1-1-2011

General Versus Specific Sentences: Automatic Identification and Application to Analysis of News Summaries

Annie Louis
University of Pennsylvania

Ani Nenkova
Univesity of Pennsylvania, nenkova@seas.upenn.edu

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-11-07.

General Versus Specific Sentences: Automatic Identification and Application to Analysis of News Summaries

Abstract

In this paper, we introduce the task of identifying general and specific sentences in news articles. Instead of embarking on a new annotation effort to obtain data for the task, we explore the possibility of leveraging existing large corpora annotated with discourse information to train a classifier. We introduce several classes of features that capture lexical and syntactic information, as well as word specificity and polarity. We then use the classifier to analyze the distribution of general and specific sentences in human and machine summaries of news articles. We discover that while all types of summaries tend to be more specific than the original documents, human abstracts contain a more balanced mix of general and specific sentences but automatic summaries are overwhelmingly specific. Our findings give strong evidence for the need for a new task in (abstractive) summarization: identification and generation of general sentences.

Comments

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-11-07.

General versus specific sentences: automatic identification and application to analysis of news summaries

Annie Louis
University of Pennsylvania
lannie@seas.upenn.edu

Ani Nenkova
University of Pennsylvania
nenkova@seas.upenn.edu

Abstract

In this paper, we introduce the task of identifying general and specific sentences in news articles. Instead of embarking on a new annotation effort to obtain data for the task, we explore the possibility of leveraging existing large corpora annotated with discourse information to train a classifier. We introduce several classes of features that capture lexical and syntactic information, as well as word specificity and polarity. We then use the classifier to analyze the distribution of general and specific sentences in human and machine summaries of news articles. We discover that while all types of summaries tend to be more specific than the original documents, human abstracts contain a more balanced mix of general and specific sentences but automatic summaries are overwhelmingly specific. Our findings give strong evidence for the need for a new task in (abstractive) summarization: identification and generation of general sentences.

1 Introduction

Sentences in written text differ in how much specific content they have. Consider the sentences in Table 1 from a news article about the Booker prize. The first one is specific and details the issues surrounding the books chosen for the award. The second sentence is general, it states that the prize is controversial but provides no details. Human-written texts are a mix of such general and specific statements. It has been observed for example that technical writing has an hour-glass like structure. The introduction starts with general content, then the text narrows down to specific content in the methods and results section. The conclusion starts with specific content but gradually becomes general (Swales and Feak, 1994).

The novel, a story of Scottish low-life narrated largely in Glaswegian dialect, is unlikely to prove a popular choice with booksellers who have damned all six books shortlisted for the prize as boring, elitist and - worst of all - unsaleable.”
...
The Booker prize has, in its 26-year history, always provoked controversy.

Table 1: General (in italics) and specific sentences

In this paper, we propose the first automatic approach for the task of distinguishing between sentences containing general or specific content and we demonstrate how the classifier can be applied to the analysis of content found in news summaries. Our work is close in spirit to the rapidly evolving area of research concerned with identifying properties of text beyond topicality and importance. Such properties are necessary for the proper interpretation of text; they include deciding if a sentence is subjective or objective (Yu and Hatzivassiloglou, 2003; Wiebe and Riloff, 2005), if it expresses positive or negative opinion (Kim and Hovy, 2004; Wilson et al., 2005), if it expresses emotion or is neutral (Aman and Szpakowicz, 2007), or if the language in the sentence is figurative or literal (Birke and Sarkar, 2006).

We present a supervised classifier for detecting general and specific sentences. Our training data comes from the Penn Discourse Treebank (PDTB), where relevant distinctions have been annotated in the larger context of discourse relation analysis. We show that classification accuracies as high as 75% can be obtained for distinguishing sentences of the two types compared with a random baseline of 50%.

We use the classifier's predictions on data for automatic summarization. We explore the question of whether general or specific sentences would be preferred in summaries compared to their inputs. Our example general sentence in Table 1 above, which mentions the controversy surrounding the Booker prize is in fact a good short summary of the topic of the article. It is unknown if such overview sentences are more desirable for summaries compared to specific sentences such as the one about booksellers.

Recently Haghighi and Vanderwende (2009) have noticed that summaries of varying specificity can be generated from the same text and that more general information can increase summarizer performance. We present the first quantitative study on a large corpus of *source texts* and their *human* and *machine* summaries. We show that summaries tend to have more specific content compared to input texts, however, system summaries have much more specific content than those written by humans. Our findings motivate further research into identification and generation of general sentences for summarization.

2 A general vs. specific sentence classifier

Our source of general and specific sentences comes from the Penn Discourse Treebank (PDTB) (Prasad et al., 2008). This corpus contains annotations for discourse relations and covers 1 million words from Wall Street Journal (WSJ) articles. Two of the discourse relations in the PDTB—Instantiation and Restatement—seem highly relevant to the distinctions we wish to make. These relations are annotated between adjacent sentences where one sentence provides only general content, the other elaborates on the general sentence providing specific information.[1] In contrast to efforts in automatic discourse processing (Marcu and Echihabi, 2001; Sporleder and Lascarides, 2008), in our work we are not interested in identifying adjacent sentences between which this relation holds. Our focus is on developing a classifier for general and specific and we simply use the annotated sentences as samples for the two classes, without preserving or exploiting any cues coming from adjacency.

The definitions of Instantiation and Restatement relations in the PDTB annotation manual are briefly listed below to motivate their choice for our study. Some examples of the relations are listed in Table 2.

INSTANTIATION: Sentence 1 evokes a set and sentence 2 describes it in further detail. It may be a set of events, reasons or a generic set of events, behaviors and attitudes. The relation involves a function which extracts the set of events from the semantics of sentence 1 and sentence 2 describes one element in the extracted set.

RESTATEMENT: The semantics of sentence 2 restates that of sentence 1. The subtypes "specification", "generalization", and "equivalence" further specify the ways in which the second sentence restates the first. In the case of specification, sentence 2 describes the situation in sentence 1 in more detail.

Our idea is to use the first sentences in these relations as general sentences and the second sentences as examples for specific sentences. The two relations are fairly frequent in the PDTB (1403 Instantiations and 2370 Restatement-Specifications) and provide a resonable training set for our classifier.

The classifier we describe in this section is built to predict for a given sentence its category as general or specific. We reiterate that although the definitions above describe the specificity of one sentence relative to the other, we do not focus on this pairwise difference in specificity for the work reported here.

2.1 Features

Based on a small development set of 10 examples each of Instantiation and Restatement-Specification, we came up with several features that distinguished between the specific and general sentences in the sample. We observed that in general sentences, strong opinion or sentiment was often expressed, providing some qualification about a person or event. In the general sentences in Table 2, we see for example the phrases "publishing sensation", "very slowly—if at all", "is significant". In a sense, general sentences appear to be more surprising, and evoke in the mind of the reader questions about some missing information or explanation. Specific sentences, on the other hand, are characterized by the use of specific proper names and numbers.

We describe below the features we developed

[1] We use only the *implicit* relations from the PDTB; ie, the sentences are not linked by an explicit discourse connective such as 'because' or 'but' that signals the relation.

Instantiations
[1] *The 40-year-old Mr. Murakami is a publishing sensation in Japan.* A more recent novel, “ Norwegian Wood ” (every Japanese under 40 seems to be fluent in Beatles lyrics), has sold more than four million copies since Kodansha published it in 1987.
[2] *Sales figures of the test-prep materials aren’t known, but their reach into schools is significant.* In Arizona, California, Florida, Louisiana, Maryland, New Jersey, South Carolina and Texas, educators say they are common classroom tools.
[3] *Despite recent declines in yields, investors continue to pour cash into money funds.* Assets of the 400 taxable funds grew by $ 1.5 billion during the last week, to $ 352.7 billion.
Specifications
[4] *By most measures, the nation’s industrial sector is now growing very slowly—if at all.* Factory payrolls fell in September.
[5] *Mrs. Hills said that the U.S. is still concerned about ‘disturbing developments in Turkey and continuing slow progress in Malaysia.*’ She didn’t elaborate, although earlier U.S. trade reports have complained of videocassette piracy in Malaysia and disregard for U.S. pharmaceutical patents in Turkey.
[6] *Alan Spoon, recently named Newsweek president said Newsweek’s ad rates would increase 5% in January.* A full, four-color page in Newsweek will cost $100,980

Table 2: Examples of general (in italics) and specific sentences from the PDTB

based on our observations. Some of our features require syntax information. We compute these using the manual parse annotations for the articles from the Penn Treebank corpus (Marcus et al., 1994).

Sentence length. We expected general sentences to be shorter than the specific ones. So we introduced two features—the number of words in the sentence and the number of nouns.

Polarity. Sentences with strong opinion such as those from examples [1], [4] and [5] are typical in the general category. Therefore, we record for each sentence the number of positive, negative and polar (not neutral) words using two lexicons—The General Inquirer (Stone et al., 1966) and the MPQA Subjectivity Lexicon (Wilson et al., 2005). We also add another set of features where each of these counts is normalized by the sentence length.

Specificity. We use two sets of features to capture specificity of words in the sentence. The first of these is based on WordNet (Miller et al., 1990) and is motivated by prior work by Resnik (1995). Simplifying Resnik (1995)’s approach, we compute a specificity measure using the hypernym relations in WordNet. For each noun and verb in our example sentences, we record the length of the path from the word to the root of the WordNet hierarchy through the hypernym relations. The longer this path, we would expect the words to be more specific. The average, min and max values of these distances are recorded individually for nouns and verbs.

Another set of measures is based on the inverse document frequency (idf) for a word w (Joho and Sanderson, 2007), defined as $\log \frac{N}{n}$. Here N is the number of documents in a large collection, and n is the number of documents that contain the word w. We use articles from one year (87,052 documents) of the New York Times (NYT) corpus (Sandhaus, 2008) to compute idf. Words not seen in the NYT corpus were treated as if they were seen once. The features for a sentence are the average, min and max idfs for words in the sentence.

NE+CD. In news articles, especially the WSJ, specific sentences are often associated with numbers and dollar amounts. So we add as features the count of numbers (identified using the part of speech), proper names and the count of dollar signs. The performance of these features, however, is likely to be genre-dependent. We also introduce another entity-related feature—the number of plural nouns. From our example sentences and the PDTB definition for Instantiation relations, we notice that plural quantities or sets are a property of general sentences.

Language models. We use one year of news articles from the NYT corpus to build unigram, bigram and trigram models. Using each model, we obtain the log probability and perplexity of the sentences to use as features. The unigram language model captures the familiarity of individual words. On the other hand, we expect the perplexity computed using higher order models to distinguish between common word transitions in the domain, and those that are unexpected and evoke surprise for a reader.

Syntax. We also noted frequent usage of qualitative words such as adjectives and adverbs in general sen-

Features	Instantiations	Specifications
NE+CD	68.6	56.1
language models	65.8	55.7
specificity	63.6	57.2
syntax	63.3	57.3
polarity	63.0	53.4
sentence length	54.0	57.2
all non-lexical	75.0	62.0
lexical (words)	74.8	59.1
all features	75.9	59.5

Table 3: Accuracy of differentiating general/specific sentences (baseline 50%)

tences. We therefore include some syntax based features: counts of the number of adjectives, adverbs, adjective phrases and adverbial phrases. We also record the number of verb phrases and their average length in words and the number of prepositional phrases. We expect that longer verb phrases would be associated with more specific sentences.

Words. We also add the count of each word in the sentence as a feature.

2.2 Results

We build two classifiers for distinguishing general and specific sentences: one trained on sentences from Instantiation relations, and one on sentences from Restatement-Specification. The first sentence in the relation was considered an example of general sentence, and the second of specific one. No pairing information was preserved or exploited.

We train a logistic regression classifier[2] with each set of features described above and evaluate the predictions using 10-fold cross validation. To also examine the effect of non-lexical features by themselves, we trained a classifier with all the other features except words. There are equal number of positive and negative examples, so the baseline random accuracy is 50%. The accuracy of predictions is shown in Table 3.

The accuracy of classifiers trained on Instantiation examples are promising and clearly better than those trained on Specifications. The highest accuracy on Instantiation-based classifier comes from combining all features, reaching 75.9% which is more than 25% absolute improvement over the baseline. The individually best class of features are the words with 74.8% accuracy, only 1% less than that obtained with all features. Combining all non-lexical features results in classification accuracies which equal the performance of word features.

Among the non-lexical features, the NE+CD class of features are the strongest individual set of predictors achieving an accuracy of 68%. Language models, syntax, polarity and WordNet/idf features are also good predictors, each outperforming the baseline by a margin of over 10% accuracy. The least indicative features are the sentence length features.

For the Specifications-based classifier, the highest performance is barely 10% above baseline. The best accuracy is obtained with a combination of all non-lexical features, 62%. In contrast to the Instantiations case, language models and entities features sets are less accurate in making the general-specific distinction on the Specification examples. Polarity is the worst set of features with only 53% accuracy, very close to random baseline performance.

A possible explanation of the difference in results from different types of training data is that in Restatement-Specification sentences, the specificity of the second sentence is only relative to that of the first. On the other hand, for Instantiation relations, there are individual characteristics related to the generality or specificity of sentences.

2.3 Feature analysis

In this section, we take a closer look at the features that most successfully distinguished specific and general sentences on the *Instantiation* dataset. Given that words were the most predictive feature class, we identified those with highest weight in the logistic regression model. Here we list those that appeared at least in 25 training examples, predictive of the two types of sentences.

General number, but, also, however, officials, some, what, prices, made, lot, business, were

Specific one, a, to, co, i, called, we, could, get, and, first, inc

Discourse connectives such as ‘but’, ‘also’ and ‘however’ show up as top indicators for general sentences. Typical for general sentences are also some vague words, such as ‘some’ and ‘lot’. Words indicative of specific sentences are ‘a’, ‘one’ and pronouns. However, a large number of other words

[2] http://www.csie.ntu.edu.tw/ cjlin/liblinear/

appear to be domain specific indicators—'officials', 'number', 'prices' and 'business' for general sentences, and 'co.', 'inc' for the specific category.

Analysis of the weights of non-lexical features showed that they conformed to our intuitions. The best performing feature set—mentions of numbers and names—are predictive of specific sentences. Plural nouns as we had expected is a property of general sentences. However, the dollar sign, which we expected is more likely with specific sentences turned out to be more frequent in the other category. As for the language model features, general sentences tended to have lower probability and higher perplexity than specific ones.

For syntax and sentiment features, we found that general sentences have greater counts of polarity words (normalized by length) and higher number of adjectives and adverbs and their phrases. At the same time, these sentences have fewer and shorter verb phrases and fewer prepositional phrases.

Next we turn to analysis of the distribution of general and specific sentences—as predicted by our classifier— in news, and summaries of news.

3 Specificity of news articles and their summaries

While it has not been known till now what proportion of summary content tends to be general or specific, the fact that there is a choice between these two types of sentences for summaries has been observed in prior work. Early work in Jing and McKeown (2000) analyzed sentences from abstracts written by people and the input sentences containing the same content and categorized the edits that humans are using for converting sentences from the input to those that they use in abstracts. They identified that two of the transformations that people employ are generalization and specification.

From the point of view of automatic systems, in recent work, Haghighi and Vanderwende (2009) developed a topic model-based summarization system which learns the topics of the input at both overall document level as well as specific subtopics. Sentences are assumed to be generated by a combination of the general and specific topics associated with the documents. However, since the preference of such sentences is not known, only heuristics were applied to choose the proportions.

Below we present the first study of levels of specificity in input texts, human and machine written summaries. Further, we study two types of human summaries—abstracts and extracts to compare the effect of the summarization method. Most importantly, an analysis of system summaries is performed to understand how the sentence selections done by automatic content selection methods compare with content selected in human summaries.

3.1 Data

We obtained news documents and their summaries from the Document Understanding Conference evaluations conducted by NIST. We use the data from 2002 for our experiments because they contain the three different types of summaries we wish to analyze—abstracts and extracts produced by people, and automatic summaries. For extracts, the person could only select complete sentences, without any modification, from the input articles.

We use data from the generic multi-document summarization task. There were 59 input sets, each containing 5 to 15 news documents on a topic. Two human-written abstracts and two extracts are available for each input. 9 automatic systems participated that year, so we have 524 automatic summaries overall. All the summaries are 200 words in length.

3.2 General and specific sentences in news

We ran our classifier (trained on Instantiation examples with all features) on the inputs and summaries from DUC. The syntactic features and part of speech tags for the summarization data were obtained using the Stanford Parser (Klein and Manning, 2003).

In Table 4, we show some example predictions. We list the three general and three specific sentences predicted with highest confidence for two inputs (d099e and d081a). Input d099e contains articles about different marathon events. The examples show that some of the sentences in these articles can indeed be very specific providing a lot of details. Others are very general, such as the sentence with the phrase 'no mystery men out on the course'.

Input d081a is a collection of articles about strikes by coal miners in different parts of the world. Here again some of the sentences are very general, 'the number appeared to be rising' and 'a number of factors'. On the opposite end, specific sentences from same topic give details about the area, time and duration of the strikes.

In the next section, we analyze the actual distribution of specific and general content in articles and their summaries for the entire DUC 2002 dataset.

3.3 Specificity analysis

For each text (input, human abstract, human extract, automatic summary), we computed the percentage of words in the text that appear in sentences predicted to be specific. The histogram for the proportion of specific content in each type of text is shown in Figure 1.

For inputs, the percentage of specific content is around 50 to 70% with a mean value of 63%. So, the majority of news articles have more specific content but the distribution is not very highly skewed.

The remaining three graphs in Figure 1 are representative of the amount of specific content in summaries for the same inputs. Human abstracts, in contrast to the inputs, are spread over a wider range of specificity levels. Some abstracts have as low as 20 to 40% specific content and others 100%. However, the number of texts around the 80% specificity level is much larger compared to the inputs and mean value increases to 67% specificity. Also, an unpaired two-sided t-test between the specificity values of inputs and the abstracts confirmed that abstracts have significantly higher specificity. The results of the test are similar for human extracts and system summaries; summaries regardless of type have more specific content compared to the inputs. However, there are striking differences when we compare the different types of summaries with each other.

Firstly, human extracts are vastly different from abstracts. Their specificity levels show a considerable shift towards the higher end. The mean specificity value for human extracts is 77%, 10% higher compared to abstractive summaries for the same inputs. Especially noticeable is the high proportion of extracts with 100% specificity. Even though, humans have created both types of summaries, we can see that the summarization method deeply influences the nature of the summary content. We also validated this finding using a t-test between the values; the higher specificity levels in human extracts compared to abstracts is statistically significant (pvalue is less than e-16).

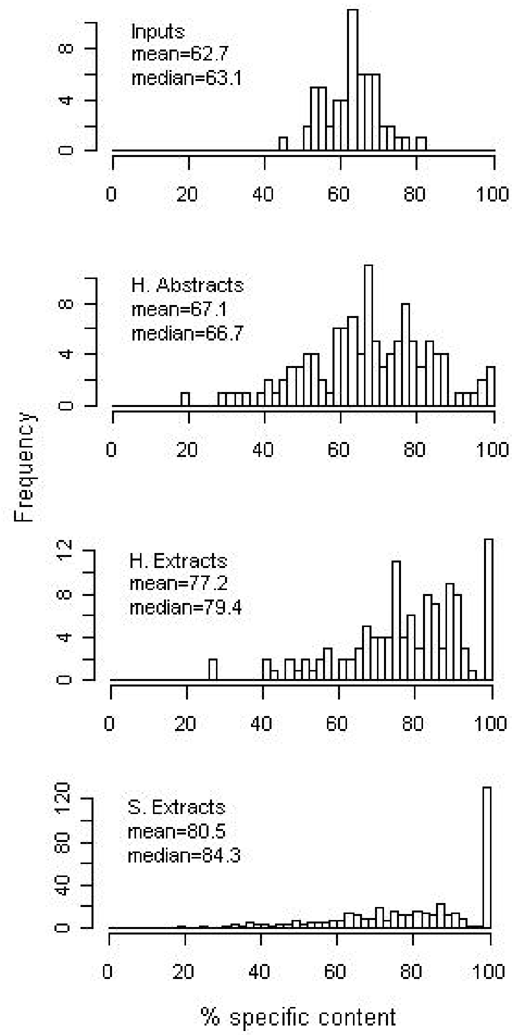

Figure 1: Proportion of specific content in inputs and summaries (H: human-written, S: system-produced)

The most surprising result for the texts we examined is the distribution of system summary specificity levels. System-produced summaries are much more specific compared to both types of human summaries. The vast majority of system summaries are concentrated at the 100% specificity mark and the mean specificity value is 80%. A t-test showed that the mean of system summary specificities is higher compared to human abstracts and inputs with very high significance (pvalues less than e-13). Their values are also higher than those in human extracts, however, this difference with extracts is not significant (pvalue of 0.0602).

In sum, we find that human summaries, while being more specific than the input, still contain a mix of general and specific sentences. Systems are heav-

d099e

[G1] This year, officials made sure there were no mystery men out on the course.

[G2] Not so, said Rosas.

[G3] The 82-year-old Mrs. Kelley adds, "I just try to stay in his shadow, but I have a hard time even doing that".

[S1] Among those chasing them will be John Treacy, the 1984 Olympic silver medalist and the third-place finisher in Boston in 1988 and 1989; Geoff Smith, the Boston winner in 1984-85; Ed Eyestone, the top-ranked U.S. marathoner; Salvador Garcia, the runner-up to Wakiihuri in last year's New York City Marathon; and Rolando Vera, who finished third in his marathon debut in Boston in 1990.

[S2] They include Ibrahim Hussein of Kenya, the 1988 winner in 2:08:43, the third-fastest time in Boston and one second ahead of Ikangaa; Abebe Mekonnen of Ethiopia, the 1989 champion in 2:09:06, Boston's eighth-best time; Douglas Wakiihuri of Kenya, the 1987 world champion and 1988 Olympic silver medalist who is making his Boston debut; and Naali, the third-place finisher in the 1990 Commonwealth Games.

[S3] Foremost among them are Kristiansen, the 35-year-old Norwegian who holds the world record of 2:21:06 and won in Boston in 1986 and 1989, and Samuelson, 33, the U.S.-record holder at 2:21:21, 1984 Olympic gold medalist and Boston champion in 1979 and 1983.

d081a

[G1] Recent reports have put the total number of strikers in the Donetsk and Kuznetsk regions at more than 112,000, but the number appeared to be rising.

[G2] That mine has also been the scene of unrest in recent months, where workers have demanded improved safety conditions.

[G3] The strike seems motivated by a number of factors.

[S1] According to the state news agency Tass, the strike has spread to the Donetsk coal basin of the Ukraine, the Vorkuta and Inta regions of northern Russia, and the Kuznetsk coal basin of western Siberia.

[S2] One United Mine Workers representative said today that the Indiana walkout was prompted by Thursday night's telecast of CBS' "48 Hours," which focused on the Pittston strike, in which 1,600 miners have been off the job since April.

[S3] "It's a victory over the system that we've had in the Soviet Union for the last 70 years, a system in which we work hard but" "get little in return," said Pyotr A. Menayev, an engineer at the Taldinski Severny open pit mine on the outskirts of Prokopyevsk.

Table 4: General and specific sentences from different inputs

ily biased toward specific content.

3.4 General sentences in human summaries

To gain some insights of the function of general sentences in summaries, we semi-automatically categorized the sentences from *human extracts* that were predicted as general with highest confidence. We used 80% as the confidence cut-off, and obtained 107 sentences for the analysis.

In Table 5, we compare some of the properties of general sentences in extracts with high confidence general sentences in the inputs and abstracts.

General sentences are often found in the beginning and ending of human extracts. 20% of all high confidence general sentences in extracts are used as the last sentences. This number is considerably higher than 4.2% general sentences from the input and 9.7% from abstracts that end a document or summary. Therefore one property of general sentences appears to be that they can be used as topic-setting opening sentences or as closing sentences that summarize the importance of the summary content. Some examples are shown in Table 6.

Another property we observed among the general sentences in extracts is the presence of direct and indirect quotations. Firstly, in news articles, quotations are used to report experiences and opinions rather than factual content and so quotations being tagged as general seems to be appropriate. A simple method of counting sentences that contain the words 'said', 'say' or 'says' shows that one-fifth of all the high confidence general sentences that appear in human extracts are quotes.

Other forms of opinion sentences also exist among those predicted to be general. We have not quantified these, however, some examples are provided in Table 6.

As we already observed in our analysis of top word features, sentences with connectives appear to be general sentences. By checking for sentences with the tokens 'but', 'although', 'however', 'though' and 'still', we identified that 23% of the general sentences in human extracts are comparisons. This fraction equals the percentage of general sentences in inputs and abstracts that are compar-

Property	**Inputs**	**H. Abs.**	**H. Ext.**
(Gen. sent.)	(2630)	(235)	(107)
First sent.	36 (1.3%)	9 (3.8%)	8 (7.4%)
Last sent.	111 (4.2%)	23 (9.7%)	22 (20.5%)
Comparison	589 (22.3%)	58 (24.6%)	25 (23.3%)
Is a quote	470 (17.8%)	7 (2.9%)	23 (21.4%)

Table 5: Analysis of high confidence general sentences in inputs and human summaries. The number of such sentences in each type of text is shown in the first line.

isons. This trend shows that general sentences that convey discourse relations may be used by humans in a constant proportion in inputs and summaries.

Our crude classification above reflects some of the cases where humans employ general sentences in their extracts. Overall, we believe that these choices are made to keep the content of extracts simple with a good mix of both types of sentences. But current systems do not appear to consider this aspect.

Another noteworthy fact is the significantly higher level of general content in human abstracts compared to extracts. Simply the nature of creating extractive summaries biases summary content to be much more specific than if the summary was produced as an abstract. This result emphasizes the need to employ generation techniques for summarization. With extractive methods, the flexibility of choosing content is always likely to be limited.

In fact, several recent efforts have focused on compression (Knight and Marcu, 2002; McDonald, 2006; Galley and McKeown, 2007) of the extracted sentences. Here, our experiments point out that specific to general transformations would be very frequent because of the more general nature of abstractive summaries. In future work, we plan to use our classifier to obtain examples of sentence mappings from extracts to abstracts, where there is a high degree of generalization in the abstract sentence. We believe that such sentences are likely to be some of the challenging cases for compression tasks. By identifying such sentences, we can study their aspects and develop techniques to expand the capability of current compression methods.

Topic-setting opening sentences:
- Lucille Ball, whose death Wednesday morning at the age of 77 will be the most widely and deeply felt show-business loss in recent memory, was the movies' greatest gift to television.

Quotes from relevant sources:
- Defense Secretary Tom King, inspecting the wreckage, said, 'It is not yet absolutely confirmed that it is a bomb, but all the evidence is quite clearly that this is an IRA atrocity.'
- Mr. Moynihan's point, in other words, is that these territories are up for grabs.

Comparison relations:
- U.S. announces suspension of all military sales and visits of Chinese military leaders, but stops short of severing diplomatic ties.
- However, Weisfeldt, who is president-elect of the heart association also cautioned that much work remains, both in improving care and encouraging people to take better care of their health.

Opinion sentences:
- What the relays demonstrate more clearly than any other athletics event is that at the Olympics the race has traditionally gone to affluent countries or to those, rather, that choose to lavish resources on athletes.
- But the people of the Soviet Union apparently concluded that the real issue was not Gorbachev himself, but the new political culture of freedom and even democracy that his reforms had created.

Closing sentences that signal impact:
- Such heavy losses suggest that the storm will be the most expensive ever for the insurance industry.
- The Booker prize has, in its 26-year history, always provoked controversy.

Table 6: Example general sentences in humans extracts

4 Conclusion

We have presented experiments on automatic identification of general and specific sentences in texts. We have used discourse relation annotations to build our training and test sets and have shown that a high accuracy classifier can be obtained using Instantiation sentences.

In addition, we have examined the predictions of the classifier on the full texts of news articles and their corresponding summaries. In particular, we compared the degree of specific content in news articles and summaries, thereby showing that summaries tend to have more specific content compared to the inputs. Further, we found that system summaries have very specific content in contrast to human summaries which have a mix of general and specific sentences. We have provided an analysis of sentence choices which can bring system summaries closer to humans. The development of generation and compression techniques to produce such general sentences would prove very useful for system development. Our study has been the first empirical analysis to quantify the specific content in human and machine-produced texts.

References

S. Aman and S. Szpakowicz. 2007. Identifying expressions of emotion in text. In *Text, Speech and Dialogue*, pages 196–205. Springer.

J. Birke and A. Sarkar. 2006. A clustering approach for nearly unsupervised recognition of nonliteral language. In *Proceedings of EACL*.

M. Galley and K. McKeown. 2007. Lexicalized markov grammars for sentence compression. In *Proceedings NAACL-HLT*.

A. Haghighi and L. Vanderwende. 2009. Exploring content models for multi-document summarization. In *Proceedings of NAACL-HLT*, pages 362–370.

H. Jing and K. McKeown. 2000. Cut and paste based text summarization. In *Proceedings of NAACL*.

H. Joho and M. Sanderson. 2007. Document frequency and term specificity. In *Proceedings of RIAO*.

S.M. Kim and E. Hovy. 2004. Determining the sentiment of opinions. In *Proceedings of COLING*, pages 1367–1373.

D. Klein and C.D. Manning. 2003. Accurate unlexicalized parsing. In *Proceedings of ACL*, pages 423–430.

K. Knight and D. Marcu. 2002. Summarization beyond sentence extraction: A probabilistic approach to sentence compression. *Artificial Intelligence*, 139(1).

D. Marcu and A. Echihabi. 2001. An unsupervised approach to recognizing discourse relations. In *Proceedings of ACL*, pages 368–375.

M.P. Marcus, B. Santorini, and M.A. Marcinkiewicz. 1994. Building a large annotated corpus of english: The penn treebank. *Computational Linguistics*, 19(2):313–330.

R. McDonald. 2006. Discriminative sentence compression with soft syntactic constraints. In *Proceedings of EACL*.

G.A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. J. Miller. 1990. Introduction to wordnet: An on-line lexical database. *International Journal of Lexicography (special issue)*, 3(4):235–312.

R. Prasad, N. Dinesh, A. Lee, E. Miltsakaki, L. Robaldo, A. Joshi, and B. Webber. 2008. The penn discourse treebank 2.0. In *Proceedings of LREC*.

P. Resnik. 1995. Using information content to evaluate semantic similarity in a taxonomy. In *Proceedings of IJCAI*, pages 448–453.

E. Sandhaus. 2008. The new york times annotated corpus. *Corpus number LDC2008T19, Linguistic Data Consortium, Philadelphia*.

C. Sporleder and A. Lascarides. 2008. Using automatically labelled examples to classify rhetorical relations: An assessment. *Natural Language Engineering*, 14:369–416.

P.J. Stone, J. Kirsh, and Cambridge Computer Associates. 1966. *The General Inquirer: A Computer Approach to Content Analysis*. MIT Press.

J. M. Swales and C. Feak. 1994. *Academic writing for graduate students: A course for non-native speakers of English*. Ann Arbor: University of Michigan Press.

J. Wiebe and E. Riloff. 2005. Creating subjective and objective sentence classifiers from unannotated texts. *Computational Linguistics and Intelligent Text Processing*, pages 486–497.

T. Wilson, J. Wiebe, and P. Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of HLT-EMNLP*, pages 347–354.

H. Yu and V. Hatzivassiloglou. 2003. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of EMNLP*, pages 129–136.