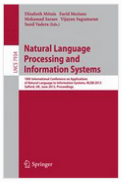
Springer Link Search



[International Conference on Application of Natural Language to Information Systems]
NLDB 2013: Natural Language Processing and Information Systems pp 126-139 | Cite as

Evaluating Syntactic Sentence Compression for Text Summarisation

Authors | Authors and affiliations

Prasad Perera, Leila Kosseim

Part of the Lecture Notes in Computer Science book series (LNCS, volume 7934)

Abstract

Evaluating Syntactic Sentence Compression for Text Summarisation

Prasad Perera and Leila Kosseim

Dept. of Computer Science & Software Engineering
Concordia University
Montreal, Canada
p_perer@encs.concordia.ca,kosseim@encs.concordia.ca

Abstract. This paper presents our work on the evaluation of syntactic based sentence compression for automatic text summarization. Sentence compression techniques can contribute to text summarization by removing redundant and irrelevant information and allowing more space for more relevant content. However, very little work has focused on evaluating the contribution of this idea for summarization. In this paper, we focus on pruning individual sentences in extractive summaries using phrase structure grammar representations. We have implemented several syntax-based pruning techniques and evaluated them in the context of automatic summarization, using standard evaluation metrics. We have performed our evaluation on the TAC and DUC corpora using the BlogSum and MEAD summarizers. The results show that sentence pruning can achieve compression rates as low as 60%, however when using this extra space to fill in more sentences, ROUGE scores do not improve significantly.

1 Introduction

Text compression has several practical applications in natural language processing such as text simplification [1], headline generation [2] and text summarization [3]. The goal of automatic text summarization is to produce a shorter version of the information contained in a text collection and produce a relevant summary [4]. In extractive summarization, sentences are extracted from the document collection and assigned a score according to a given topic/query relevance [5] or some other metric to determine how important it is to the final summary. Summaries are usually bound by a word or sentence limit and within these limits, the challenge is to extract and include as much relevant information as possible. However, since the sentences are not processed or modified, they may contain phrases that are irrelevant or may not contribute to the targeted summary. As an example, consider the following topic, query and sentence (1) [1],

Topic: *Southern Poverty Law Center*

Query: *Describe the activities of Morris Dees and the Southern Poverty Law Center*

(1) *Since co-founding the Southern Poverty Law Center in 1971, Dees has wielded the civil lawsuit like a buck knife, carving financial assets out of hate group leaders who inspire followers to beat, burn and kill.*

In sentence (1), some phrases could be dropped without losing much information relevant to the query. Possible shorter forms of the sentence include :

[1] All examples are taken from the TAC 2008 or DUC 2007 corpora.

(1c1) *Since co-founding the Southern Poverty Law Center in 1971, Dees has wielded the civil lawsuit ~~like a buck knife~~, carving financial assets out of hate group leaders who inspire followers to beat, burn and kill.*

(1c2) *~~Since co-founding the Southern Poverty Law Center in 1971,~~ Dees has wielded the civil lawsuit like a buck knife, carving financial assets out of hate group leaders who inspire followers to beat, burn and kill.*

(1c3) *~~Since co-founding the Southern Poverty Law Center in 1971,~~ Dees has wielded the civil lawsuit ~~like a buck knife,~~ carving financial assets out of hate group leaders ~~who inspire followers to beat, burn and kill.~~*

In principle, sentence compression should improve automatic extractive text summarization by removing redundant and less relevant information within sentences and thus preserve space to include more useful information into length-limited summaries. However to the best of our knowledge, very little previous work has focused on measuring the contribution of specific sentence compression techniques as a means to improve summary content.

2 Previous Work on Sentence Compression

Previous sentence compression methods have relied on different techniques ranging from machine learning and classifier based (eg. [3]), syntactic pruning (based on complete parses or shallow parses) (eg. [6–9]) to keyword based (eg. [10]) techniques.
One early approach to sentence pruning focused on removing inessential phrases in extractive summaries based on an analysis of human written abstracts [9, 6]. In their work, the authors have used a syntactic parser to identify different types of phrases which are present in the original sentences but not in human written simplified sentences. These phrases were used to train a Naive Bayes Classifier to decide how likely a phrase is to be removed from a sentence. For evaluation, they have compared their compressed sentences to those compressed by humans and acheived a 78.1% overall success rate but have noted a low success rate for removing adjectives, adverbs and verb phrases. However, the effect on summary content was not indicated.
Another interesting work is that of [3] who proposed a noisy channel model technique based on the hypothesis that there exists a shorter original sentence (s) and the existing longer sentence (t) was formed by adding optional phrases. Given the long string t and every pair of (t, s), the probability $P(t \mid s)$ represents the likelihood of arriving at the long string t, when s is expanded. Their model was designed considering two key features: preserving grammaticality and preserving useful information. In order to calculate the probabilities, they have used context free grammar parses of sentences and a word based bi-gram estimation model. They have evaluated their system using the Ziff-Davis corpus and have showed that their approach could score similar compression rates compared to human written compressed texts but importance and grammaticality are slightly lower than human-written texts. On the other hand, [11] introduced semantic features to improve a decision tree based classification. Here, the authors used Charniak's parser [12] to generate syntactic trees and incorporated semantic information using WordNet [13]. The evaluation showed a slight improvement in importance of information preserved in shortened sentences. But again, the effect on summarization was not noted.
[14] points out that text compression could be seen as a problem of finding a global optimum by considering the compression of the whole text/document. The authors used syntactic trees of each pairs of long and short sentences to define rules to deduce shorter syntactic trees out of original syntactic trees. They also used the Ziff-Davis corpus for their evaluation as well as

human judgment. They evaluated their technique based on importance and grammaticality of sentences and the results were lower compared to the scores of the human written abstractions. Similarly, [15] describes the use of integer linear programming model to infer globally optimal compressions while adhering to linguistically motivated constraints and show improvement in automatic and human judgment evaluations. [16] have also described an approach on syntactic pruning based on trasformed dependency trees and a linear integer model. The authors have transformed the dependency trees into graphs where the nodes represents nouns and verbs and these transformed dependency trees are trimmed based on the results of an integer linear programming model that decides the importance of each subtree. Their evaluation has shown an improvement compared to the language model based compression techniques.

The previous work described above were evaluated intrinsically by comparing their results to human generated summmaries. A few previous work did however measure sentence compression extrinsically for the purpose of text summarization. In particular, [10] took a conservative approach and used a list of keyword phrases to identify less significant parts of the text and remove them from long sentences. The keyword list was implemented in an adhoc fashion and was used to omit specific terms. They have evaluated their pruning techniques within their summarization system CLASSY [17] with DUC 2005 [18], and showed an improvement in ROUGE scores. In their participation to the DUC 2006 [19] automatic summarization track, their system placed among the top three based on ROUGE scores.
In contrast, [7] used complete dependency parses and applied pruning rules based on grammatical structures. They used specific grammatical filters including prepositional complements of verbs, subordinate clauses, noun appositions and interpolated clauses. They have achieved a compression rate of 74% while retaining grammaticality and readability of text. In [20] the authors also used syntactic structures and applied linguistically motivated filtering to simplify sentences. Using the TIPSTER [21] corpus, they identified syntactic patterns which were absent from human-written summaries compared to the original corpus and defined a trimming algorithm consisting of removing sub-trees of grammatical phrase structures while traversing through a complete parsed tree structure. They have evaluated their pruning technique on the DUC 2003 summarization task and showed an improvement in ROUGE scores compared to uncompressed length-limited summaries. Finally, [8] describes the sentence compression module of their text summarization system, based on syntactic level sentence pruning. They have implemented a module of compression which filters adverbial modifiers and relative clauses from original sentences to achieve text compression. Their evaluations were performed using the DUC 2007 summarization track and have showed an improvement in ROUGE scores after applying their compression technique to their summarization system.

As described above, most previous work have evaluated their sentence compression technique intrinsically against human generated compressed sentences. Very few (notably [7, 8, 20, 10]) have evaluated them extrinsically as part of a summarization system but the exact contribution of each technique to the summary content has not been measured.

3 Pruning Heuristics

To evaluate syntactic sentence pruning methods in the context of automatic text summarization, we have implemented several syntax-based heuristics and have evaluated them with standard summarization benchmarks. We took as input a list of extracted sentences ranked by their relevance score as generated by an automatic summarizer. We then performed a complete parse of these sentences, and applied various syntax-based pruning approaches to each tree node to determine whether to prune or not a particular sub-tree. The pruned sentences were then included in the final summary in place of the original sentences and evaluated for content against the given model summaries. Three basic sentence compression approaches were attempted: syntax-driven pruning, syntax and relevancy based pruning, and relevancy-driven syntactic pruning. Let us describe each approach in detail.

3.1 Syntax-Driven Pruning

Our first approach to sentence pruning was based solely on syntactic simplifications. After parsing the extracted sentences deemed relevant by the summarizer, we tried to remove specific sub-trees regardless of their computed relevance to the query/topic. Here, the rationale was that specific syntactic structures by default carry secondary informative content, hence removing them should not decrease the content of the summary significantly. These pruning heuristics are based on the work of [20] and [7] (adapted to English). Specially, we removed: relative clauses, adjective and adverbial phrases, conjuncted clauses as well as specific types of prepositional phrases. Let us describe each heuristic:

Pruning Relative Clauses A relative clause modifies a noun or noun phrase and is connected to the noun by a relative pronoun, a relative adverb, or a zero relative. As such, they act as adjectival phrases that provide additional information about the noun it modifies. As an example, consider the following sentence:

(2) *"It's over", said Tom Browning, an attorney for Newt Gingrich, <u>who was not present at Thursday's hearing.</u>*

Pruning the sub tree structure headed by *who*, which represents a relative clause, results in a shortened sentence.

Pruning Adjective Phrases An ajective phrase is a word, phrase, or sentence element that enhances, limits or qualifies the meaning of a noun phrase. As complementary phrases, they can often be dropped from a sentence without loosing the main content of the sentence. As an example:

(3) *Mark Barton, the <u>44-year-old</u> day trader at the center of Thursday's <u>bloody</u> rampage, was described by neighbors in the Atlanta suburb of Morrow as a <u>quiet, churchgoing</u> man who worked all day on his computer.*

The phrases *44-year-old*, *bloody* and *quiet, churchgoing* are suitable candidates to be pruned from the original phrase structure.

Pruning Adverbial Phrases An adverbial phrase is a word, phrase, or sentence element that modifies a verb phrase. For example:

(4) *So <u>surely</u> there will be a large number of people who <u>only</u> know us for Yojimbo.*

Here, the phrases *surely* and *only* provide additional information regarding their associated verb phrases, but can often be dropped without affecting the content of the sentence significantly.

Pruning Trailing Conjuncted Verb Phrases Conjunctions may be used to attach several types of phrases. In our corpora, verb phrases (VPs) are often conjuncted and the second VP is typically shorter and contains secondary information. For example, consider the following sentence:

(5) *The Southern Poverty Law Center has accumulated enough wealth in recent years to embark on a major construction project and to have assets totaling around $100 million.*

Based on our corpus analysis, we developed a heuristic that removes trailing conjuncted VPs.

Pruning Prepositional Phrases Prepositional phrases (PP) are used to modify noun phrases, verb phrases or complete clauses. Pruning PPs can be done, but with caution. Indeed, some PPs do contain secondary information which can be removed without hindering the grammar or the semantics of the sentence; while other types of PPs do contain necessary information. Consider the following example:

(6) *In the Public Records Office in London archivists are creating a catalog of all British public records regarding the Domesday Book of the 11th century.*

Here, the prepositional phrase, *In the Public Records Office in London* is attached to the entire clause; while, *of all British public records* and *of the 11th century* are attached to the nouns *catalog* and *Domesday Book*. PPs attached to NPs often act as noun modifiers and as a consequence can be pruned like any adjective phrase. In addition, PPs attached to an entire clause often present complementary information that can also be removed. On the other hand, PPs can be attached to verb phrases, as in:

(7) *Australian Prime Minister John Howard today defended the governments decision to go ahead with uranium mining on development and environmental grounds.*

where *with uranium mining* and *on development and environmental grounds* are attached to *go ahead*. PPs that modify verb phrases should be pruned with caution as they may be part of the verb's frame and required to understand the verb phrase. In that case, removing them would likely loose the meaning of the sentence. PPs attached to VPs that are positioned after the head verb are therefore not pruned. However, PPs attached to VPs that are positioned prior to the verb are considered less likely to be mandatory and are removed. Removing PPs based solely on syntactic information will likely make mistakes. PPs that do not contain necessary information may be kept, and vice-versa. However, the purpose of this heuristic is to prune as cautiously as possible. Sections 3.2 and 3.3 describe heuristics that take semantics into account.

3.2 Syntax and Relevancy based Pruning

The danger with syntax only based pruning is that it may remove sub-trees that do contain relevant information for the summary. In order to avoid this, we toned down our syntax based heuristics described in Section 3.1, by measuring the relevancy of the sub-tree to prune and only remove it if it is below a certain threshold. In the case of query-based summarization, we used the cosine similarity between the tf-idf values of the candidate sub-tree to prune and the topic/query. Specifically, if the syntax based heuristic consider that a sub-tree is a candidate for pruning but its similarity with the topic & query is above some threshold, we do not prune it on the grounds that it seems to have relevant content.

3.3 Relevancy-Driven Syntactic Pruning

Our previous techniques (see Sections 3.1 and 3.2) focused on keeping the sentence grammaticality as much as possible by driving the pruning based on syntax. Next we took an approach to prune sentences focusing less on preserving grammaticality and more on preserving relevant information. Our last approach focused on finding irrelevant information within a sentence and remove its embedding sub-tree. Specifically, we parse the extracted sentences as before and for each sub-tree except for noun phrases, verb phrases or individual words, we calculate its cosine similarity with the topic/query based on tf-ifd values. Sub-trees below a certain threshold are pruned; the others are kept. We do not allow pruning of noun phrases, verb phrases and individual words in order to preserve a minimal grammaticality; all other phrase types, are however possible candidates for pruning. For example, consider the following scenario:

Topic: *Turkey and the European Union*

Query: *What positive and negative developments have there been in Turkey's efforts to become a formal member of the European Union?*

(8) *Turkey had been asking for three decades to join the European Union but its demand was turned away by the European Union in December 1997 that led to a deterioration of bilateral relations.*

Here, Sentence 8 is the original candidate extracted from the corpus. Its parse tree generated by the Stanford Parser [22] is shown is Figure 1, with the relevancy score indicated in bold. For example, the sub-tree rooted by the SBAR *(that led to a deterioration of bilateral relations)* was computed to have a relevance of 0.0 with the topic and the query. All sub-trees rooted at a node whose relevance is smaller than some threshold value are pruned. If we set $t = 0$ (i.e. any relevance with topic/query will be considered useful), the above sentence would therefore be compressed as:

(8c) *Turkey had been asking ~~for three decades~~ to join the European Union but its demand was turned away by the European Union ~~in December 1997 that led to a deterioration of bilateral relations.~~*

4 Evaluation

To evaluate our pruning techniques extrinsically for the purpose of summary generation, we used two standard text corpora available for summarization: the Text Analysis Conference (TAC) 2008 [23], which provides a text corpus created from blogs and the Document Understanding Conference (DUC) 2007 [18] which provides a text corpus of news articles. To ensure that our results were not tailored to one specific summarizer, we used two different systems: BlogSum [24], an automatic summarizer based on discourse relations and MEAD [25], a generalized automatic summarization system. In order to generate syntactic trees for our experiment, we used the Stanford Parser [22]. To evaluate each compression technique, we generated summaries without any compression and compared the results based on two metrics: compression rates and ROUGE scores for content evaluation.

4.1 Evaluation of Compression Rates

To measure the compression rate of each technique, we first created summaries using BlogSum and MEAD, setting a limit of 250 words per summary, then applied each sentence pruning heuristic independently to generate different sets of summaries.

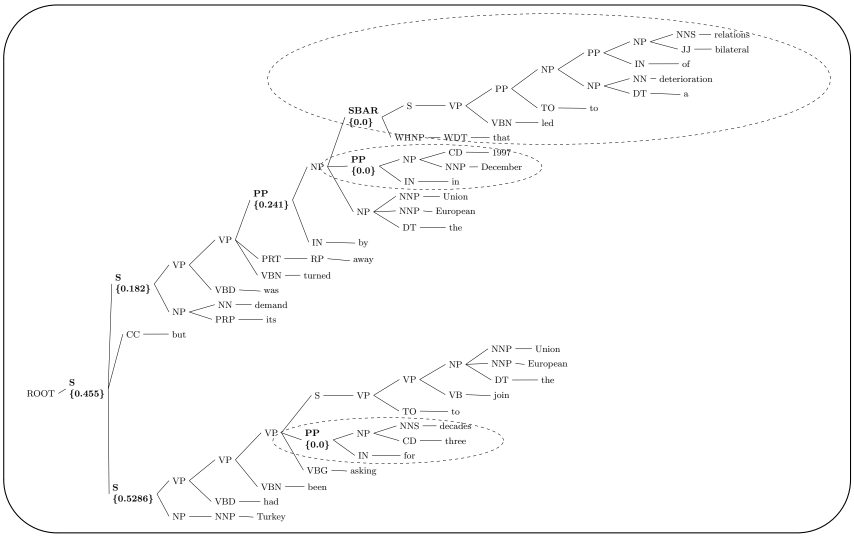


Fig. 1. Dependency Phrase Structure For Sentence 8.

Syntax-Driven Pruning Table 1 shows the compression rates achieved by each heuristic for both summarizers and both datasets. As Table 1 shows, with both datasets, apart from the combined approach, the highest sentence compression was achieved by preposition based pruning (PP pruning); while the lowest compressions were observed with relative clause (RC), adverbial phrases (Adv) and trailing conjuncted verb phrases (TC-VP) pruning. This is not surprising as PPs are a priori more frequent than the other syntactic constructions. Also not surprisingly, the combined approach which applies all pruning heuristics achieved the highest compression rate in both datasets reaching about 73% to 75% compression rates.

	BlogSum				**MEAD**			
	TAC 2008		**DUC 2007**		**TAC 2008**		**DUC 2007**	
	No of. Words	**Compression Rate**	**No of. Words**	**Compression Rate**	**No of. Words**	**Compression Rate**	**No of. Wordss**	**Compression Rate**
Original	11272	100.0%	10648	100.0%	11759	100.0%	11186	100.0%
Adv Pruning	10804	95.8%	10422	97.9%	11491	97.7%	10973	98.1%
RC Pruning	10803	95.8%	10309	96.8%	11273	95.9%	10708	95.7%
TC-VP Pruning	10887	96.6%	10271	96.5%	11530	98.0%	10789	96.4%
Adj Pruning	10430	92.5%	9897	93.0%	11225	95.4%	10391	92.9%
PP Pruning	9349	83.0%	8442	79.3%	10359	76.7%	8584	76.3%
Combined	8170	72.5%	7995	75.1%	9799	83.3%	8143	72.8%

Table 1. Sentence Compression Rates of Syntax-Driven Pruning.

Syntax and Relevancy Based Pruning Table 2 shows the compression rate achieved by each heuristic using the syntax and relevancy based pruning. As the results show, with both datasets, the compression effect of each heuristic has been toned down, but the relative ranking of the heuristics are the same. This seems to imply that each type of syntactic phrase is as likely to contain irrelevant information; and one particular construction should not be privileged for pruning purposes. Overall, when all pruning heuristics are combined, the relevancy factor reduces the pruning by about 8 to 11% (from 73-75% to 82-86%) .[2]

	BlogSum				MEAD			
	TAC 2008		DUC 2007		TAC 2008		DUC 2007	
	No of. Words	Compression Rate	No of. Words	Compression Rate	No of. Words	Compression Rate	No of. Wordss	Compression Rate
Original	11272	100.0%	10648	100.0%	11759	100.0%	11186	100.0%
Adv Pruning	10869	96.4%	10435	98.0%	11526	98.0%	11030	98.6%
RC Pruning	11100	98.4%	10575	99.3%	11495	97.7%	10988	98.2%
TC-VP Pruning	10887	96.6%	10478	98.4%	11644	99.0%	10969	98.1%
Adj Pruning	11111	98.6%	10085	94.7%	11287	96.0%	10535	94.2%
PP Pruning	10261	91.0%	9834	92.3%	10976	93.3%	9754	87.2%
Combined	9234	82.0%	9170	86.1%	10361	88.1%	9178	82.0%

Table 2. Sentence Compression Rates of Syntax and Relevancy Based Pruning

Relevancy-Driven Syntactic Pruning Table 3 shows the results of the compression rate achieved by relevancy-driven syntactic pruning. The relevancy-driven syntactic pruning has achieved a higher compression rate than syntax and relevancy based pruning. Table 4 shows

	BlogSum				MEAD			
	TAC 2008		DUC 2007		TAC 2008		DUC 2007	
	No Of. Words	Compression Rate	No Of. Words	Compression Rate	No Of. Words	Compression Rate	No Of. Words	Compression Rate
Original	11272	100.0%	10648	100.0%	11759	100.0%	11186	100.0%
Relevancy-Driven	7457	66.1%	7879	74.0%	7122	60.6%	6801	69.0%

Table 3. Sentence Compression Rates of Relevancy-Driven Syntactic Pruning.

the types of syntactic structures that were removed by the relevancy-driven pruning and their relative frequencies. As the results shows, the most frequent syntactic structures removed were PPs and the least were adverbial phrases (Adv). This result correlates with our syntax-driven pruning as we achieved similar individual compression rates for these phrase structures.

4.2 Evaluation of Content

Compression rate is interesting, but not at the cost of pruning useful information. In order to measure the effect of the pruning strategies on the content of summaries, we have ran the same experiments again but this time we have calculated the F-measures of the ROUGE scores (R-2 and R-SU4). In principle, pruning sentences should shorten summaries thus allowing us to fill the summary with new relevant sentences and hence improve its overall content. In order to evaluate the effect of sentence compression on this, we first created summaries with a word limit of 250 and then created two summaries: one without filling the summary with extra content to reach the 250 word limit and one with filling with new content to reach the 250 word limit. We calculated

[2] The reduction rate is of course proportional to the relevancy threshold used (see Section 3.2). In this experiment, we set the threshold to be the most conservative (t = 0), hence keeping everything that has any relevance to the topic/query.

	BlogSum				MEAD			
	TAC 2008		DUC 2007		TAC 2008		DUC 2007	
	No of. Phrases	Relative Frequency	No of. Phrases	Relative Frequency	No of. Phrases	Relative Frequency	No of. Phrases	Relative Frequency
PP Pruning	395	50.5%	402	62.4%	177	42.3%	408	63.6%
Other	189	24.1%	136	29.3%	157	31.6%	149	30.1%
RC Pruning	94	12.0%	56	8.7%	44	10.5%	59	9.2%
Adj Pruning	75	9%	35	5.4%	26	6.2%	20	3.1%
Adv Pruning	29	3.7%	15	2.4%	14	3.3%	5	1.0%
Total	782	100%	644	100%	418	100%	641	100%

Table 4. Syntactic Phrase Structures Removed by Relevancy-Driven Pruning.

ROUGE scores for both set of summaries. Again to avoid any bias, we created summaries using both the BlogSum and MEAD systems and both datasets.

Syntax-Driven Pruning Tables 5 and 6 show the results obtained with and without content filling respectively. Table 5 show a drop in ROUGE score for both summarization systems and both datasets. This goes against our hypothesis that by default specific syntactic constructions can be removed without losing much content. In addition, when filling the summary with extra sentences, ROUGE scores do seem to improve (as shown in Table 6); however Pearson's χ^2 and t-tests show that this difference is not statistically significant. What is more surprising is that this phenomenon is true for the combined heuristics, but also for each individual pruning heuristic.

	BlogSum				MEAD			
	TAC 2008		DUC 2007		TAC 2008		DUC 2007	
	R-2	R-SU4	R-2	R-SU4	R-2	R-SU4	R-2	R-SU4
Original	0.074	0.112	0.088	0.141	0.040	0.063	0.086	0.139
Adv Pruning	0.074	0.113	0.089	0.143	0.039	0.063	0.086	0.139
RC Pruning	0.072	0.109	0.087	0.140	0.039	0.062	0.085	0.138
TC-VP Pruning	0.073	0.111	0.088	0.140	0.040	0.063	0.085	0.137
Adj Pruning	0.068	0.108	0.084	0.140	0.038	0.063	0.080	0.136
PP Pruning	0.065	0.103	0.072	0.121	0.035	0.056	0.069	0.117
Combined	0.060	0.100	0.074	0.128	0.034	0.056	0.068	0.121

Table 5. Content Evaluation of Compressed Summaries with Syntax-Driven Pruning (Without Filling).

Syntax and Relevancy Based Pruning Recall that the syntax-driven pruning did not consider the relevancy of the sub-tree to prune. When we do take the relevancy to account; surprisingly the ROUGE scores do not improve significantly either. Tables 7 and 8 show the ROUGE scores of the compressed summaries based on syntax and relevancy without filling (Table 7) and with content filling (Table 8). Again any semblance of improvement is not statistically significant.

Relevancy-Driven Pruning Table 9 shows the results of relevancy-driven pruning with and without filling and compares them to the original summaries. Again the results are surprisingly low. This last approach was still not able to improve ROUGE scores significantly.

	BlogSum				**MEAD**			
	TAC 2008		**DUC 2007**		**TAC 2008**		**DUC 2007**	
	R-2	**R-SU4**	**R-2**	**R-SU4**	**R-2**	**R-SU4**	**R-2**	**R-SU4**
Original	0.074	0.112	0.088	0.141	0.040	0.063	0.086	0.140
Adv Pruning	0.075	0.114	0.090	0.143	0.044	0.063	0.087	0.140
RC Pruning	0.073	0.111	0.088	0.141	0.039	0.062	0.086	0.140
TC-VP Pruning	0.073	0.111	0.089	0.141	0.040	0.062	0.086	0.139
Adj Pruning	0.075	0.110	0.085	0.142	0.038	0.063	0.082	0.140
PP Pruning	0.070	0.131	0.079	0.131	0.035	0.058	0.076	0.127
Combined	0.065	0.139	0.065	0.139	0.035	0.060	0.077	0.135

Table 6. Content Evaluation of Compressed Summaries with Syntax-Driven Pruning (With Filling).

	BlogSum				**MEAD**			
	TAC 2008		**DUC 2007**		**TAC 2008**		**DUC 2007**	
	R-2	**R-SU4**	**R-2**	**R-SU4**	**R-2**	**R-SU4**	**R-2**	**R-SU4**
Original	0.074	0.112	0.088	0.141	0.040	0.063	0.086	0.139
Adv Pruning	0.074	0.113	0.089	0.143	0.039	0.063	0.087	0.140
RC Pruning	0.073	0.110	0.088	0.141	0.039	0.062	0.086	0.139
TC-VP Pruning	0.073	0.111	0.088	0.141	0.040	0.063	0.085	0.138
Adj Pruning	0.070	0.110	0.086	0.142	0.038	0.063	0.082	0.138
PP Pruning	0.072	0.110	0.086	0.137	0.039	0.062	0.079	0.129
Combined	0.069	0.110	0.085	0.140	0.038	0.061	0.078	0.132

Table 7. Content Evaluation of Compressed Summaries with Syntax-Driven with Relevancy Pruning (Without Filling).

4.3 Discussion

Although the results of the compression rate were inline with previous work [7, 20], we were surprised at the results of the content evaluation. However this might explain why, to our knowledge, so little work can be found in the literature on the evaluation of syntactic sentence pruning for summarization. Our pruning heuristics could of course be fine-tuned to be more discriminating. We could, for example, use verb frames or lexical-grammatical rules to prune PPs; but we do not foresee a significant increase in ROUGE scores. The relevance measure that we used (see Section 3.3) could also be experimented with, but again, we do not expect much increase from that end. Using a better performing summarizer might also be a possible avenue of investigation to provide us with better input sentences and better "filling" sentences after compression.

5 Conclusion and Future Work

In this paper, we have described our experiments on syntactic based sentence pruning applied to automatic text summarization. We have defined three types of pruning techniques based on complete syntactic parses: a first technique based solely on syntax, a second technique that tones down the syntactic pruning by taking relevancy into account and third technique that is driven by relevancy. These techniques were applied to the sentences extracted by two different summarizers to generate compressed summaries and evaluated on the TAC-2008 and DUC-2007 benchmarks. According to results, these pruning techniques generate a compression rate between 60% to 88% which is inline with previous work [7, 20]. However, when using the extra space to include additional sentences, the content evaluation does not show a significant improvement in

	BlogSum				MEAD			
	TAC 2008		DUC 2007		TAC 2008		DUC 2007	
	R-2	R-SU4	R-2	R-SU4	R-2	R-SU4	R-2	R-SU4
Original	0.074	0.112	0.088	0.141	0.040	0.063	0.086	0.140
Adv Pruning	0.075	0.114	0.090	0.143	0.040	0.063	0.087	0.140
RC Pruning	0.072	0.111	0.088	0.141	0.040	0.062	0.086	0.140
TC-VP Pruning	0.074	0.111	0.089	0.141	0.040	0.062	0.086	0.139
Adj Pruning	0.072	0.111	0.086	0.142	0.038	0.063	0.085	0.141
PP Pruning	0.072	0.111	0.088	0.141	0.040	0.062	0.084	0.135
Combined	0.071	0.112	0.088	0.145	0.037	0.062	0.085	0.141

Table 8. Content Evaluation of Compressed Summaries with Syntax-Driven with Relevancy Pruning (With Filling).

	BlogSum				MEAD			
	TAC 2008		DUC 2007		TAC 2008		DUC 2007	
	R-2	R-SU4	R-2	R-SU4	R-2	R-SU4	R-2	R-SU4
Original	0.074	0.112	0.088	0.141	0.040	0.063	0.086	0.139
Relevancy-Driven Without Filling	0.065	0.100	0.077	0.125	0.034	0.055	0.066	0.110
Relevancy-Driven With Filling	0.068	0.106	0.083	0.135	0.033	0.060	0.078	0.128

Table 9. Content Evaluation of Compressed Summaries with Relevancy-Driven Syntactic Pruning (With and Without Filling).

ROUGE scores.
As future work, we are planning to move to a manual human evaluation, as [3] and [11] did in their work. We are interested to find out if human assesssors agree with ROUGE scores, and thus we need to re-think our syntactic approach or if a human evaluation does consider the condensed summaries to be more informative than the original ones, hence putting aside ROUGE measures for the task.

Acknowledgments

This project was financially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC). The authors would also like to thank the anonymous reviewers for their valuable comments.

References

1. Chandrasekar, R., Doran, C., Srinivas, B.: Motivations and Methods for Text Simplification. Proceedings of COLING '96, Copenhagen (1996) 1041–1044
2. Dorr, B., Zajic, D., Schwartz, R.: Hedge Trimmer: A Parse-and-Trim Approach to Headline Generation. In: Proceedings of the HLT-NAACL Workshop on Text Summarization. (2003) 1–8
3. Knight, K., Marcu, D.: Summarization beyond sentence extraction: A probabilistic approach to sentence compression. Artificial Intelligence **139**(1) (July 2002) 91–107
4. Hahn, U., Mani, I.: The Challenges of Automatic Summarization. IEEE Computer
5. Murray, G., Joty, S., Ng, R.: The University of British Columbia at TAC 2008. Proceedings of TAC 2008, Gaithersburg, Maryland USA (2008)
6. Jing, H.: Sentence Reduction for Automatic Text Summarization. In: Proceedings of the Sixth Conference on Applied Natural Language Processing, Seattle (April 2000) 310–315

7. Gagnon, M., Da Sylva, L.: Text Compression by Syntactic Pruning. Proceedings of the 19^{th} International Conference on Advances in Artificial Intelligence, Quebec City, Springer-Verlag (2006) 312–323
8. Jaoua, M., Jaoua, F., Belguith, L.H., Hamadou, A.B.: Évaluation de l'impact de l'intgration des étapes de filtrage et de compression dans le processus d'automatisation du résumé. In: Résumé automatique de documents. Volume 15 of Document numérique, Lavoisier (2012) 67–90
9. Jing, H., McKeown, K.R.: Cut and Paste Based Text Summarization. Proceedings of NAACL-2000, Seattle (2000) 178–185
10. Conroy, J.M., Schlesinger, J.D., O'Leary, D.P., Goldstein, J.: Back to Basics: CLASSY 2006. In: Proceedings of the HLT-NAACL 2006 Document Understanding Workshop, New York City (2006)
11. Nguyen, M.L., Phan, X.H., Horiguchi, S., Shimazu, A.: A New Sentence Reduction Technique Based on a Decision Tree Model. International Journal on Artificial Intelligence Tools **16**(1) (2007) 129–138
12. McClosky, D., Charniak, E., Johnson, M.: Effective Self-Training for Parsing. Proceedings of HLT-NAACL'06, New York (2006) 152–159
13. Fellbaum, C.: WordNet: An Electronic Lexical Database. The MIT Press (May 1998)
14. Le Nguyen, M., Shimazu, A., Horiguchi, S., Ho, B.T., Fukushi, M.: Probabilistic Sentence Reduction Using Support Vector Machines. Proceedings of COLING'04, Geneva (August 2004) 743–749
15. Clarke, J., Lapata, M.: Global Inference for Sentence Compression an Integer Linear Programming Approach. Journal of Artificial Intelligence Research (JAIR) **31**(1) (March 2008) 399–429
16. Filippova, K., Strube, M.: Dependency Tree Based Sentence Compression. In: Proceedings of the Fifth International Natural Language Generation Conference. INLG '08, Stroudsburg, PA, USA (2008) 25–32
17. Schlesinger, J.D., O'Leary, D.P., Conroy, J.M.: Arabic/English Multi-document Summarization with CLASSY: The Past and the Future. Proceedings of CICLing'08, Haifa, Springer-Verlag (2008) 568–581
18. Dang, H.T.: DUC 2005: Evaluation of Question-focused Summarization Systems. In: Proceedings of the Workshop on Task-Focused Summarization and Question Answering, Sydney (2006) 48–55
19. Dang, H.T.: Overview of DUC 2006. In: Proceedings of the HLT-NAACL 2006 Document Understanding Workshop. (2006)
20. Zajic, D.M., Dorrand, B.J., Lin, J., Schwartz, R.: Multi-candidate Reduction: Sentence Compression as a Tool for Document Summarization Tasks. Information Processing and Management **43**(6) (November 2007) 1549–1570
21. Harman, D., Liberman, M.: TIPSTER Complete. Linguistic Data Consortium (LDC), Philadelphia (1993)
22. Marneffe, M.C.D., Manning, C.D.: The Stanford typed dependencies representation. In: Proceedings of the Workshop on Cross-Framework and Cross-Domain Parser Evaluation. CrossParser '08, Manchester (2008) 1–8
23. Dang, H., Owczarzak, K.: Overview of the TAC 2008 Update Summarization Task. In: Proceedings of the Text Analysis Conference. TAC 2008, Gaithersburg (2008)
24. Mithun, S.: Exploiting Rhetorical Relations in Blog Summarization. In: Proceedings of the 23rd Canadian Conference on Advances in Artificial Intelligence. AI'10, Ottawa, Springer-Verlag (2010) 388–392
25. Radev, D., Allison, T., Blair-Goldensohn, S., Blitzer, J., Çelebi, A., Dimitrov, S., Drabek, E., Hakim, A., Lam, W., Liu, D., Otterbacher, J., Qi, H., Saggion, H., Teufel, S., Topper, M., Winkel, A., Zhang, Z.: MEAD - A platform for multidocument multilingual text summarization. In: Proceedings of LREC-2004, Lisbon, Portugal (May 2004)