A Bottom-Up Parsing Model of Local Coherence Effects

Emily Morgan (emily@ling.ucsd.edu)
Department of Linguistics, 9500 Gilman Drive #108
La Jolla, CA 92093, USA

Frank Keller (keller@inf.ed.ac.uk)
Mark Steedman (steedman@inf.ed.ac.uk)
School of Informatics, 10 Crichton Street
Edinburgh EH8 9AB, UK

Abstract

Human sentence processing occurs incrementally. Most models of human processing rely on parsers that always build connected tree structures. But according to the theory of Good Enough parsing (Ferreira & Patson, 2007), humans parse sentences using small chunks of local information, not always forming a globally coherent parse. This difference is apparent in the study of local coherence effects (Tabor, Galantucci, & Richardson, 2004), wherein a locally plausible interpretation interferes with the correct global interpretation of a sentence. We present a model that accounts for these effects using a wide-coverage parser that captures the idea of Good Enough parsing. Using Combinatory Categorial Grammar, our parser works bottom-up, enforcing the use of local information only. We model the difficulty of processing a sentence in terms of the probability of a locally coherent reading relative to the probability of the globally coherent reading of the sentence. Our model successfully predicts psycholinguistic results.

Keywords: sentence processing; parsing complexity; local coherence; Good Enough parsing; Combinatory Categorial Grammar

Introduction

A major topic of inquiry in cognitive science is the process by which people produce and comprehend sentences. Human sentence processing is known to proceed incrementally: people construct syntactic and semantic interpretations gradually as a sentence unfolds, rather than waiting until after the whole sentence has been received. But although we know that syntactic information becomes available progressively while comprehending a sentence, it is still an open question to what extent decisions made early in the parsing process can constrain later decisions.

One phenomenon that can shed light on this question is local coherence effects. Local coherence effects arise when a sentence includes a substring with a plausible local interpretation that is incompatible with the global interpretation. (In other words, the interpretation is merely locally coherent, but not globally coherent.) A typical example (from Tabor, Galantucci, & Richardson, 2004) is:

(1) **A/R**: The coach smiled at the player tossed a frisbee.

A typical reader, seeing this sentence for the first time, will find it difficult to understand and will likely judge it to be ungrammatical. But this difficulty is unexpected in light of similar sentences:

(2) **U/R**: The coach smiled at the player thrown a frisbee.

(3) **A/U**: The coach smiled at the player who was tossed a frisbee.

(4) **U/U**: The coach smiled at the player who was thrown a frisbee.

These four sentences, all intended to be close paraphrases of one another, illustrate a puzzle: while the majority of readers reject (1), they accept (3) and (4), with mixed results for (2). These sentence differ on two dimensions: the past participle can be Ambiguous (such as *tossed*, which can be a past participle or a past tense form) or Unambiguous (such as *thrown*), and the relative clause can be Reduced (without *who was*) or Unreduced (with *who was*). Neither of these alternations generally changes the grammaticality of a sentence, so we would naively predict that if (4) is acceptable, then (1) is as well. Our challenge is to explain why this naive prediction is wrong. Intuitively, it seems that the local coherence of the substring *the player tossed a frisbee* in (1) as a plausible complete sentence is distracting from its globally correct interpretation as an object with a relative clause.

Tabor, Galantucci, and Richardson demonstrate the existence of local coherence effects as a psycholinguistic phenomenon in two different studies: in the first, they find increased reading times at the ambiguous past participle in (1). They present subjects with sentences from 20 sets of sentences like those seen above and measure reading times for each word using self-paced reading. In this methodology, longer reading times are taken to indicated increased processing difficulty. As expected based on previous studies (e.g. Ferreira & Clifton, 1986), they find substantially increased reading times for the Reduced cases as compared to the Unreduced cases, both on the past participle (e.g. *tossed*) and on the following word. Moreover, they find an unexpected interaction between Ambiguity and Reducedness: while the A/U reading times are not significantly different from the U/U reading times, the A/R reading times are substantially increased relative to the U/R reading times. This superadditive difficulty of the A/R condition is the signature of a local coherence effect.

In the second experiment, Tabor, Galantucci, and Richardson replicate the first using a grammaticality judgement task.

Stellan Ohlsson and Richard Catrambone, eds., *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, 1559–1564. Portland, 2010.

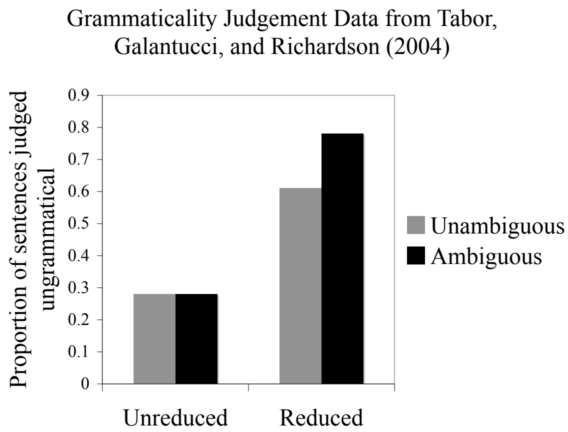


Figure 1: Grammaticality judgement data from Tabor, Galantucci, and Richardson (2004). The signature of a local coherence effect is the superadditive proportion of ungrammatical judgements in the Ambiguous/Reduced condition.

They find decreased acceptance of Reduced sentences as grammatical, with an interaction between Ambiguity and Reducedness such that A/R sentences are judged unacceptable superadditively often (see Figure 1). Once again, decreased acceptability judgements are taken to indicate processing difficulty.

Note that sentences in the A/R condition are not just standard garden path sentences. In a standard garden path sentence, the disambiguating information comes after the reader has already been led astray. In contrast, in sentences such as (1), the disambiguating information comes at the beginning of the sentence. Thus the reader in theory already knows that *tossed* cannot be a past tense form and must be a past participle. Yet despite that, these sentences cause processing difficulty.

A model of human sentence processing should be able to predict the difficulty of sentences with local coherence effects. However, most existing models cannot. In particular, most standard theories of parsing assume that that all accrued knowledge from the parsing process is taken into account at all times. Models following this assumption can straightforwardly account for standard garden paths because there is nothing inconsistent about initially misinterpreting a sentence before having access to the disambiguating information. But these models cannot take the same position in accounting for local coherence effects: when the disambiguating information has already been seen and *smiled* has already been recognized as the main verb of the sentence, they cannot entertain the inconsistent possibility that *tossed* is also a main verb. Computational implementations of wide-coverage parsers generally also make this assumption of global consistency (e.g. Roark, 2001; Sturt, Costa, Lombardo, & Frasconi, 2003; Demberg & Keller, 2008). For many applications, this assumption may be convenient. But for a parser to be credible as a model of human sentence processing, it must be able to predict these psycholinguistic effects, which requires relaxing this assumption.

An alternate theory of sentence processing is Ferreira and colleagues' *Good Enough* (GE) parsing. Ferreira and Patson (2007) describe GE parsing:

> People compute local interpretations that are sometimes inconsistent with the overall sentence structure, indicating that the comprehension system tries to construct interpretations over small numbers of adjacent words whenever possible and can be lazy about computing a more global structure and meaning.

The GE theory of parsing asserts that people do not construct full representations for sentences the majority of the time. Rather, they construct just enough to complete the task at hand, only constructing a further representation if necessary. Moreover, because people base their first-pass constructions on local information and generally construct only partial parse trees, these partial parses may contradict one another. A GE parsing account can thus easily account for local coherence effects. We will develop a computational model of why local coherence effects arise within the framework of GE parsing.

Previous Models of Local Coherence Effects

Two models have previously attempted to account for local coherence effects: Levy (2008) uses a noisy-channel model to argue that because there is uncertainty in linguistic input, the parse of a sentence should be modeled as a probability distribution over a set of candidate sentences (including the intended sentence and its near-neighbors). Given such a probability distribution, the effect of reading each word can be modeled and quantified in terms of a belief update. Levy predicts that a larger change in beliefs will correspond to greater processing difficulty and longer reading times. This in turn predicts local coherence effects because the rarer sentences provoke larger changes in belief.

Levy's model only considers fully connected and grammatical (partial) parses as candidates, thus it does not capture the intuition of GE parsing. An additional limitation of the model is that due to the computational load of calculating near-neighbors, it has only been implemented using a toy Probabilistic Context Free Grammar (PCFG), rather than a richer, wide-coverage language model.

The other previously existing model of local coherence effects comes from Bicknell and Levy (2009). They again model local coherence effects as arising from belief updates. Specifically, they model them as the consequences of an update from a bottom-up prior belief to a posterior belief that takes top-down information into account. They thus predict processing difficulty in the case of locally coherent substrings because the bottom-up statistics make strong predictions about the category of the substrings, which are then contradicted by top-down information.

This model begins to capture the idea of GE parsing by looking at substrings of different lengths. However, it has no way to integrate the information it receives from these different substring lengths because evaluating these substrings is post hoc, not an actual part of the parsing process. Additionally, like Levy's (2008) model, it has only been implemented using a toy PCFG.

Thus there is currently no general, wide-coverage model of human parsing that implements a GE parsing strategy. Computational models of local coherence effects have instead had to account for the phenomenon indirectly, either through a noisy channel model or by predicting the effects without actually simulating the parsing process, and have been confined to parsing with small toy grammars. We will develop a model to address these shortcomings.

A New Model of Local Coherence Effects

Our goal is to model the process by which local coherence effects emerge as the result of Good Enough parsing, within the context of a wide-coverage parser. In the example sentence *The coach smiled at the player tossed a frisbee*, our intuition is that processing difficulty arises from the locally coherent reading of *the player tossed a frisbee*, which distracts from the globally coherent reading. Our model will capture this intuition by using a strictly bottom-up parser to remove the top-down influence of non-local constraints.

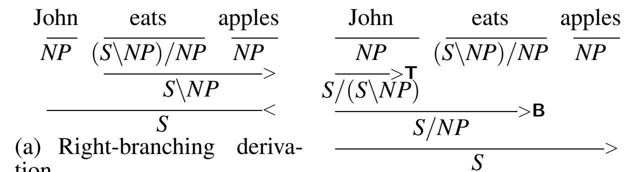
Strictly bottom-up parsing is frequently rejected as a plausible model for human parsing because, it is claimed, it does not allow for incremental interpretation. The standard argument says that a clause can only be interpreted when it is seen in full (i.e., at the end of a constituent). But in a strictly right-branching language, this means that nothing can be interpreted until the very end of the sentence because only then is any constituent completed.

To overcome this objection, our parser uses the Combinatory Categorial Grammar (CCG) formalism to represent linguistic structures. CCG was specifically designed to allow for incremental bottom-up parsing by using a more flexible notion of constituents.

Combinatory Categorial Grammar

Combinatory Categorial Grammar is a grammar formalism based on Categorial Grammar (CG). We base our description of it here on Steedman (2000).

CCG revolves around functional categories and rules for combining them. Categories can be either functions or arguments and are defined recursively: Base categories such as *S* and *NP* represent arguments. Functions are of the form α/β or $\alpha\backslash\beta$, where α and β are categories. To the right of the slash is the argument of the function, and to the left is its result. The direction of the slash indicates the directionality of composition: / means the argument is to the right and \ means the argument is to the left. An English verb phrase, for example, will have the category $S\backslash NP$, indicating that it combines with an NP on its left and results in a sentence. We also allow a finite set of features on our base categories, such



(a) Right-branching derivation
(b) Left-branching derivation

Figure 2: Right- and left-branching CCG derivations for the sentence *John eats apples*. $(S\backslash NP)/NP$ is the CCG category for a transitive verb. Without type-raising, *eats* can only combine with *apples*, yielding the typical right-branching derivation in (a). With type-raising, *John* can combine immediately with *eats*, yielding the left-branching derivation in (b).

as person, number, and gender on *NP*s. These are notated as e.g. $NP[3sf]$.

A CCG derivation uses rules to combine categories. Pure CG relies on two rules, named $>$ and $<$, to combine categories:

(5) $X/Y \quad Y \rightarrow X \quad (>)$

(6) $Y \quad X\backslash Y \rightarrow X \quad (<)$

CCG introduces further combinatory rules that allow for more flexible notions of constituency than other grammar formalisms. In particular, it includes two lexical *type-raising* rules, named $>\mathbf{T}$ and $<\mathbf{T}$:

(7) $X \rightarrow T/(T\backslash X) \quad (>\mathbf{T})$

(8) $X \rightarrow T\backslash(T/X) \quad (<\mathbf{T})$

In these rules—which are here shown in the derivation, but in fact operate in the lexicon—T can be any lexical category taking X as argument. For instance, we could use $>\mathbf{T}$ to type-raise NP to $S/(S\backslash NP)$. Applying this rule limits the other categories the NP can combine with. Intuitively, we can think of the output of this rule as similar to an *NP* with nominative case-marking. It specifies not just that the word or phrase in question is a noun, but that it is a subject which must combine with a predicate.

These type raising rules allow us to parse a sentence incrementally by forming nontraditional constituents, leading to left-branching derivations (see Figure 2). CCG thus allows each new word of the input to be incorporated into the existing constituent structure as it is encountered, which makes incremental bottom-up parsing possible.

The Model

We take a bottom-up CCG parser as the basis of our model of human sentence processing. In order to predict processing difficulty caused by local coherence effects, we need a linking hypothesis to specify the relation between the parser output and psycholinguistic measures such as grammaticality judgements or reading times. Our linking hypothesis should embody the theory of Good Enough parsing, focusing on interpretations of local substrings.

We adapt a model proposed by Jurafsky (1996) to predict garden path effects. To make graded predictions, rather than categorical distinctions, we will adopt a probabilistic framework, and consider the probabilities of various substrings of a sentence. In particular, we could consider either the inside probability $\mathrm{P}(S \rightarrow \text{substring})$ (alternately written as $\mathrm{P}(\text{substring} \mid S)$) or the inverse probability $\mathrm{P}(S \mid \text{substring})$. We do not know of a computationally tractable way to calculate $\mathrm{P}(S \mid \text{substring})$ from our parser. Calculating the inside probability, on the other hand, is a fundamental part of the parsing process. It is most parsimonious to base our model on the inside probabilities that are already being calculated.

Our intuition is that if an incorrect interpretation of a substring is highly plausible relative to the correct interpretation of the sentence, then it will cause processing difficulty. In a sentence such as *The coach smiled at the player tossed a frisbee*, the substring that we expect to cause difficulty is the locally coherent substring *the player tossed a frisbee*. We thus consider the ratio:

$$\frac{\mathrm{P}(S \rightarrow \textit{the player tossed a frisbee})}{\mathrm{P}(S \rightarrow \textit{The coach smiled at the player tossed a frisbee})}$$

In this case, the ratio will be high because *The player tossed a frisbee* is a relatively likely sentence. In the other three cases, the ratio will be low because none of the following are very plausible sentences:

(9) the player thrown a frisbee

(10) the player who was tossed a frisbee

(11) the player who was thrown a frisbee

Although in theory this ratio could be as low as 0, in practice this does not occur because there is generally some (low probability) way to parse each phrase as a sentence. We take this ratio as a measure of processing difficulty.

Implementation

We implement our model using a Combinatory Categorial Grammar parser based on the Cocke-Kasami-Younger (CKY) algorithm. This algorithm was originally developed for Context Free Grammars and uses dynamic programing to parse from the bottom up: given a sentence, it first calculates the probabilities of all ways to generate each word using a rule $X \rightarrow \textit{word}$. For each potential pair of categories X_1 and X_2 that could have generated adjacent words w_1 and w_2, it then calculates the probabilities of all ways to generate that pair using a rule $X_3 \rightarrow X_1 X_2$. This allows us to calculate the inside probability $\mathrm{P}(X_3 \rightarrow w_1 w_2)$. Continuing iteratively, we can calculate the inside probabilities of all substrings of the sentence.

We used a modified version of the StatOpenCCG parser, developed by Christodoulopoulos (2008), which is itself an extension of the OpenCCG parser (White, 2008). StatOpenCCG implements a statistical version of the CKY algorithm which operates using a generative head-dependency model over CCG categories: From the parent (starting with a ROOT node), a head is generated with a certain probability. Then its sisters are generated with probability conditioned on the head category, the sister's direction from the head, and whether it is adjacent to the head. Although the number of CCG categories is theoretically infinite, our parser is constrained to only use categories that have appeared in the training data set. With this constraint, the runtime of the parser is bounded by $O(n^3)$. The parser has been trained on sections 1 through 22 of the CCGbank (Hockenmaier, 2003), a CCG version of the Penn treebank.

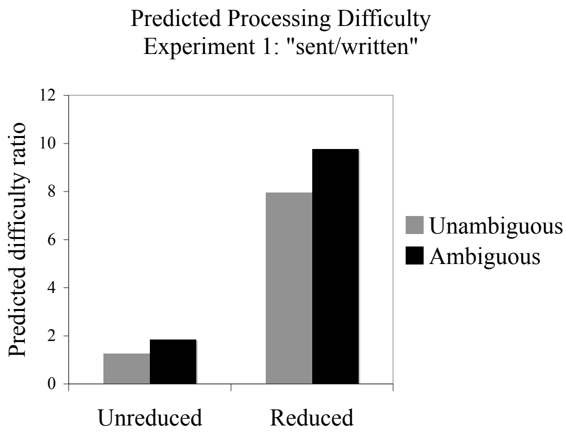
Our experiments use two different lexicons. The first lexicon is that taken from sections 1 through 22 of the CCGbank. However, this lexicon is too small to parse the majority of the sentences we wish to consider. To obtain a larger lexicon, we parsed six months of the New York Times (comprising approximately 50 million word tokens) taken from the Gigaword corpus (Graff, 2003). Sentences from the corpus were passed through the RASP tokenizer (Briscoe, Carroll, & Watson, 2006) and then parsed using the C&C CCG parser (Curran, Clark, & Bos, 2007). This state-of-the-art parser obtains labelled precision of 84.8% and labelled recall of 84.5% on section 23 of the CCGbank. It is extremely fast and provides the best parse accuracy from a CCG parser, making it convenient for obtaining large amounts of data to construct a larger lexicon. (However, it is not a cognitively plausible parser, as it relies on its supertagger and other cognitively implausible tricks to speed its parsing.) From this parsed sample, we extracted the lexicon for use in the StatOpenCCG parser (with the statistical parsing model over categories trained as before on CCGbank data). Although this lexicon of course contains quite a few errors, we verify that it nonetheless parses our test sentences correctly, placing the correct parses among the top results.

Experiments

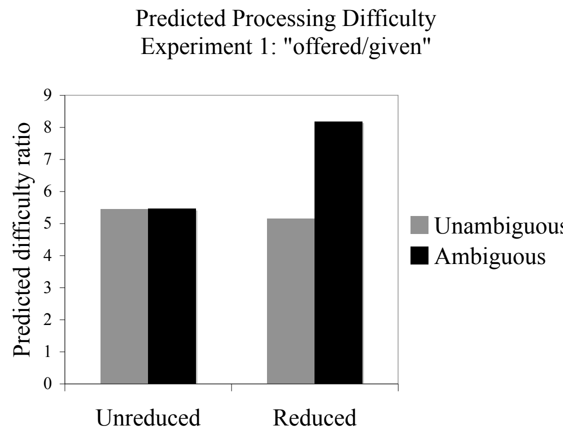
We present two sets of experiments in which we test our model against the results from Tabor, Galantucci, and Richardson (2004). The first uses a small but high-quality lexicon to parse two test cases. The second uses a larger, error-ridden lexicon to parse a larger set of sentences. Recall that Tabor, Galantucci, and Richardson's (2004) study used 20 sets of sentences like those in (1)–(4).

Experiment 1: Test Cases using the CCGbank Lexicon

Because CCGbank is derived from a human-annotated treebank, the quality of the lexicon it yields is high. Nevertheless, it is small in comparison to human lexicons, and the passive relative constructions we are investigating are sparsely represented. In fact, the CCGbank lexicon contains only two words which are unambiguous ditransitive passive participles (i.e., $(S[\text{pss}]\backslash NP)/NP$ but not $(S[\text{dcl}]\backslash NP)/NP$—where [pss] indicates a past participle used in a passive construction, and [dcl] indicates a declarative sentence). These two words are



(a)



(b)

Figure 3: Results from Experiment 1, two test cases using the high-quality CCGbank lexicon. In both sets of sentences, the A/R case displays the correct pattern of superadditive difficulty.

written and *given*. Using these words, we construct two sentence sets, based on sentences used by Tabor, Galantucci, and Richardson:

(12) He questioned a congressman (who was) sent/written a letter.

(13) He addressed the woman (who was) offered/given a beer.

All words in these sentences are in the CCGbank lexicon. We parse them using our high-quality lexicon.

Results For these sentences, we obtain the predicted ratios:

$$\frac{\mathrm{P}(S \rightarrow \text{locally coherent substring})}{\mathrm{P}(S \rightarrow \text{whole sentence})}$$

Results are in Table 1 and Figure 3. We compare our results to the grammaticality judgements from Tabor, Galantucci, and Richardson (see Figure 1).

As we see in Figure 3(a), the set of sentences (12) displays the correct pattern of superadditive difficulty in the A/R case. While there is little difference in difficulty between the A/U and U/U conditions, there is a marked increase to the U/R condition, and a superadditive increase to the A/R condition. This mirrors the pattern seen in Tabor, Galantucci, and Richardson's grammaticality judgements.

We see the same superadditive pattern of difficulty in our results for the set of sentences (13), shown in Figure 3(b). Somewhat surprisingly, the U/R condition is in fact predicted to be marginally easier than the Unreduced sentences in this set. This may be because *given* is an extremely common word. Although it is unambiguous in that it cannot be a past tense, it is in fact a highly ambiguous word, with 18 entries in the CCGbank lexicon. For instance, it can serve as a preposition, as in *Given the weather, I will stay inside today*. Regardless of this slight puzzle, the A/R case displays the correct pattern of superadditive difficulty.

Table 1: Predicted difficulty ratios from all experiments, alongside grammaticality judgements from Tabor, Galantucci, and Richardson (2004).

Type	TG&R	Exp1: written	Exp1: given	Exp2
U/U	.28	1.27	5.45	5.74
A/U	.28	1.85	5.46	8.46
U/R	.61	7.96	5.16	11.60
A/R	.78	9.76	8.18	12.34

Experiment 2: Using the Gigaword Lexicon

Using the Gigaword lexicon, we are able to parse 13 out of the 20 sentences in the Tabor study. (Sentences were excluded only if their past participles were not present in the lexicon. All other vocabulary items are present.) We standardize all sentences to begin with a pronoun. Additionally, for the sake of parsing efficiency, we do not include the *by* phrases that give the agent of the sentence. We further shorten two sentence sets in ways that do not affect the target part of the sentence.

Results Results from Experiment 2 are shown in Table 1 and Figure 4. We compare our results to the grammaticality judgements from Tabor, Galantucci, and Richardson (see Figure 1). We find the correct trend of difficulties, with the A/R condition most difficult, followed by U/R, followed by the two Unreduced cases. We do not find the exact pattern of superadditive difficulty in the A/R case, due to the fact that the A/U case is in fact predicted to be much more difficult than the U/U case, in contrast to the grammaticality ratings. Because the Gigaword lexicon is very error-prone, it is difficult

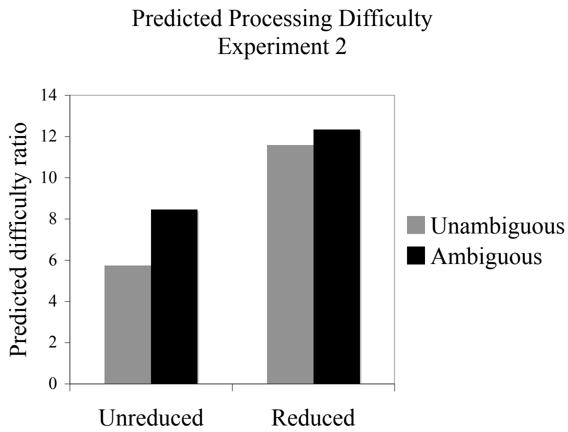


Figure 4: Experiment 2 results. We find the expected pattern of difficulty, but, due to the inflated predicted difficulty of the U/R case, do not see superadditive difficulty in the A/R case.

to draw any firm conclusions from this quirk in our results. However, we note that the A/R case is correctly predicted to be substantially more difficult than either of the Unreduced cases.

Conclusion

We have presented a model of local coherence effects using a wide-coverage bottom-up Combinatory Categorial Grammar parser. Our model can accurately predict which sentences humans will have difficulty in processing; specifically, it predicts the local coherence effects found by Tabor, Galantucci, and Richardson (2004). Our results support the psycholinguistic plausibility of CCG and the Good Enough theory of parsing by demonstrating that a parser that uses bottom-up local information can both perform well as a wide-coverage parser and predict specific psycholinguistic results.

Interestingly, the architecture of our version of the GE parser differs from Ferreira's original proposal. Ferreira (2003) proposes that GE parsing occurs via two separate strategies: one "algorithmic" and one "heuristic". In contrast, our parser does not include this separation: all analyses, both local and global, are produced by a uniform algorithm, and all are heuristically evaluated using the parsing model. This integration of strategies is a strength of our model, as it demonstrates how local coherence effects could emerge naturally as an inherent part of the parsing process.

In future work, we would like to make not just sentence-level predictions but word-by-word reading time predictions. Given that we have an entire parse chart, such predictions should be possible. We are currently choosing inside probabilities from two cells in the parse chart to compare, based on outside knowledge of where processing difficulty is likely to arise. We could do something similar for every cell in the chart, considering the inside probability of the substring it spans relative to the probability of the sentence as a whole. With word by word predictions, we could model reading time data as well as grammaticality judgement data. Such a model would be applicable to a wide range of psycholinguistic data beyond local coherence effects.

Acknowledgments

This work was supported by EU IST Cognitive Systems IP FP6-2004-IST-4-27657 "Paco-Plus".

References

Bicknell, K., & Levy, R. (2009). A model of local coherence effects in human sentence processing as consequences of updates from bottom-up prior to posterior beliefs. In *Proceedings of North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT) 2009 Conference* (pp. 665–673). Boulder, CO: Association for Computational Linguistics.

Briscoe, E., Carroll, J., & Watson, R. (2006). The second release of the RASP system. In *Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions.* Sydney, Australia: Association for Computational Linguistics.

Christodoulopoulos, C. (2008). *Creating a natural logic inference system with Combinatory Categorial Grammar*. Master's thesis, University of Edinburgh.

Curran, J. R., Clark, S., & Bos, J. (2007). Linguistically motivated large-scale NLP with C&C and Boxer. In *Proceedings of the ACL 2007 Demonstrations Session (ACL-07 demo)* (pp. 29–32). Morristown, NJ: Association for Computational Linguistics.

Demberg, V., & Keller, F. (2008). A psycholinguistically motivated version of TAG. In *Proceedings of the 9th International Workshop on Tree Adjoining Grammars and Related Formalisms* (pp. 25–32). Tübingen.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*, 164–203.

Ferreira, F., & Clifton, C., Jr. (1986). The independence of syntactic processing. *Journal of Memory and Language*, *25*, 348–368.

Ferreira, F., & Patson, N. D. (2007). The 'Good Enough' approach to language comprehension. *Language and Linguistics Compass*, *1*(1–2), 71–83.

Graff, D. (2003). *English Gigaword.* Linguistic Data Consortium, Philadelphia. (DVD)

Hockenmaier, J. (2003). *Data and models for statistical parsing with Combinatory Categorial Grammar*. Doctoral dissertation, University of Edinburgh.

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science: A Multidisciplinary Journal*, *20*(2), 137–194.

Levy, R. (2008). A noisy-channel model of rational human sentence comprehension under uncertain input. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP).* Morristown, NJ: Association for Computational Linguistics.

Roark, B. (2001). Probabilistic top-down parsing and language modeling. *Computational Linguistics*, *29*(2), 249–276.

Steedman, M. (2000). *The syntactic process*. Cambridge, MA: The MIT Press.

Sturt, P., Costa, F., Lombardo, V., & Frasconi, P. (2003). Learning first-pass structural attachment preferences with dynamic grammars and recursive neural networks. *Cognition*, *88*, 133–169.

Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language*, *50*, 355–370.

White, M. (2008). *Open CCG: The OpenNLP CCG library.* (http://openccg.sourceforge.net/ [Online; accessed 27-July-2009])