PCFGs with Syntactic and Prosodic Indicators of Speech Repairs

John Hale[a] **Izhak Shafran**[b] **Lisa Yung**[c]
Bonnie Dorr[d] **Mary Harper**[de] **Anna Krasnyanskaya**[f] **Matthew Lease**[g]
Yang Liu[h] **Brian Roark**[i] **Matthew Snover**[d] **Robin Stewart**[j]

[a] Michigan State University; [b,c] Johns Hopkins University; [d] University of Maryland, College Park; [e] Purdue University

[f] UCLA; [g] Brown University; [h] University of Texas at Dallas; [i] Oregon Health & Sciences University; [j] Williams College

Abstract

A grammatical method of combining two kinds of speech repair cues is presented. One cue, prosodic disjuncture, is detected by a decision tree-based ensemble classifier that uses acoustic cues to identify where normal prosody seems to be interrupted (Lickley, 1996). The other cue, syntactic parallelism, codifies the expectation that repairs continue a syntactic category that was left unfinished in the reparandum (Levelt, 1983). The two cues are combined in a Treebank PCFG whose states are split using a few simple tree transformations. Parsing performance on the Switchboard and Fisher corpora suggests that these two cues help to locate speech repairs in a synergistic way.

1 Introduction

Speech repairs, as in example (1), are one kind of disfluent element that complicates any sort of syntax-sensitive processing of conversational speech.

(1) and [the first kind of invasion of] the first type of privacy seemed invaded to me

The problem is that the bracketed *reparandum* region (following the terminology of Shriberg (1994)) is approximately repeated as the speaker "repairs" what he or she has already uttered. This extra material renders the entire utterance ungrammatical—the string would not be generated by a correct grammar of fluent English. In particular, attractive tools for natural language understanding systems, such as Treebank grammars for written corpora, naturally lack appropriate rules for analyzing these constructions.

One possible response to this mismatch between grammatical resources and the brute facts of disfluent speech is to make one look more like the other, for the purpose of parsing. In this separate-processing approach, reparanda are located through a variety of acoustic, lexical or string-based techniques, then excised before submission to a parser (Stolcke and Shriberg, 1996; Heeman and Allen, 1999; Spilker et al., 2000; Johnson and Charniak, 2004). The resulting parse tree then has the reparandum re-attached in a standardized way (Charniak and Johnson, 2001).

An alternative strategy, adopted in this paper, is to use the same grammar to model fluent speech, disfluent speech, and their interleaving.

Such an integrated approach can use syntactic properties of the reparandum itself. For instance, in example (1) the reparandum is an *unfinished* noun phrase, the repair a *finished* noun phrase. This sort of phrasal correspondence, while not absolute, is strong in conversational speech, and cannot be exploited on the separate-processing approach. Section 3 applies metarules (Weischedel and Sondheimer, 1983; McKelvie, 1998a; Core and Schubert, 1999) in recognizing these correspondences using standard context-free grammars.

At the same time as it defies parsing, conversational speech offers the possibility of leveraging prosodic cues to speech repairs. Sec-
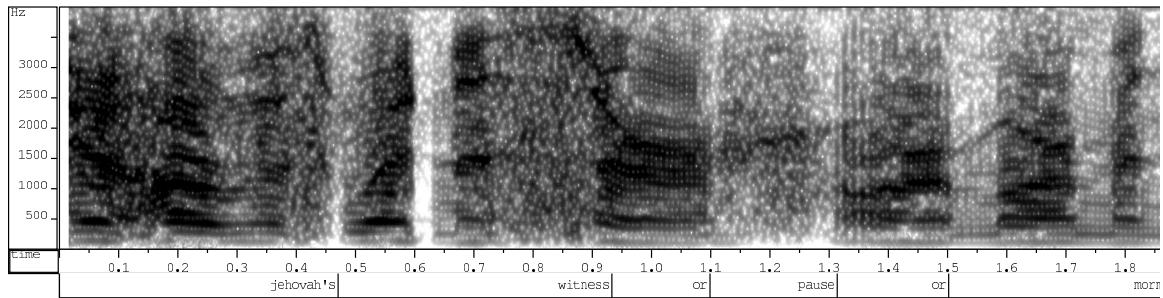
Figure 1: The pause between two *or*s and the glottalization at the end of the first makes it easy for a listener to identify the repair.

tion 2 describes a classifier that learns to label prosodic breaks suggesting upcoming disfluency. These marks can be propagated up into parse trees and used in a probabilistic context-free grammar (PCFG) whose states are systematically split to encode the additional information.

Section 4 reports results on Switchboard (Godfrey et al., 1992) and Fisher EARS RT04F data, suggesting these two features can bring about independent improvements in speech repair detection. Section 5 suggests underlying linguistic and statistical reasons for these improvements. Section 6 compares the proposed grammatical method to other related work, including state of the art separate-processing approaches. Section 7 concludes by indicating a way that string- and tree-based approaches to reparandum identification could be combined.

2 Prosodic disjuncture

Everyday experience as well as acoustic analysis suggests that the syntactic interruption in speech repairs is typically accompanied by a change in prosody (Nakatani and Hirschberg, 1994; Shriberg, 1994). For instance, the spectrogram corresponding to example (2), shown in Figure 1,

(2) the jehovah's witness or [or] mormons or someone

reveals a noticeable pause between the occurrence of the two *or*s, and an unexpected glottalization at the end of the first one. Both kinds of cues have been advanced as explanations for human listeners' ability to identify the reparandum even before the repair occurs.

Retaining only the second explanation, Lickley (1996) proposes that there is no "edit signal" per se but that repair is cued by the absence of smooth

formant transitions and lack of normal juncture phenomena.

One way to capture this notion in the syntax is to enhance the input with a special disjuncture symbol. This symbol can then be propagated in the grammar, as illustrated in Figure 2. This work uses a suffix ~+ to encode the perception of abnormal prosody after a word, along with phrasal -BRK tags to decorate the path upwards to reparandum constituents labeled EDITED. Such
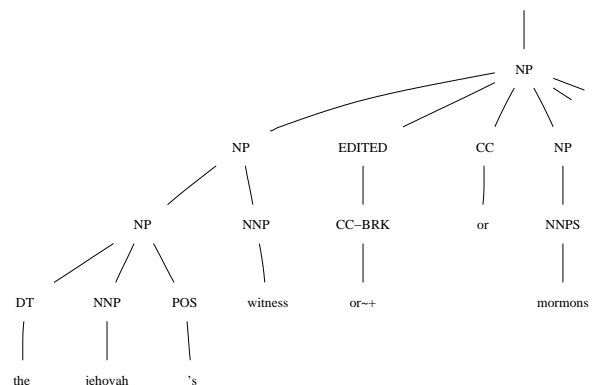
Figure 2: Propagating *BRK*, the evidence of disfluent juncture, from acoustics to syntax.

disjuncture symbols are identified in the ToBI labeling scheme as break indices (Price et al., 1991; Silverman et al., 1992).

The availability of a corpus annotated with ToBI labels makes it possible to design a break index classifier via supervised training. The corpus is a subset of the Switchboard corpus, consisting of sixty-four telephone conversations manually annotated by an experienced linguist according to a simplified ToBI labeling scheme (Ostendorf et al., 2001). In ToBI, degree of disjuncture is indicated by integer values from 0 to 4, where a value of 0 corresponds to clitic and 4 to a major phrase break. In addition, a suffix p denotes perceptually disfluent events reflecting, for example,

162

hesitation or planning. In conversational speech the intermediate levels occur infrequently and the break indices can be broadly categorized into three groups, namely, 1, 4 and p as in Wong et al. (2005).

A classifier was developed to predict three break indices at each word boundary based on variations in pitch, duration and energy associated with word, syllable or sub-syllabic constituents (Shriberg et al., 2005; Sonmez et al., 1998). To compute these features, phone-level time-alignments were obtained from an automatic speech recognition system. The duration of these phonological constituents were derived from the ASR alignment, while energy and pitch were computed every 10ms with *snack*, a public-domain sound toolkit (Sjlander, 2001). The duration, energy, and pitch were post-processed according to stylization procedures outlined in Sonmez et al. (1998) and normalized to account for variability across speakers.

Since the input vector can have missing values such as the absence of pitch during unvoiced sound, only decision tree based classifiers were investigated. Decision trees can handle missing features gracefully. By choosing different combinations of splitting and stopping criteria, an ensemble of decision trees was built using the publicly-available IND package (Buntine, 1992). These individual classifiers were then combined into ensemble-based classifiers.

Several classifiers were investigated for detecting break indices. On ten-fold cross-validation, a bagging-based classifier (Breiman, 1996) predicted prosodic breaks with an accuracy of 83.12% while chance was 67.66%. This compares favorably with the performance of the supervised classifiers on a similar task in Wong et al. (2005). Random forests and hidden Markov models provide marginal improvements at considerable computational cost (Harper et al., 2005).

For speech repair, the focus is on detecting disfluent breaks. The precision and recall trade-off on its detection can be adjusted using a threshold on the posterior probability of predicting "p", as shown in Figure 3.

In essence, the large number of acoustic and prosodic features related to disfluency are encoded via the ToBI label 'p', and provided as additional observations to the PCFG. This is unlike previous work on incorporating prosodic information (Gre-
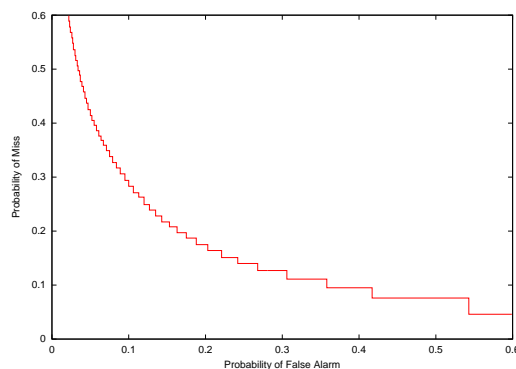

Figure 3: DET curve for detecting disfluent breaks from acoustics.

gory et al., 2004; Lease et al., 2005; Kahn et al., 2005) as described further in Section 6.

3 Syntactic parallelism

The other striking property of speech repairs is their parallel character: subsequent repair regions 'line up' with preceding reparandum regions. This property can be harnessed to better estimate the length of the reparandum by considering parallelism from the perspective of syntax. For instance, in Figure 4(a) the unfinished reparandum noun phrase is repaired by another noun phrase – the syntactic categories are parallel.

3.1 Levelt's WFR and Conjunction

The idea that the reparandum is syntactically parallel to the repair can be traced back to Levelt (1983). Examining a corpus of Dutch picture descriptions, Levelt proposes a bi-conditional well-formedness rule for repairs (WFR) that relates the structure of repairs to the structure of conjunctions. The WFR conceptualizes repairs as the conjunction of an unfinished reparandum string (α) with a properly finished repair (γ). Its original formulation, repeated here, ignores optional interregna like "er" or "I mean."

Well-formedness rule for repairs (WFR) A repair $\langle \alpha \, \gamma \rangle$ is well-formed if and only if there is a string β such that the string $\langle \alpha\beta \; and^* \; \gamma \rangle$ is well-formed, where β is a completion of the constituent directly dominating the last element of α. (*and* is to be deleted if that last element is itself a sentence connective)

In other words, the string α is a prefix of a phrase whose completion, β—if it were present—would

render the whole phrase $\alpha\beta$ grammatically conjoinable with the repair γ. In example (1) α is the string 'the first kind of invasion of', γ is 'the first type of privacy' and β is probably the single word 'privacy.'

This kind of conjoinability typically requires the syntactic categories of the conjuncts to be the same (Chomsky, 1957, 36). That is, a rule schema such as (2) where X is a syntactic category, is preferred over one where X is not constrained to be the same on either side of the conjunction.

$$X \rightarrow X \, \mathrm{Conj} \, X \qquad (2)$$

If, as schema (2) suggests, conjunction does favor like-categories, and, as Levelt suggests, well-formed repairs are conjoinable with finished versions of their reparanda, then the syntactic categories of repairs ought to match the syntactic categories of (finished versions of) reparanda.

3.2 A WFR for grammars

Levelt's WFR imposes two requirements on a grammar

- distinguishing a separate category of 'unfinished' phrases

- identifying a syntactic category for reparanda

Both requirements can be met by adapting Treebank grammars to mirror the analysis of McKelvie[1] (1998a; 1998b). McKelvie derives phrase structure rules for speech repairs from fluent rules by adding a new feature called `abort` that can take values true and false. For a given grammar rule of the form

$$A \rightarrow B \, C$$

a metarule creates other rules of the form

$$A \, [\texttt{abort} = Q] \rightarrow$$
$$B \, [\texttt{abort} = \textit{false}] \, C \, [\texttt{abort} = Q]$$

where Q is a propositional variable. These rules say, in effect, that the constituent A is aborted just in case the last daughter C is aborted. Rules that don't involve a constant value for Q ensure that the same value appears on parents and children. The

---

[1]McKelvie's metarule approach declaratively expresses Hindle's (1983) Stack Editor and Category Copy Editor rules. This classic work effectively states the WFR as a program for the Fidditch deterministic parser.

WFR is then implemented by rule schemas such as (3)

$$X \rightarrow X \, [\texttt{abort} = \textit{true}] \, (\mathrm{AFF}) \, X \qquad (3)$$

that permit the optional interregnum AFF to conjoin an unfinished X-phrase (the reparandum) with a finished X-phrase (the repair) that comes after it.

3.3 A WFR for Treebanks

McKelvie's formulation of Levelt's WFR can be applied to Treebanks by systematically recoding the annotations to indicate which phrases are unfinished and to distinguish matching from non-matching repairs.

3.3.1 Unfinished phrases

Some Treebanks already mark unfinished phrases. For instance, the Penn Treebank policy (Marcus et al., 1993; Marcus et al., 1994) is to annotate the lowest node that is unfinished with an `-UNF` tag as in Figure 4(a).

It is straightforward to propagate this mark upwards in the tree from wherever it is annotated to the nearest enclosing EDITED node, just as `-BRK` is propagated upwards from disjuncture marks on individual words. This percolation simulates the action of McKelvie's $[\texttt{abort} = \textit{true}]$. The resulting PCFG is one in which distributions on phrase structure rules with 'missing' daughters are segregated from distributions on 'complete' rules.

3.4 Reparanda categories

The other key element of Levelt's WFR is the idea of conjunction of elements that are in some sense the same. In the Penn Treebank annotation scheme, reparanda always receive the label EDITED. This means that the syntactic category of the reparandum is hidden from any rule which could favor matching it with that of the repair. Adding an additional mark on this EDITED node (a kind of daughter annotation) rectifies the situation, as depicted in Figure 4(b), which adds the notation `-childNP` to a tree in which the unfinished tags have been propagated upwards. This allows a Treebank PCFG to represent the generalization that speech repairs tend to respect syntactic category.

4 Results

Three kinds of experiments examined the effectiveness of syntactic and prosodic indicators of
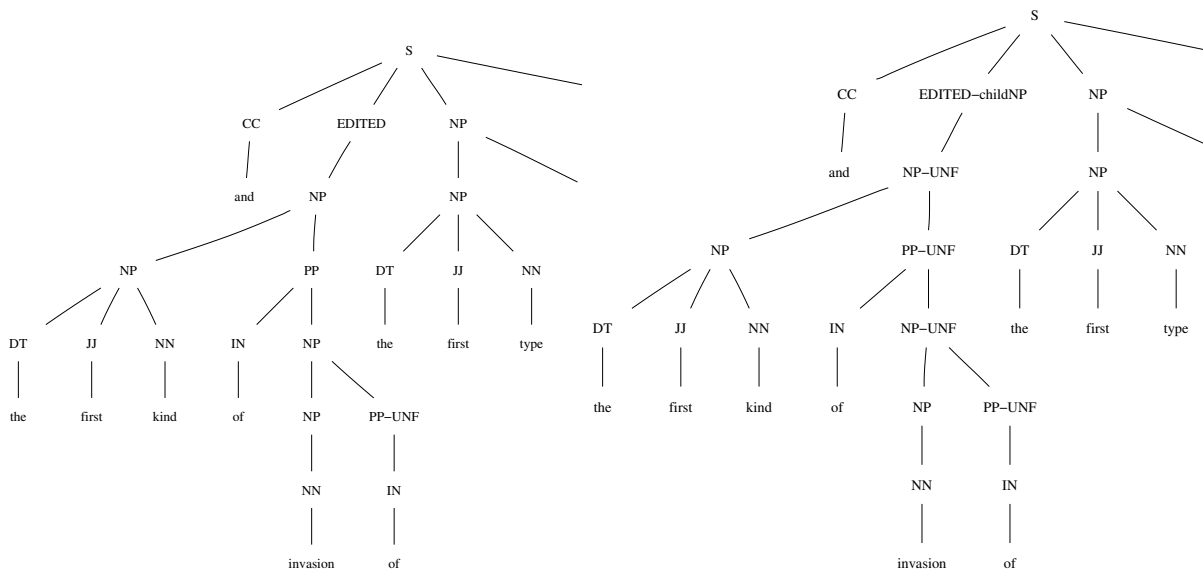
(a) The lowest unfinished node is given. (b) -UNF propagated, daughter-annotated Switchboard tree

Figure 4: Input (a) and output (b) of tree transformations.

speech repairs. The first two use the CYK algorithm to find the most likely parse tree on a grammar read-off from example trees annotated as in Figures 2 and 4. The third experiment measures the benefit from syntactic indicators alone in Charniak's lexicalized parser (Charniak, 2000). The tables in subsections 4.1, 4.2, and 4.3 summarize the accuracy of output parse trees on two measures. One is the standard Parseval F-measure, which tracks the precision and recall for all labeled constituents as compared to a gold-standard parse. The other measure, EDIT-finding F, restricts consideration to just constituents that are reparanda. It measures the per-word performance identifying a word as dominated by EDITED or not. As in previous studies, reference transcripts were used in all cases. A check ($\sqrt{}$) indicates an experiment where prosodic breaks where automatically inferred by the classifier described in section 2, whereas in the ($\times$) rows no prosodic information was used.

4.1 CYK on Fisher

Table 1 summarizes the accuracy of a standard CYK parser on the newly-treebanked Fisher corpus (LDC2005E15) of phone conversations, collected as part of the DARPA EARS program. The parser was trained on the entire Switchboard corpus (ca. 107K utterances) then tested on the 5368-utterance 'dev2' subset of the Fisher data. This test set was tagged using MX-

POST (Ratnaparkhi, 1996) which was itself trained on Switchboard. Finally, as described in section 2 these tags were augmented with a special prosodic break symbol if the decision tree rated the probability a ToBI 'p' symbol higher than the threshold value of 0.75.

Annotation	Break index	Parseval F	EDIT F
none	$\times$	66.54	22.9
	$\sqrt{}$	66.08	26.1
daughter annotation	$\times$	66.41	29.4
	$\sqrt{}$	65.81	31.6
-UNF propagation	$\times$	67.06	31.5
	$\sqrt{}$	66.45	34.8
both	$\times$	69.21	40.2
	$\sqrt{}$	67.02	40.6

Table 1: Improvement on Fisher, MXPOSTed tags.

The Fisher results in Table 1 show that syntactic and prosodic indicators provide different kinds of benefits that combine in an additive way. Presumably because of state-splitting, improvement in EDIT-finding comes at the cost of a small decrement in overall parsing performance.

4.2 CYK on Switchboard

Table 2 presents the results of similar experiments on the Switchboard corpus following the

train/dev/test partition of Charniak and Johnson (2001). In these experiments, the parser was given correct part-of-speech tags as input.

Annotation	Break index	Parseval F	EDIT F
none	×	70.92	18.2
	√	69.98	22.5
daughter annotation	×	71.13	25.0
	√	70.06	25.5
-UNF propagation	×	71.71	31.1
	√	70.36	30.0
both	×	71.16	41.7
	√	71.05	36.2

Table 2: Improvement on Switchboard, gold tags.

The Switchboard results demonstrate independent improvement from the syntactic annotations. The prosodic annotation helps on its own and in combination with the daughter annotation that implements Levelt's WFR.

4.3 Lexicalized parser

Finally, Table 3 reports the performance of Charniak's non-reranking, lexicalized parser on the Switchboard corpus, using the same test/dev/train partition.

Annotation	Parseval F	EDIT F
baseline	83.86	57.6
daughter annotation	80.85	67.2
-UNF propagation	81.68	64.7
both	80.16	70.0
flattened EDITED	82.13	64.4

Table 3: Charniak as an improved EDIT-finder.

Since Charniak's parser does its own tagging, this experiment did not examine the utility of prosodic disjuncture marks. However, the combination of daughter annotation and -UNF propagation does lead to a better grammar-based reparandum-finder than parsers trained on flattened EDITED regions. More broadly, the results suggest that Levelt's WFR is synergistic with the kind of head-to-head lexical dependencies that Charniak's parser uses.

5 Discussion

The pattern of improvement in tables 1, 2, and 3 from *none* or *baseline* rows where no syntac-

tic parallelism or break index information is used, to subsequent rows where it is used, suggest why these techniques work. Unfinished-category annotation improves performance by preventing the grammar of unfinished constituents from being polluted by the grammar of finished constituents. Such purification is independent of the fact that rules with daughters labeled EDITED-childXP tend to also mention categories labeled XP further to the right (or NP and VP, when XP starts with S). This preference for syntactic parallelism can be triggered either by externally-suggested ToBI break indices or grammar rules annotated with -UNF. The prediction of a disfluent break could be further improved by POS features and N-gram language model scores (Spilker et al., 2001; Liu, 2004).

6 Related Work

There have been relatively few attempts to harness prosodic cues in parsing. In a spoken language system for VERBMOBIL task, Batliner and colleagues (2001) utilize prosodic cues to dramatically reduce lexical analyses of disfluencies in a end-to-end real-time system. They tackle speech repair by a cascade of two stages – identification of potential interruption points using prosodic cues with 90% recall and many false alarms, and the lexical analyses of their neighborhood. Their approach, however, does not exploit the synergy between prosodic and syntactic features in speech repair. In Gregory et al. (2004), over 100 real-valued acoustic and prosodic features were quantized into a heuristically selected set of discrete symbols, which were then treated as pseudo-punctuation in a PCFG, assuming that prosodic cues function like punctuation. The resulting grammar suffered from data sparsity and failed to provide any benefits. Maximum entropy based models have been more successful in utilizing prosodic cues. For instance, in Lease et al. (2005), interruption point probabilities, predicted by prosodic classifiers, were quantized and introduced as features into a speech repair model along with a variety of TAG and PCFG features. Towards a clearer picture of the interaction with syntax and prosody, this work uses ToBI to capture prosodic cues. Such a method is analogous to Kahn et al. (2005) but in a generative framework.

The TAG-based model of Johnson and Charniak (2004) is a separate-processing approach that rep-

resents the state of the art in reparandum-finding.

Johnson and Charniak explicitly model the crossed dependencies between individual words in the reparandum and repair regions, intersecting this sequence model with a parser-derived language model for fluent speech. This second step improves on Stolcke and Shriberg (1996) and Heeman and Allen (1999) and outperforms the specific grammar-based reparandum-finders tested in section 4. However, because of separate-processing the TAG channel model's analyses do not reflect the syntactic structure of the sentence being analyzed, and thus that particular TAG-based model cannot make use of properties that depend on the phrase structure of the reparandum region. This includes the syntactic category parallelism discussed in section 3 but also predicate-argument structure. If edit hypotheses were augmented to mention particular tree nodes where the reparandum should be attached, such syntactic parallelism constraints could be exploited in the reranking framework of Johnson et al. (2004).

The approach in section 3 is more closely related to that of Core and Schubert (1999) who also use metarules to allow a parser to switch from speaker to speaker as users interrupt one another. They describe their metarule facility as a modification of chart parsing that involves copying of specific arcs just in case specific conditions arise. That approach uses a combination of longest-first heuristics and thresholds rather than a complete probabilistic model such as a PCFG.

Section 3's PCFG approach can also be viewed as a declarative generalization of Roark's (2004) EDIT-CHILD function. This function helps an incremental parser decide upon particular tree-drawing actions in syntactically-parallel contexts like speech repairs. Whereas Roark conditions the expansion of the first constituent of the repair upon the corresponding first constituent of the reparandum, in the PCFG approach there exists a separate rule (and thus a separate probability) for each alternative sequence of reparandum constituents.

7 Conclusion

Conventional PCFGs can improve their detection of speech repairs by incorporating Lickley's hypothesis about interrupted prosody and by implementing Levelt's well-formedness rule. These benefits are additive.

The strengths of these simple tree-based techniques should be combinable with sophisticated string-based (Johnson and Charniak, 2004; Liu, 2004; Zhang and Weng, 2005) approaches by applying the methods of Wieling et al. (2005) for constraining parses by externally-suggested brackets.

References

L. Breiman. 1996. Bagging predictors. *Machine Learning*, 24(2):123–140.

W. Buntine. 1992. Tree classication software. In *Technology 2002: The Third National Technology Transfer Conference and Exposition*, Baltimore.

E. Charniak and M. Johnson. 2001. Edit detection and parsing for transcribed speech. In *Proceedings of the 2nd Meeting of the North American Chapter of the Association for Computational Linguistics*, pages 118–126.

E. Charniak. 2000. A maximum-entropy-inspired parser. In *Proceedings of NAACL-00*, pages 132–139.

N. Chomsky. 1957. *Syntactic Structures*. Anua Linguarum Series Minor 4, Series Volume 4. Mouton de Gruyter, The Hague.

M. G. Core and L. K. Schubert. 1999. A syntactic framework for speech repairs and other disruptions. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, pages 413–420.

J. J. Godfrey, E. C. Holliman, and J. McDaniel. 1992. SWITCHBOARD: Telephone speech corpus for research and development. In *Proceedings of ICASSP*, volume I, pages 517–520, San Francisco.

M. Gregory, M. Johnson, and E. Charniak. 2004. Sentence-internal prosody does not help parsing the way punctuation does. In *Proceedings of North American Association for Computational Linguistics*.

M. Harper, B. Dorr, J. Hale, B. Roark, I. Shafran, M. Lease, Y. Liu, M. Snover, and L. Yung. 2005. Parsing and spoken structural event detection. In *2005 Johns Hopkins Summer Workshop Final Report*.

P. A. Heeman and J. F. Allen. 1999. Speech repairs, intonational phrases and discourse markers: modeling speakers' utterances in spoken dialog. *Computational Linguistics*, 25(4):527–571.

D. Hindle. 1983. Deterministic parsing of syntactic non-fluencies. In *Proceedings of the ACL*.

M. Johnson and E. Charniak. 2004. A TAG-based noisy channel model of speech repairs. In *Proceedings of ACL*, pages 33–39.

M. Johnson, E. Charniak, and M. Lease. 2004. An improved model for recognizing disfluencies in conversational speech. In *Proceedings of Rich Transcription Workshop*.

J. G. Kahn, M. Lease, E. Charniak, M. Johnson, and M. Ostendorf. 2005. Effective use of prosody in parsing conversational speech. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 233–240.

M. Lease, E. Charniak, and M. Johnson. 2005. Parsing and its applications for conversational speech. In *Proceedings of ICASSP*.

W. J. M. Levelt. 1983. Monitoring and self-repair in speech. *Cognitive Science*, 14:41–104.

R. J. Lickley. 1996. Juncture cues to disfluency. In *Proceedings the International Conference on Speech and Language Processing*.

Y. Liu. 2004. *Structural Event Detection for Rich Transcription of Speech*. Ph.D. thesis, Purdue University.

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

M. Marcus, G. Kim, M. A. Marcinkiewicz, R. MacIntyre, A. Bies, M. Ferguson, K. Katz, and B. Schasberger. 1994. The Penn Treebank: Annotating Predicate Argument Structure. In *Proceedings of the 1994 ARPA Human Language Technology Workshop*.

D. McKelvie. 1998a. SDP – Spoken Dialog Parser. ESRC project on Robust Parsing and Part-of-speech Tagging of Transcribed Speech Corpora, May.

D. McKelvie. 1998b. The syntax of disfluency in spontaneous spoken language. ESRC project on Robust Parsing and Part-of-speech Tagging of Transcribed Speech Corpora, May.

C. Nakatani and J. Hirschberg. 1994. A corpus-based study of repair cues in spontaneous speech. *Journal of the Acoustical Society of America*, 95(3):1603–1616, March.

M. Ostendorf, I. Shafran, S. Shattuck-Hufnagel, L. Carmichael, and W. Byrne. 2001. A prosodically labelled database of spontaneous speech. In *Proc. ISCA Tutorial and Research Workshop on Prosody in Speech Recognition and Understanding*, pages 119–121.

P. Price, M. Ostendorf, S. Shattuck-Hufnagel, and C. Fong. 1991. The use of prosody in syntactic disambiguation. *Journal of the Acoustic Society of America*, 90:2956–2970.

A. Ratnaparkhi. 1996. A maximum entropy part-of-speech tagger. In *Proceedings of Empirical Methods in Natural Language Processing Conference*, pages 133–141.

B. Roark. 2004. Robust garden path parsing. *Natural Language Engineering*, 10(1):1–24.

E. Shriberg, L. Ferrer, S. Kajarekar, A. Venkataraman, and A. Stolcke. 2005. Modeling prosodic feature sequences for speaker recognition. *Speech Communication*, 46(3-4):455–472.

E. Shriberg. 1994. *Preliminaries to a Theory of Speech Disfluencies*. Ph.D. thesis, UC Berkeley.

H. F. Silverman, M. Beckman, J. Pitrelli, M. Ostendorf, C. Wightman, P. Price, J. Pierrehumbert, and J. Hirshberg. 1992. ToBI: A standard for labeling English prosody. In *Proceedings of ICSLP*, volume 2, pages 867–870.

K. Sjlander, 2001. *The Snack sound visualization module*. Royal Institute of Technology in Stockholm. http://www.speech.kth.se/SNACK.

K. Sonmez, E. Shriberg, L. Heck, and M. Weintraub. 1998. Modeling dynamic prosodic variation for speaker verification. In *Proceedings of ICSLP*, volume 7, pages 3189–3192.

Jörg Spilker, Martin Klarner, and Günther Görz. 2000. Processing self-corrections in a speech-to-speech system. In Wolfgang Wahlster, editor, *Verbmobil: Foundations of speech-to-speech translation*, pages 131–140. Springer-Verlag, Berlin.

J. Spilker, A. Batliner, and E. Nöth. 2001. How to repair speech repairs in an end-to-end system. In R. Lickley and L. Shriberg, editors, *Proc. of ISCA Workshop on Disfluency in Spontaneous Speech*, pages 73–76.

A. Stolcke and E. Shriberg. 1996. Statistical language modeling for speech disfluencies. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 405–408, Atlanta, GA.

R. M. Weischedel and N. K. Sondheimer. 1983. Meta-rules as a basis for processing ill-formed input. *American Journal of Computational Linguistics*, 9(3-4):161–177.

M. Wieling, M-J. Nederhof, and G. van Noord. 2005. Parsing partially bracketed input. Talk presented at Computational Linguistics in the Netherlands.

D. Wong, M. Ostendorf, and J. G. Kahn. 2005. Using weakly supervised learning to improve prosody labeling. Technical Report UWEETR-2005-0003, University of Washington Electrical Engineering Dept.

Q. Zhang and F. Weng. 2005. Exploring features for identifying edited regions in disfluent sentences. In *Proceedings of the Nineth International Workshop on Parsing Technologies*, pages 179–185.