# Empirical Evidence for Intention-based Discourse Segmentation

Diane J. Litman
AT&T Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974
diane@research.att.com

Rebecca J. Passonneau
Department of Computer Science
Columbia University
New York, NY 10027
becky@cs.columbia.edu

## 1 Introduction

Each utterance of a discourse either bears a semantic relation to a preceding utterance, or constitutes the onset of a new semantic unit. Thus, a critical task in discourse understanding is determining how to relate each new utterance to the current representation of the discourse. Sequences of semantically related utterances are referred to as *segments*. The discourse intentions of the speaker provide one basis for determining which utterances belong within one segment [Grosz and Sidner, 1986]. As discussed in [Passonneau and Litman, 1993], we are conducting an empirical study of the relation between discourse segments and intentions. We have established that naive subjects can reliably identify the same discourse segment boundaries, using a commonsense notion of speaker intention as the segmentation criterion. Briefly, Section 2 describes our study, and Section 3 presents our results that agreement among subjects on discourse segment boundaries is highly statistically significant. Taking these results as a starting point, we ask in Section 4 whether subjects also agree on the intentions they associate with segments. Examination of our data suggests that subjects do agree on the semantic labels they associate with segments. In Section 5 we discuss the question of how relations among segments are recognized by looking at a class of cases in which an earlier suspended segment is resumed (i.e., discourse *pops*). We hypothesize that in spontaneous oral discourse, relations among segments are often not directly signaled, but must be inferred.

## 2 The Study

Our corpus consists of 20 narrative monologues about the same movie (about 14,000 words total), taken from Chafe's "Pear Stories" [Chafe, 1980]. Seven subjects per narrative were presented with a verbatim transcript,[1] such that each line of the transcript corresponded to one prosodic phrase (sentence-final or phrase-final contour, see Chafe [1980] for details). Subjects were instructed to identify sequential chunks,[2] each representing a single intention. Subjects were also instructed to describe the speaker intention for each discourse segment. Intention was explained in common sense terms and by example. Subjects were restricted to placing boundaries between prosodic phrases. An excerpt from the instructions is shown in Figure 1.

Figure 2 illustrates a portion of an intention-based segmentation produced by 7 subjects. Distinct subjects are indicated by letters of the alphabet. Prosodic phrases are numbered sequentially; the

[1] We eliminated visually distracting material such as pause locations and durations.

[2] Grosz and Hirschberg [1992] previously conducted an empirical study of hierarchical, intention-based segmentation. We have looked at a simpler linear intention-based segmentation task. Our pilot study as well as the work of Rotondo [1984] indicated that more complex segmentation tasks were too cumbersome given our average narrative length. Hearst [1993] also examines linear segmentation, based on a notion of topic change.

...
You should think of each movie narration as resulting from many decisions made by the speaker about what to do next. You will be asked to evaluate what the speaker was doing at each point ... Read through the transcript and draw a horizontal line across the page between complete text lines (utterances) where you think the speaker started doing something new.
...
In the wide left hand margin, say in abbreviated form what the speaker is doing. ... [Here] is an example of how to proceed. You are free to use any criteria in deciding what the narrator of your transcript is *doing*.

| | |
|---|---|
| `speaker recommends` | Well it's really a great movie, |
| `movie` | really beautiful scenery. |
| | You should see it, |
| | I recommend it, |
| | I really do. |
| | The first part of the movie just sets up |

Figure 1: Excerpt from Instructions

first field of the phrase number indicates sentence-final contour, and the second indicates phrase-final contour. At each potential boundary site, i.e., between each pair of prosodic phrases, the number and identity of subjects who classified the site as a boundary is indicated. The segmentation shown in Figure 2 contains 1 boundary proposed by all 7 subjects, 1 boundary proposed by 5 subjects, and 2 boundaries proposed by 1 subject.

1 SUBJECT (g)
13.1 Because he$_i$'s looking at the girl.
1 SUBJECT (f)
14.1 [.75] {ZERO-PRONOUN$_i$} Falls over,
5 SUBJECTS (a, b, c, d, e)
14.2 [1.5 [1.35] uh] there's no conversation in this movie.
15.1 [.6] There's sounds,
15.2 you know,
15.3 like the birds and stuff,
15.4 but there.. the humans beings in it don't say anything.
7 SUBJECTS (a, b, c, d, e, f, g)
16.1 [1.0] He$_i$ falls over,

Figure 2: Portion of Segmentation from Narrative 6

| Subject | Annotation of Narrator's Intention |
|---|---|
| a | Digression to describe sound track |
| b | No verbal communication [i.e., speaker describes lack thereof] |
| c | Describes that it is a silent movie with only nature sounds |
| d | Speaker describes sound techniques used in movie |
| e | Explain that there is no speaking in movie |

Figure 3: Segment spanning 14.2 through 15.4

# 3 Discourse Segment Boundaries

In [Passonneau and Litman, 1993], we show that our subjects agree with one another at levels that are statistically significant, thus demonstrating the reliability of intention as a segmentation criterion. *Percent agreement* is defined in [Gale *et al.*, 1992] as the ratio of observed agreements with the majority opinion to possible agreements with the majority opinion. We use percent agreement to measure the ability of subjects to agree with one another on whether there is a segment boundary between two adjacent prosodic phrases. We find that the average agreement across the 20 narratives on the status of all potential boundary locations is 89% (with a range from 82%-92%). We then use Cochran's test [Cochran, 1950] to determine if these levels of agreement are statistically significant. Cochran's test compares the observed number of subjects placing a boundary at every potential site with the number predicted by a random distribution; it is assumed that the total number of boundaries assigned by any one subject is given by that subject's actual performance. The greater the difference from randomness, the more unlikely is the observed distribution. For the 20 narratives, the probabilities of the observed boundary distributions ranged from $p = .1 \times 10^{-6}$ to $p \leq .6 \times 10^{-9}$, all very highly significant.

We also show why we consider boundaries agreed upon by a majority of subjects to be empirically validated. By partioning Cochran's statistic, we find the threshold for significance across all subjects and all narratives to be when at least 4 of 7 subjects agree. Using this threshold, we can derive a single discourse segmentation for each narrative. For the excerpt in Figure 2, this gives two empirically validated boundaries, represented as ordered pairs of prosodic phrases: (14.1,14.2) and (15.4,16.1).

## 4 Discourse Segment Intention

The 5 prosodic phrases from 14.2 through 15.4 constitute a segment, which we represent as [14.2,15.4]. From Figure 2, we see that 5 different subjects identified exactly this segment. Inspection of subjects' descriptions of speaker intention shows that in such cases of agreement on segments, subjects also generally agree on the narrator's intentions for the segment. Figure 3 presents the intentions attributed to the narrator by the 5 subjects for [14.2,15.4]. The 5 subjects who agreed on the segment (a-e) all indicate that the speaker's intentions pertain to describing the audio characteristics of the movie. The other 2 subjects (f,g) included one or two preceding phrases in the segment (cf. Figure 2). Subject f characterized the segment purpose as *techniques used in the movie*, and thus identifies essentially the same intentional structure. Subject g, who began the segment with 13.1, characterized the intention as *when the boy falls no one else cares about him*. Subject g thus not only identifies a different segment boundary, but also a different overall purpose.

Presumably, the 5 subjects depicted in Figure 3 abstract from the fact that the three full clauses in the segment all refer to auditory characteristics, as signaled by the lexical items *conversation* in the first clause, *sounds* in the second, and *say* in the third. Here we have generalized further from the subjects annotations to note that each asserts something about the movie's auditory character. In general, for segments delimited by high agreement boundaries, a single formulation of speaker purpose can be generalized from the data provided by the 7 subjects. We believe that such data provides evidence that when asked to, subjects perform the same kinds of abstraction across related utterances described in [Polanyi, 1988] and elsewhere.

## 5 Discourse Segment Pops

In Figure 2, one of the main characters of the movie (a boy on a bicycle) is referred to in phrases 13.1 and 14.1. The speaker suspends her description of the boy's activities throughout the next segment ([14.2,15.4]), but resumes reference to the boy in the first utterance of the following segment (16.1) using a third person definite pronoun subject (*he*). This illustrates a specific class of discourse pops, in which a segment resumes an earlier *suspended segment*. In particular, this class of *resumption segments* begins with an utterance in which a third person definite pronoun refers to an entity that is not in the focus space [Grosz and Sidner, 1986] associated with the *intervening segment*. Processing a clause that signals this type of discourse pop involves several tasks. Resolving the pronoun in the initial clause of the resumption segment requires shifting the attentional state [Grosz and Sidner, 1986], since the active focus space (corresponding to the intervening segment) does not contain a representation of the referent. This shift depends on recognizing the termination of the intervening segment, and a continuation relation between the resumption and suspended segments, so that the entities in focus in the suspended segment are again in focus for the resumption segment.

There are 8 discourse pops of the type in Figure 2 in the 10 narratives that we have coded for referential relations (coding described in [Passonneau, 1993]). These 8 examples exhibit various structural and semantic relations to the presumed discourse model. For example, they contain intervening segments that provide more detail, provide general background, or are digressions. In 7 cases, the resumption segment begins with a word that can function as a cue word,[3] but in 4 cases the cue word is *and*, a word whose discourse usage is hard to distinguish and which provides very

[3] We assume that different uses of cue phrases can be discriminated; cf. [Hirschberg and Litman, 1993].

little semantic information. The cue words in the remaining 3 cases are *so, all right* and *then*, none of which clearly signal attentional change [Grosz and Sidner, 1986; Hirschberg and Litman, 1993]. Non-lexical signals (e.g., *uh, tsk,* false starts) precede the initial clause of the resumption segment in 5 cases, and relatively long pauses (> 1.5 sec.) in 6 (cf. [Hirschberg and Grosz, 1992]). This suggests that in spontaneous oral discourse, instead of giving explicit indicators of the structural and semantic relations among segments, speakers provide non-lexical and pausal cues to breaks in segmental structure, relying on the hearer to infer the abstract structural and semantic relations. For example, the semantics of the predication *fall over* at the onset of the resumption segment is arguably very dissimilar to the predications occurring in the intervening segment, possibly supporting the inference that the new clause is unrelated to the intervening segment. In contrast, the two clauses which end the suspended segment (at 14.1) and begin the resumption segment (at 16.1) are semantically identical and structurally parallel, supporting the inference that 16.1 resumes the segment containing 14.1.

## 6 Conclusion

Our study establishes that a naive notion of speaker intention serves as a reliable criterion for identifying discourse segments. Qualitative analysis of annotations of speaker intention supports the conclusion that where subjects agree on segment boundaries, they also agree on the segment's intention. Using the empirically validated segments, we can begin to ask specific questions about the relation between segments and their abstract representation in an evolving discourse model. We note that for spontaneous oral narrative, discourse pops may not be explicitly signaled by cue words, and that structural and semantic relations among distinct segments may instead require inference. In [Passonneau and Litman, 1993], we directly address how explicit devices such as pauses, cue words and referential noun phrases correlate with segmental structure in order to posit constraints between surface structure choices and intentional and segmental structure.

## References

[Chafe, 1980] W. L. Chafe. *The Pear Stories: Cognitive, Cultural and Linguistic Aspects of Narrative Production.* Ablex Publishing Corporation, Norwood, NJ, 1980.

[Cochran, 1950] W. G. Cochran. The comparison of percentages in matched samples. *Biometrika*, 37:256–266, 1950.

[Gale *et al.*, 1992] W. Gale, K. W. Church, and D. Yarowsky. Estimating upper and lower bounds on the performance of word-sense disambiguation programs. In *Proc. of ACL*, pages 249–256, Newark, Delaware, 1992.

[Grosz and Sidner, 1986] B. J. Grosz and C. L. Sidner. Attention, intentions and the structure of discourse. *Computational Linguistics*, 12:175–204, 1986.

[Hearst, 1993] M. A. Hearst. TextTiling: A quantitative approach to discourse segmentation. Technical Report 93/24, Sequoia 2000 Technical Report, University of California, Berkeley, 1993.

[Hirschberg and Grosz, 1992] J. Hirschberg and B. Grosz. Intonational features of local and global discourse structure. In *Proc. of Darpa Workshop on Speech and Natural Language*, 1992.

[Hirschberg and Litman, 1993] J. Hirschberg and D. Litman. Empirical studies on the disambiguation of cue phrases. *Computational Linguistics*, 19, 1993.

[Passonneau and Litman, 1993] R. Passonneau and D. Litman. Intention-based segmentation: Human reliability and correlation with linguistic cues. In *Proc. of the ACL*, 1993.

[Passonneau, 1993] R. J. Passonneau. Coding scheme and algorithm for identification of discourse segment boundaries on the basis of the distribution of referential noun phrases. Technical report, Columbia University, 1993.

[Polanyi, 1988] L. Polanyi. A formal model of the structure of discourse. *Journal of Pragmatics*, 12:601–638, 1988.

[Rotondo, 1984] J. A. Rotondo. Clustering analysis of subject partitions of text. *Discourse Processes*, 7:69–88, 1984.