BRIDGES: REGIONS BETWEEN DISCOURSE SEGMENTS

Nanette M. Veilleux

Department of Mathematics and Computer Science Department
Simmons College, Boston, MA, USA
veilleux@simmons.edu

ABSTRACT

A spoken dialogue or discourse can be globally organized into coherent discourse segments in which local salience and coherence properties apply. This paper looks at the regions between these discourse segments in American English spontaneous speech dialogue. These bridge regions do not form a coherent discourse segment themselves. When fluent, their utterances belong to neither the previous nor subsequent discourse segment. Often the utterances are disfluent. Dialogue turns in these bridge regions are characteristically shorter, on average, both in duration and number of words. Moreover, there are fewer pronouns, on average, in bridge turns, compared to the entire dialogue.

1. INTRODUCTION

In Grosz and Sidner’s 1986 paper on Discourse Segmentation and in Grosz et al.’s 1995 Centering Theory synopsis [1][2], Grosz and her colleagues outline a 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. In earlier work, Grosz and Sidner [1] proposed a theory of discourse structure that describes three components: a linguistic structure, an intentional state, and an attentional state. Furthermore, they explicitly distinguish between local and global coherence. While the earlier work focused primarily on global discourse structure, in later work, the centering framework was proposed to describe local structure.

The concepts of salience and coherence are central to these theories of discourse structure. Salience is relative among entities in a discourse rather than an absolute property of any entity. In addition, a discourse entity may gain or lose salience as the discourse progresses. Entities in a discourse will have a greater or lesser salience at some specified point in the discourse depending on the structure and state of the discourse, both locally and globally, at that point. Discourse segmentation and centering provide a framework to describe the evolving structure and state of the discourse, thus providing a “snapshot” of the discourse state at any point.

Globally, entities are more salient if they are contained (referenced directly or indirectly) in the current discourse segment, and can lose salience once a discourse segment is ended. Locally, discourse entities are most salient if they adhere to the constraints outlined in centering theory. In general, entities placed in certain grammatical contexts (e.g. subject position), or expressed as certain referring expressions (e.g. as pronouns) will have greater salience.

Coherence is also relative. A discourse is most coherent when the cognitive load requirement for the listener is less than it would be given other possible renditions, interpretations or structures. The cognitive load is lighter when more salient entities are evoked and is heavier when entities which are less salient, perhaps completely new, are mentioned. A coherent discourse maintains a relatively low cognitive load on the hearer by introducing new items in a structured, somewhat predictable way, as compared to a discourse that is less coherent.

1.1. Global Issues in Discourse Structure

On a global level, topics are introduced to a discourse and are modeled as discourse segment blocks (GS66) pushed onto the top of a first-in, last-out stack (a push action). The discourse segment block contains all entities named or evoked as part of the discussion of that topic. A discourse stack contains all blocks that are currently active and therefore grows and shrinks, from the top, as the discourse evolves. While a discourse segment is at the top of the stack, its entities are the most salient. The segment remain active, on the stack, as part of the discourse's attentional state, until the speaker moves to another topic. At this point, the current topic block may be dispensed with, modeled as being "popped off" the top of the stack and the next topic segment is pushed on in its place (a pop-push). Alternatively, the next topic may contribute to the current topic. In this case, the current discourse segment, and all the entities it has named or evoked, is not dispensed with. Instead, the new, contributing segment is pushed onto the stack, on top of the current topic (a push). The new segment becomes the current topic and its entities are the most salient. However, the previous topic is still within the discourse, and is returned to when the sub-topic is finished. At this completion point, the third possible stack-action occurs: removing a sub-topic segment (a pop) and a return to segment of a previous, not-yet-finished topic that now appears at the top of the stack. The now-current segment's entities are restored in salience, without the need for a full reintroduction.

For example, the following excerpt from the WBUR FM radio news corpus illustrates a discourse segment (A), followed by a sub-segment push(B), then pop(C).

A. In nineteen seventy-six, Democratic Governor Michael Dukakis fulfilled a campaign promise to de-politicize...
judicial appointment. He named Republican Edward Hennessy to head the State Supreme Judicial Court. 

B. For Hennessy, it was another step along a distinguished career that began as a trial lawyer and led to an appointment as associate Supreme Court Justice in nineteen seventy-one. …

Despite the state's massive budget deficit, Hennessy recently urged colleagues in the bar association not to retreat from these hard won gains.

Attorney C. Haskel Kassler chairs the Judicial Nominating Council, eighteen attorneys and laypeople charged with screening applicants for vacancies on the bench. Usually the J.N.C. refers three nominees to the Governor. His top choice is rated by bar associations and grilled by the Governor's executive council. Kassler says, unlike the Federal Supreme Court, there's no litmus test on particular issues that Massachusetts high court nominees must pass.

Example 1: Discourse segmentation of an FM Radio News broadcast.

Segment A. introduces the main issue to be presented in the news story: the appointment of a new head of the Supreme Judicial Court. Segment B digresses to a description of the current SJC head. In the discourse segmentation framework, this is modeled by pushing the discourse segment where entities that describe or are related to this individual (Hennessy) are most salient onto the stack. However, at the end of this segment (B), the broadcaster concludes this digression and resumes the discussion about the SJC appointment at hand. Discourse segment B is popped off the stack, revealing the discourse segment initiated with Paragraph A. Therefore, segment C is able to use referents to "nominees" and "bench" without the need to fully re-introduce the specific organization, the SJC.

The next example, cited by Nakatani, (Nakatani96) from an unrehearsed monologue on Sigmund Freud's personal life shows a similar case in a different speaking style:

A. so Freud had a few affairs with Fliese so big deal you know what I'm saying he knocked up Minnie Bernais he was married to Martha and knocked up his sister-in-law.

B. and they gave her hey - she had an abortion in one of these [clap]

C. alright he was human too alright.

Example 2: Discourse segmentation of a spontaneous speech monologues, from Nakatani [4].

As in the FM Radio news example above, [Freud, Fliese, Minne and Martha] are salient discourse entities in the first segment (A). After the digression to Martha's abortion (B), modeled as a push, then a pop at the end of this short segment, entities that were salient in (A) regain their salience. This is demonstrated by the ability of the speaker to felicitously use "he" in discourse segment (C) to refer to the discourse entity (Sigmund Freud) that was salient in (A). Again, re-introduction is not necessary, since the segment that contains Freud, etc. remains on the stack and is revealed after (B) is completed.

In practice, the discourse segmentation is often recognized retrospectively. A researcher labeling the discourse segments from text is best able to determine which utterances belong to a current block, a sub-segment block or a new block when she has access to the entire discourse. A participant in the discourse may also need a lag time in order to make this determination. Grosz and Hirschberg [3] found that intonational phrasing might cue breaks between discourse segments in professionally read speech. However, analysis on spontaneous, participant-motivated dialogues indicates that the push/pop may not be a clean as monologues or narratives suggest.

2. INTER-SEGMENTAL REGIONS: BRIDGES.

One corpus of spontaneous speech dialogues is the Callhome database. These dialogues contain participant-motivated telephone conversations between two acquaintances who have their own goals for the discourse. Although the quality of telephone speech introduces artifacts, it precludes non-speech communication. For a variety of reasons outside the study paradigm, participants tended to choose a conversational partner with whom they are not in regular contact. This serendipitously limits the context of immediately shared knowledge and therefore number of entities that are salient enough to both participants to need no introduction.

During the investigation and segmentation of several Callhome dialogues, areas that seemed between discourse segments, henceforth referred to as bridge segments, were discovered. In the following example from the Callhome telephone dialogue corpus, there appears to be a negotiation region before a new discourse segment is introduced. The discourse segment, which is in the process of ending, concerns speaker A's daughter who is an author. Speaker B attempts to ask about speaker A's sons (unsuccessfully) and then changes the subject to the difficult time (a then-recent bombing in her home city of Jerusalem, made universally salient by international press coverage).

A: Here you know she has a second masters in English as a second language
B: I know ah
A: so it became a book
B: Is it th- that's so wonderful. And the boys?
A: (laugh) It is exciting
B: uch it's exciting is not the word
B: It's really wonderful
B: It's such a happy news, you know
A: well
B: We had such (a) hard time here this morning
A: it's it's different
B: We had such a difficult time here this morning
A: I can imagine

Example 3: Inter-discourse segment region, bridging two discourse segments.

In theory, the cognitive load is less when the discourse is evolving locally, e.g. within a discourse segment, and is heavier as pops and pushes happen to execute topic shifts on the global level. In this case, as new topics (the daughter's second masters, the boys, the bombing) are pushed and popped from the stack in rapid succession, these negotiation regions...
can be seen to exhibit incoherent behavior. It is not clear whether these regions have characteristic prosodic or other patterns that allow a participant to follow the details of the discourse less carefully until it sorts itself out. For example, shorter turns with less semantic content could be used as place markers or introductions until the two participants have ascertained that they are synchronized on the next topic.

2.1. Characteristics of Bridge Regions

In this paper, four text-related factors were examined in four Callhome dialogues to see if they have a particular behavior in the areas of major discourse segment shifts:

- number of words/turn compared to the dialogue's average
- duration of turns, compared to the dialogue's average
- frequency of proper nouns
- frequency of pronouns

Transcriptions of segments of four American English Callhome dialogues were examined. Each was long enough to contain at least two major topics of discussion. After the transcriptions were segmented, the regions at these discourse segment boundaries were examined. Since local discourse segmentation may be difficult and therefore unreliable to label, only segments where the participants clearly shifted to another main (unrelated) topic of discussion were examined here. For this reason, the discourse segment actions examined here are pop-push types, or explicitly marked pops. The alternative discourse action, pushing a subordinate topic onto a stack is more ambiguous to pinpoint. For example, a conversation between two jazz musicians pushes a segment about a collaboration onto the stack while they are discussing the larger topic of how fit to play one participant is.

B. Well yeah I'm playing, sure.
A: so your chops are up?
B: Well no they're not up but I'm playing it's I just it's very hard to get my chops back in form but I'm going to try. I'll tell you what's happening (breath) uh
first of all you knew that we had &Conti here last year.
A: yeah but I haven't talked to hi- yes I did talk to Count, yeah, mhm.
B: (( ))
A: What did he say?
B: well he had said he had a good time but he said you didn't play much.
A: [laugh]
B: [laugh]
B: Well no, because uh my chops were down. My chops were down so I just conducted the band. He has the video of the night we played here you and whenever you can get together with him sometime have him --

Example 4: Ambiguous discourse structure.

In the last sentence "he" ("he has the video") refers to the common acquaintance (Count), leaving some ambiguity about whether there are two discourse segments actually present. Theoretically, once the segment about Count is popped, a pronoun should not be a sufficient referring expression. For these reasons, push discourse segments are not used in the present study.

On examination, the regions between the major discourse segments (between the pop and the next push) exhibit behavior that can be categorized into three main types:

1) Wide (several utterances or turns) but fluent. For example, in the bridging utterances in the Callhome dialogue above (Example 3), where the two participants are speaking about different topics.

2) Wide, but disfluent: several utterances that contain fragments, disfluencies and non-speech acts, in between discourse segments. In the example below, speaker B concludes a discussion about what she's been doing and speaker A introduces another discourse segment (about B's family). Between this pop and subsequent push are four turns, several non-speech acts, and incoherent speech that lasts for nearly 5 seconds.

Example 5: A wide, but disfluent region between two discourse segments.

3) Narrow: This type represents canonical discourse segment shift where one segment is ended and another one immediately begins, sometimes introduced with a question: Here a discourse segment about a common friend who recently had surgery is popped and a discourse segment about Japan begins:

In the wide bridge regions, both fluent and non-fluent, the two discourse participants have not agreed to a topic of discussion. Here, one participant may be talking about a previous topic while the other might be talking about something else (as in the first Callhome dialogue). Alternatively, the bridge areas may have a sequence of incomplete phrases, disfluencies or short questions while one member tries to figure out what the other is talking about, in order to move the conversation to that topic.

3. RESULTS

The wide regions, both fluent and disfluent, contain sequences of short utterances. In the fluent regions, these might be questions or some other indication that one speaker is trying to adjust to a topic change introduced by the other. In this case, these regions would be expected to exhibit shorter turn lengths, since neither speaker is actually on topic yet. In the disfluent bridge regions, many turns contain non-speech acts or disfluencies. Tables 1 and 2 compare the average number of words in a turn and the turn duration (sec.) in the bridge regions and in the entire dialogue.
Table 1: Average length of a turn, in words, for the entire dialogue and for bridge regions.

<table>
<thead>
<tr>
<th>dialogue</th>
<th>whole dialogue</th>
<th>bridge regions</th>
<th>% diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>13.56</td>
<td>3.07</td>
<td>77</td>
</tr>
<tr>
<td>2</td>
<td>10.76</td>
<td>6.125</td>
<td>43</td>
</tr>
<tr>
<td>3</td>
<td>6.22</td>
<td>3.44</td>
<td>45</td>
</tr>
<tr>
<td>4</td>
<td>4.19</td>
<td>1.66</td>
<td>60</td>
</tr>
</tbody>
</table>

Table 2: Average length of a turn, in seconds, for the entire dialogue and for the bridge regions.

<table>
<thead>
<tr>
<th>dialogue</th>
<th>whole dialogue</th>
<th>bridge regions</th>
<th>% diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.11</td>
<td>1.42</td>
<td>65</td>
</tr>
<tr>
<td>2</td>
<td>4.23</td>
<td>2.20</td>
<td>48</td>
</tr>
<tr>
<td>3</td>
<td>2.31</td>
<td>1.43</td>
<td>38</td>
</tr>
<tr>
<td>4</td>
<td>1.427</td>
<td>.98</td>
<td>31</td>
</tr>
</tbody>
</table>

Table 3: Average number of proper nouns in a turn for the entire dialogue.

<table>
<thead>
<tr>
<th>dialogue</th>
<th>whole dialogue</th>
<th>bridge regions</th>
<th>% diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.35</td>
<td>0.07</td>
<td>81</td>
</tr>
<tr>
<td>2</td>
<td>0.53</td>
<td>0.50</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>0.32</td>
<td>.03</td>
<td>92</td>
</tr>
<tr>
<td>4</td>
<td>0.08</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 4: Average number of pronouns in a turn for the entire dialogue.

<table>
<thead>
<tr>
<th>dialogue</th>
<th>whole dialogue</th>
<th>bridge regions</th>
<th>% diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.68</td>
<td>0.55</td>
<td>68</td>
</tr>
<tr>
<td>2</td>
<td>1.23</td>
<td>0.62</td>
<td>51</td>
</tr>
<tr>
<td>3</td>
<td>0.98</td>
<td>0.28</td>
<td>74</td>
</tr>
<tr>
<td>4</td>
<td>0.63</td>
<td>0.24</td>
<td>61</td>
</tr>
</tbody>
</table>

4. CONCLUSIONS

Finding factors that indicate upcoming shifts can be helpful in automatically detecting segmentation. This study shows that bridge discourse segments contain shorter turns, both in duration and in the number of words. In addition, they contain fewer pronouns, although this may be a consequence of containing shorter turns. If an automatic system has access to turn information, measuring the length of turns may indicate the presence of an up-coming discourse shift. Of course, in a running application, enough data must be collected to obtain the average turn length for the current dialogue before determining whether a turn is relatively shorter. Moreover, a short turn itself is not diagnostic since turn length within discourse segments is variable. However, this factor may be useful when combined with others.

In addition, these investigations may offer insight into human cognitive requirements during topic switches. Global discourse shifts are points in a conversation where local coherence fails. Short, frequent turns where little information is transmitted may serve a practical purpose such as lessening the cognitive load while the participants adjust to a new topic, replete with its own set of discourse entities. This has implications in human-computer interactions where the automatic system may require a global discourse shift.

Future investigation will attempt to determine a more precise definition and taxonomy for these bridge regions in spontaneous speech by segmenting more Callhome dialogues and investigating this phenomenon in other corpora. Once the endpoints of a bridge region are more rigorously understood, a typical window (number of turns or duration based) over which the factors discussed here and others can be monitored to automatically detect bridge regions. Furthermore, other factors will be examined, including whether disfluencies are more frequent in bridge regions and whether there are prosodic cues, such as lower amplitude, during these inter-discourse segment periods.

5. REFERENCES