# Automatically Learning Cognitive Status for Multi-Document Summarization of Newswire

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

## Abstract

Machine summaries can be improved by using knowledge about the cognitive status of news article referents. In this paper, we present an approach to automatically acquiring distinctions in cognitive status using machine learning over the forms of referring expressions appearing in the input. We focus on modeling references to people, both because news often revolve around people and because existing natural language tools for named entity identification are reliable. We examine two specific distinctions—whether a person in the news can be assumed to be known to a target audience (hearer-old vs hearer-new) and whether a person is a major character in the news story. We report on machine learning experiments that show that these distinctions can be learned with high accuracy, and validate our approach using human subjects.

## 1 Introduction

Multi-document summarization has been an active area of research over the past decade (Mani and Maybury, 1999) and yet, barring a few exceptions (Daumé III et al., 2002; Radev and McKeown, 1998), most systems still use shallow features to produce an extractive summary, an age-old technique (Luhn, 1958) that has well-known problems. Extractive summaries contain phrases that the reader cannot understand out of context (Paice, 1990) and irrelevant phrases that happen to occur in a relevant sentence (Knight and Marcu, 2000; Barzilay, 2003). Referring expressions in extractive summaries illustrate this problem, as sentences compiled from different documents might contain too little, too much or repeated information about the referent.

Whether a referring expression is appropriate depends on the location of the referent in the hearer's mental model of the discourse—the referent's *cognitive status* (Gundel et al., 1993). If, for example, the referent is unknown to the reader at the point of mention in the discourse, the reference should include a description, while if the referent was known to the reader, no descriptive details are necessary.

Determining a referent's cognitive status, however, implies the need to model the intended audience of the summary. Can such a cognitive status model be inferred automatically for a general readership? In this paper, we address this question by performing a study with human subjects to confirm that reasonable agreement on the distinctions can be achieved between different humans (cf. §5). We present an automatic approach for inferring what the typical reader is likely to know about people in the news. Our approach uses machine learning, exploiting features based on the form of references to people in the input news articles (cf. §4). Learning cognitive status of referents is necessary if we want to ultimately generate new, more appropriate references for news summaries.

### 1.1 Cognitive status

In human communication, the wording used by speakers to refer to a discourse entity depends on their *communicative goal* and their beliefs about *what listeners already know*. The speaker's goals and beliefs about the listener's knowledge are both a part of a cognitive/mental model of the discourse.



*Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP)*, pages 241–248, Vancouver, October 2005. 

Cognitive status distinctions depend on two parameters related to the referent—*a)* whether it already exists in the hearer's model of the discourse, and *b)* its degree of salience. The influence of these distinctions on the form of referring expressions has been investigated in the past. For example, centering theory (Grosz et al., 1995) deals predominantly with local salience (local attentional status), and the givenness hierarchy (information status) of Prince (1992) focuses on how a referent got in the discourse model (e.g. through a direct mention in the current discourse, through previous knowledge, or through inference), leading to distinctions such as discourse-old, discourse-new, hearer-old, hearer-new, inferable and containing inferable. Gundel et al. (1993) attempt to merge salience and givenness in a single hierarchy consisting of six distinctions in cognitive status (in focus, activated, familiar, uniquely identifiable, referential, type-identifiable).

Among the distinctions that have an impact on the form of references in a summary are the *familiarity* of the referent:

**D.** Discourse-old vs discourse-new

**H.** Hearer-old vs hearer-new

and its global salience[1]:

**M.** Major vs minor

In general, initial (discourse-new) references to entities are longer and more descriptive, while subsequent (discourse-old) references are shorter and have a purely referential function. Nenkova and McKeown (2003) have studied this distinction for references to people in summaries and how it can be used to automatically rewrite summaries to achieve better fluency and readability.

The other two cognitive status distinctions, whether an entity is central to the summary or not (major or minor) and whether the hearer can be assumed to be already familiar with the entity (hearer-old vs hearer-new status), have not been previously studied in the context of summarization. There is a tradeoff, particularly important for a short summary, between what the speaker wants to convey and how much the listener needs to know. The hearer-old/new distinction can be used to determine whether a description for a character is required from the listener's perspective. The major/minor distinction plays a role in defining the communicative goal, such as what the summary should be about and which characters are important enough to refer to by name.

[1]The notion of global salience is very important to summarization, both during content selection and during generation on initial references to entities. On the other hand, *in focus* or *local attentional state* are relevant to anaphoric usage during subsequent mentions.

### 1.2 Hearer-Old vs Hearer-New

Hearer-new entities in a summary should be described in necessary detail, while hearer-old entities do not require an introductory description. This distinction can have a significant impact on overall length and intelligibility of the produced summaries. Usually, summaries are very short, 100 or 200 words, for input articles totaling 5,000 words or more. Several people might be involved in a story, which means that if all participants are fully described, little space will be devoted to actual news. In addition, introducing already familiar entities might distract the reader from the main story (Grice, 1975). It is thus a good strategy to refer to an entity that can be assumed hearer-old by just a title + last name, e.g. *President Bush*, or by full name only, with no accompanying description, e.g. *Michael Jackson*.

### 1.3 Major vs Minor

Another distinction that human summarizers make is whether a character in a story is a major or a minor one and this distinction can be conveyed by using different forms of referring expressions. It is common to see in human summaries references such as *the dissident's father*. Usually, discourse-initial references solely by common noun, without the inclusion of the person's name, are employed when the person is not the main focus of a story (Sanford et al., 1988). By detecting the cognitive status of a character, we can decide whether to name the character in the summary. Furthermore, many summarization systems use the presence of named entities as a feature for computing the importance of a sentence (Saggion and Gaizaukas, 2004; Guo et al., 2003). The ability to identify the major story characters and use only them for sentence weighting can benefit such systems since only 5% of all people mentioned in the input are also mentioned in the summaries.

## 2 Why care about people in the news?

News reports (and consequently, news summaries) tend to have frequent references to people (in DUC data - see §3 for description - from 2003 and 2004, there were on average 3.85 references to people per 100-word human summary); hence it is important for news summarization systems to have a way of modeling the cognitive status of such referents and a theory for referring to people.

It is also important to note that there are differences in references to people between news reports and human summaries of news. Journalistic conventions for many mainstream newspapers dictate that initial mentions to people include a minimum description such as their role or title and affiliation. However, in human summaries, where there are greater space constraints, the nature of initial references changes. Siddharthan et al. (2004) observed that in DUC'04 and DUC'03 data[2], news reports contain on average one appositive phrase or relative clause every 3.9 sentences, while the human summaries contain only one per 8.9 sentences on average. In addition to this, we observe from the same data that the average length of a first reference to a named entity is 4.5 words in the news reports and only 3.6 words in human summaries. These statistics imply that human summarizers do compress references, and thus can save space in the summary for presenting information about the events. Cognitive status models can inform a system when such reference compression is appropriate.

## 3 Data preparation: the DUC corpus

The data we used to train classifiers for these two distinctions is the Document Understanding Conference collection (2001–2004) of 170 pairs of document input sets and the corresponding human-written multi-document summaries (2 or 4 per set). Our aim is to identify every person mentioned in the 10 news reports and the associated human summaries for each set, and assign labels for their cognitive status (hearer old/new and major/minor). To do this, we first preprocess the data (§3.1) and then perform the labeling (§3.2).

[2]The data provided under DUC for these years includes sets of about 10 news reports, 4 human summaries for each set, and the summaries by participating machine summarizers.

### 3.1 Automatic preprocessing

All documents and summaries were tagged with BBN's IDENTIFINDER (Bikel et al., 1999) for named entities, and with a part-of-speech tagger and simplex noun-phrase chunker (Grover et al., 2000). In addition, for each named entity, relative clauses, appositional phrases and copula constructs, as well as pronominal co-reference were also automatically annotated (Siddharthan, 2003). We thus obtained coreference information (cf. Figure 1) for each person in each set, across documents and summaries.

| **Andrei Sakharov** | |
|---|---|
| *Doc 1:* | [IR] laureate Andrei D. Sakharov [CO] Sakharov [CO] Sakharov [CO] Sakharov [CO] Sakharov [PR] his [CO] Sakharov [PR] his [CO] Sakharov [RC] who acted as an unofficial Kremlin envoy to the troubled Transcaucasian region last month [PR] he [PR] He [CO] Sakharov |
| *Doc 1:* | [IR] Andrei Sakharov [AP] , 68 , a Nobel Peace Prize winner and a human rights activist , [CO] Sakharov [IS] a physicist [PR] his [CO] Sakharov |

Figure 1: Example information collected for *Andrei Sakharov* from two news report. 'IR' stands for 'initial reference', 'CO' for noun co-reference, 'PR' for pronoun reference, 'AP' for apposition, 'RC' for relative clause and 'IS' for copula constructs.

The tools that we used were originally developed for processing single documents and we had to adapt them for use in a multi-document setting. The goal was to find, for each person mentioned in an input set, the list of all references to the person in both input documents and human summaries. For this purpose, all input documents were concatenated and processed with IDENTIFINDER. This was then automatically post-processed to mark-up coreferring names and to assign a unique canonical name (unique id) for each name coreference chain. For the coreference, a simple rule of matching the last name was used, and the canonical name was the "First-Name LastName" string where the two parts of the name could be identified [3]. Concatenating all documents assures that the same canonical name will be assigned to all named references to the same person.

[3]Occasionally, two or more different people with the same last name are discussed in the same set and this algorithm would lead to errors in such cases. We did keep a list of first names associated with the entity, so a more refined matching model could be developed, but this was not the focus of this work.

The tools for pronoun coreference and clause and apposition identification and attachment were run separately on each document. Then the last name of each of the canonical names derived from the IDENTIFINDER output was matched with the initial reference in the generic coreference list for the document with the last name. The tools that we used have been evaluated separately when used in normal single document setting. In our cross-document matching processes, we could incur more errors, for example when the general coreference chain is not accurate. On average, out of 27 unique people per cluster identified by IDENTIFINDER, 4 people and the information about them are lost in the matching step for a variety of reasons such as errors in the clause identifier, or the coreference.

### 3.2 Data labeling

Entities were automatically labeled as hearer-old or new by analyzing the syntactic form that human summarizers used for initial references to them. The labeling rests on the assumption that the people who produced the summaries used their own model of the reader when choosing appropriate references for the summary. The following instructions had been given to the human summarizers, who were not professional journalists: "To write this summary, assume you have been given a set of stories on a news topic and that your job is to summarize them for the general news sections of the Washington Post. Your audience is the educated adult American reader with varied interests and background in current and recent events." Thus, the human summarizers were given the freedom to use their assumptions about what entities would be generally hearer-old and they could refer to these entities using short forms such as (1) title or role+ last name or (2) full name only with no pre- or post-modification. Entities that the majority of human summarizers for the set referred to using form (1) or (2) were labeled as hearer-old. From the people mentioned in human summaries, we obtained 118 examples of hearer-old and 140 examples of hearer-new persons - 258 examples in total - for supervised machine learning.

In order to label an entity as major or minor, we again used the human summaries—entities that were mentioned *by name* in at least one summary were labeled *major*, while those not mentioned by name in any summary were labeled *minor*. The underlying assumption is that people who are not mentioned in any human summary, or are mentioned without being named, are not important. There were 258 major characters who made it to a human summary and 3926 minor ones that only appeared in the news reports. Such distribution between the two classes is intuitively plausible, since many people in news articles express opinions, make statements or are in some other way indirectly related to the story, while there are only a few main characters.

## 4 Machine learning experiments

The distinction between hearer-old and hearer-new entities depends on the readers. In other words, we are attempting to automatically infer which characters would be hearer-old *for the intended readership of the original reports*, which is also expected to be the intended readership of the summaries. For our experiments, we used the WEKA (Witten and Frank, 2005) machine learning toolkit and obtained the best results for hearer-old/new using a support vector machine (SMO algorithm) and for major/minor, a tree-based classifier (J48). We used WEKA's default settings for both algorithms.

We now discuss what features we used for our two classification tasks (cf. list of features in table 1). Our hypothesis is that features capturing the frequency and syntactic and lexical forms of references are sufficient to infer the desired cognitive model.

Intuitively, pronominalization indicates that an entity was particularly salient at a specific point of the discourse, as has been widely discussed in attentional status and centering literature (Grosz and Sidner, 1986; Gordon et al., 1993). Modified noun phrases (with apposition, relative clauses or premodification) can also signal different status.

In addition to the syntactic form features, we used two months worth of news articles collected over the web (and independent of the DUC collection we use in our experiments here) to collect unigram and bigram lexical models of first mentions of people. The names themselves were removed from the first mention noun phrase and the counts were collected over the premodifiers only. One of the lexical features we used is whether a person's description contains any of the 20 most frequent description words from our web corpus. We reasoned that these frequent de-

| | | | |
|---|---|---|---|
| 0,1: | Number of references to the person, including pronouns (total and normalized by feature 16) | 2,3: | Number of times apposition was used to describe the person(total and normalized by feature 16) |
| 4,5: | Number of times a relative clause was used to describe the person (total and normalized by 16) | 6: | Number of times the entity was referred to by name after the first reference |
| 7,8: | Number of copula constructions involving the person (total and normalized by feature 16) | 9,10: | Number of apposition, relative clause or copula descriptions (total and normalized by feature 16) |
| 11,12,13: | Probability of an initial reference according to the bigram model (av.,max and min of all initial references) | 14: | Number of top 20 high frequency description words (from references to people in large news corpus) present in initial references |
| 15: | Proportion of first references containing full name | 16: | Total number of documents containing the person |
| 17,18: | Number of appositives or relative clause attaching to initial references (total and normalized by feature 16) | | |

Table 1: List of Features provided to WEKA.

scriptors may signal importance; the full list is:

*president, former, spokesman, sen, dr, chief, coach, attorney, minister, director, gov, rep, leader, secretary, rev, judge, US, general, manager, chairman.*

Another lexical feature was the overall likelihood of a person's description using the bigram model from our web corpus. This indicates whether a person has a role or affiliation that is frequently mentioned. We performed 20-fold cross validation for both classification tasks. The results are shown in Table 2 (accuracy) and Table 3 (precision/recall).

### 4.1 Major vs Minor results

For major/minor classification, the majority class prediction has 94% accuracy, but is not a useful baseline as it predicts that *no* person should be mentioned by name and all are minor characters. J48 correctly predicts 114 major characters out of 258 in the 170 document sets. As recall appeared low, we further analyzed the 148 persons from DUC'03 and DUC'04 sets, for which DUC provides four human summaries. Table 4 presents the distribution of recall taking into account *how many* humans mentioned the person by name in their summary (originally, entities were labeled as main if *any* summary had a reference to them, cf. §3.2). It can be seen that recall is high (0.84) when all four humans consider a character to be major, and falls to 0.2 when only one out of four humans does. These observations reflect the well-known fact that humans differ in their choices for content selection, and indicate that in the automatic learning is more successful when there is more human agreement.

In our data there were 258 people mentioned by name in at least one human summary. In addition, there were 103 people who were mentioned in at least one human summary using only a common noun reference (these were identified by hand, as common noun coreference cannot be performed reliably enough by automatic means), indicating that 29% of people mentioned in human summaries are not actually named. Examples of such references include *an off duty black policeman*, *a Nigerian born Roman catholic priest*, *Kuwait's US ambassador*. For the purpose of generating references in a summary, it is important to evaluate how many of these people are correctly classified as minor characters. We removed these people from the training data and kept them as a test set. WEKA achieved a testing accuracy of 74% on these 103 test examples. But as discussed before, different human summarizers sometimes made different decisions on the form of reference to use. Out of the 103 referent for which a non-named reference was used by a summarizer, there were 40 where other summarizers used named reference. Only 22 of these 40 were labeled as minor characters in our automatic procedure. Out of the 63 people who were not named in *any* summary, but mentioned in at least one by common noun reference, WEKA correctly predicted 58 (92%) as minor characters. As before, we observe that when human summarizers generate references of the same form (reflecting consensus on conveying the perceived importance of the character), the machine predictions are accurate.

We performed feature selection to identify which are the most important features for the classification task. For the major/minor classification, the important features used by the classifier were the number of documents the person was mentioned in (feature 16), number of mentions within the document set (features 1,6), number of relative clauses (feature

4,5) and copula (feature 8) constructs, total number of apposition, relative clauses and copula (feature 9), number of high frequency premodifiers (feature 14) and the maximum bigram probability (feature 12). It was interesting that presence of apposition did not select for either major or minor class. It is not surprising that the frequency of mention within and across documents were significant features—a frequently mentioned entity will naturally be considered important for the news report. Interestingly, the syntactic form of the references was also a significant indicator, suggesting that the centrality of the character was signaled by the journalists by using specific syntactic constructs in the references.

| | Major/Minor | Hearer New/Old |
|---|---|---|
| WEKA | 0.96 (J48) | 0.76 (SMO) |
| Majority class prediction | 0.94 | 0.54 |

Table 2: Cross validation *testing* accuracy results.

| | Class | Precision | Recall | F-measure |
|---|---|---|---|---|
| SMO | hearer-new | 0.84 | 0.68 | 0.75 |
| | hearer-old | 0.69 | 0.85 | 0.76 |
| J48 | major-character | 0.85 | 0.44 | 0.58 |
| | minor-character | 0.96 | 0.99 | 0.98 |

Table 3: Cross validation *testing* P/R/F results.

| Number of summaries containing the person | Number of examples | Number and % recalled by J48 |
|---|---|---|
| 1 out of 4 | 59 | 15 (20%) |
| 2 out of 4 | 35 | 20 (57%) |
| 3 out of 4 | 29 | 23 (79%) |
| 4 out of 4 | 25 | 21 (84%) |

Table 4: J48 Recall results and human agreement.

### 4.2 Hearer Old vs New Results

The majority class prediction for the hearer-old/new classification task is that no one is known to the reader and it leads to overall classification accuracy of 54%. Using this prediction in a summarizer would result in excessive detail in referring expressions and a consequent reduction in space available to summarize the news events. The SMO prediction outperformed the baseline accuracy by 22% and is more meaningful for real tasks.

For the hearer-old/new classification, the feature selection step chose the following features: the number of appositions (features 2,3) and relative clauses (feature 5), number of mentions within the document set (features 0,1), total number of apposition, relative clauses and copula (feature 10), number of high frequency premodifiers (feature 14) and the minimum bigram probability (feature 13). As in the minor-major classification, the syntactic choices for reference realization were useful features.

We conducted an additional experiment to see how the hearer old/new status impacts the use of apposition or relative clauses for elaboration in references produced in human summaries. It has been observed (Siddharthan et al., 2004) that on average these constructs occur 2.3 times *less* frequently in human summaries than in machine summaries. As we show, the use of postmodification to elaborate relates to the hearer-old/new distinction.

To determine when an appositive or relative clause can be used to modify a reference, we considered the 151 examples out of 258 where there was at least one relative clause or apposition describing the person in the input. We labeled an example as positive if *at least* one human summary contained an apposition or relative clause for that person and negative otherwise. There were 66 positive and 85 negative examples. This data was interesting because while for the majority of examples (56%) all the human summarizers agreed not to use postmodification, there were very few examples (under 5%) where all the humans agreed to postmodify. Thus it appears that for around half the cases, it should be obvious that no postmodification is required, but for the other half, human decisions go either way.

Notably, none of the hearer-old persons (using test predictions of SMO) were postmodified. Our cognitive status predictions cleanly partition the examples into those where postmodification is not required, and those where it might be. Since no intuitive rule handled the remaining examples, we added the testing predictions of hearer-old/new and major/minor as features to the list in Table 1, and tried to learn this task using the tree-based learner J48. We report a testing accuracy of 71.5% (majority class baseline is 56%). There were only three useful features—the predicted hearer-new/old status, the number of high frequency premodifiers for that person in the input (feature 14 in table 1) and the average number of postmodified initial references in the input documents (feature 17).

## 5 Validating the results on current news

We tested the classifiers on data different from that provided by DUC, and also tested human consen-

sus on the hearer-new/old distinction. For these purposes, we downloaded 45 clusters from one day's output from Newsblaster[4]. We then automatically compiled the list of people mentioned in the machine summaries for these clusters. There were 107 unique people that appeared in the machine summaries, out of 1075 people in the input clusters.

### 5.1 Human agreement on hearer-old/new

A question arises when attempting to infer hearer-new/old status: Is it meaningful to generalize this across readers, seeing how dependent it is on the world knowledge of individual readers?

To address this question, we gave 4 American graduate students a list of the names of people in the DUC human summaries (cf. §3), and asked them to write down for each person, their country/state/organization affiliation and their role (writer/president/attorney-general etc.). We considered a person hearer-old to a subject if they correctly identified both role and affiliation for that person. For the 258 people in the DUC summaries, the four subjects demonstrated 87% agreement ($\kappa = 0.74$)[5].

Similarly, they were asked to perform the same task for the Newsblaster data, which dealt with contemporary news[6], in contrast with the DUC data that contained news from the the late 80s and early 90s. On this data, the human agreement was 91% ($\kappa = 0.78$). This is a high enough agreement to suggest that the classification of national and international figures as hearer old/new across *the educated adult American reader with varied interests and background in current and recent events* is a well defined task. This is not necessarily true for the full range of cognitive status distinctions; for example Poesio and Vieira (1998) report lower human agreement on more fine-grained classifications of definite descriptions.

### 5.2 Results on the Newsblaster data

We measured how well the models trained on DUC data perform with current news labeled using human judgment. For each person who was mentioned in the automatic summaries for the Newsblaster data, we compiled one judgment from the 4 human subjects: an example was labeled as hearer-new if two or more out of the four subjects had marked it as hearer new. Then we used this data as *test data*, to test the model trained solely on the DUC data. The classifier for hearer-old/hearer-new distinction achieved 75% accuracy on Newsblaster data labeled by humans, while the cross-validation accuracy on the automatically labeled DUC data was 76%. These numbers are very encouraging, since they indicate that the performance of the classifier is stable and does not vary between the DUC and Newsblaster data. The precision and recall for the Newsblaster data are also very similar for those obtained from cross-validation on the DUC data:

| Class | Precision | Recall | F-Measure |
|---|---|---|---|
| Hearer-old | 0.88 | 0.73 | 0.80 |
| Hearer-new | 0.57 | 0.79 | 0.66 |

### 5.3 Major/Minor results on Newsblaster data

For the Newsblaster data, no human summaries were available, so no direct indication on whether a human summarizer will mention a person in a summary was available. In order to evaluate the performance of the classifier, we gave a human annotator the list of people's names appearing in the machine summaries, together with the input cluster and the machine summary, and asked which of the names on the list would be a suitable keyword for the set (keyword lists are a form of a very short summary). Out of the 107 names on the list, the annotator chose 42 as suitable for descriptive keyword for the set.

The major/minor classifier was run on the 107 examples; only 40 were predicted to be major characters. Of the 67 test cases that were predicted by the classifier to be minor characters, 12 (18%) were marked by the annotator as acceptable keywords. In comparison, of the 40 characters that were predicted to be major characters by the classifier, 30 (75%) were marked as possible keywords. If the keyword selections of the annotator are taken as ground truth, the automatic predictions have precision and recall of 0.75 and 0.71 respectively for the *major* class.

## 6 Conclusions

Cognitive status distinctions are important when generating summaries, as they help determine both

[4] http://newsblaster.cs.columbia.edu

[5] $\kappa$ (kappa) is a measure of inter-annotator agreement over and above what might be expected by pure chance (See Carletta (1996) for discussion of its use in NLP). $\kappa = 1$ if there is perfect agreement between annotators and $\kappa = 0$ if the annotators agree only as much as you would expect by chance.

[6] The human judgments were made within a week of the news stories appearing.

what to say and how to say it. However, to date, no one has attempted the task of inferring cognitive status from unrestricted news.

We have shown that the hearer-old/new and major/minor distinctions can be inferred using features derived from the lexical and syntactic forms and frequencies of references in the news reports. We have presented results that show agreement on the *familiarity* distinction between educated adult American readers with an interest in current affairs, and that the learned classifier accurately predicts this distinction. We have demonstrated that the acquired cognitive status is useful for determining which characters to name in summaries, and which named characters to describe or elaborate. This provides the foundation for a principled framework in which to address the question of how much references can be shortened without compromising readability.

## References


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

D. Bikel, R. Schwartz, and R. Weischedel. 1999. An algorithm that learns what's in a name. *Machine Learning*, 34:211–231.

J. Carletta. 1996. Assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics*, 22(2):249–254.

H. Daumé III, A. Echihabi, D. Marcu, D.S. Munteanu, and R. Soricut. 2002. GLEANS: A generator of logical extracts and abstracts for nice summaries. In *Proceedings of the Second Document Understanding Conference (DUC 2002)*, pages 9 – 14, Philadelphia, PA.

P. Gordon, B. Grosz, and L. Gilliom. 1993. Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, 17:311–347.

H.P. Grice. 1975. Logic and conversation. In P. Cole and J.L. Morgan, editors, *Syntax and semantics*, volume 3, pages 43–58. Academic Press.

B. Grosz and C. Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 3(12):175–204.

B. Grosz, A. Joshi, and S. Weinstein. 1995. Centering: A framework for modelling the local coherence of discourse. *Computational Linguistics*, 21(2):203–226.

C. Grover, C. Matheson, A. Mikheev, and M. Moens. 2000. Lt ttt: A flexible tokenization toolkit. In *Proceedings of LREC'00*.

J. Gundel, N. Hedberg, and R. Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language*, 69:274–307.

Y. Guo, X. Huang, and L. Wu. 2003. Approaches to event-focused summarization based on named entities and query words. In *Document Understanding Conference (DUC'03)*.

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

H. P. Luhn. 1958. The automatic creation of literature abstracts. *IBM Journal of Research and Development*, 2(2):159–165.

I. Mani and M. Maybury, editors. 1999. *Advances in Automatic Text Summarization*. MIT Press, Cambridge, Massachusetts.

A. Nenkova and K. McKeown. 2003. References to named entities: a corpus study. In *Proceedings of HLT/NAACL 2003*.

C. D. Paice. 1990. Constructing literature abstracts by computer: techniques and prospects. *Inf. Process. Manage.*, 26(1):171–186.

M. Poesio and R. Vieira. 1998. A corpus-based investigation of definite description use. *Computational Linguistics*, 24(2):183–216.

E. Prince. 1992. The zpg letter: subject, definiteness, and information status. In S. Thompson and W. Mann, editors, *Discourse description: diverse analyses of a fund raising text*, pages 295–325. John Benjamins.

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

H. Saggion and R. Gaizaukas. 2004. Multi-document summarization by cluster/profile relevance and redundancy removal. In *Document Understanding Conference (DUC04)*.

A. Sanford, K. Moar, and S. Garrod. 1988. Proper names as controllers of discourse focus. *Language and Speech*, 31(1):43–56.

A. Siddharthan, A. Nenkova, and K. McKeown. 2004. Syntactic simplification for improving content selection in multi-document summarization. In *Proceedings of the 20th International Conference on Computational Linguistics (COLING 2004)*, pages 896–902, Geneva, Switzerland.

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

I .Witten and E. Frank. 2005. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, San Francisco.