Application of the Centering Framework in Spontaneous Dialogues

Nanette Veilleux, Ph.D.
Simmons College, 300 The Fenway, Boston, MA, USA 02115

ABSTRACT

Spontaneous speech poses problems for automatic systems. While many investigators are making progress in recognition and dialogue processing, spontaneous speech also raises interesting problems for a deeper level of discourse modeling. Here, the discourse segmentation (Grosz and Sidner, 1986) and the centering frameworks (Grosz, Joshi and Weinstein, 1995) are used to track the evolution of local and global attentional states in a spontaneous speech dialogue. Several issues characteristic of this corpus are discussed: the informality of sentence structure, the use of pronouns, and the role of prosodic cues. Empirical analysis indicates that the most useful domain for assigning discourse centers is the clause in spontaneous speech, rather than the utterance. Speakers appear to have little constraint on inserting references to discourse participants, but these assume center status only when used as active agents in the discourse. Furthermore, pitch accents can be used to promote the salience of discourse entities, allowing prominent entities to become backward-looking centers in subsequent clauses.

1. Introduction

Spontaneous speech poses problems for automatic systems, including speech recognition and understanding. While many investigators are making progress in recognition and dialogue processing, spontaneous speech also raises interesting problems for a deeper level of discourse modeling. Here, the discourse segmentation (Grosz and Sidner, 1986) and the centering frameworks (Grosz, Joshi and Weinstein, 1995) are used to track the evolution of local and global attentional states in a spontaneous speech dialogue.

In their 1995 synopsis, Grosz et. al. (GJW95), outline a centering framework that describes the mechanism of maintaining coherence in a discourse. They point out that a discourse is not a random selection of sentences, but must also exhibit coherence, both locally and globally. To do so, the conversant must use strategies to connect phrases, signal changes of topic or return to previous topics. Grosz et. al. propose a framework of ranking entities, called “centers”, in an utterance. Each utterance has a set of ranked forward-looking centers (Cf), and a single backward-looking center (Cb). Within a discourse segment, the highest-ranking Cf that appears in the next utterance becomes that next utterance’ s Cb. The ranking is a partial ordering (there are forward-looking centers of equal rank) and is determined in the context of the discourse segment, not from the isolated utterance. Relevant factors in determining which center is the Cb are syntactic position in the current and previous utterance (Subjects > Objects > Others) and pronounization (if any entity appears as a pronoun, the Cb will). A particular Cb can therefore be continued from one utterance to the next (it remains not only the Cb but also the highest-ranking Cf), merely retained (it’s the Cb but not the highest-ranked Cf) or the Cb can be shifted to a different discourse entity (optimally, the Cf of the previous utterance).

For example, the following discourse continues the Cb (Anna’, a previously mentioned author and the speaker’s daughter, whom the listener knows) by using a pronoun as the subject.

(1) what she was aiming to is to get to the Israeli teacher’s organization
• Cb = Anna ("she")
• Cf = {Anna, Israel Teacher’s Organization}
(2) so she could speak more about her book
• Cb = Anna, CONTINUE
• Cf = {Anna, book1}
(3) and then she has got another one that she is writing now.
• Cb = Anna, CONTINUE
• Cf = {Anna, book2 ("another one")}

The next utterance gives an example of a shift in the Cb.

(4) You see she wrote the other one
• Cb = Anna, CONTINUE
• Cf = {Anna("she"), listener ("you"), book1("the one")}
(5) and it’s been really a success
• Cb = book1 ("it"), SHIFT
• Cf = {book1}

Grosz et.al. maintain that there is a preference for sequences of continuation over sequences of retention, and a preference for sequences of retention over sequences of shifting.

Recently, a database of conversational telephone speech (Callhome) has allowed us to examine a spontaneous, unplanned (in fact, undirected in any way) dialogue such as the one used in the examples above. Although telephone speech introduces acoustic flaws, this media prevents the significant use of non-auditory acts (facial and other gestures). Participants must rely on speech to make themselves understood. In addition, participants in Callhome know each other and direct the conversation for their own purposes, rather than for external goals (assigned discussion topics or tasks, as in Switchboard or Trains/Maps). Callhome dialogues present a natural dialogue between two people who want to communicate information.

2. Issues in spontaneous Speech

In this work, the evolution of forward and backward centers was examined in a Callhome dialogue segment representing 3.34 minutes of a spontaneous speech between two participants who knew each other well. There were a total of 87

∗ This name has been changed to preserve the anonymity of participants.
clauses (although not complete) and 67 turns. This data was also prosodically labeled using a subset of the ToBI labeling system (Silverman et al., 1992). High level syntactic phrasing was hand-labeled in the course of the centering analysis. The speech data contained two distinct top-level discourse segments. The examples (1-5) above were selected from the first segment, which revolved around a book written by one speaker's daughter. This second segment involved a then-recent bombing in Jerusalem where the second speaker lives.

Several issues arose stemming from the characteristics of spontaneous speech and each will be discussed in turn. Briefly, this spontaneous speech contained many more pronouns than typical written text and the sentence structure was informal. In addition, frequent references to discourse participants using subjective pronouns ("I"/"you") and the use of discourse cues and filler phrases added ambiguity in tracking the evolution of centers. Evidence is also presented that indicates that participants are able to affect the ranking of forward-looking centers by using prominence (pitch accents). In addition, statistics on the length of discourse segments, and sequences of continuation, shifting and retention of centers in this dialogue segment are examined.

2.1 Informal Sentence Structure

This paper empirically explores the consequences of applying the centering theory framework to spontaneous conversations. In the process, differences between spontaneous speech and text raise several issues. In spontaneous speech, clauses are often well formed, but complete sentence boundaries are ambiguous and not necessarily bounded by speaker turns. This raises questions about the domain of centering. Since Grosz et al. define forward and backward centers over each utterance, issues arise when it is not clear what constitutes an utterance. The transcription (1 – 5) above represents only one side of the conversation (locally, the primary speaker). In addition, the examples above were segmented and the disfluencies were removed for illustrative purposes. The actual conversation between speakers A and B was transcribed (intonational phrases marked with ‘[]’): as:

A: Well this is what I wanted to tell you | what she was aiming to | is to get to the uh Israeli teacher’s organization so she could speak more about her book and then she has got another one that she is writing now |
B: Isn’t she something |
A: You see she wrote the one | and its been really a success |
B: Wonderful |
A: Yeah, she spoke to them | and the ministry of education has approved it | so it will be available |

Note that neither a turn nor an intonational phrase is suitable as the domain over which centers are defined. For example, A's "Yeah, she spoke ... available" contains more than one Cb (Anna ("she"), Ministry, the book ("it")). On the other hand, the intonational phase "is to get...is writing now" is also too long (Cb = Anna ("she"), organization) but the intonational phrase "what...aiming to" is too short. Instead this work uses single clauses as the domain of centering in place of Grosz et. al.'s utterances. While this motivates the segmentation found in the first example, it may not satisfy constraints in the last turn, reproduced below with center assignments.

(6) and she – Yeah, she spoke to them

- Cb = Anna ("she")
- Cf = [Anna ("she"), Israel Teachers' Organization ("them")]

(7) and the ministry of education has approved it

- Cb = book1 ("it")
- Cf = [book1 ("it"), ministry]

(8) so it will be available

- Cb = Cf = book1 ("it")

Notice the shift from Cb (6) = Anna to Cb (7) = book1. However, book1 is not mentioned in clause (6) and therefore can not be in the set of Cf (6). Since the Cb is theoretically selected from this set of previous utterance Cf's, this suggests that clauses may still be too restricted as centering domains. However, Grosz et al. noted similar occurrences and postulated a constrained shift back to a previously centered entity. In the dialogue examined here, 18 of the 33 Cb shifts were shifts to discourse entities not directly mentioned in the preceding main clause. In fact, the re-introduction of Israel Teacher's Organization ("them" in (6)) after six clauses is not very tightly constrained. This reference is less coherent than others, and may represent the tolerance in spontaneous speech for less coherent patterns.

2.2 Extensive Use of Pronouns

Another feature of spontaneous speech is that participants use pronouns much more liberally than might appear in text. For example, a randomly selected 138 consecutive word (one speaker) sample from Callhome contained 28 (or 20.3%) pronouns (excluding possessives). In contrast, data from the million-word Brown Corpus of English text contains 6.56% pronouns (4.75% in the informative-style sub-corpus and 11.94% in the imaginative-style subset.) In the same Callhome dialogue, participant A mentions her daughter's name only twice during a 2.95 minute discourse segment involving this daughter. Spontaneous speech utterances, with their abundance of pronouns, can potentially lead to ambiguities in determining a unique and locally derived Cb, possibly leading to a degradation in coherence or an increase in inference load. (Note: in (4) above, all entities are pronominalized: "she", "you" and "one").

2.3 References to Discourse Participants

Another aspect of the use of pronouns in spontaneous speech is the global salience of "I" and "you", usually references to the dialogue participants (although these pronouns can also appear in quotes within the dialogue, for example, when referring to a reported conversation.) Subjective pronouns occur frequently without preface in the previous utterance. Although this situation can be understood as a constrained shift to a globally salient entity (as described above), assigning these subjective pronouns Cb status can produce a dialogue containing frequent sequences of shifts. This violates the tendency noted by Grosz et al. for a preference for sequences of continued Cb's. For example, note the sequence of shifts that occurs with the following centering assignments in the segment of the same discourse between speakers A and B (especially (11 and 12b)).
To satisfy this ambiguity, the following rule is proposed:

- Ch = book1 (directly: "book", indirectly: "title")
- Cf = \{book1, title of book1\}

(10) A: it's uh f- oh oh uh folklore.

- Ch = book1 ("it"), CONTINUE
- Cf = \{book1\}

(11) B: Oh, I have to buy that book.

- Ch = speaker B ("I"), SHIFT
- [alternate: Ch = book1 (book), RETAIN ]
- Cf = \{1 (speaker B), book1\}

(12a) A: It's World F- folklore.

- Ch = book1 ("it"), CONTINUE from (7)
- Cf = \{book1, title of book1\}

(12b) A: I'll send you a pamphlet.

- Ch = speaker A ("you"), SHIFT
- [alternate: Ch = book1 ("pamphlet"), RETAIN]
- Cf = \{speaker A ("I"), speaker B ("you"), book1\}
  (indirectly as "pamphlet")

(12c) A: You'll see.

- Ch = speaker B ("you"), SHIFT
- [alternate: FILLER]
- Cf = \{speaker B ("you")\}

(12d) A: Well, she … [interrupted by B]

- Ch = Anna ("she"), attempt to SHIFT

In statement (11), "I" would be the assigned Ch using Grosz et al.'s pronoun rule, since it is the only pronoun in the clause. This represents a SHIFT in Ch from the Folklore book under discussion to one of the discourse participants, speaker B. However, this assignment is at odds with the tendency for the subject in the previous utterance (10); the book) to be retained as the next Ch. In (12b), a similar situation occurs. Here, the pronoun rule requires that one of the pronominalized entities (here "I" or "you") be assigned as the Ch. However, "pamphlet" refers to only entity and subject of the previous utterance (the Folklore book). Again, the preference for a Ch continued or retained from the previous utterance is at odds with the indiscriminate ability to re-introduce certain pronouns ("I" and "you") at any point in the discourse. Assigning center status to speaker B ("you'll see") in (12c) is even less satisfying since "you" has little informative status and acts as a filler or turn-holding phrase.

On the other hand, note the role of speaker B ("me", "I") in the next utterance (as speaker B interrupts A at (12d)),

(13a) B: She has to uh she has to uh autograph it for me.

- Ch = speaker B ("me") RETAIN from (12c)
- [alternate: Anna or book SHIFT]
- Cf = \{Anna ("she"), book1 ("it"), speaker B ("me")\}

(13b) B: and I have to have it.

- Ch = speaker B ("I"), CONTINUE
- Cf = \{speaker B ("I"), book1 ("it")\}

Here, repeated references to speaker B as an agent in the dialogue (as opposed to the passive (12c) ("you'll see") reference) suggest that the shifts to e.g. speaker B are justified. To satisfy this ambiguity, the following rule is proposed:

PARTICIPANT RULE: references to discourse participants are assigned Ch status if they are present as agents in a sequence of clauses

Hence, book would be assigned Ch status (RETAIN) in (11) and (12b) but speaker B would become the Ch in (13).

2.4 Discourse cues, fillers

Another class of phenomenon that arises in spontaneous speech are words and phrases that do not direct the evolution of information transmission. Such phrases perform other higher-level functions, such as holding the floor, arousing the listener's interest and indicating interest to the primary speaker. The "you'll see" in (12c) serves such a function. Other phrases, such as "That is fascinating.", "Remarkable", "Yeah", "I'll tell you." are similarly used as filler in the dialogue. These filler phrases do not directly effect the evolution of the discourse centers although such phrases may be influencing the drift of the conversation on a higher level. Since these phrases are frequent and widely distributed, the question of whether they serve to interrupt sequences of Ch continuation, shift or retention arises.

The tables below list the average length of sequences of each type of Ch evolution. Table 1 shows that there were indeed more clauses that continued the previous Ch (47) and the sequences of continuing clauses were in fact longer on average (1.74) than sequences of shifted Ch on average (1.45). There were many fewer clauses where the Ch was retained and these sequences seemed to be shorter on average than either sequences of continued or shifted Ch's. This may arise from the difficulty in determining whether a Ch is retained or continued, since the distinction is based on whether the entity is also the highest ranked Ch. An exhaustive set of factors for determining the Ch rank remains an open research question, blurring the distinction. In the next section, the role of pitch accents in spontaneous speech is examined in relationship to Ch ranking.

Table 1: Length of phrase sequences, allowing filler phrases to interrupt.

<table>
<thead>
<tr>
<th></th>
<th># phrases</th>
<th># sequences</th>
<th>Average</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>continue</td>
<td>47</td>
<td>26</td>
<td>1.8</td>
<td>0.86</td>
</tr>
<tr>
<td>shift</td>
<td>33</td>
<td>23</td>
<td>1.4</td>
<td>0.65</td>
</tr>
<tr>
<td>retain</td>
<td>11</td>
<td>9</td>
<td>1.2</td>
<td>0.42</td>
</tr>
</tbody>
</table>

Table 2 lists the average sequence length when filler phrases and discontinuities are not considered as interruptions to sequences. These filler phrases affected seven of the 26 (27%) Ch-continued sequences and four of the 23 (17%) Ch-shifted sequences. This does not indicate a marked preference for inserting fillers in areas where the dialogue is volatile, that is, where the centers are shifting. Instead there is evidence that fillers occur more frequently in stable segments of dialogue. However, more data is necessary to substantiate this view. As expected, all types of sequences are longer on average. Sequences of continued Ch's are 42% longer and sequences of shifted Ch's are 23% longer on average.

Table 2: Length of phrase sequences, not allowing filler phrases to interrupt.

<table>
<thead>
<tr>
<th></th>
<th># sequences</th>
<th>Average</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>continue</td>
<td>19</td>
<td>2.5</td>
<td>1.27</td>
</tr>
<tr>
<td>shift</td>
<td>19</td>
<td>1.7</td>
<td>1.02</td>
</tr>
<tr>
<td>retain</td>
<td>8</td>
<td>1.4</td>
<td>0.99</td>
</tr>
</tbody>
</table>
2.5 Prominence and Cf Ranking

As mentioned above, Grosz et. al. noted a tendency for the subject in an utterance to be the most highly ranked Cf, therefore the Cb in the next utterance. In spontaneous speech examples, this tendency seems to be frequently disregarded. Using (4 -5) as an example and disregarding the "you" (Participant Rule), it appears that the most highly ranked entity in (4) does not become the Cb in (5). The ranking in (4) is based on the syntactic position (SUBJECT > OBJECT) since the use of pronouns doesn't differentiate Anna ("she") from the book ("the one"). Here utterances (4 - 5) are reproduced, capitalizing words with prominence (pitch accented) and marking the intonational phrases ( | ).

(4) You SEE | she WROTE the ONE |
  • Cf = Anna, CONTINUE
  • Cf = {Anna("she"), book1 ("the one")}
(5) and IT'S been REALLY a SUCCESS |
  • Cf = book1 ("it"), SHIFT
  • Cf = {book1}

If prominence contributes to Cf ranking in spoken language, "the ONE" is promoted ahead of "she" and become the most likely candidate for Cb in the next utterance. In the test dialogue examined here, 33 Cb-shift events were assigned. Of these, 18 new Cb entities were not present either directly or indirectly in the previous main clause. Of the remaining 14 shift events, where a reference to the Cb did occur in the previous clause, only two were subjects of that previous clause and therefore the highest-ranking Cf. One of these was also pitch accent and the other was not. Eleven of the remaining 12 entities that became Cb's despite secondary ranking in the Cf set were pitch accent (see Table 3).

<table>
<thead>
<tr>
<th>Pitch accent:</th>
<th>TRUE</th>
<th>FALSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ranking Cf</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRUE</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>FALSE</td>
<td>11</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 3: Features of shifted Cb entity, as they appeared as Cf's in previous clause.

This suggests that pitch accenting is promoting entities to a higher ranking in the set of forward-looking centers than would be accounted for using syntactic position alone. To account for the effect of prosodic prominence on Cf entities, the following rule is proposed:

PROMINENCE RULE: Entities with prosodic prominence are more highly ranked in the set of forward-looking centers than those without prominence.

3. Conclusions

Several insights into the application of centering theory can be gained from the examination of spontaneous speech. First, the difficulties of assigning the backward-looking center become clear when the unrestricted re-introduction of pronominal referents to globally salient discourse participants occurs. It becomes clear that not all pronouns serve the role as centers in an utterance. Frequently, "I" and "you" are used for higher-level discourse purposes, such as judging or creating interest. Other phrases are also not relevant to center assignment as serve as fillers in the course of the evolution of a discourse. By eliminating such "red herrings" from consideration, the assignment of center status and ranking is simplified.

Second, the domain of centering is clarified. Utterances in dialogues are not well defined and clauses are better suited as the domain of centering assignments. However, the use of subordinate clauses was not examined closely and remains an open research question.

Empirical evidence from spontaneous speech extends the understanding of factors that contribute to the ranking of forward-looking centers. Prosodic prominence can serve to promote a discourse entity that is in subject position to the highest rank in the Cf set. This primary position may not be exclusive. It has been left to future work to determine if more than one entity can share the highest rank, allowing the speaker to dynamically chose the path through the dialogue at these branch points.

The length of Cb continuation, shift and retention sequences supports the finding that speakers generate longer sequences of continuation, presumably because both speaker and listener prefer them. The role of filler phrases in these sequences suggests that less volatile points in the dialogue (continuation rather than shifting of Cb's) support more frequent filler phrases. However, more work on the frequency and type of filler is necessary to extend any predictive conclusions about this data.

4. References: