Towards CST-Enhanced Summarization

Zhu Zhang
School of Information
University of Michigan
Ann Arbor, MI 48109
zhuzhang@umich.edu

Sasha Blair-Goldensohn
School of Information
University of Michigan
Ann Arbor, MI 48109
sashabg@umich.edu

Dragomir R. Radev
School of Information and
Department of EECS
University of Michigan
Ann Arbor, MI 48109
radev@umich.edu

Abstract

In this paper, we propose to enhance the process of automatic extractive multi-document text summarization by taking into account cross-document structural relationships as posited in Cross-document Structure Theory (CST). An arbitrary multi-document extract can be CST-enhanced by replacing low-salience sentences with other sentences that increase the total number of CST relationships included in the summary. We show that CST-enhanced summaries outperform their unmodified counterparts using the relative utility evaluation metric. We also show that the effect of a CST relationship on an extract depends on its type.

Introduction

Text summarization

Text summarization is "... the process of distilling the most important information from one a source (or sources) to produce an abridged version for a particular user (or users) and task (or tasks)" (Mani & Maybury 1999).

One distinguishes single-document summarization (SDS) and multi-document summarization (MDS). MDS, which our approach will be focusing on, is much more complicated than SDS in nature. Besides the obvious difference in input size, several other factors account for the complication, e.g.:

- Multiple documents might come from different sources, be written by different authors, and therefore have different styles, although they are topically related. This means that a summarizer cannot make the same coherence assumption that it can for a single article.

- Multiple documents might come out of different time frames. Therefore an intelligent summarizer has to take care of the temporal information and try to maximize the overall temporal cohesiveness of the summary.

- Descriptions of the same event may differ in perspective, or even conflict with one another. The summarizer should provide a mechanism to deal with issues of this kind.

We make the distinction between information-extraction-vs. sentence-extraction-based summarizers. The former, such as SUMMONS (Radev & McKeown 1998), rely on

an information extraction system to extract very specific aspects of some events and generate abstracts thereof. This approach can produce good summaries but is usually knowledge intensive and domain dependent. Sentence extraction techniques (Luhn 1958; Radev, Jing, & Budzikowska 2000), on the other hand, compute a score for each sentence based on certain features and output the most highly ranked sentences. This approach is conceptually straightforward and usually domain independent, but the summaries produced by it often need further revision to be more smooth and coherent.

Centroid-based summarization and MEAD

Centroid-based summarization is a method of multi-document summarization. It operates on a cluster of documents with a common subject (the cluster may be produced by a Topic Detection and Tracking, or TDT, system). A cluster centroid, a collection of the most important words from the whole cluster, is built. The centroid is then used to determine which sentences from individual documents are most representative of the entire cluster.

MEAD is a publicly available toolkit for multi-document summarization (Radev, Jing, & Budzikowska 2000; MEAD 2002). It is based on sentence extraction. For each sentence in a cluster of related documents, MEAD computes a number of features and uses a linear combination of them to determine what sentences are most salient.

In this paper, we will investigate the effect of inter-document rhetorical relationships on summary extraction. Two potential summaries for the same cluster are illustrated in Figure 1. As we can see, there is an obvious contradiction between sentences 2 of both summaries (which are obviously from different sources). The eventual summary is well justified to either include both of them (to provide an all-sided view of the event, e.g.) or to include only one of them (to accommodate readers' bias over sources, e.g.).

The structure of the paper will be the following. First we highlight some related work; second we introduce the CST theory (Radev 2000) that we developed and formalize our framework; third, we present a CST identification user study; then we propose the idea of enhancing summarization using CST relationships and present corresponding experimental results; we conclude the paper with some discussion of future work.

| S1: A Gulf Air Airbus A320 carrying 143 people from Cairo, Egypt, to Bahrain crashed today in the Persian Gulf. |
| S2: A320 has a good accident record; the crash in the Persian Gulf today that killed 143 people was the aircraft's *fourth* air disaster. |
| S3: ... |
| S1: A Gulf Air Airbus A320 carrying 143 people from Cairo, Egypt, to Bahrain crashed today in the Persian Gulf. |
| S2: A320 has been involved in *six* accidents, including Wednesday's. |
| S3: ... |

Figure 1: Two sample summaries

Related work

Our work bears a strong connection to Rhetorical Structure Theory (RST) (Mann & Thompson 1988), which is a comprehensive functional theory of text organization. RST offers an explanation of the coherence of texts by positing the existence of coherent relations among text spans. Most relations consist of one or more *nuclei* (the more central components of a rhetorical relation) and zero or more *satellites* (the supporting component of the relation). An example of a RST relation is "evidence", which is decomposed into a nucleus (a claim) and a satellite (text that supports the claim).

RST is intentionally limited to single documents as it is based on the notion of deliberate writing. In contrast, Cross-document Structure Theory (CST), the theory that we propose, will attempt to discover rhetorical structure in sets of related textual documents. Unlike RST, we cannot rely on the deliberateness of writing. We can, however, make use of some observations of structure across documents that, while clearly not deliberate in the RST sense, can be quite predictable and useful.

A recent representative work on RST and its applications is (Marcu 1997). He proposes a first-order formalization of the high-level, rhetorical structure of text, and provides theoretical analysis and empirical comparison of four algorithms for automatic derivation of text structures. A set of empirically motivated algorithms are designed for rhetorical parsing, i.e., determining the elementary textual units of a text, hypothesizing rhetorical relations that hold among these units, and eventually deriving the discourse structure of the text. Most relevant to our work, Marcu explores using RST for summarization. The basic idea of the discourse-based summarization algorithm is to induce a partial ordering on the importance of the units in a text based on the text structure, and output the most important units according to a certain threshold.

Another significant piece of work that inspired ours is (Salton *et al.* 1997). The authors generate intra-document semantic hyperlinks (between passages of a document which are related by a lexical similarity higher than a certain threshold) and characterize the structure of the text based on the intra-document linkage pattern. They represent each single document in the form of text relationship maps. A text summary is generated by selectively extracting important paragraphs from the text, more specifically, by automatically identifying the important paragraphs in a text relationship map and traversing the selected nodes in text order or along a certain path. The assumption underlying their technique is that highly "bushy" nodes are more likely to contain information central to the topic of the article.

All summarization techniques described above are limited to one single document. What we present in this paper is a more general framework for multi-document summarization using cross-document rhetorical relationships. We get better summaries by taking these relationships into account.

Cross-document Structure Theory

We propose Cross-document Structure Theory (CST), which enables multi-document summarization through identification of cross-document rhetorical relationships within a cluster of related documents. The proposed taxonomy for CST relationships can be found in Table 1. Notice that some CST relationships, such as *identity*, are symmetric (*multinuclear*, in RST terms), while some other ones, such as *subsumption*, do have directionality, i.e., they have *nucleus* and *satellite*. Some of the relationships are direct descendents of those used in (Radev & McKeown 1998). However, in CST, the relationships are domain-independent.

Whereas Marcu relied on "cue phrases" in implementing algorithms to discover the valid RST "trees" for a single document, such a technique is not very plausible for discovering CST "links" between documents. For instance, the "cue phrase" "*although statement X, statement Y*" might indicate the RST relationship "concession" in some circumstances. Marcu is able to use these phrases for guidance because of the conventions of writing and the valid assumption that authors tend to write documents using certain rhetorical techniques. However, in the case of multiple documents and CST inter-document relationships (links), we cannot expect to encounter a reliable analog to the cue phrase. This is because separate documents, even when they are related to a common topic, are generally not written with an overarching structure in mind. Particularly in the case of news, we are most often looking at articles which are written by different authors working from partially overlapping information as it becomes available. So, except in cases of explicit citation, we cannot expect to find a static phrase in one document which reliably indicates a particular relationship to some phrase in another document.

How, then, to approach the problem of discovering CST relationships in a set of documents? We present in a later section an exploratory experiment, in which human subjects were asked to find these relationships over a multi-document news cluster.

Formalization of the problem
Extractive summarizer

Suppose we have a document cluster C, which contains documents d_1 through d_l; each document d_i entails a list of sentences s_1 through s_{n_i}. The set of all sentences in the cluster is defined as S.

An extractive summarizer E produces an extract S', such that $S' \subset S$. Technically, an extract is simply a condensed representation of a summary, i.e., there is a one-to-one map-

Relationship	Description	Text span 1 (S1)	Text span 2 (S2)
Identity	The same text appears in more than one location	Tony Blair was elected for a second term today.	Tony Blair was elected for a second term today.
Equivalence (Paraphrase)	Two text spans have the same information content	Derek Bell is experiencing a resurgence in his career.	Derek Bell is having a "comeback year."
Translation	Same information content in different languages	Shouts of "Viva la revolucion!" echoed through the night.	The rebels could be heard shouting, "Long live the revolution".
Subsumption	S1 contains all information in S2, plus additional information not in S2	With 3 wins this year, Green Bay has the best record in the NFL.	Green Bay has 3 wins this year.
Contradiction	Conflicting information	There were 122 people on the downed plane.	126 people were aboard the plane.
Historical Background	S1 gives historical context to information in S2	This was the fourth time a member of the Royal Family has gotten divorced.	The Duke of Windsor was divorced from the Duchess of Windsor yesterday.
Citation	S2 explicitly cites document S1	Prince Albert then went on to say, "I never gamble."	An earlier article quoted Prince Albert as saying "I never gamble."
Modality	S1 presents a qualified version of the information in S2, e.g., using "allegedly"	Sean "Puffy" Combs is reported to own several multimillion dollar estates.	Puffy owns four multimillion dollar homes in the New York area.
Attribution	S1 presents an attributed version of information in S2, e.g. using "According to CNN,"	According to a top Bush advisor, the President was alarmed at the news.	The President was alarmed to hear of his daughter's low grades.
Summary	S1 summarizes S2.	The Mets won the Title in seven games.	After a grueling first six games, the Mets came from behind tonight to take the Title.
Follow-up	S1 presents additional information which has happened since S2	102 casualties have been reported in the earthquake region.	So far, no casualties from the quake have been confirmed.
Indirect speech	S1 indirectly quotes something which was directly quoted in S2	Mr. Cuban then gave the crowd his personal guarantee of free Chalupas.	"I'll personally guarantee free Chalupas," Mr. Cuban announced to the crowd.
Elaboration / Refinement	S1 elaborates or provides details of some information given more generally in S2	50% of students are under 25; 20% are between 26 and 30; the rest are over 30.	Most students at the University are under 30.
Fulfillment	S1 asserts the occurrence of an event predicted in S2	After traveling to Austria Thursday, Mr. Green returned home to New York.	Mr. Green will go to Austria Thursday.
Description	S1 describes an entity mentioned in S2	Greenfield, a retired general and father of two, has declined to comment.	Mr. Greenfield appeared in court yesterday.
Reader Profile	S1 and S2 provide similar information written for a different audience.	The Durian, a fruit used in Asian cuisine, has a strong smell.	The dish is usually made with Durian.
Change of perspective	The same entity presents a differing opinion or presents a fact in a different light.	Giuliani criticized the Officer's Union as "too demanding" in contract talks.	Giuliani praised the Officer's Union, which provides legal aid and advice to members.

Table 1: CST relationships and examples

ping between extracts and summaries. We will not distinguish these two terms from now on.

The summarizer E can be characterized by the following components:

1. A scoring algorithm A_S that computes a numeric score, which is a function of a number of features, for each sentence. Specifically, $score(s_i) = A_S(f_{1i}, f_{2i}, \ldots f_{ki})$, where f_1 through f_k are the features of each sentence.

2. A re-ranker R that adjusts sentence scores by looking at some other (usually global) information, such as lexical similarity or CST relationships between pair of sentences. Specifically, $score(s_i) = score(s_i) + \Delta(S)$, where the adjustment is determined is by certain global information with regard to S. Notice that Δ can be negative.

3. A compression ratio r, such that $0 \leq r \leq 1$.

4. A ranking algorithm A_R that selects the highest-score sentences, such that $N_{S'} = \lceil N_S \cdot r \rceil$ where N_S is the number of sentences in the original text and $N_{S'}$ is the number of sentences in the extract.

CST connectivity

For any extract S', we can define a connectivity matrix M, the elements of which are defined as:

$$m_{ij} = \begin{cases} 1 & : & \text{connectivity condition holds} \\ 0 & : & \text{otherwise} \end{cases}$$

The CST connectivity of the extract S' is defined as

$$\chi = \sum_{i=1}^{N_{S'}} \sum_{j=1}^{N_{S'}} m_{ij}$$

.

Depending on our purposes, we could define different connectivity conditions for the elements in the connectivity matrix to be equal to 1. In our study, we define the condition as existence of a certain relationship with agreement strength higher than a certain threshold.

CST identification

In this section, we present an experiment in which subjects were asked to analyze a set of documents using the set of proposed relationships in Table 1. We then present the experimental results and consider the implications for further work in CST.

Experiment 1: establishing CST relationships

The experiment which we conducted required subjects to read a set of news articles and write down the inter-document relationships which they observed. Specifically, the 11 articles were on the subject of an airplane crash of a flight from Egypt to Bahrain in August 2000. They were written by several different news organizations and retrieved from online news web sites in the days following the accident.

The subjects were eight graduate students and one professor. The instructions specified five article pairs comprised of random pairings from within the eleven articles mentioned

above. No article was included in more than two pairs. For each pair, the subjects were instructed to first read the articles carefully. They were then instructed to look for and note down any occurrences of relationships like those in Figure 1. (Subjects were also provided with the examples shown in Figure 1 to illustrate each relationship type.) It was stated in the instructions that the relationships comprised only a "proposed" list, and were not to be considered exhaustive. Subjects were invited to make up new relationship types if they observed cross-document relationships which did not correspond to those in Table 1.

Although subjects were given examples of the proposed relationships at the sentence level, the instructions also explicitly stated that it was possible for a relationship to hold with one or both text spans being more than one sentence long. There was no provision for subjects to mark text spans shorter than a full sentence. Subjects were instructed not to pay attention to possible *intra*-document rhetorical relationships. Also, subjects were instructed that it was possible for more than one relationship to exist across the same pair of text spans, and to note down as many relationships as they observed for each pair of text spans.

Results

A summary of the raw results of the experiment is shown in Table 2. The relationships are presented in descending order of observed frequency. On average, each subject identified approximately 45 occurrences of the proposed relationships. The relationships "Elaboration/Refinement," "Equivalence," and "Description" were identified most frequently. Other relationships, such as "Translation," "Citation," and "Summary," were observed either never or only by one subject. Although subjects were encouraged in the study instructions to name new relationships, none did so.

Relationship Type	Sum	Average
Elaboration / Refinement	85	9.44
Equivalence	70	7.78
Description	64	7.11
Historical Background	44	4.89
Follow-up	42	4.67
Subsumption	39	4.22
Contradiction	31	3.22
Attribution	15	1.67
Identity	7	0.78
Indirect speech	6	0.67
Fulfillment	4	0.44
Modality	2	0.22
Summary	1	0.11
Reader Profile	1	0.11
Change of Perspective	1	0.11
Translation	0	0.00
Citation	0	0.00
Total	**415**	**45.44**

Table 2: Identifications of CST relationships by type

Table 3 describes the sentence pairs for which judges noted relationships. The total number of sentence pairs for all five article pairs assigned was 4579, which is $\sum_{i=1}^{5} n_i \times m_i$, where i is the number of the article pair, n is

the number of sentences in the first article in the pair, and m is the number of sentences in the second article in the pair.

Judges Finding a Relationship	Number of Sentence Pairs
No Judges	4,291
One Judge	200
Multiple Judges	88

Table 3: Sentence pairs by number of judges marking a CST relationship

As can be seen in Table 3, there are 88 sentence pairs for which multiple judges identify at least one CST relationship. Table 4 describes the breakdown of these 88 pairs in terms of inter-judge agreement. Although subjects were permitted to mark more than one relationship per sentence pair, they are counted as "in agreement" here if at least one of the relations they marked agrees with one of the relations marked by another judge.

Discrete Relationship Types Observed	Judges in Agreement	Sentences
Only one	All	16
More than one	At least two	35
More than one	None	37

Table 4: Judge agreement on relationship types among sentence pairs marked by multiple judges

Observations

Because our data comes from observations about a subset of document pairs from a single news cluster, it would clearly be premature to make conclusions about the natural frequencies of these relationships based on the data in Table 2. Nonetheless, we can at least speculate that human subjects are capable of identifying some subset of these relationships when reading news articles.

We obviously need more data before we can say if the lack of identifications for those unobserved / rarely observed relationships is because of a true lack of frequency or some other factor. For instance, some of the proposed relationship names, like "modality," may not be intuitive enough for judges to feel comfortable identifying them, even when examples are given.

However, the most encouraging data concerns the relatively high level of overlap when multiple judges made an observation for a sentence. In 51 of 88 cases where more than one judge marked a sentence pair, at least two judges concurred about at least one relationship holding for the pair. Although approximately two-thirds of the marked pairs were marked by only one judge, the overall data sparseness makes this ratio less discouraging. Apparently, before we can attempt to build automated means of detecting CST links, we must have a better understanding of what (if any) empirical properties reliably indicate CST relationships.

Another key step is to gather further data. In order to do so, an automated markup tool in the style of Alembic Workbench (Day *et al.* 1997) or SEE (Lin 2001) would be

extremely helpful. Not only is there a great deal of transcription (and associated possibilities for error) involved in running this experiment on paper, but a number of subjects expressed the belief that an automated tool would allow them to provide better and more consistent data.

CST-enhanced summarization
Motivation

In our experiments with CST-enhanced summarization, we try to explore the effects of different CST relationships on the quality of output summary. A motivating example is illustrated in Figure 2. As the first step, a 10-sentence extract (denoted by the bold circles) is generated by MEAD from a 3-article cluster. The original CST connectivity is 1, as is denoted by the relationship arc between sentence 44 in article 81 and sentence 10 in article 87. Now we try to increase the CST connectivity of the summary to 2 by including sentence 46 in article 81 (which has a CST relationship with sentence 26 in article 41) and drop sentence 22 in article 87 (the lowest ranked sentence in the original extract). Based on linguistic intuition, we postulate the following hypotheses:

Hypothesis 1 Enhancing the CST connectivity of a summary will significantly affect its quality as measured by relative utility.

Hypothesis 2 The effect of the enhancement will be dependent on the type of CST relationship added into the summary.
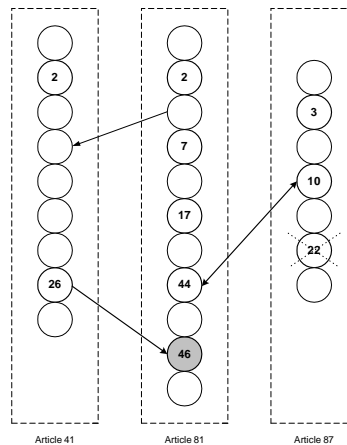

Figure 2: Graph representation of a sample cluster

Experiment 2: utility-based evaluation

The setup We produce summaries using MEAD and CST-enhanced version of it and compare the quality of the summaries. A significant difference in terms of summary quality will indicate the effect of CST enhancement.

First we need to measure the quality of summaries. We don't want to have human judges look at a number of summaries and determine how good they are, which is too tedious and too expensive. Therefore we decided to resort

to sentence utility judgement. More specifically, we had 9 judges read the 11 articles in the same cluster used in the CST identification user study and assign utility scores (0 to 10) to each sentence. We then average the utility score for each sentence respectively and use the total utility of a summary as a proxy of its quality measure. This way we can control the amount of the work human judges have to do and reuse the utility judgement through a large number of experiments.

To run the experiments, we used the re-ranker module in MEAD to adjust the content of a summary. In the default version of MEAD, the re-ranker only looks at the lexical similarity between the sentence under consideration and the sentences already in the summary and drop the current sentence if it is too similar to those already in the summary. To enhance the summary using CST relationship, we ended up implementing a new re-ranker which includes new sentences into the summary according to the notion of increasing CST connectivity and dropping the lowest-ranked sentences correspondingly.

The evaluation metrics we use are the same as those used in (Radev, Jing, & Budzikowska 2000). All the summary performance numbers calculated in the experiments are relative utility scores.

The algorithm   The algorithm is described in Algorithm 1. With this rather naive algorithm, we expect to show the feasibility of CST-enhancement, instead of claiming that this is "the" best way to do it.

In the actual experiments, we used the following parameters:

- Total number of clusters: 10

- Compression ratio: 10% through 30%,

- Total number of CST relationships: 17

- Threshold to hypothesize a CST relationship: 1

Experiments, results and analysis   We also consider the distinction between incoming and outgoing relationships, in the sense of the directionality of the added relationship according to user judgement (e.g., in Figure 2, incorporating sentence 46 in article 81 is enhancement by outgoing relationship). The intuition behind this is obviously that some CST relationships do have directionality and the directionality could potentially make some difference on the resultant summary.

First we look at the overall effect of CST enhancement by comparing the average utility of all CST-enhanced summaries and the utility of baseline summaries (Table 5). As we can see, overall, CST enhancement significantly improves the utility of resultant summary in both incoming and outgoing scenarios. This justifies our first hypothesis.

	p-value	Sign
Incoming	0.008	+
Outgoing	0.026	+

Table 5: Average effect of CST enhancement (The p-values are for two-tailed pairwise T-tests), *average* case.

Algorithm 1 Experiment algorithm for CST-enhanced summarization

for all clusters c **do**
   for all compression ratios r **do**
      Produce baseline summary S for cluster c at compression ratio r using MEAD
      Compute relative utility measure for S
      for all possible CST relationship R **do**
         S_E=CST_Enhance(S,R,c)
         Compute relative utility measure for S_E
      end for
   end for
end for

Function CST_Enhance(S,R,c)
$S_E = S$
Initialize list L to null
for all sentences s in c but not in S **do**
   if s has CST relationship R with any sentence in S **then**
      Add s into L
   end if
end for
if L not null **then**
   if $NumberOfElements(L) > 1$ **then**
      Randomly choose a sentence s from L
   else
      Use the only sentence s in L
   end if
   s_0 = lowest ranked sentence in S_E with no CST relationship to other sentences
   Add sentence s into summary S_E
   Drop s_0 from S_E
end if
Return S_E

Does the compression ratio matter? Although we didn't experiment with long summaries ($r > 0.3$) since we believe those wouldn't make much practical sense, we still expect to observe some preliminary patterns in the data. Table 6 gives the comparison of pre-enhancement and post-enhancement summary performance by compression ratio. There is a tendency that, although not strong enough to be conclusive, CST-enhancement is relatively more helpful for shorter summaries.

		10%	20%	30%
Incoming	pre-enhancement	0.242	0.287	0.322
	post-enhancement	0.246	0.293	0.325
	difference	0.004	0.006	0.003
Outgoing	pre-enhancement	0.242	0.287	0.322
	post-enhancement	0.247	0.292	0.323
	difference	0.005	0.005	0.001

Table 6: Effects of CST enhancement at different compression ratios

To show how we tested our second hypothesis, we present the results by different CST relationship types in Table 7. The observations are the following:

- It appears that some CST relationships have no effect on the quality of enhanced summaries. This is not necessarily true, though, because all the unchanged utility num-

bers are due to the sparseness of user judgement data, in other words, the enhancement algorithm couldn't find any occurrence of these relationships to enhance the baseline summary.

- Most relationships that do affect the enhanced summary are "positive" ones, in the sense that incorporating them into the baseline summary significantly increases relative utility.

- Two CST relationships, *historical background* and *description*, stand out as potential "negative" ones, in that incorporaing them reduces the utility of post-enhancement summary to some extent.

CST relationship	Incoming		Outgoing	
	p-value	Sign	p-value	Sign
Identity	1.000	N	1.000	N
Equivalence	0.007	+	0.032	+
Translation	1.000	N	1.000	N
Subsumption	0.011	+	0.006	+
Contradiction	0.044	+	0.002	+
Historical	0.131	-	0.491	-
Citation	1.000	N	1.000	N
Modality	1.000	N	1.000	N
Attribution	0.049	+	0.255	+
Summary	1.000	N	1.000	N
Follow-up	0.002	+	0.055	+
Indirect speech	0.057	+	0.172	+
Elaboration	0.004	+	0.006	+
Fulfillment	1.000	N	1.000	N
Description	0.008	-	0.102	-
Reader Profile	1.000	N	1.000	N
Change of Perspective	1.000	N	1.000	N

Table 7: Effects of different CST relationships (The p-values are for two-tailed pairwise T-tests), *breakdown-by-type* case.

In both the "average" case and "break-down-by-type" case, the importance of directionality doesn't emerge. As can be again explained by the sparseness of user judgement data, this is disappointing but not surprising.

Conclusions and future work

In this paper, we showed that taking CST relationships into account affects the quality of extractive summaries. Moreover, enhancement by adding different types of CST relationships has different effects on resulting summaries.

In the long run, we foresee the following directions in future work:

- The proposed CST relationships (as shown in Table 1) need more refinement. A reasonably standardized taxonomy should be in place and act as ground for future research along this line.

- Currently, all CST relationships are based on human judgement. To automatically enhance summaries in the light of CST, we need to be able to automatically parse CST relationships. The first step is to build a CST-annotated corpus.

- It seems both theoretically and empirically interesting to find more about the connection between RST and CST.

Maybe first identifying the intra-document RST relationships can help CST parsing?

- Having showed that a relatively naive CST-enhancement algorithm does help in improving the quality of extractive summarization, we need to find a more intelligent version of it and to incorporate it into the MEAD summarizer.

- At this moment, either sentence utility scores by user judgement or sentence ranking scores by MEAD are really "inherent utility", in the sense that $score(s_i)$ doesn't depend on $score(s_j)$. It might be interesting and useful, however, to pursue the notion of "conditional utility" of sentences (thus acknowledging that the utility of one sentence depends on the utilities of other sentences), which could potentially influence the output of the summarizer.

References

Day, D.; Aberdeen, J.; Hirschman, L.; Kozierok, R.; Robinson, P.; and Vilain, M. 1997. Mixed-initiative development of language processing systems.

Lin, C.-Y. 2001. See - summary evaluation environment. WWW site, URL: http://www.isi.edu/ cyl/SEE/.

Luhn, H. 1958. The Automatic Creation of Literature Abstracts. *IBM Journal of Research Development* 2(2):159–165.

Mani, I., and Maybury, M., eds. 1999. *Advances in automatic text summarization*. Cambridge, MA: MIT Press.

Mann, W. C., and Thompson, S. A. 1988. Rhetorical Structure Theory: towards a functional theory of text organization. *Text* 8(3):243–281.

Marcu, D. 1997. *The Rhetorical Parsing, Summarization, and Generation of Natural Language Texts*. Ph.D. Dissertation, Department of Computer Science, University of Toronto.

MEAD. 2002. Mead documentation. WWW site, URL: http://www.clsp.jhu.edu/ws2001/groups/asmd/meaddoc.ps.

Radev, D. R., and McKeown, K. R. 1998. Generating natural language summaries from multiple on-line sources. *Computational Linguistics* 24(3):469–500.

Radev, D. R.; Jing, H.; and Budzikowska, M. 2000. Centroid-based summarization of multiple documents: sentence extraction, utility-based evaluation, and user studies. In *ANLP/NAACL Workshop on Summarization*.

Radev, D. 2000. A common theory of information fusion from multiple text sources, step one: Cross-document structure. In *Proceedings, 1st ACL SIGDIAL Workshop on Discourse and Dialogue*.

Salton, G.; Singhal, A.; Mitra, M.; and Buckley, C. 1997. Automatic text structuring and summarization. *Information Processing & Management* 33:193–207.