The Association between Subject Matter and Discourse Segmentation

Susan Haller and Timothy Fossum

Computer Science Department
University of Wisconsin - Parkside
Kenosha, Wisconsin 53141-2000
{*haller, fossum*}@cs.uwp.edu

Abstract

We present an empirical study of the relationship between subject matter and the perception of discourse segment boundaries in dialogue. In this study, we presented subjects with dialogues that had the same discourse structure, instantiated with content from different domains. They were asked to note where these dialogues started to digress. Our results indicate that there is a signi£cant association between dialogue subject matter and subjects' perceptions of discourse segment boundaries. These results have implications for the design of interactive systems.

Introduction

A *discourse segment* is a unit of text with an identi£able purpose or intention (Grosz & Sidner 1986). Computational models for discourse segmentation have played an important role in research on discourse processing and have inßuenced the design of systems that interact with humans.

The computational models for discourse segmentation described in the literature are domain-independent. In this study, we show that subject matter signi£cantly affects human perception of discourse segmentation, even though the interactions have the same rhetorical structure. This result has important implications for the design of interactive systems.

In the next section, we give a brief overview to work related to discourse segmentation. We follow this with a description of an algorithm that we have implemented to automatically identify potential segment boundaries in a system that gives driving route advice. In our problem statement section, we hypothesize that human perception of discourse segment boundaries depends on dialogue subject matter. The next two sections describe the design and results of our study. In our conclusions we discuss what our results suggest for the design of interactive systems.

Related Work

Early approaches to analyzing discourse segmentation focussed on careful analyses of corpora to develop segmentation schemes. These approaches produced organizations based on discourse features such as coherence relations

(Polyani 1988), the attentional, intentional, and linguistic structure of the discourse (Grosz & Sidner 1986), a hierarchy of £xed schemata (McKeown 1985), and a hierarchical organization of *rhetorical structures* (Mann & Thompson 1988). These approaches relied on the subjective judgments of researchers.

More recently, researchers have conducted empirical studies of discourse segmentation. Several investigations have focussed on the ability of human coders to agree with each other on segmentations of text according to a speci£c model (Nakatani, Hirschberg, & Grosz 1995; Moser & Moore 1995). Others have asked subjects to select discourse segment boundaries given a choice of potential boundary sites (Hearst 1994; Litman & Passonneau 1995). While these studies have considered different types of corpora (news stories, spontaneous narrative, and task-oriented dialogues, for example), the underlying assumption has been that domain-independent techniques exist for segmenting discourse that can be operationalized.

From the perspective of natural language generation, it is well understood how text structure reßects relationships between pieces of domain content and how text structures can be built up into segments (McKeown 1985; Moore & Pollack 1992; Young & Moore 1994). However, the relationships that generation researchers look for and propose (attribution and constituency, for example) are domain-independent.

In this paper, we report results that suggest that domain-related features that we have not characterized yet have a signi£cant effect on subjects' perceptions of discourse segmentation.

Automatically Detecting Discourse Segment Boundaries

In previous work, which we sketch here, we designed and implemented an interactive text planner that detects when a participant's questions lead the discussion across a discourse segment boundary (Haller 1994; 1996). Following Grosz and Sidner(Grosz & Sidner 1986), we assign a purpose to each discourse segment. When there is a change of discussion purpose, we call this a *digression*.

We based our work on a discourse theory that uses a uniform conception of sentence topics and structurally higher-

p_1 leads to $cont_{p_1} = \{LexHypo1.1\}$; p_2, the phrase actor containing *"ausgebaut"* as VERBPARTPASSIVE, results in $cont_{p_2} = \{LexAmbigHypo1.1, LexAmbigHypo1.2\}$. This leads to two different interpretation contexts (two tuples), one with the *upgrade* and the other with the *removal* reading.[4] Since we consider semantic interpretation as a search problem to link the conceptual correlates of the content words spanning the semantically interpretable subgraph – composed of *"Speicher"* *(memory)* and *"ausgebaut"* *(removed/upgraded)* – a search for appropriate conceptual relations is conducted in the domain knowledge base. It retrieves the relations REMOVE-PATIENT and UPGRADE-PATIENT for MEMORY.1-01 with regard to REMOVE.2-01 and UPGRADE.2-02, respectively. The resulting assertions are added to new interpretation contexts, viz. (UPGRADE.2-02 UPGRADE-PATIENT MEMORY.1-01)$_{NewHypo1.1}$ and (REMOVE.2-01 REMOVE-PATIENT MEMORY.1-01)$_{NewHypo1.2}$. These contexts form the context set $cont_{p_3}$ and are linked to the phrase actor p_3 that contains the entire dependency graph for the input sentence (cf. Figure 5, left, lower part). Note that a single syntactic structure is associated with two different semantic interpretations.

As the input text is incrementally analyzed, each of the interpretation results resides in *terminal contexts*, i.e., ones that have no children. Finally, each sentence is assigned its set of terminal contexts as sentential reading(s).

Selection. If multiple semantic interpretations exist after the analysis of a sentence has been completed, we apply several assessment heuristics to accumulate different sources of evidence for selecting a preferred reading: syntactic coverage, ranking of semantic interpretations (e.g., PP interpretations receive a higher weight than genitives), and the potential of particular assertions to foster additional, reasonable inferences.

The weight of a context is determined by iteratively adding the weight of the current terminal context and its ancestor contexts in the context graph. Selecting the best reading(s) for a sentence then boils down to the selection of the terminal context(s) with maximal weight. This can be considered as a brute-force mechanism for removing possibly contradictory sets of assumptions from the underlying knowledge base (for more sophisticated ways of dealing with contradiction identification and advanced belief revision using contexts, cf. Martins & Shapiro (1988)).

Conclusions

We have introduced a formal context mechanism for representing and reasoning about ambiguities which occur at different levels of text analysis. An important feature of our approach is the clear separation between knowledge levels: part-of-speech and structural ambiguities are handled at the syntax level (by phrase actors), while lexical polysemy and

sentence-level semantic ambiguities are dealt with at the context level (by multiple interpretation contexts). The dynamic context extension mechanism which reflects the incremental processing strategy of the input text further constrains the satisfiability of additional assertional axioms and supports the disambiguation of the alternative readings.

A crucial feature of this methodology is its clean embedding into the terminological reasoning mechanisms underlying text understanding. Ambiguities are represented as disjunctions of logical axioms (Buvač, Buvač, & Mason 1995), tentatively assumed to hold in their local contexts. This implies that all interpretation alternatives be enumerated explicitly (cf. Reyle (1993) for a different approach using semantic underspecification as a technique to cope with scoping ambiguities of quantifiers, an issue not touched upon here). The use of contexts as a formal vehicle for encapsulating alternative readings and reasoning about their resolution distinguishes our work from more ambitious uses of contexts, e.g., to detect contradictory information within or between different contexts for the purpose of belief revision as discussed by Martins & Shapiro (1988).

References

Buvač, S.; Buvač, V.; and Mason, I. A. 1995. Metamathematics of contexts. *Fundamenta Informaticae* 23(3):263–301.

Church, K., and Patil, R. 1982. Coping with syntactic ambiguity or how to put the block in the box on the table. *American Journal of Computational Linguistics* 8(3/4):139–149.

Ferrari, G. 1997. Types of contexts and their role in multimodal communication. *Computational Intelligence* 13(3):414–426.

Hahn, U., and Romacker, M. 2000. Content management in the ITSYNDIKATE system: How technical documents are automatically transformed to text knowledge bases. *Data & Knowledge Engineering* 35(2):137–159.

Hahn, U.; Bröker, N.; and Neuhaus, P. 2000. Let's PARSETALK: Message-passing protocols for object-oriented parsing. In Bunt, H., and Nijholt, A., eds., *Advances in Probabilistic and other Parsing Technologies*. Dordrecht, Boston: Kluwer. 177–201.

Martins, J. P., and Shapiro, S. C. 1988. A model for belief revision. *Artificial Intelligence* 35(1):25–79.

McCarthy, J. 1993. Notes on formalizing context. In *Proceedings of the IJCAI'93*, 555–560.

Reyle, U. 1993. Dealing with ambiguities by underspecification. *Journal of Semantics* 10:123–179.

Romacker, M.; Markert, K.; and Hahn, U. 1999. Lean semantic interpretation. In *Proceedings of the IJCAI'99*, 868–875.

Sowa, J. F. 1995. Syntax, semantics, and pragmatics of contexts. In *Conceptual Structures: Applications, Implementation and Theory. Proceedings of the ICCS'95*, 1–15.

Strube, M., and Hahn, U. 1999. Functional centering: grounding referential coherence in information structure. *Computational Linguistics* 25(3):309–344.

Tabossi, P. 1989. What's in a context? In Gorfein, D. S., ed., *Resolving Semantic Ambiguity*. New York: Springer. 25–39.

Wiebe, J.; Hirst, G.; and Horton, D. 1996. Language use in context. *Communications of the ACM* 39(1):102–111.

Woods, W. A., and Schmolze, J. G. 1992. The KL-ONE family. *Computers & Mathematics with Applications* 23(2/5):133–177.

---

[4]The initial contexts *LexAmbigHypo2.1* and *LexAmbigHypo-2.2* are abandoned because of syntactic reasons. The VERBPART-PERFECT reading of *"ausgebaut"* cannot be bound by any of the preceding word actors in the sentence.

order discourse topics (van Kuppevelt 1992). During a dialogue, our text planner selects content for each response to a question based on what £ts best into a single overarching text plan. As the system formulates each text plan for a response, it must attach it to an overarching text plan as a subplan. If the plan is attached successfully, the system registers it as a contribution to satisfying the discussion purpose. If the system cannot incorporate the text plan into this overarching text plan, it marks the user's question and its response to it (the text plan) as a digression.

The system's text plan operators are based on RST (introduced above). In RST, each rhetorical structure relates some essential textual content – *the nucleus* – to one or more supporting units of content – called *satellites*. There are two categories of rhetorical structures based on the effects of using a particular structure to relate units of selected content. A *presentational rhetorical relation* is used to combine textual units in a way that positively affects the attitude of the listener. Presentational relations behave like speech acts that advise or persuade rather than just inform. In contrast, a *subject-matter rhetorical relation* is used only to convey content and its relationship to other text content.

In our system, the new content of a system response can only be incorporated into the overarching text plan with a subject-matter relation if the content is the satellite in a subject-matter relation, and the nucleus is one of: (a) a primitive speech act, (b) content introduced with a presentational relation, or (c) a text span that uses a presentational relation. In contrast, if the system formulates a response that incorporates the new content as the satellite in a subject-matter relation that has a nucleus that is either (a) content that has already been included using another subject-matter relation or (b) a text span that uses another subject-matter relation, the system marks the user's question and its plan for responding to it as a digression.

Figure 1 gives an example of this process, based on a particular interaction with the system that is not shown. In response to the user's initial question, the system posts a discourse goal and formulates an initial text plan that consists solely of speech act1. In this example, after our system performs the speech act, the user asks a question. To answer it, the text planner augments speech act1 with content1 using presentational rhetorical relation *PR1*. After this portion of the text plan is executed, the user asks another question that the system answers by including content2, using the presentational rhetorical relation *PR2* to augment the text span covered by *PR1*. Another question from the user requires the planner to add content3 using a subject-matter relation, *SR1*, to augment content1. Finally, another question from the user requires the planner to try to attach content4 to the plan. Since the system must use subject-matter relation *SR2* to augment content3 to do this, it marks the text plan used to convey content4 as a digression from the discourse purpose.

We de£ne *dialogue structure* to be the actual sequence of exchanges that occur during the interaction and the incremental formulation of the text plan. The same text plan (that is, a plan with the same structure of rhetorical relationships) might have different dialogue structures determined by the
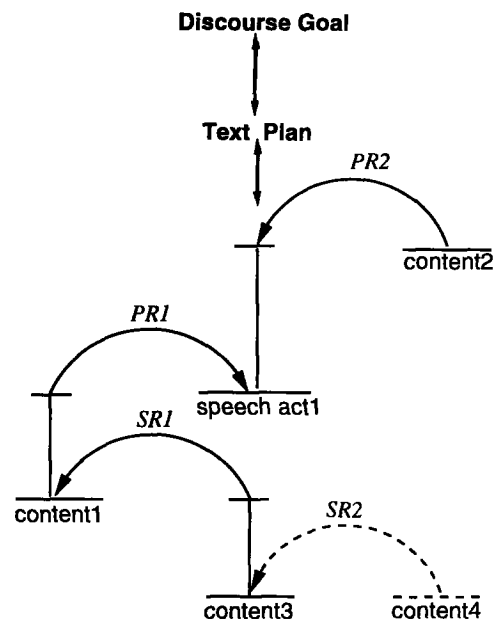

Figure 1: Example Text Plan Illustrating a Digression

user's focus of interest in asking questions. As an example, the text plan in Figure 1 could have been built following a different dialogue structure. Following the presentation of speech act1 and content1 using relation *PR1*, the user's next question might prompt the system to convey content3 using relation *SR1*. This could be followed by a question that requires content2 to be incorporated using relation *PR2*.

Statement of Problem

As described above, our system identi£es digressions based solely on rhetorical structure, independent of subject domain. We are interested in whether humans perceive digressions independently of subject domain. We hypothesize that subject matter has a signi£cant in²uence on human perception of where digressions occur. If this is true, the design of interactive systems must take this into account, as we discuss more fully in our conclusion.

Method

We £rst recorded three actual interactions between a human participant and our system, with three different dialogue structures (called A, B, and C) but the same structure of rhetorical relationships. All three dialogues were domain-speci£c, concerning a subject (the human participant) interacting with our system to get driving route advice. We will call this the *driving* domain. In each of the three dialogues, our system identi£ed a digression using the method described above.

We cast the three original dialogues from the driving domain into two other subject domains, health and auto £nance, preserving the rhetorical structure of the original dialogues. Since our system identi£es digression based on rhetorical structure, it would identify the digression at the

same location in each dialogue structure regardless of the domain. In all, the three subject domains were:

1. health (a patient/doctor interaction)
2. driving (described above)
3. £nance (a client/£nancial consultant interaction)

We gave untrained human subjects (hereafter called *raters*) these dialogues, asking each rater to identify the site in the dialogues where a digression occurs (if at all). We expected that our raters, when given interactions having the same dialogue structure but different subject matter, would identify digression sites differently.

As an example, here is one of the original dialogues – dialogue structure A – in the driving domain (subject domain number 2). H is the human participant and S is our system. The digression identi£ed by our system is marked with a dagger (†).

```
H: Should I take Maple or Sheridan to
   go to the Eastern Hills Mall?
S: You should take Maple.
H: Why?
S: Maple avoids heavy traffic.
H: Why?
S: There are fewer businesses on Maple
   than on Sheridan.
H: † Why is traffic heavy now?
S: Since it's rush hour.
H: How long does rush hour usually
   last?
S: About two hours.
```

Here is the same dialogue structure A, but cast in the £nance domain (subject domain number 3). H is planning on getting a new car, and S is a £nancial consultant.

```
H: Should I buy or lease a car?
S: You should buy a car.
H: Why?
S: Buying avoids penalty payments at
   the end of the lease.
H: Why?
S: Since you drive more than 15,000
   miles a year.
H: † Why do leases have these penal-
   ties?
S: Since the dealer will want to resell
   the car at the end of the lease.
H: What if I want to buy the car at the
   end of the lease?
S: Leases usually have an option to
   buy.
```

A third dialogue – not shown – was created using dialogue structure A but cast in the health domain (subject domain 1).

Similarly, for each of the other two dialogue structures, B and C, we cast driving-domain interactions into each of the other two subject domains, health and £nance. In total, we created nine structure/subject dialogues, using the notation "A1" to represent dialogue structure A (illustrated above) with dialogue topic 1 (health) and so on.

We devised an instrument that could be administered to our raters. The £rst page of the instrument consisted of instructions that identi£ed the purpose of the instrument:

We are interested in how people are able to identify when a digression has occurred in an interaction between two people.

The instructions identi£ed a digression as

... when one person begins to stray from the main subject of the interaction, that is, they begin to talk about something that gets away from the goal or purpose of the interaction.

The instruction page then indicated that three short interactions were to follow, and asked the rater to mark, for each interaction, the location in the dialogue

... where the interaction begins to digress (in [the rater's] opinion) from the purpose of the interaction (which is given in the £rst question of the interaction).

If the rater thought that there was no digression for a particular dialogue, the rater was asked to leave the page unmarked.

The next three pages of the instrument consisted of the three dialogues to be rated. To ensure that each rater saw all three possible dialogue structures (A, B, and C) and all three subject domains (health, driving, and £nance), the dialogue combinations were distributed into three groups:

Group	structure/content
X	A1, B2, C3
Y	B1, C2, A3
Z	C1, A2, B3

In order to obtain ratings that were large enough to justify statistical signi£cance, we clustered the responses for each dialogue structure. We used our system to divide each of the dialogue structures into three partitions: the partition prior to the digression identi£ed by our system (called the "before" segment), the partition consisting of the exchange at the location of digression (called the "at" segment), and the partition following the digression exchange (called the "after" segment). If a rater identi£ed no digression for a particular dialogue, this was considered to be in the "after" partition, since "no digression" was considered to be a digression occurring in£nitely "after".

Sixty untrained student volunteers in an undergraduate setting were used as raters, divided randomly into three groups (X, Y, and Z above) of twenty. Each of the raters was given a packet consisting of the instruction page and the three dialogues to be rated. All the ratings were collected and tabulated by subject domain and by partition ("before", "at", or "after"),

Carletta (Carletta 1996) argues that untrained raters (naive coders) are suitable for subjective rating, as we have done. We rely on raters consistently interpreting the instruments' written instructions, all of which were the same. If untrained raters were unable to interpret or operationalize the instructions, we would expect insigni£cant differences in their ratings.

Statistical Results

The table below gives the number of raters, among all rating groups, who selected a digression "before", "at", and "after" the location identi£ed by our system, separated into the three different subject domains.

Domain	Digression ratings		
	before	at	after
Health	19	10	31
Driving	29	18	13
Finance	22	19	19

We conclude that perception of where a digression begins is not independent of subject domain ($\chi^2 = 13.36$, df $= 4$, $p < .01$, two-tailed).

Conclusion

Our results show that human perception of discourse segment boundaries is sensitive to the subject domain. Consequently, interactive systems based on domain-independent techniques cannot reliably identify discourse segment boundaries consistent with human perception of these boundaries.

While we do not have a theory of handling digressions, we note that our present system does attempt to refocus the dialogue as part of its response to a digressive question. For example, in response to H's question in the sample dialogue from the driving domain: Why is traffic heavy now?, our system responds with: Since it's rush hour, but it also adds: As I was saying, you should take Maple. We removed these refocusing statements for the purpose of this study.

We describe three contexts in which interactive systems may need to identify discourse segment boundaries consistently with humans:

Resource Limitations Resource limitations (seat time, network bandwidth, knowledge capacity) may make unconstrained dialogue unrealistic. For example, a heavily used interactive airline reservation system that is able to detect when its interaction with a user is no longer contributing to the purpose of the interaction may be able to refocus the user as to the interaction's purpose.

Goal-Directed Systems A system may have its own agenda. For example, a tutoring system that is able to detect digressions can help students to stick with the study topic. An on-line problem resolution system in an industrial environment may need to keep a user focussed on the problem at hand.

Interactive Fit An interactive system may identify discourse segment boundaries, but inconsistently with humans. If the human perceives the discourse segment boundary to be later than ("after") the site identified by the system, the human may think that the system's behavior is impatient. If the human perceives the discourse segment boundary to be earlier than ("before") the site identified by the system, the human may think that the system's behavior is pointless.

A possible scale for subject domains

The data suggests that the more personal the subject matter is to a human subject, the longer it takes for the subject to perceive that the discussion is digressing. In the driving domain, 78% of our raters perceived the dialogue to digress "before" or "at" the location identified by our system. In contrast, in the health domain, more than half of our raters perceived the dialogue to digress "after" (including "no digression"). The results for the financial domain appear to be somewhere in between.

This suggests that our domain-independent algorithm for identifying discourse segment boundaries may be appropriate in factual domains with little personal long-term relevance to humans.

We have shown that untrained human raters perceive discourse segment boundaries differently based on the discourse subject. Consequently, designers of interactive systems should take subject matter into account when determining whether a conversation is purposeful. Moreover, systems that are expected to be able to present substantially different types of information – or information of different personal relevance to a user – will need this same capability.

Acknowledgments

References

Carletta, J. 1996. Assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics* 181–188.

Grosz, B. J., and Sidner, C. L. 1986. Attention, intentions, and the structure of discourse. *Computational Linquistics* 175–204.

Haller, S. 1994. Recognizing digressive questions during interactive generation. *The Seventh International Natural Language Generation Workshop* 181–188.

Haller, S. 1996. Planning text about plans interactively. *International Journal of Expert Systems* 85–112.

Hearst, M. 1994. Multi-paragraph segmentation of expository text. *Proceedings of the Thirty-second Annual Meeting of the Asscociation for Computational Linguistics*.

Litman, D., and Passonneau, R. 1995. Combining multiple knowledge sources for discourse segmentation. *Proceedings of the Thirty-third Annual Meeting of the Asscociation for Computational Linguistics* 108–115.

Mann, W. C., and Thompson, S. A. 1988. Rhetorical structure theory: Towards a functional theory of text organization. *TEXT* 243–281.

McKeown, K. 1985. *Text Generation: Using Discourse Strategies and Focus Constraints to Generate Natural Language Text*. Cambridge University Press.

Moore, J., and Pollack, M. 1992. A problem for RST: The need for multi-level discourse analysis. *Computational Linguistics*.

Moser, M., and Moore, J. 1995. Using discourse analysis and automatic text generation to study discourse cue usage. *AAAI Technical Report: Empirical Methods in Discourse Interpretation and Generation, #SS-95-06* 92–98.

Nakatani, C.; Hirschberg, J.; and Grosz, B. 1995. Discourse structure in spoken language: Studies on speech

corpora. *AAAI Technical Report: Empirical Methods in Discourse Interpretation and Generation, #SS-95-06* 106–112.

Polyani, L. 1988. A formal model of discourse structure. *Journal of Pragmatics* 601–638.

van Kuppevelt, J. 1992. About a uniform conception of S- and D- topics. *Proceedings of the Prague Conference on Functional Approaches to Language Description.*

Young, M., and Moore, J. 1994. DPOCL: A principled approach to discourse planning. *The Seventh International Natural Language Generation Workshop* 13–20.