL*, pages 64–72.

Bibliography V

Marr, D. (1971).
Simple memory: A theory for archicortex.
Philosophical Transactions of the Royal Society (London) B, 262:23–81.

McClelland, J. L., McNaughton, B. L., and O'Reilly, R. C. (1995).
Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory.
Psychological Review, 102:419–457.

Murdock, B. (1982).
A theory for the storage and retrieval of item and associative information.
Psychological Review, 89:609–626.

Bibliography VI

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)*, pages 2125–2140, Mumbai, India.

Parsons, T. (1990).
Events in the Semantics of English.
MIT Press.

Philip, W. (1995).
Event quantification in the acquisition of universal quantification.
PhD thesis, Univeristy of Massachusetts.

Bibliography VII

Schuler, W. (2014).
Sentence processing in a vectorial model of working memory.
In *Fifth Annual Workshop on Cognitive Modeling and Computational Linguistics (CMCL 2014)*.

Thomason, R. H. (1980).
A model theory for propositional attitudes.
Linguistics and Philosophy, 4:47–70.

van Schijndel, M. and Schuler, W. (2013).
An analysis of frequency- and recency-based processing costs.
In *Proceedings of NAACL-HLT 2013*. Association for Computational Linguistics.