Project Summary

RI: EAGER: Incremental semantic sentence processing models

Humans are a successful species in large part because they can pass knowledge about the world to one another using linguistic explanations. These explanations can be quite complex, involving generalizations about classes of objects and events that introduce multiple scoped quantifiers and anaphora. Accurate models of how these complex relationships are decoded from natural language could further our understanding of how the brain works, and may someday allow non-programmer domain experts to explain desired products, goals and constraints to machines.

But current broad-coverage sentence processing models are focused primarily on modeling syntax, and make unrealistic assumptions that word sequences are generated without any continuity of referential meaning or any preferences among possible coreference and quantifier scope orderings. The proposed work will develop a more human-like semantic processing model by augmenting an existing incremental parser with a graphical dependency-based adaptation of discourse representation structures, to derive probabilities for generating these dependencies.

The model will be evaluated by comparing its ability to predict existing data sets of self-paced reading times (Bachrach et al., 2009) and eye-tracking fixation durations (Kennedy et al., 2003) with that of a PCFG model based on the same syntactic representation. In addition to this fitting task, the model will also be evaluated as a component of a natural language interface system using standard evaluations of parsing (Collins, 1997), coreference (Pradhan et al., 2011) and quantifier scope disambiguation (Manshadi et al., 2013).

Intellectual merit: The proposed work will be the first to integrate parsing, quantifier scope disambiguation, and coreference resolution into a single incremental probability model. More generally, the proposed work will explore basic questions about the nature of language and thought by connecting established algorithmic-level models of cued association in working memory to current purely computational-level descriptions of the logical structure of natural language sentences. The proposed incremental semantic processing model, along with annotation and evaluation scripts and corrected syntactic and semantic annotations of Wikipedia text, will be distributed free online under open-source and Creative Commons licenses.

Broader impact: A semantic processing model based on the distributional similarity of semantic predicates may help improve parsing, coreference resolution and quantifier scope disambiguation of other kinds of text, and models of semantic processing of explanatory text may facilitate the development of natural language human-machine interfaces for non-programmer domain experts.

This project involves interdisciplinary integration of psycholinguistic theories of human sentence processing into an already interdisciplinary dialogue between linguistics and computer scientists at OSU. Students involved in this project will be trained in computational language modeling and computational memory modeling. Part of this project will support regular reading group meetings to give students outside the project a broader perspective in algorithmic-level memory and processing models in addition to computational-level theories of linguistic semantics. This work will figure prominently in a new computational psycholinguistics course being developed by the principal investigator. The incremental semantic processing model developed in this work will also be introduced as a hands-on unit on deep semantics, memory, and ambiguity resolution in sentence processing in the undergraduate honors psycholinguistics course taught by the principal investigator. Course materials for graduate and undergraduate courses will be made publicly available on the PI's web site.

RI: EAGER: Incremental semantic sentence processing models

1 Introduction

Humans are a successful species in large part because they can pass knowledge about the world to one another using linguistic explanations. Explanatory text can be found in a variety of sources, including encyclopedia articles, textbook chapters, and product descriptions. These explanations can be quite complex, involving generalizations about classes of objects and events that introduce multiple scoped quantifiers and anaphora. For example, the Simple English Wikipedia sentence:

(1) *Most humans have two arms each, coming out of their body just below the neck.*

has two quantifiers, *most* and *two*, and two anaphoric references to variables scoped by these quantifiers, *their* and *the neck*, the latter of which is a bridging anaphor referring to the neck of each human. Accurate models of how these complex relationships are decoded from natural language could further our understanding of how the brain works, and may someday allow non-programmer domain experts to explain desired products, goals and constraints to machines.

But current broad-coverage sentence processing models are focused primarily on modeling syntax, in particular using PCFG surprisal (Hale, 2001; Levy, 2008). This predictor is derived from probabilistic context-free grammars (PCFGs) estimated from syntactically annotated corpora (e.g. Marcus et al., 1993) to calculate incremental estimates of the probability of each word in a sentence. These estimates have been used to predict reading times (Roark et al., 2009; Fossum and Levy, 2012) and isolate the effects of memory operations (Demberg and Keller, 2008; Wu et al., 2010; van Schijndel et al., 2013b) and of various kinds of subcategorization information (van Schijndel et al., 2014) on reading times. These estimates have also been used to guide disfluency detection (Miller, 2009a,b) and filter machine translation output (Schwartz et al., 2011).

Yet despite their syntactic sophistication, PCFG models make unrealistic assumptions that word sequences are generated without any continuity of referential meaning or any preferences among possible coreference and quantifier scope orderings. For example, sentence pairs like:

(2) *Miners often bring a small bird$_i$. The canary$_i$ breathes the same air as the miners.*

have been shown to be difficult to process (Almor, 1999; Almor and Eimas, 2008) because the anaphor *the canary* is more specific than the antecedent *a small bird*. This phenomenon seems to arise from probability distributions over word choice in anaphora, but effects like this would not be predicted by PCFG probabilities because PCFGs assume each word is conditionally independent of its preceding words given only its category. Subtle frequency effects like this would therefore be attributed to noise, making memory effects and other non-frequency effects harder to detect.

The proposed work will develop a more human-like semantic processing model by augmenting an existing incremental parser (Schuler et al., 2010; van Schijndel et al., 2013a) with a graphical dependency-based adaptation of discourse representation structures (Kamp, 1981; Kamp and Reyle, 1993; Schuler and Wheeler, 2014). The proposed semantic processing model will define complete semantic dependency representations of sentences, including quantifier scope and coreference relationships (even those that cross sentence boundaries). The model will then exploit the graphical nature of these dependency representations by estimating the probability of each analysis as the product of the probabilities of its component dependencies, based on the distributional similarity of each dependency's source predicate to the other predicates connected to its destination. This deep semantic dependency representation will also be used to develop semi-automatically annotated training and test corpora for this model.

It is anticipated that the proposed incremental semantic processing model will be useful in broad-coverage psycholinguistic experiments as a means to estimate the effects of the brain's ability to predict future events based on the frequency of past events. These estimated frequency effects can then be removed in analyses of naturalistic psycholinguistic datasets to look for non-frequency effects of memory load, interference and other processing constraints. This is important because previous attempts to isolate memory costs in naturalistic datasets (Demberg and Keller, 2008; Wu et al., 2010; van Schijndel et al., 2013b) have found only weak and sometimes negative costs for recall when controlling for frequency using PCFG surprisal, suggesting that the positive memory cost observed using constructed stimuli (e.g. Chen et al., 2005) may have arisen from frequency-based confounds relating to the semantic or pragmatic strangeness of the constructed sentences. Toward this end, the model will be evaluated by comparing its ability to predict existing data sets of self-paced reading times (Bachrach et al., 2009) and eye-tracking fixation durations (Kennedy et al., 2003) with that of a PCFG model based on the same underlying syntax. In addition to this fitting task, the model will also be evaluated as a component of a natural language interface system using standard evaluations of parsing (Collins, 1997), coreference (Pradhan et al., 2011) and quantifier scope disambiguation (Manshadi et al., 2013).

1.1 Intellectual Merit

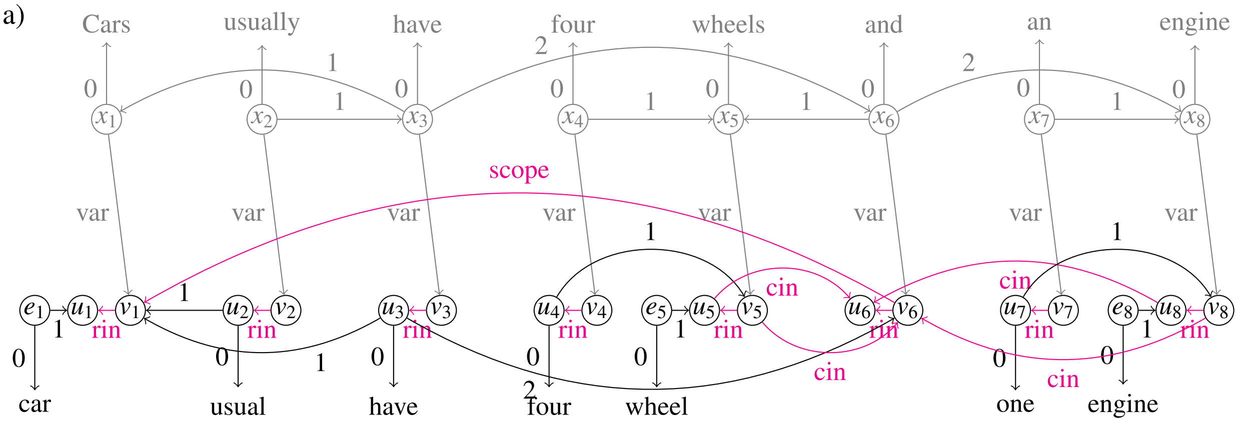
The proposed work will be the first to integrate parsing, quantifier scope disambiguation, and coreference resolution into a single incremental probability model. More generally, the proposed work will explore basic questions about the nature of language and thought by connecting established algorithmic-level models of cued association in episodic memory to current purely computational-level descriptions of the logical structure of natural language sentences. The proposed incremental semantic processing model, along with annotation and evaluation scripts and corrected syntactic and semantic annotations of Wikipedia text, will be distributed free online under open-source and Creative Commons licenses.

1.2 Broader Impacts

The brain is a prediction machine; it hides much of its mechanism behind a heavily optimized capacity to anticipate events based on frequencies of past events. The proposed semantic processing model is intended to account for as many of these frequency effects as possible in order to help uncover the latent mechanistic effects of working memory that remain. A semantic processing model based on the distributional similarity of semantic predicates may help improve parsing, coreference resolution and quantifier scope disambiguation of other kinds of text, and models of semantic processing of explanatory text may facilitate the development of natural language human-machine interfaces for non-programmer domain experts.

This project involves interdisciplinary integration of psycholinguistic theories of human sentence processing into an already interdisciplinary dialogue between linguistics and computer scientists at OSU. Students involved in this project will be trained in computational language modeling and computational memory modeling. Part of this project will support regular reading group meetings to give students outside the project a broader perspective in algorithmic-level memory and processing models in addition to computational-level theories of linguistic semantics. This work will figure prominently in a new computational psycholinguistics course being developed by the principal investigator. The incremental semantic processing model developed in this work will also be introduced as a hands-on unit on deep semantics, memory, and ambiguity resolution in sentence processing in the undergraduate honors psycholinguistics course taught by the PI. Course materials for graduate and undergraduate courses will be made publicly available on the PI's web site.

a)



b) (usual (λ_{u_1} car u_1) (λ_{v_1} (four (λ_{u_5} wheel u_5) (λ_{v_5} have v_1 v_5)) $\wedge$ (one (λ_{u_8} engine u_8) (λ_{v_8} have v_1 v_8))))

Figure 1: (a) Syntactic and semantic dependencies for the sentence, *Cars usually have four wheels and an engine.* Syntactic dependencies between lexical items and 'var' dependencies from lexical items to variables are shown in gray; semantic argument dependencies are shown in black; restriction and conjunction inheritance and scope dependencies are highlighted. (b) Translated logical form (Schuler and Wheeler, 2014).

2 Proposed Model

Most existing text processing systems use a modular pipeline or toolchain architecture (Manning et al., 2014) which feeds results of lower-level processes to higher-level processes with no information propagating back. In order to produce a single probability to use as a predictor, the proposed model is instead 'interactive' in that it calculates probability of parsing decisions based on coreference decisions and vice versa. This interactivity is thought to be cognitively plausible (Marslen-Wilson, 1975; Tanenhaus et al., 1995; Brown-Schmidt et al., 2002) but introduces some engineering challenges, which will be addressed in the proposed research.

The proposed incremental semantic probability model is defined over a graphical adaptation of Discourse Representation Theory (DRT; Kamp, 1981; Kamp and Reyle, 1993). This graphical adaptation uses vertices to define discourse referents, numbered argument dependencies to define predicates, and 'scope' dependencies to define nesting structure in place of DRT boxes (Schuler and Wheeler, 2014). An example dependency graph for the sentence, *Cars usually have four wheels and an engine,* is shown in Figure 1.

This simple semantic graph structure will be associated with with categorial grammar derivations (Nguyen et al., 2012) and incrementally decoded in a probabilistic left-corner processing model, adapted from Schuler et al. (2010) and van Schijndel et al. (2013a). Decoding is divided into a constant set of probabilistic decisions about which semantic dependencies to construct at each word. First, a numbered dependency is constructed between discourse referents for predicate and argument categories. Then an antecedent discourse referent is chosen and connected to the current referent by an inheritance dependency. Then a scope referent is chosen and connected to the current referent by a scope dependency (usually it recently precedes the current word, corresponding to in-situ scoping). Then a word is chosen, depending on the results of the above decisions. Probabilities for each of these decisions are multiplied together to obtain probability estimates for combined structures.

Models for these decisions will be trained on existing syntax and coreference corpora (Marcus et al., 1993; Pradhan et al., 2011), and on a preliminary dataset for quantifier scope developed as part of this project. These models will be trained using logistic regression with features for category labels, graph paths, and predicate-argument dependency contexts as defined in syntax-based distributional semantic representations

(Levy and Goldberg, 2014). Data sparsity, especially with a relatively small quantifier scope dataset, will be addressed through the addition of regularization terms to the logistic models, and if necessary through the use of backoff and smoothing models with simpler features, using parameters trained on held-out data.

3 Plan of Work

The proposed work will consist of the following tasks:

1. development of a deep-semantic sentence processing model;
2. development of a hand-corrected semantically annotated corpus;
3. evaluation of this model on parsing, coreference, and quantifier scope disambiguation tasks; and
4. evaluation of this model on a fixation duration prediction task;

3.1 Development of a deep-semantic sentence processing model

The proposed model will extend an existing incremental parser (van Schijndel et al., 2013b), which calculates a beam at each time step using a best-first search through variable values from hypotheses at the previous time step. The model will be extended to include referents associated with each hypothesis, extending the best-first search to also cover values for variables over semantic dependencies.

Because anaphora may co-refer with any referent in a semantic dependency graph, the implementation will need to store entire graphs on the beam of hypothesized analyses. An interactive parsing and coreference model will therefore need to relax the usual dynamic-programming state merging of a Viterbi-style algorithm, reducing the task to a beam search through the entire exponential search space. This relaxation may seem risky, but preliminary results on an incremental graph-based parser (van Schijndel and Schuler, 2013) show that a similarly relaxed system can achieve statistically significant improvements on an eye-tracking fixation duration prediction task without substantially changing the model's runtime performance.

In order to minimize the storage requirements of this model, semantic dependency graphs associated with hypothesized store states can be efficiently stored together in a unified hash map, which may be cleared at the end of each sentence. Referents in each incomplete category of each store state will then be represented by unique integers, and dependencies will be represented as cued associations in the map from referents and dependency labels to other referents. This will bound the size of the map by the product of the beam width multiplied by the total number of time steps. This method will also be used to store reachable referents in hash maps cued by each referent, merging reachable sets when referents are connected.

If time permits it would be interesting to replace the referent hash map with a more human-like model (Marr, 1971; Anderson et al., 1977; Murdock, 1982; McClelland et al., 1995; Howard and Kahana, 2002) in which cued associations between vectorial states are stored as outer products in an association matrix. This representation has been shown to produce natural recency (decay) effects in a center-embedding task (Schuler, 2014). This modification is therefore expected to produce a natural recency effect for old referents due to interference with new cues stored at the capacity of the association matrix. To prevent the matrix from becoming saturated, this natural model would then use Howard and Kahana (2002) updates, in which targets of new cue vectors are cleared from the association matrix.

This model development will be facilitated by using existing C++ classes for incremental parsing, and is expected to require approximately 8 RA months.

3.2 Development of a hand-corrected semantically annotated corpus

High quality corpora annotated with quantifier scopings are rare. Corpora based on newspaper text (Higgins and Sadock, 2003) contain few examples of complex quantifier scope, and corpora based on more technical

```
(S (N-n1@1 They)
   (V-aN (R-aN usually)
         (V-aN (V-aN-bN have)
               (N (D an)
                  (N-aD (A-aN internal-combustion)
                        (N-aD-s@1 engine))))))
```

Figure 2: Nguyen et al. (2012) generalized categorial grammar tree for the sentence *They usually have an internal-combustion engine,* augmented with restriction inheritance and scope dependencies for coreference and quantifier scope relations.

texts (Manshadi et al., 2011) are relatively small and narrowly defined (e.g. constrained to regular expression programming problems). The proposed work will instead annotate explanatory text from English Wikipedia and Simple English Wikipedia, which has a richer quantifier structure than newspaper text, but is still general enough to be broadly usable.

The model will be trained and evaluated on curated examples of explanatory text from English Wikipedia and Simple English Wikipedia articles annotated with cued-association dependencies for semantic arguments, inheritance, and quantifier scope as described in Section 2. This curated corpus will consist of about 15,000 hand-annotated English Wikipedia sentences adapted from existing corpora (Ytrestøl et al., 2009) and 5,000 Simple English Wikipedia sentences automatically annotated using existing parser (Petrov and Klein, 2007; Nguyen et al., 2012), coreference resolver (Raghunathan et al., 2010), and quantifier scope disambiguation (Schuler and Wheeler, 2014) tools. The English Wikipedia sentences are then automatically reannotated, and the Simple English Wikipedia sentences are hand corrected by trained linguists. This curation will be facilitated by existing tools for viewing graphical representations for complex explanatory text, including explicit dependencies for quantifier scope, coreference, extraposition or heavy shift (e.g. *a different* t_i *day [than Tuesday]*$_i$), and bridging anaphora (e.g. *[Most humans]*$_i$ *have two arms, coming out of the body below the neck* $t_{of\text{-}i}$), which are common in explanatory text. These graphical representations, and categorial grammar projections of these graphs (Nguyen et al., 2012), provide a readable representation for linguistic annotators, which can be directly translated into conventional lambda calculus logical forms (Schuler and Wheeler, 2014). These curated semantic dependency annotations will then be used to train an existing non-interactive parser, coreference resolver, and scope disambiguator tools to automatically annotate a complete Wikipedia dump file, similar to the approach used to automatically generate the HPSG WikiWoods corpus (Flickinger et al., 2010). The curated and automatically annotated corpora will then be used to define an interactive incremental probability model, which will be compared to the non-interactive toolchain on linguistic accuracy and as a predictor of eye-tracking fixation durations (Kennedy et al., 2003).

Syntactic annotations will use an HPSG-like categorial grammar representation that directly maps to the cued-association semantics of Schuler and Wheeler (2014) and can be easily learned by a latent-variable PCFG (Nguyen et al., 2012). Semantic argument and inheritance dependencies will be deterministically derived from Nguyen et al. (2012) generalized categorial grammar trees, annotated with nested parentheses as in the Penn Treebank (Marcus et al., 1993), as shown in Figure 2. Nodes in these trees are labeled with categories, each of which consists of a part-of-speech tag for the head word ('N' for nouns, 'V' for verbs, 'D' for determiners, 'A' for adjectives, etc.) followed by a list of zero or more syntactic arguments unsatisfied within that category, each consisting of an operator ('-a' and '-b' for preceding and succeeding arguments, '-c' and '-d' for preceding and succeeding conjuncts, etc.) followed by another category for the argument. These trees can be automatically translated into semantic argument and inheritance dependencies using an existing set of inference rules (Schuler and Wheeler, 2014). Restriction inheritance dependencies arising from anaphoric pronouns and other coreferences are annotated using '-n' operators followed by zero or more digits for the sentence number (zero if the antecedent is in the current sentence) followed by two digits encoding the word number (no sentences longer than 100 words have been observed in the corpus). Scope dependencies are annotated similarly using '-s' operators.

About 1,500 sentences of Simple English Wikipedia articles have already been annotated by student volunteers. Correction of automatically parsed syntax trees in this format takes approximately two minutes per sentence, with an average sentence length of about ten words. In order to speed annotation, only one annotator will be allocated to each sentence, except for a randomly selected shared subset in which corpus annotation accuracy will be estimated using kappa agreement measures. Pilot results for automatic annotation of a small (300 sentence) initial gold-standard test corpus from Simple English Wikipedia show an 87% accuracy using the Petrov and Klein (2007) parser with the Nguyen et al. (2012) generalized categorial grammar, trained out-of-domain on the Wall Street Journal section of the Penn Treebank (Marcus et al., 1993), with a sentence accuracy rate of almost 50%. Pilot curation of these Simple English Wikipedia sentences by student volunteers suggests that hand correction of this kind of data is indeed relatively straightforward, with most errors attributable to modifier- and conjunction-scoping ambiguities which can be corrected by dragging strings of brackets in a text editor.

Based on pilot data, this automatic annotation and hand correction is expected to require approximately 3 RA months distributed over the length of the project, focusing mainly on quantifier scope and related phenomena like negation.

3.3 Evaluation on parsing, coreference, and quantifier scope disambiguation tasks

The proposed semantic dependency model calculates dependency probabilities based on compatibility (similarity) of predicate contexts from chains of inheritance dependencies. Since ambiguous attachments will differ in these dependencies, and therefore in these predicate contexts, the model will naturally be able to distinguish attachments based on selectional preferences of heads and dependents. For example, in processing the sentence:

(3) *They closed the building until the police arrived.*

in which the prepositional phrase *until the police arrived* may attach either to *closed* or to *building*, the model will compare the predicate context $\{until_1\}$, for the first argument of *until*, with the predicate context $\{close_0\}$, for the closing eventuality, and with the predicate context $\{building_1, close_2\}$, for the referent of *the building*. In this example, the similarity to $\{close_0\}$ is expected to be greater.

As with parsing attachments, the proposed semantic dependency model will also assign different probabilities to coreference decisions based on the similarity of predicate contexts of inheritance chains. This means the processing model will also function as an incremental probabilistic coreference resolver. For example, in processing the sentence pair:

(4) *People were scared of comets. They did not know what comets were.*

the model will compare the predicate context $\{know_1\}$, for the first argument of the knowing eventuality, with the predicate context $\{people_1, scare_2\}$ for the referent of *people* and $\{comet_1, scare_1\}$ for the referent of *comets*. In this example, the similarity to $\{people_1, scare_2\}$ is expected to be greater.

The linguistic accuracy of the model will then be evaluated on parsing and coreference resolution tasks, matching to gold-standard annotations in the English Wikipedia and Simple English Wikipedia corpora developed in Section 3.2. Significance of improvement over a non-interactive toolchain baseline will be evaluated using whole-sentence accuracy or bootstrap resampling on dependency accuracy.

The scope portion of the generated semantic graphs will also be evaluated on a standard quantifier scope disambiguation task (Manshadi et al., 2013), similar to previous work supervised by the principal investigator (Schuler and Wheeler, 2014).

If time permits, the model will also be evaluated on phenomenon-specific tasks to ascertain weaknesses in coverage. These will include examples of unbounded dependencies, similar to (Rimell et al., 2009; Nivre et al., 2010) but using explanatory text data collected as described in Section 3.2, and examples of extraposition and heavy shift, which seem to be very common in Wikipedia explanatory text. Corrected corpora and evaluation scripts for these experiments may also be offered as shared tasks at upcoming SemEval meetings.

This is expected to require approximately 3 RA months, consisting mostly of modifications to the model following evaluations on development data.

3.4 Evaluation of this model on a fixation duration prediction task

The final evaluation of the model will be on psycholinguistic accuracy in predicting reading time and fixation durations, similar to previous experiments supervised by the principal investigator (Wu et al., 2010; van Schijndel et al., 2013b; van Schijndel and Schuler, 2013) and others (Demberg and Keller, 2008; Boston et al., 2008; Roark et al., 2009; Fossum and Levy, 2012). This will involve constructing a credible baseline including commonly used predictors for sentence position, word length and whether the previous word was fixated, as well as latent-variable PCFG surprisal and *n*-gram surprisal (e.g. based on Google *n*-grams with Kneser-Ney smoothing), then comparing this baseline with and without surprisal based on the proposed semantic processing model, trained as described in Section 3.2. The model will be evaluated on existing datasets, including the Dundee corpus (Kennedy et al., 2003), the Bachrach et al. (2009) corpus, and recently-collected MEG data for spoken narratives obtained from Brian Murphy at Queen's University Belfast.

If a tighter fit to these observations is obtained using this semantic processing model, the model will then be used as a stronger control to isolate effects of integration operations in human sentence processing (operations in which long-distance dependencies are finally integrated; Gibson, 2000). This is important because previous attempts to isolate integration cost in naturalistic datasets (Demberg and Keller, 2008; Wu et al., 2010; van Schijndel et al., 2013b) have found only weak and sometimes negative integration cost when controlling for frequency using PCFG surprisal, suggesting that the positive integration cost observed using constructed stimuli (e.g. Chen et al., 2005) may have arisen from frequency-based confounds relating to the semantic or pragmatic strangeness of the constructed sentences. This experiment will compare the fit of the stronger model with and without additional predictors for embedding depth (Wu et al., 2010) and integration cost calculated by the depth of the store in a left-corner parse (van Schijndel et al., 2013b).

This is expected to require approximately 1 RA month, consisting mostly of modifications to the model following evaluations on development data.

If these predictors produce statistically significantly better fits, this will suggest that syntactic dependencies nested in center embedded sentences are observable in reading times, fixation durations, or neural oscillations, which will potentially create a means by which theories about human dependency structures and parsing strategies can be directly empirically compared. For example, robust observable memory effects could help distinguish serial and parallel theories of sentence processing if such effects are observed when dispreferred analyses of ambiguous sentences contain embeddings but preferred analyses do not. This kind of model could also be used to test surprisal as an account of anaphor choice, as a generalization of the Information Load Hypothesis (Almor, 1999). However, since they depend on a series of successful prior results, these experiments are left for future work.

4 Results from Prior NSF Support

Award 0447685, "CAREER: Integrating Denotational Meaning into Probabilistic Language Models," Schuler

This project explored the use of interactive interpretation (using model-theoretic extensions) in spoken language interface. This research produced: a simple formulation of incomplete constituents for incremental interpretation based on the left-right dual of an existing left-corner transform (Johnson, 1998); an empirical result that the vast majority of the Penn Treebank can be recognized within attested human memory bounds of three to four elements in the formulation predicted by this model, published in the journal *Computational Linguistics* (Schuler et al., 2010); a simple formulation of a memory-bounded incremental parser as a special case of an existing factored time-series model, in particular a Hierarchic Hidden Markov Model (Murphy and Paskin, 2001); a simple formulation of this model using a right-corner transform defined on derived probabilistic context-free grammars (PCFGs), yielding equivalent performance to the original PCFG (Schuler, 2009); an implementation (and proof-by-existence) of a real-time interactive speech interpreter achieving better accuracy than a trigram model compiled from an environment model, published in the journal *Computational Linguistics* (Schuler et al., 2009); a simple formulation of speech repair based on incomplete constituents derived from a right-corner transform, and empirical results that this model is better than equivalent (unsmoothed and unlexicalized) models (Miller and Schuler, 2008; Miller, 2009a,b); and empirical results that syntactic surprisal derived from the HHMM parser described above correlates with observed reading times (Wu et al., 2010).

References

Almor, A. (1999). Noun-phrase anaphora and focus: The informational load hypothesis. *Psychological Review*, 106(4):748–765.

Almor, A. and Eimas, P. D. (2008). Focus and noun phrase anaphors in spoken language comprehension. *Language and cognitive processes*, 23(2):201–225.

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.

Bachrach, A., Roark, B., Marantz, A., Whitfield-Gabrieli, S., Cardenas, C., and Gabrieli, J. D. (2009). Incremental prediction in naturalistic language processing: An fMRI study.

Boston, M. F., Hale, J. T., Kliegl, R., Patil, U., and Vasishth, S. (2008). Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam Sentence Corpus. *Journal of Eye Movement Research*, 2(1):1–12.

Brown-Schmidt, S., Campana, E., and Tanenhaus, M. K. (2002). Reference resolution in the wild: Online circumscription of referential domains in a natural interactive problem-solving task. In *Proceedings of the 24th Annual Meeting of the Cognitive Science Society*, pages 148–153, Fairfax, VA.

Chen, E., Gibson, E., and Wolf, F. (2005). Online syntactic storage costs in sentence comprehension. *Journal of Memory and Language*, 52(1):144–169.

Collins, M. (1997). Three generative, lexicalised models for statistical parsing. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics (ACL'97)*.

Demberg, V. and Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, 109(2):193–210.

Flickinger, D., Oepen, S., and Ytrestøl, G. (2010). Wikiwoods: Syntacto-semantic annotation for english wikipedia. In Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., Rosner, M., and Tapias, D., editors, *LREC*. European Language Resources Association.

Fossum, V. and Levy, R. (2012). Sequential vs. hierarchical syntactic models of human incremental sentence processing. In *Proceedings of CMCL 2012*. Association for Computational Linguistics.

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In *Image, language, brain: Papers from the first mind articulation project symposium*, pages 95–126, Cambridge, MA. MIT Press.

Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. In *Proceedings of the second meeting of the North American chapter of the Association for Computational Linguistics*, pages 159–166, Pittsburgh, PA.

Higgins, D. and Sadock, J. M. (2003). A machine learning approach to modeling scope preferences. *Computational Linguistics*, 29(1):73–96.

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

Johnson, M. (1998). Finite state approximation of constraint-based grammars using left-corner grammar transforms. In *Proceedings of COLING/ACL*, pages 619–623, Montreal, Canada.

Kamp, H. (1981). A theory of truth and semantic representation. In Groenendijk, J. A. G., Janssen, T. M. V., and Stokhof, M. B. J., editors, *Formal Methods in the Study of Language: Mathematical Centre Tracts 135*, pages 277–322. Mathematical Center, Amsterdam.

Kamp, H. and Reyle, U. (1993). *From Discourse to Logic*. Kluwer, Dordrecht.

Kennedy, A., Pynte, J., and Hill, R. (2003). The Dundee corpus. In *Proceedings of the 12th European conference on eye movement*.

Levy, O. and Goldberg, Y. (2014). Dependency-based word embeddings. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014, June 22-27, 2014, Baltimore, MD, USA, Volume 2: Short Papers*, pages 302–308.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3):1126–1177.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J. R., Bethard, S., and McClosky, D. (2014). The stanford corenlp natural language processing toolkit. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014, June 22-27, 2014, Baltimore, MD, USA, System Demonstrations*, pages 55–60.

Manshadi, M., Allen, J. F., and Swift, M. (2011). A corpus of scope-disambiguated English text. In *Proceedings of ACL*, pages 141–146.

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.

Marcus, M. P., Santorini, B., and Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: the Penn Treebank. *Computational Linguistics*, 19(2):313–330.

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

Marslen-Wilson, W. D. (1975). Sentence perception as an interactive parallel process. *Science*, 189(4198):226–228.

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.

Miller, T. (2009a). Improved syntactic models for parsing speech with repairs. In *Proceedings of the North American Association for Computational Linguistics*, Boulder, CO.

Miller, T. (2009b). Word buffering models for improved speech repair parsing. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 737–745, Singapore. Association for Computational Linguistics.

Miller, T. and Schuler, W. (2008). A syntactic time-series model for parsing fluent and disfluent speech. In *Proceedings of the 22nd International Conference on Computational Linguistics (COLING'08).*

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

Murphy, K. P. and Paskin, M. A. (2001). Linear time inference in hierarchical HMMs. In *Proceedings of Neural Information Processing Systems*, pages 833–840, Vancouver, BC, Canada.

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.

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.

Petrov, S. and Klein, D. (2007). Improved inference for unlexicalized parsing. In *Proceedings of NAACL HLT 2007*, pages 404–411, Rochester, New York. Association for Computational Linguistics.

Pradhan, S., Ramshaw, L., Marcus, M., Palmer, M., Weischedel, R., and Xue, N. (2011). Conll-2011 shared task: Modeling unrestricted coreference in ontonotes. In *Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task*, CONLL Shared Task '11, pages 1–27, Stroudsburg, PA, USA. Association for Computational Linguistics.

Raghunathan, K., Lee, H., Rangarajan, S., Chambers, N., Surdeanu, M., Jurafsky, D., and Manning, C. D. (2010). A multi-pass sieve for coreference resolution. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, EMNLP 2010, 9-11 October 2010, MIT Stata Center, Massachusetts, USA, A meeting of SIGDAT, a Special Interest Group of the ACL*, pages 492–501.

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

Roark, B., Bachrach, A., Cardenas, C., and Pallier, C. (2009). Deriving lexical and syntactic expectation-based measures for psycholinguistic modeling via incremental top-down parsing. *Proceedings of the 2009 Conference on Empirical Methods in Natural Langauge Processing*, pages 324–333.

Schuler, W. (2009). Parsing with a bounded stack using a model-based right-corner transform. In *Proceedings of NAACL/HLT 2009*, NAACL '09, pages 344–352, Boulder, Colorado. Association for Computational Linguistics.

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

Schuler, W., AbdelRahman, S., Miller, T., and Schwartz, L. (2010). Broad-coverage incremental parsing using human-like memory constraints. *Computational Linguistics*, 36(1):1–30.

Schuler, W. and Wheeler, A. (2014). Cognitive compositional semantics using continuation dependencies. In *Third Joint Conference on Lexical and Computational Semantics (*SEM'14)*.

Schuler, W., Wu, S., and Schwartz, L. (2009). A framework for fast incremental interpretation during speech decoding. *Computational Linguistics*, 35(3):313–343.

Schwartz, L., Callison-Burch, C., Schuler, W., and Wu, S. T. (2011). Incremental syntactic language models for phrase-based translation. In *The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference, 19-24 June, 2011, Portland, Oregon, USA*, pages 620–631.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., and Sedivy, J. C. E. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268:1632–1634.

van Schijndel, M., Exley, A., and Schuler, W. (2013a). A model of language processing as hierarchic sequential prediction. *Topics in Cognitive Science*, 5(3):522–540.

van Schijndel, M., Nguyen, L., and Schuler, W. (2013b). An analysis of memory-based processing costs using incremental deep syntactic dependency parsing. In *Proc. of CMCL 2013*. Association for Computational Linguistics.

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

van Schijndel, M., Schuler, W., and Culicover, P. W. (2014). Frequency effects in the processing of unbounded dependencies. In *Proc. of CogSci 2014*. Cognitive Science Society.

Wu, S., Bachrach, A., Cardenas, C., and Schuler, W. (2010). Complexity metrics in an incremental right-corner parser. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL'10)*, pages 1189–1198.

Ytrestøl, G., Flickinger, D., and Oepen, S. (2009). Extracting and annotating wikipedia sub-domains. towards a new escience community resource. In *Proceedings of the Seventh International Workshop on Treebanks and Linguistic Theories*.