Proceedings of DiSS'05, Disfluency in Spontaneous Speech Workshop.
10–12 September 2005, Aix-en-Provence, France, pp. 59-63.

Promotion of Disfluency in Syntactic Parallelism

Andrew A. Cooper & John T. Hale

Michigan State University, East Lansing, Michigan, USA

Abstract

The development of a disfluency-robust speech parser requires some insight into where disfluencies occur in spontaneous spoken language. This corpus study deals with one syntactic variable which is predictive of disfluency location: syntactic parallelism.

A formal definition of syntactic parallelism is used to show that syntactic parallelism is indeed predictive of disfluency.

1. Introduction

Disfluencies (such as 'uh', 'um', and repairs) are a significant challenge in the implementation of parsers and other tools dealing with spontaneous spoken language.

One avenue to the development of disfluency-robust tools is to treat them probabilistically, an approach that has met with success elsewhere in syntax. This approach requires some knowledge of where disfluencies occur in spontaneous speech.

Syntactic parallelism is a variable which (as we will show below) interacts with disfluency. Intuitively, syntactic parallelism is the reuse of a syntactic structure in proximity.

The results presented here are from the Switchboard corpus.[1]

1.1. Outline

The paper is organized as follow. First, section 2 defines the disfluency types of interest. Then section 3 proposes an explicit definition of syntactic parallelism. Two hypotheses based on this definition are set out. Section 4 presents several methods for evaluating these hypotheses in the parsed section of the Switchboard corpus. Section 5 presents the results of experiments using these methods, and section 6 discusses their significance. Sections 7 interprets the results, considering their implications both for the theory of disfluency and for robust parsing. Section 8 suggests possible directions for future study.

2. Disfluency

Disfluency is generally treated as a phenomenon of spontaneous speech in which the speaker makes agrammatical utterance such as "pauses, fillers ('um's and 'er's), repetitions, speech repairs, and fresh starts" [7]. One element common to all types of disfluency is their potentially-universal distribution[2]—disfluencies can and do occur everywhere.

Because they are not marked in the Switchboard corpus, pauses and fresh starts are not considered here, but fillers, repetitions, and self-repairs are.

2.1. Fillers

Fillers are words which are semantically empty and seem to serve as placeholders in speech. An example from the corpus:

> I'm sure we have to have <u>uh</u> permits

2.2. Edit-type disfluencies

We treat repetitions and self-repairs as subtypes of *edit-type disfluencies*. These are characterized by the speaker's attempt at a fluent utterance and subsequent correction of it. An example from the corpus:

> <u>the first kind of invasion of</u> the first type of privacy

3. Syntactic Parallelism

Linguistic theories differ in their characterizations of syntactic parallelism. The definition proposed here attempts to formalize both proximity (in section 3.1) and structural similarity (in section 3.2).

The concepts used in this definition are: c-command, like category of nodes, and depth.

3.1. Lobes

First, a domain on which to describe the parallelism is required. This domain will be called **lobes**.

> Two phrases are said to be lobes of a parallelism if they are of the same category and one c-commands the other[3].

This definition is the basis of **Hypothesis I**:

> A lobe is more likely to contain a disfluency if the other lobe in the parallelism contains a disfluency.

Hypothesis I was tested using methods described in section 4.1, with results reported in section 5.1 and discussed in section 6.1.

3.2. Parallel Constituents

Having defined the parallelism coarsely, it is possible to describe parallelism between the substructures of one lobe and the substructures of the other.

A constituent (say, X1) **is parallel to** another constituent (say, X2) if:

> A lobe Y1 contains X1 and a lobe Y2 contains X2; and X1 performs a parallel grammatical function in Y1 as X2 does in Y2.

Formalizing the notion of 'grammatical function' is of course controversial. One option is to follow Chomsky [1] in the idea that grammatical function corresponds to configurational position. The **parallel grammatical function** requirement is then approximated[4] by:

---

[1] The Switchboard corpus consists of 2400 telephone conversations among 543 speakers [4].

[2] Universal, but not uniform—this study finds that the uniform-distribution hypothesis is, in all cases, much less than 0.1% likely to derive the results presented in tables 1, 2 and 5.

[3] A node A is said to c-command a node B if every node dominating A also dominates B.

[4] This approximation is based on the conjecture (unproved to the authors' knowledge) that, in a binary-branching

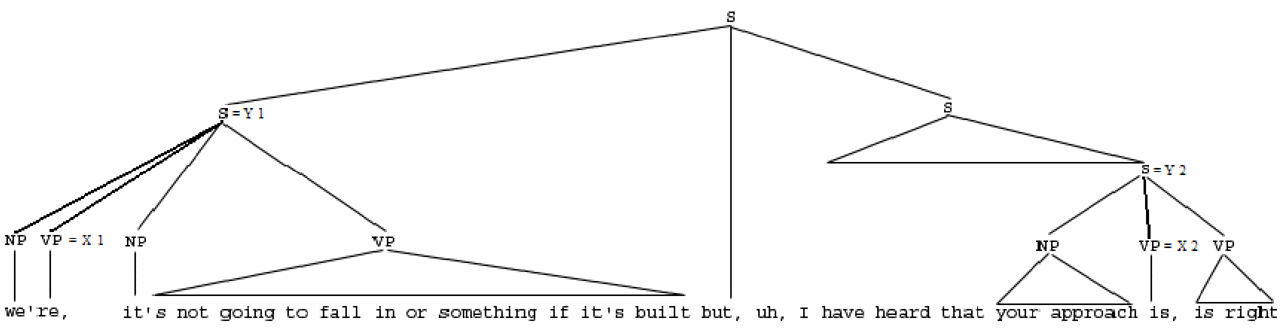

Figure 2: Two parallel S lobes with parallel disfluent VP constituents.

The same number of nodes intervene between X1 and Y1 as between X2 and Y2.

This conception of parallel grammatical function is the basis of **Hypothesis II**:

Disfluencies occur in parallel positions.

A (partially-treed) example in which disfluencies occur in parallel positions is given in Figure 2.

Hypothesis II was tested using methods described in section 4.2, with results reported in section 5.2 and discussed in section 6.2.

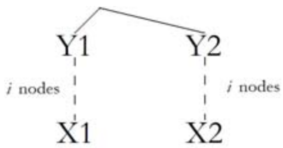

Figure 1: Y1 and Y2 are lobes; X1 and X2 are parallel constituents.

4. Methods

A sample of seventeen conversations was inspected by eye to test Hypothesis I. Only parallelisms resulting from conjunction[5] were counted. The results of this inspection justified further examination.

4.1. Examination of Lobes - Hypothesis I

Hypothesis I, namely that parallel lobes tend to have similar disfluency containment, was examined using `tgrep2`, a tree search program[6]. This examination again considered only conjunction-related parallelism in the sample of seventeen conversations mentioned above.

4.1.1. Editing Treebank

To test Hypothesis I, the Treebank parse files were edited by hand to make conjoined top-level sentences sisters under a common node called ‘TOP’. This was necessary to have `tgrep2` recognize top-level conjunctions.

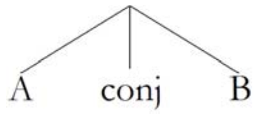

Figure 3: Hypothesis I structure searched for with `tgrep2`.

4.2. Examination of Parallel Constituents - Hypothesis II

Hypothesis II, namely that disfluencies tend to occur in parallel positions, was confirmed using both `tgrep2` and `TIGERSearch`[7], another tree search program.

4.2.1. tgrep2

The `tgrep2` examination of Hypothesis II covered a sample of 78 conversations among 86 speakers[8], constituting 12% of the entire corpus.

For the `tgrep2` examination, the Treebank files were automatically edited to remove the “,” tag, which obscured some c-command relations.

The queries used in this examination were designed to consider each node and determine its disfluency status, then find all parallel constituents and determine their disfluency statuses. Since `tgrep2` cannot calculate depths, 4400 queries enumerating every possible depth up to 10 were used.

4.2.2. TIGERSearch

The `TIGERSearch` examination of Hypothesis II covered all 650 Treebank files. Prior editing and enumeration over categories and depths were unnecessary because of `TIGERSearch`’s more powerful query language.

4.3. Statistical tests

The chi-square test of distributional significance and the phi statistic were used to examine the data obtained from both experiments.

The chi-square test measures the likelihood that the axes of a contingency table are independent. Higher values of χ^2 indicate a smaller likelihood of independence.

The phi-statistic measures how much of the variance along one axis of a contingency table is explained by the variance of the other axis. Higher values of ϕ indicate a greater percentage of variation which is so explained.

5. Results

The results described above in general confirmed both research hypotheses.

grammar obeying reasonably formalized X-bar restrictions, it is impossible for two nodes of the same category and depth from a common ancestor to occupy different configurational positions.

[5] Treebank treats conjuncts as sisters.

[6] `tgrep2` was developed by Douglas Rohde of the Department of Brain and Cognitive Science at Massachusettes Institute of Technology, accessible at `http://tedlab.mit.edu/~dr/Tgrep2/`.

[7] `TIGERSearch` is a project at the Universität Stuttgart, accessible at `http://www.ims.uni-stuttgart.de/projekte/TIGER/TIGERSearch`.

[8] No single speaker participated in more than 6 conversations.

5.1. Hypothesis I

The results of testing Hypothesis I are presented in contingency table 1. Each cell contains the number of parallelisms with lobes of the indicated disfluency status. Here 'disfluent' means that the lobe contains one or more disfluencies, and 'fluent' means that the lobe contains no disfluencies at all.

Table 1: Disfluency status of conjoined lobes, obtained with `tgrep`. A significant distribution.

N=821 df=1	χ^2=61.25, p<.001 ϕ=0.273	Lobe 2	
		Disfluent	Fluent
Lobe 1	Disfluent (Expected) % of total	150 (99.2) 18.3	126 (176.8) 15.3
	Fluent (Expected) % of total	145 (195.8) 17.7	400 (349.2) 48.7

5.1.1. Other results

The majority (68%) of parallelisms due to conjunction were top-level sentential conjunctions. An even stronger majority (91%) of parallelisms with a disfluent lobe were top-level sentential conjunctions.

5.2. Hypothesis II

5.2.1. tgrep2

The results of testing Hypothesis II with `tgrep2` are presented in contingency table 2. Each cell contains the number of parallelisms with lobes of the indicated disfluency status. Here 'disfluent' means that the constituent is the site of a disfluency, and 'fluent' means that the constituent is fluent. 'Constituent 1' is the linearly first constituent.

Table 2: Disfluency status of parallel constituents, obtained with `tgrep` . A significant distribution.

N=20267 df=1	χ^2=461.29, p<.001 ϕ=0.151	Constituent 2	
		Disfluent	Fluent
Constituent 1	Disfluent (Expected) % of total	635 (307.1) 3.1	1242 (1569.9) 6.1
	Fluent (Expected) % of total	2681 (3008.9) 13.2	15709 (15381.1) 77.5

There are potentially several constituents parallel to any given node, with potentially different disfluency statuses. That possibility is excluded here; in order to be counted as 'fluent', a node had to be parallel to no disfluent nodes. Such an exclusion is conservative because it tends to reduce the number of matches with two fluent nodes in favor of matches with one fluent and one disfluent node.

5.2.2. TIGERSearch

The results of the `TIGERSearch` examination are given in contingency tables 3, 4 and 5. Each cell contains the number of pairs of parallel constituents with the indicated disfluency status. 'fluent' and 'disfluent' have the same meanings as above. 'Constituent 1' here is the constituent in the c-commanding node; 'Constituent 2' is the constituent which is c-commanded.

Table 3: Filler-type disfluency status. A significant distribution.

N=111844 df=1	χ^2=899.19, p<.001 ϕ=0.089	Constituent 2	
		Disfluent	Fluent
Constituent 1	Disfluent (Expected) % of total	434 (118.4) 0.4	2745 (3060.6) 2.4
	Fluent (Expected) % of total	3732 (4047.6) 3.3	104933 (104617.4) 93.8

Table 4: Edit-type disfluency status. A significant distribution.

N=111844 df=1	χ^2=286.7, p<.001 ϕ=0.051	Constituent 2	
		Disfluent	Fluent
Constituent 1	Disfluent (Expected) % of total	355 (153.2) 0.3	3179 (3380.8) 2.8
	Fluent (Expected) % of total	4495 (4696.8) 4.0	103815 (103613.2) 92.8

Table 5: Disfluency status (aggregate). A significant distribution.

N=111844 df=1	χ^2=286.7, p<.001 ϕ=0.034	Constituent 2	
		Disfluent	Fluent
Constituent 1	Disfluent (Expected) % of total	789 (541.2) 0.7	5924 (6171.9) 4.9
	Fluent (Expected) % of total	8227 (8474.9) 7.4	96904 (96656.2) 86.6

Contingency table 3 contains the data for filler-type disfluencies alone. Contingency table 4 contains the data for edit-type disfluencies alone. Contingency table 5 contains the aggregate data, excluding matches in which one parallel constituent was the site of an edit-type disfluency and the other was the site of a filler-type disfluency. This omission is conservative, since it biases against matches in which both parallel constituents are disfluent.

6. Discussion

In general, the results confirm the research hypotheses. These results extend Shriberg [9], who found that sentence-initial disfluencies and sentence-medial disfluencies have a high cooccurrence rate.

The results here also show that disfluency in the Switchboard corpus (both filler-type and edit-type) is responsive to a syntactic variable, extending the findings of Fox & Jasperson [3], who argue that repair responds to several syntactic variables.

6.1. Hypothesis I

Hypothesis I was confirmed in conjoined lobes; that is, *conjoined lobes tend to have the same disfluency status*. The ϕ-value reported indicates that about 7% of the variation in disfluency status in conjoined lobes is due to parallelism.

6.2. Hypothesis II

Hypothesis II was confirmed in both experiments testing it; that is, *syntactically parallel constituents tend to have the same disfluency status*. Given the several conservative assumptions made in calculating ϕ-values, it is possible that the effects observed are stronger than reported here.

6.2.1. tgrep2

The tgrep2 examination of Hypothesis II confirmed it. The ϕ-value reported indicates that about 2% of the variation in disfluency status is due to parallelism.

6.2.2. TIGERSearch

The TIGERSearch examination of Hypothesis II confirmed it. The ϕ-values reported indicate that very little of the variation in disfluency when considered in the aggregate is due to parallelism, but about 1% of the variation in filler-type disfluency and edit-type disfluency is due to parallelism.

That different types of disfluency are impacted distinctly by parallelism likely reflects the differences in distribution and characteristics of different types of disfluency observed by Shriberg [9], McKelvie [7] and others.

6.3. Sources of Error

Possible error sources are the original Switchboard corpus and the Treebank annotation scheme. The exclusion of some types of disfluency also introduces some error.

6.3.1. Switchboard

As Shriberg [9] and McKelvie [7] note, the Switchboard's transcriptions include some errors. Shriberg reports that in a small sample, her transcriptions disagreed with the Switchboard's on 25% of disfluent turns. Most of the discrepancies were due to miscategorization. Since this study treats fillers as one category and all other disfluency types annotated in the Treebank scheme as another, it is largely immune to error resulting from mistranscriptions. Only in the rare event that a filler were misheard by the transcriber as an edit or vice versa, or in the rarer case that a disfluency were transcribed when one did not occur would the error introduced be non-conservative.

6.3.2. Treebank

The Treebank annotation scheme introduces some additional sources of potential error.

In particular, the way Treebank treats repairs obscures constituency relations, as the reparandum is treated as sister to the repair.

A similar problem results from Treebank's treatment of sentential adjuncts, which are treated as sisters, again obscuring constituency.

Finally, and most critically, the Treebank is not parsed according to a binary X-bar grammar. This means that the critical conjecture upon which the definition of syntactic parallelism in section 3.2 is based is not strictly applicable to the corpus. In some cases, such as a VP like *give the dog a bath*, where *the dog* and *a bath* are treated as sister NPs, this will tend to falsely increase the number of 'parallel' constituents. However, since less than half of occurrences of each category are disfluent (McKelvie [7] gives a 6% rate of edit-type disfluency in the MAPTASK corpus) such false hits will come overwhelmingly in the fluent-disfluent, disfluent-fluent, and fluent-fluent cells of the contingency tables. Consider the effect of a false hit in each case:

Case I: Fluent-disfluent (bottom-left cell)

A false hit here artificially increases the number of observed hits. Since in each contingency table the expected number of hits is already greater than the number of observed hits, false hits here reduce the contribution of this cell to χ^2. Thus a false hit in case I is a conservative error.

In table 5, for example, the observed number in case I is 8227. If 227 of these were false hits, the true number would be 8000, which is farther from the expected value of 8474.9. The increase in χ^2 due to eliminating the false hits would be 19.36.

Case II: Disfluent-fluent (top-right cell)

As in case I, false hits reduce the contribution of this cell to χ^2. Thus a false hit in case II is a conservative error.

For example, in table 5, if 224 false hits were counted, the true number of observed hits would be 5700, which is farther from the expected value of 6171.9. The increase in χ^2 due to eliminating the false hits would be 26.12.

Case III: Fluent-fluent (bottom-right cell)

A false hit here artificially increases the number of observed hits. In each contingency table the expected number of hits is less than the number of observed hits, so the false hits artificially increase the contribution to χ^2 of this cell, a non-conservative error. But a look at the numbers in the tables suggests that the observed number case III could indeed be much lower and the distribution remain significant.

In table 5, the contribution to χ^2 of case III is only 0.64. Any decrease in the observed value up to 496 hits would not increase the total χ^2 for the table.

The results seen here are in fact stronger than necessary for application to parsing. For left-to-right parsing, one needs only to use a disfluency to predict subsequent disfluencies. Even true hits in case I (a fluent constituent followed by a parallel, disfluent constituent) are irrelevant to this prediction. In the same way, case III is entirely irrelevant to parsing, as it measures only the base rate of disfluency. True hits in case II are the only possible counterexamples to disfluency's predictiveness in parallel syntactic contexts, and the results of section 5 find that they occur at a unexpectedly low rate.

7. Interpretations

These results may be read to argue against the conventional interpretation of disfluency as only a phenomenon of production errors. If this view were correct, one would expect the second lobe of a parallelism to be *more* fluent than the first, since the speaker has already had a chance to work out the difficulties in a particular construction. But this is the opposite of what is observed in the present study. Often the speaker is disfluent in the second lobe, just as in the first.

There are a number of possible interpretations along these lines.

One might be that disfluency is prosodically and pragmatically salient enough that a speaker inserts disfluencies in the second lobe of a parallelism to maintain similarity with a disfluent first lobe. This conclusion would explain the stronger effect of conjunction, a kind of intentional parallelism, on disfluency. It would also explain why fillers, which are more likely to have pragmatic use, are more strongly affected by parallelism.

Another argument might be that speakers simply do not improve their production abilities, even in the short term. The speaker might be in a similar mental state, hence likely to make similar errors, at the second lobe as he was at the first. A production analogue of Steiner's [10] Iteration Model might be appropriate to this interpretation.

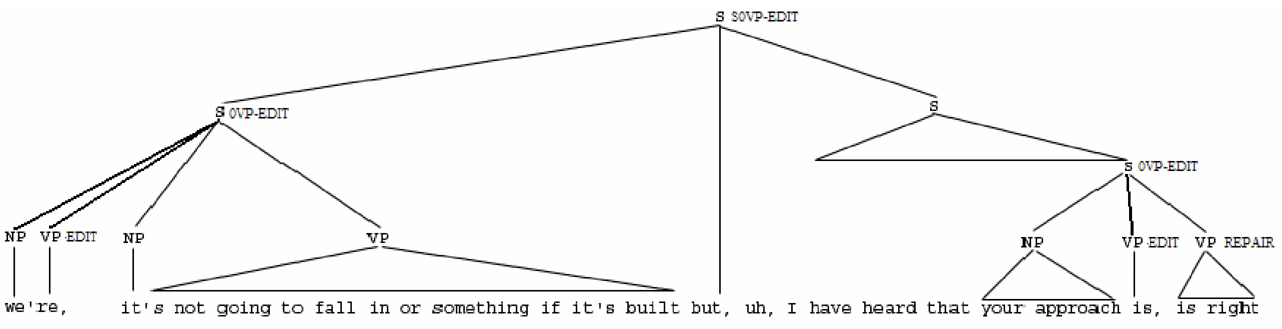

Figure 4: A possible parsing scheme.

8. Further Work

As noted above, the χ^2 statistic is fairly robust to the sorts of error encountered here. The ϕ-statistic does not fare as well, and further investigation with more powerful measures will be necessary to more precisely quantify the effect observed here.

To see how these numbers might be employed, consider the sentence in Figure 2 again. Having parsed the first S and noted that its VP is the site of an edit-type disfluency, the parser could add a 0VP-EDIT feature to that S, indicating that a descendant at depth 0 was an edited VP. Upon reading 'but', the parser would build the top-level S and propagate the feature S0VP-EDIT, indicating that it dominates an S with a zero-depth edited VP descendant. The parser then passes the 0VP-EDIT feature to all S descendants of the top-level S, thus predicting the second disfluent VP and its repair. Figure 4 is a schematic of this parse.

Such an arrangement might be implemented by interacting phrase structure schema:

1. S[0VP-EDIT] → NP VP[EDIT] VP[REPAIR]
2. S[0VP-EDIT] → NP VP[EDIT]
3. S[0VP-EDIT] → NP VP

Rule 1 would have a higher probability than rule 2, representing the greater likelihood of a repaired disfluency. Rule 3, representing case II (disfluent-fluent), would have the lowest probability.

Even if not explicitly coded in the Treebank, such rules may be automatically derivable from extant annotations.

9. Conclusion

Studied with the aid of a formal definition of syntactic parallelism, the Switchboard data suggest that syntactic parallelism is predictive of disfluency. With some further investigation, this predictive power could be incorporated into a parser to enable more efficient and accurate parsing of spontaneous speech.

In addition, the interaction of parallelism and disfluency has theoretical ramifications. It calls into question the treatment of disfluencies as solely the result of production problems. It also provides more evidence that disfluency is partly syntactic in nature.

10. Acknowledgements

Thanks to Dan Jurafsky and the participants in the 2004 Midwest Computational Linguistics Colloquium for their comments on an earlier iteration of this project.

11. References

[1] Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, Massachusetts: MIT Press.

[2] Clark, Herbert and Jean Fox Tree. 2002. Using uh and um in spontaneous speech. *Cognition*, 84:73-111.

[3] Fox, Barbara, and Robert Jasperson. 1995. *A syntactic exploration of repair in English conversation*. In Davis, Philip W., ed. Alternative Linguistics: Descriptive and Theoretical Modes. pp. 77-134. Amsterdam: Benjamins.

[4] Godfrey, John, Edward Holliman, and Jane McDaniel. 1992. Switchboard: Telephone speech corpus for research and development. *Proc IEEE ICASSP*. volume 1, pp. 517-520.

[5] Koenig, Esther, Wolfgang Lezius, and Holger Voorman. 2004. *TIGERSearch user's manual*.

[6] Marshall, Mitchell, Beatrice Santorini, and Mary Ann Marcinkewicz. 1993. Building a large annotated corpus of English: the Penn Treebank. *Computational Linguistics*. 19:313-330.

[7] McKelvie, David. 1998. The syntax of disfluency in spontaneous spoken language. Technical Report HCRC/RP-95, Human Communications Research Centre.

[8] Rohde, Douglas. 2001. *Tgrep2 manual, version 1.11*

[9] Shriberg, Elizabeth. 1994. *Preliminaries to a theory of speech disfluencies*. Ph.D. thesis, University of Berkeley, California.

[10] Steiner, Ilona. 2003. Parsing syntactic redundancies in coordinate structures. *Proc EuroCogSci 03*. p. 443