Syntactic Simplification for Improving Content Selection in Multi-Document Summarization

Advaith Siddharthan, **Ani Nenkova** and **Kathleen McKeown**
Columbia University
Computer Science Department
{advaith|ani|kathy}@cs.columbia.edu

Abstract

In this paper, we explore the use of automatic syntactic simplification for improving content selection in multi-document summarization. In particular, we show how simplifying parentheticals by removing relative clauses and appositives results in improved sentence clustering, by forcing clustering based on central rather than background information. We argue that the inclusion of parenthetical information in a summary is a reference-generation task rather than a content-selection one, and implement a baseline reference rewriting module. We perform our evaluations on the test sets from the 2003 and 2004 Document Understanding Conference and report that simplifying parentheticals results in significant improvement on the automated evaluation metric *Rouge*.

1 Introduction

Syntactic simplification is an NLP task, the goal of which is to rewrite sentences to reduce their grammatical complexity while preserving their meaning and information content. Text simplification is a useful task for varied reasons. Chandrasekar et al. (1996) viewed text simplification as a preprocessing tool to improve the performance of their parser. The PSET project (Carroll et al., 1999), on the other hand, focused its research on simplifying newspaper text for aphasics, who have trouble with long sentences and complicated grammatical constructs. We have previously (Siddharthan, 2002; Siddharthan, 2003) developed a shallow and robust syntactic simplification system for news reports, that simplifies relative clauses, apposition and conjunction. In this paper, we explore the use of syntactic simplification in multi-document summarization.

1.1 Sentence Shortening for Summarization

It is interesting to survey the literature in *sentence shortening*, a task related to syntactic simplification. Grefenstette (1998) proposed the use of sentence shortening to generate telegraphic texts that would help a blind reader (with a text-to-speech software) skim a page in a manner similar to sighted readers. He provided eight levels of telegraphic reduction. The first (the most drastic) generated a stream of all the proper nouns in the text. The second generated all nouns in subject or object position. The third, in addition, included the head verbs. The least drastic reduction generated all subjects, head verbs, objects, subclauses and prepositions and dependent noun heads. Reproducing from an example in his paper, the sentence:

> Former Democratic National Committee finance director Richard Sullivan faced more pointed questioning from Republicans during his second day on the witness stand in the Senate's fund-raising investigation.

got shortened (with different levels of reduction) to:

- Richard Sullivan Republicans Senate.
- Richard Sullivan faced pointed questioning.
- Richard Sullivan faced pointed questioning from Republicans during day on stand in Senate fund-raising investigation.

Grefenstette (1998) provided a rule based approach to telegraphic reduction of the kind illustrated above. Since then, Jing (2000), Riezler et al. (2003) and Knight and Marcu (2000) have explored statistical models for sentence shortening that, in addition, aim at ensuring grammaticality of the shortened sentences.

These sentence-shortening approaches have been evaluated by comparison with human-shortened sentences and have been shown to compare favorably. However, the use of sentence shortening for the multi-document summarization task has been largely unexplored, even though intuitively it appears that sentence-shortening can allow more important information to be included in a summary. Recently, Lin (2003) showed that statistical sentence-shortening approaches like Knight and Marcu (2000) do *not* improve content selection in summaries. Indeed he reported that syntax-based sentence-shortening resulted in significantly worse content selection by their extractive summarizer NeATS. Lin (2003) concluded that pure syntax-based compression does not improve overall summarizer performance, even though the compression algorithm performs well at the sentence level.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2004		**00-00-2004 to 00-00-2004**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Syntactic Simplification for Improving Content Selection in Multi-Document Summarization	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Columbia University,Department of Computer Science,2960 Broadway,New York,NY,10027-6902**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**7**	

1.2 Simplifying Syntax for Summarization

A problem with using statistical sentence-shortening for summarization is that syntactic form does not always correlate with the importance of the information contained within. As a result, syntactic sentence shortening might get rid of important information that should be included in the summary. In contrast, the syntactic simplification literature deals with syntactic constructs that can be interpreted from a rhetorical perspective. In particular, appositives and non-restrictive relative clauses are considered *parentheticals* in RST (Mann and Thompson, 1988). Their role is to provide background information on entities, and to relate the entity to the discourse. Along with restrictive relative clauses, their inclusion in a summary should ideally be determined by a reference generating module, not a content selector. It is thus more likely that the removal of appositives and relative clauses will impact content-selection than the removal of adjectives and prepositional phrases, as attempted by sentence shortening. It is precisely this hypothesis that we explore in this paper.

1.3 Outline

We describe our sentence-clustering based summarizer in the next section, including our experiments on using simplification of parentheticals to improve clustering in §2.1. We evaluate our summarizer in §3 and then describe our reference regenerator in §4. We present a discussion of our approach in §5 and conclude in §6.

2 The Summarizer

We use a sentence-clustering approach to multi-document summarization (similar to *multigen* (Barzilay, 2003)), where sentences in the input documents are clustered according to their similarity. Larger clusters represent information that is repeated more often across input documents; hence the size of a cluster is indicative of the importance of that information. For our current implementation, a representative (simplified) sentence is selected from each cluster and these are incorporated into the summary in the order of decreasing cluster size.

A problem with this approach is that the clustering is not always accurate. Clusters can contain spurious sentences, and a cluster's size might then exaggerate its importance. Improving the quality of the clustering can thus be expected to improve the content of the summary. We now describe our experiments on syntactic simplification and sentence clustering. Our hypothesis is that simplifying parenthetical units (relative clauses and appositives) will improve the performance of our clustering algorithm, by preventing it from clustering on the basis of background information.

2.1 Simplification and Clustering

We use *SimFinder* (Hatzivassiloglou et al., 1999) for sentence clustering and its similarity metric to evaluate cluster quality; *SimFinder* outputs similarity values (*simvals*) between 0 and 1 for pairs of sentences, based on word overlap, synonymy and n-gram matches. We use the average of the simvals for each pair of sentences in a cluster to evaluate a quality-score for the cluster. Table 1 below shows the quality-scores averaged over all clusters when the original document set is and is not preprocessed using our syntactic simplification software (described in §2.2). We use 30 document sets from the 2003 Document Understanding Conference (see §3.1 for description). For each of the experiments in table 1, *SimFinder* produced around 1500 clusters, with an average cluster size beween 3.6 and 3.8.

	Orig	Simp-Paren	Simp-Conj
Av. quality-score	0.687	0.722	0.686
Std. deviation (σ)	0.130	0.112	0.126

Table 1: Syntactic Simplification and Clustering

Table 1 shows that removing parentheticals results in a 5% relative improvement in clustering. This improvement is significant at confidence $t = 95\%$ as determined by the *difference in proportions test* (Snedecor and Cochran, 1989). Further, the standard deviation for the performance of the clustering decreases by around 2%. This suggests that removing parentheticals results in better and more robust clustering. As an example of how clustering improves, our simplification routine simplifies:

> PAL, which has been unable to make payments on dlrs 2.1 billion in debt, was devastated by a pilots' strike in June and by the region's currency crisis, which reduced passenger numbers and inflated costs.

to:

> PAL was devastated by a pilots' strike in June and by the region's currency crisis.

Three other sentences also simplify to the extent that they represent PAL being hit by the June strike. The resulting cluster (with quality score=0.94) is:

1. PAL was devastated by a pilots' strike in June and by the region's currency crisis.
2. In June, PAL was embroiled in a crippling three-week pilots' strike.
3. Tan wants to retain the 200 pilots because they stood by him when the majority of PAL's pilots staged a devastating strike in June.

4. In June, PAL was embroiled in a crippling three-week pilots' strike.

On the other hand, splitting conjoined clauses does not appear to aid clustering[1]. This indicates that the improvement from removing parentheticals is not because shorter sentences might cluster better (as *SimFinder* controls for sentence length, this is anyway unlikely). For confirmation, we performed one more experiment—we deleted words at random, so that the average sentence length for the modified input documents was the same as for the inputs with parentheticals removed. This actually made the clustering worse (av. quality score of 0.637), confirming that the improvement from removing parentheticals was not due to reduced sentence length. These results demonstrate that the parenthetical nature of relative clauses and appositives makes their removal useful.

Improved clustering, however, need not necessarily translate to improved content selection in summaries. We therefore also need to evaluate our summarizer. We do this in §3, but first we describe the summarizer in more detail.

2.2 Description of our Summarizer

Our summarizer has four stages—preprocessing of original documents to remove parentheticals, clustering of the simplified sentences, selecting of one representative sentence from each cluster and deciding which of these selected sentences to incorporate in the summary.

We use our syntactic simplification software (Siddharthan, 2002; Siddharthan, 2003) to remove parentheticals. It uses the LT TTT (Grover et al., 2000) for POS-tagging and simple noun-chunking. It then performs apposition and relative clause identification and attachment using shallow techniques based on local context and animacy information obtained from WordNet (Miller et al., 1993).

We then cluster the simplified sentences with *SimFinder* (Hatzivassiloglou et al., 1999). To further tighten the clusters and ensure that their size is representative of their importance, we post-process them as follows. *SimFinder* implements an incremental approach to clustering. At each incremental step, the similarity of a new sentence to an existing cluster is computed. If this is higher than a threshold, the sentence is added to the cluster. There is no backtracking; once a sentence is added to a cluster, it cannot be removed, even if it is dissimilar to all the sentences added to the cluster in the future. Hence, there are often one or two sentences that have low similarity with the final cluster. We remove these with a post-process that can be considered equivalent to a back-tracking step. We redefine the criteria for a sentence to be part of the final cluster such that it has to be similar (simval above the threshold) to *all* other sentences in the final cluster. We prune the cluster to remove sentences that do not satisfy this criterion. Consider the following cluster and a threshold of 0.65. Each line consists of two sentence ids (*P[sent_id]*) and their simval.

P37	P69	0.9999999999964279
P37	P160	0.8120098824183786
P37	P161	0.8910485867563762
P37	P176	0.8971370325713883
P69	P160	0.8120098824183786
P69	P161	0.8910485867563762
P69	P176	0.8971370325713883
P160	P161	**0.2333051325617611**
P160	P176	**0.0447901658343020**
P161	P176	0.7517636285580539

We mark all the lines with similarity values below the threshold (in bold font). We then remove as few sentences as possible such that these lines are excluded. In this example, it is sufficient to remove $P160$. The final cluster is then:

P37	P69	0.9999999999964279
P37	P161	0.8910485867563762
P37	P176	0.8971370325713883
P69	P161	0.8910485867563762
P69	P176	0.8971370325713883
P161	P176	0.7517636285580539

The result is a much tighter cluster with one sentence less than the original. This pruning operation leads to even higher similarity scores than those presented in table 1.

Having pruned the clusters, we select a representative sentence from each cluster based on *tf*idf*. We then incorporate these representative sentences into the summary in decreasing order of their cluster size. For clusters with the same size, we incorporate sentences in decreasing order of *tf*idf*. Unlike *multigen* (Barzilay, 2003), which is generative and constructs a sentence from each cluster using information fusion, we implement extractive summarization and select one (simplified) sentence from each cluster. We discuss the scope for generation in our summarizer in §4 and §6.

3 Evaluation

We present two evaluations in this section. Our system, as described in the previous section, was entered for the DUC'04 competition. We describe how it fared in §3.3. We also present an evaluation over a larger data set to show that syntactic simplification of parenthetical units significantly improves

[1] In this example, splitting subordination helps as sentence 3 yields *the majority of PAL's pilots staged a devastating strike in June.* However, averaged over the entire DUC'03 data set, there is no net improvement from splitting conjunction.

content selection (§3.4). But first, we describe our data (§3.1) and the evaluation metric *Rouge* (§3.2).

3.1 Data

The Document Understanding Conference (DUC) has been run annually since 2001 and is the biggest summarization evaluation effort, with participants from all over the world. In 2003, DUC put special emphasis on the development of automatic evaluation methods and also started providing participants with multiple human-written models needed for reliable evaluation. Participating generic multi-document summarizers were tested on 30 event-based sets in 2003 and 50 sets in 2004, all 80 containing roughly 10 newswire articles each. There were four human-written summaries for each set, created for evaluation purposes. In DUC'03, the task was to generate 100 word summaries, while in DUC'04, the limit was changed to 665 bytes.

3.2 Evaluation Metric

We evaluated our summarizer on the DUC test sets using the *Rouge* automatic scoring metric (Lin and Hovy, 2003). The experiments in Lin and Hovy (2003) show that among n-gram approaches to scoring, *Rouge-1* (based on unigrams) has the highest correlation with human scores. In 2004, an additional automatic metric based on longest common subsequence was included (*Rouge-L*), that aims to overcome some deficiencies of *Rouge-1*, such as its susceptibility to ungrammatical keyword packing by dishonest summarizers[2]. For our evaluations, we use the *Rouge* settings from DUC'04: stop words are included, words are Porter-stemmed, and all four human model summaries are used.

3.3 DUC'04 Evaluation

We entered our system as described above for the DUC'04 competition. There were 35 entries for the generic summary task, including ours. At 95% confidence levels, our system was significantly superior to 23 systems and indistinguishable from the other 11 (using *Rouge-L*). Using *Rouge-1*, there was one system that was significantly superior to ours, 10 that were indistinguishable and 23 that were significantly inferior. We give a few *Rouge* scores from DUC'04 in figure 2 below for comparison purposes. The 95% confidence intervals for our summarizer are +-0.0123 (*Rouge-1*) and +-0.0130 (*Rouge-L*).

3.4 Benefits from Syntactic Simplification

Table 3 below shows the *Rouge-1* and *Rouge-L* scores for our summarizer when the text is and is not simplified to remove parentheticals. The data for this evaluation consists of the 80 document sets from DUC'03 and DUC'04. We did not use data from previous years as these included only one human model-summary and *Rouge* requires multiple models to be reliable.

Summarizer	Rouge-1	Rouge-L
Our Summarizer	0.3672	0.3804
Best Summarizer	0.3822	0.3895
Median Summarizer	0.3429	0.3538
Worst Summarizer	0.2419	0.2763
Av. of Human Summarizers	0.4030	0.4202

Table 2: Rouge Scores for DUC'04 competition.

Summarizer	Rouge-1	Rouge-L
With simplification	0.3608	0.3839
Without simplification	0.3398	0.3643

Table 3: Rouge Scores for DUC'03 and '04 data.

The improvement in performance when the text is preprocessed to remove parenthetical units is significant at 95% confidence limits. When compared to the 34 other participants of DUC'04, the simplification step raises our clustering-based summarizer from languishing in the bottom half to being in the top third and statistically indistinguishable from the top system at 95% confidence (using *Rouge-L*).

4 Reference Regeneration

As the evaluations above show, preprocessing text with syntactic simplification significantly improves content selection for our summarizer. This is encouraging; however, our summarizer, as describe so far, generates summaries that contain no parentheticals (appositives or relative clauses), as these are removed from the original texts prior to summarization. We believe that the inclusion of parenthetical information about entities should be treated as a reference generation task, rather than a content selection one. Our analysis of human summaries suggests that people select parentheticals to improve coherence and to aid the hearer in identifying referents and relating them to the discourse. A complete treatment of parentheticals in reference regeneration in summaries is beyond the scope of this paper, the emphasis of which is content-selection, rather than coherence. We plan to address this issue elsewhere; in this paper, we restrict ourselves to describing a baseline approach to incorporating parentheticals in regenerated references to people in summaries.

4.1 Including Parentheticals

Our text-simplification system (Siddharthan, 2003) provides us with with a list of all relative clauses, appositives and pronouns that attach to/co-refer

[2]More detail on the Rouge evaluation metrics can be obtained online from http://www.isi.edu/∼cyl/papers/ROUGE-Working-Note-v1.3.1.pdf

with every entity. We used a named entity tagger (Wacholder et al., 1997) to collect all such information for every person. The processed references to the same people across documents were aligned using the named entity tagger canonic name, resulting in tables similar to those shown in figure 1.

Abdullah Ocalan

APW19981106.1119: [IR] Abdullah Ocalan; [AP] leader of the outlawed Kurdistan Worker 's Party; [CO] Ocalan;

APW19981104.0265: [IR] Kurdish rebel leader Abdullah Ocalan; [RC] who is wanted in Turkey on charges of heading a terrorist organization; [CO] Ocalan; [RC] who leads the banned Kurdish Workers Party , or PKK , which has been fighting for Kurdish autonomy in Turkey since 1984; [CO] Ocalan; [CO] Ocalan; [CO] Ocalan;

APW19981113.0541: [IR] Abdullah Ocalan; [AP] leader of Kurdish insurgents; [RC] who has been sought for years by Turkey; [CO] Ocalan; [CO] Ocalan; [CO] Ocalan; [PR] He; [CO] Ocalan; [CO] Ocalan; [PR] his; [CO] Ocalan; [CO] Ocalan; [CO] Ocalan; [PR] his; [CO] Ocalan; [CO] Ocalan; [AP] a political science dropout from Ankara university in 1978;

APW19981021.0554: [IR] rebel leader Abdullah Ocalan; [PR] he; [CO] Ocalan;

Figure 1: Example information collected for entities in the input. The canonic form of the named entity is shown in bold and the input article id in italic. IR stands for "initial reference", CO for subsequent noun co-reference, PR for pronoun reference, AP for apposition and RC for relative clause.

We automatically post-edited our summaries using a modified version of the module described in Nenkova and McKeown (2003). This module normalizes references to people in the summary, by introducing them in detail when they are first mentioned and using a short reference for subsequent mentions; these operations were shown to improve the readability of the resulting summaries.

Nenkova and McKeown (2003) avoided including parentheticals due to both the unavailability of fast and reliable identification and attachment of appositives and relative clauses, and theoretical issues relating to the selection of the most suitable parenthetical unit in the new summary context. In order to ensure a balanced inclusion of parenthetical information in our summaries, we modified their initial approach to allow for including relative clauses and appositives in initial references.

We made use of two empirical observations made by Nenkova and McKeown (2003) based on human summaries: a first mention is very likely to be modified in some way (probability of 0.76), and subsequent mentions are very unlikely to be post-modified (probability of 0.01–0.04). We therefore only considered incorporating parentheticals in first mentions. We constructed a set consisting of appositives and relative clauses from initial references in the input documents and an empty string option (for the example in figure 1, the set would be $\{$ "leader of the outlawed Kurdistan Worker's Party", "who is wanted in Turkey on charges of heading a terrorist organization",“ leader of Kurdish insurgents", "who has been sought for years by Turkey", $\epsilon\}$). We then selected one member of the set randomly for inclusion in the initial reference. A more sophisticated approach to the treatment of parentheticals in reference regeneration, based on lexical cohesion constraints, is currently underway.

4.2 Evaluation

We repeated the evaluations on the 80 document sets from DUC'03 and DUC'04, using our simplification+clustering based summarizer with the reference regeneration component included. The results are shown in the table below. At 95% confidence, the difference in performance is not significant.

Summarizer	Rouge-1	Rouge-L
Without reference rewrite	0.3608	0.3839
With reference rewrite	0.3599	0.3854

Table 4: Rouge scores for DUC'03 and '04 data.

This is an interesting result because it suggests that rewriting references does not adversely affect content selection. This might be because the extra words added to initial references are partly compensated for by words removed from subsequent references. In any case, the reference rewriting can significantly improve readability, as shown in the examples in figures 2 and 3. We are also optimistic that a more focused reference rewriting process based on lexical-cohesive constraints and information-theoretic measures can improve *Rouge* content-evaluation scores as well as summary readability.

5 Surface Analysis of Summaries

Table 5 compares the average sentence lengths of our summaries (after reference rewriting) with those of the original news reports, human (model) summaries and machine summaries generated by the participating summarizers at DUC'03 and '04.

These figures confirm various intuitions about human vs machine-generated summaries—machine summaries tend to be based on sentence extraction;

Before:
Pinochet was placed under arrest in London Friday by British police acting on a warrant issued by a Spanish judge. Pinochet has immunity from prosecution in Chile as a senator-for-life under a new constitution that his government crafted. Pinochet was detained in the London clinic while recovering from back surgery.
After:
Gen. Augusto Pinochet, the former Chilean dictator, was placed under arrest in London Friday by British police acting on a warrant issued by a Spanish judge. Pinochet has immunity from prosecution in Chile as a senator-for-life under a new constitution that his government crafted. Pinochet was detained in the London clinic while recovering from back surgery.

Figure 2: First three sentences from a machine generated summary before/after reference regeneration.

many have an explicitly encoded preference for long sentences (assumed to be more informative); humans tend to select information at a sub-sentential level. As a result, human summaries contain on average shorter sentences than the original, while machine summaries contain on average longer sentences than the original. Interestingly, our summarizer, like human summarizers, generates shorter sentences than the original news text.

News Reports	Human Summaries	Other Machine Summaries	Our Summaries
21.43	17.43	28.75	19.16

Table 5: Av. sentence lengths in 80 document sets from DUC'03 and '04.

Equally interesting is the distribution of parentheticals. The original news reports contain on average one parenthetical unit (appositive or relative clause) every 3.9 sentences. The machine summaries contain on average one parenthetical every 3.3 sentences. On the other hand, human summaries contain only one parenthetical unit per 8.9 sentences on average.

In other words, human summaries contain fewer parenthetical units per sentence than the original reports; this appears to be a deliberate attempt at including more events and less background information in a summary. Machine summaries tend to contain on average more parentheticals than the original reports. This is possibly an artifact of the preference for longer sentences, but the data suggests that 100 word machine summaries use up valuable space by presenting unnecessary background information.

Our summaries contain one parenthetical unit every 10.0 sentences. This is closer to human summaries than to the average machine summary, again suggesting that our approach of treating the inclusion of parentheticals as a reference generation task is justified.

Before:
Turkey has been trying to form a new government since a coalition government led by Yilmaz collapsed last month over allegations that he rigged the sale of a bank. Ecevit refused even to consult with the leader of the Virtue Party during his efforts to form a government. Ecevit must now try to build a government. Demirel consulted Turkey's party leaders immediately after Ecevit gave up.
After:
Turkey has been trying to form a new government since a coalition government led by Prime Minister Mesut Yilmaz collapsed last month over allegations that he rigged the sale of a bank. Premier-designate Bulent Ecevit refused even to consult with the leader of the Virtue Party during his efforts to form a government. Ecevit must now try to build a government. President Suleyman Demirel consulted Turkey's party leaders immediately after Ecevit gave up.

Figure 3: First four sentences from another machine summary before/after reference regeneration.

6 Conclusions and Future Work

We have demonstrated that simplifying news reports by removing parenthetical information results in better sentence clustering and consequently better summarization. We have further demonstrated that using a reference rewriting module to introduce parentheticals as a post-process does not significantly affect the score on an automated content-evaluation metric; indeed we believe that a more sophisticated rewriting module might indeed improve performance on content selection. In addition, the summaries produced by our summarizer closely resemble human summaries in surface features such as average sentence length and the distribution of relative clauses and appositives.

The results in this paper might be useful to generative approaches to summarization. It is likely that the improved clustering will make operations like information fusion (Barzilay, 2003; Dalianis and Hovy, 1996) within clusters more reliable. We plan to examine whether this is indeed the case.

We feel that the performance of our summarizer is encouraging (it performs at 90% of human performance as measured by *Rouge*) as it is conceptually very simple—it selects informative sentences from the largest clusters and does not contain any theoretically inelegant optimizations, such as excluding overly long or short sentences.

Our approach of extracting out parentheticals as a pre-process also provides a framework for reference rewriting, by allowing the summarizer to select

background information independently of the main content. We believe that there is a lot of research left to be carried out in generating references in open domains and will address this issue in future work.

7 Acknowledgements

The research reported in this paper was partially supported through grants from the NSF KDD program, the DARPA TIDES program (contract N66001-00-1-8919) and an NSF ITR (award 0325887).

References

Regina Barzilay. 2003. *Information Fusion for Multidocument Summarization: Paraphrasing and Generation*. Ph.D. thesis, Columbia University, New York.

John Carroll, Guido Minnen, Darren Pearce, Yvonne Canning, Siobhan Devlin, and John Tait. 1999. Simplifying English text for language impaired readers. In *Proceedings of the 9th Conference of the European Chapter of the Association for Computational Linguistics (EACL'99)*, pages 269–270, Bergen, Norway.

Raman Chandrasekar, Christine Doran, and Bangalore Srinivas. 1996. Motivations and methods for text simplification. In *Proceedings of the 16th International Conference on Computational Linguistics (COLING '96)*, pages 1041–1044, Copenhagen, Denmark.

Hercules Dalianis and Eduard Hovy. 1996. Aggregation in natural language generation. In G. Adorni and M. Zock, editors, *Trends in natural language generation: an artificial intelligence perspective*, pages 88–105. Springer Verlag, Berlin.

Gregory Grefenstette. 1998. Producing intelligent telegraphic text reduction to provide an audio scanning service for the blind. In *Intelligent Text Summarization, AAAI Spring Symposium Series*, pages 111–117, Stanford, California.

Claire Grover, Colin Matheson, Andrei Mikheev, and Marc Moens. 2000. LT TTT - A flexible tokenisation tool. In *Proceedings of Second International Conference on Language Resources and Evaluation*, pages 1147–1154, Athens,Greece.

Vasileios Hatzivassiloglou, Judith Klavans, and Eleazar Eskin. 1999. Detecting text similarity over short passages: exploring linguistic feature combinations via machine learning. In *Proceedings of empirical methods in natural language processing and very large corpora (EMNLP'99)*, MD, USA.

Hongyan Jing. 2000. Sentence simplification in automatic text summarization. In *Proceedings of the 6th Applied Natural Language Processing Conference (ANLP'00*, Seattle, Washington.

Kevin Knight and Daniel Marcu. 2000. Statistics-based summarization — step one: Sentence compression. In *Proceeding of The 17th National Conference of the American Association for Artificial Intelligence (AAAI-2000)*, pages 703–710.

Chin-Yew Lin and Eduard Hovy. 2003. Automatic evaluation of summaries using n-gram co-occurrence statistics. In *Proceedings of the Human Language Technology Conference (HLT-NAACL 2003*, Edmonton, Canada.

Chin-Yew Lin. 2003. Improving summarization performance by sentence compression - a pilot study. In *In Proceedings of the Sixth International Workshop on Information Retrieval with Asian Languages (IRAL 2003)*, Sapporo, Japan.

William Mann and Sandra Thompson. 1988. Rhetorical Structure Theory: Towards a functional theory of text organization. *Text*, 8(3):243–281.

George A. Miller, Richard Beckwith, Christiane D. Fellbaum, Derek Gross, and Katherine Miller. 1993. Five Papers on WordNet. Technical report, Princeton University, Princeton, N.J.

A. Nenkova and K. McKeown. 2003. References to named entities: a corpus study. In *Proceedings of NAACL-HLT'03*, pages 70–72.

Stefan Riezler, Tracy H. King, Richard Crouch, and Annie Zaenen. 2003. Statistical sentence condensation using ambiguity packing and stochastic disambiguation methods for lexical-functional grammar. In *Proceedings of the 3rd Meeting of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL'03)*, Edmonton, Canada.

Advaith Siddharthan. 2002. Resolving attachment and clause boundary ambiguities for simplifying relative clause constructs. In *Proceedings of the Student Workshop, 40th Meeting of the Association for Computational Linguistics (ACL'02)*, pages 60–65, Philadelphia, USA.

Advaith Siddharthan. 2003. *Syntactic simplification and Text Cohesion*. Ph.D. thesis, University of Cambridge, UK.

George Snedecor and William Cochran. 1989. *Statistical Methods*. Iowa State University Press, Ames, IA.

N. Wacholder, Y. Ravin, and M. Choi. 1997. Disambigaution of names in text. In *Proceedings of the Fifth Conference on Applied NLP*, pages 202–208.