Who Said What to Whom?
Capturing the Structure of Debates

Rianne Kaptein, Maarten Marx and Jaap Kamps
University of Amsterdam, The Netherlands
kaptein@uva.nl, maartenmarx@uva.nl, kamps@uva.nl

ABSTRACT

Transcipts of meetings are a document genre characterized by a complex narrative structure. The essence is not only what is said, but also by who and to whom. This paper investigates whether we can use semantic annotions like the speaker in order to capture this debate structure, and as well as the related content of debate. The structure is visualized in a graph, while the content is condensed into word clouds, that are created using a parsimonious language model. Evaluation shows that both tools adequately capture the structure and content of the debate at an aggregated level.

Categories and Subject Descriptors:
 H.3.3 **[Information Search and Retrieval]**: Search process

General Terms: Experimentation, Measurement, Performance

1. INTRODUCTION

Meeting notes of parliamentary debates are documents which contain lots of structure. This structure is often implicit in layout and reserved words. But since meetings tend to occur regularly and are repeated for long periods of time, this structure is often (semi)formalized. This makes these documents suitable for automatic semantic annotation efforts resulting in an added XML structure, allowing for focused retrieval, entity retrieval, aggregated search and yielding new ways of browsing, mining and summarizing these documents.

The notes of a one day meeting of the Dutch parliament tend to be quite long, typically between 50 and 80 pages two column PDF. For instance, the meeting of September 18, 2008 that took the whole day, consisted of 624 speeches with a total of 74.068 words, all within one topic. It can be hard to find the information you are looking for in such a long document. Using the semantic annotations we can create tools that give a quick first impression of the debate. In this paper we describe a method to visualize the structure of a debate in a graph, and summarize speeches into word clouds.

2. DATA

Our data consists of the notes of meetings of the Dutch parliament of the last 20 years. Everything that is being said in the meeting is transcribed, keeping the content, but making it grammatically correct and pleasant to read. In our data every word spoken in parliament is annotated with 1) the speaker, 2) her party at the time of speaking, 3) her role/function in parliament and 4) the iso-date [1].
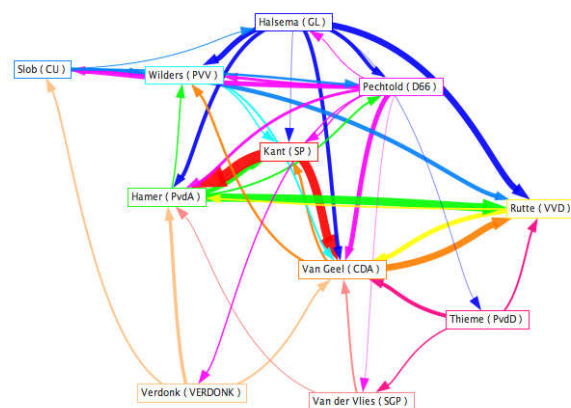
Figure 1: Debate Structure

We will use this annotated textual content to get a higher level view on the data.

3. STRUCTURE AND CONTENT SUMMARIZATION

In a debate it is of interest to see who is interrupting who. This information is implicitly available in the proceedings because we know who is speaking and who is interrupting the speaker. In the Dutch context these interruptions are usually attacks. In Figure 1 we show a graph visualizing the debate structure, where a person is depicted as a node, and an interruption as an arrow between the person who is interrupting and the current speaker. The size of the arrow is representative for the number of interruptions. So, for example *Kant (SP)* from the opposition frequently interrupts governing parties, i.e *Hamer (PvdA)* 24 times and *Van Geel (CDA)* 18 times. The figure gives a high level summary of the structure of the debate.

Besides knowing who is interrupting who, we also want to summarize the content of the interruptions, as well as the speeches. We want to remove the usual stopwords, but we also have to exclude corpus specific stopwords, such as *parliament* and *president*. Furthermore, there are words that will be common and not informative in all interruptions on a certain person, e.g. the name of that person. To filter out all these non-informative words, we use a parsimonious language model [2]. The parsimonious language model concentrates the probability mass on fewer words than a standard language model. The model automatically removes both common stopwords and corpus specific stopwords, and words that are men-

Table 1: Word cloud examples (translated from Dutch)

Log-Likelihood	Parsimonious
animals	animals
that	budget memorandum
budget memorandum	bio
bio	industry
I	animal welfare
animal welfare	purchasing power
industry	earth
the	businesses
of	cattle feed
purchasing power	lnv

tioned occasionally in the document

Instead of using the complete collection to estimate background probabilities, we take a much smaller unit, i.e. the text of the debate. We assume this document is still long and diverse enough to estimate well the probabilities of the frequently occurring (corpus specific) stopwords. Moreover, we will be able to identify words that are used relatively more frequent in a speech or interruption than in the complete debate.

For each node in the graph, we concatenated the text of the speech of a person. For each arrow from node A to node B in the diagram, we concatenated the text of all interruptions of person A during the speech of person B. We use a unigram language model to estimate probabilities and generate word clouds, where we assume that the most probable words are the most informative. The parsimonious probabilities are estimated using *Expectation-Maximization*:

$$\text{E-step:} \quad e_t = tf(t,S) \cdot \frac{\lambda P(t|S)}{\lambda P(t|S) + (1 - \lambda)P(t|D)}$$

$$\text{M-step:} \quad P(t|S) = \frac{e_t}{\sum_t e_t}, \text{i.e. normalize the model}$$

where S is either a speech or a set of interruptions, and D is the complete debate i.e. our background model. In the initial E-step, maximum likelihood estimates are used. For λ we use a value of 0.01. In the M-step the words that receive a probability below our threshold of 0.0001 are removed from the model. In the next iteration the probabilities of the remaining words are again normalized. The iteration process stops after a fixed number of iterations.

A problem here is that some of the interruptions are very short, maybe only one sentence. To improve the quality of the word clouds of interruptions, we use a two-step parsimonious model, running the words of the interruptions through the parsimonious model one more time, but with different background probabilities. The graph suggests two choices for the 'interruption specific background model': 1) all interruptions made by person A on everyone in the graph and 2) all interruptions on person B by everyone in the graph. We have chosen the first, thereby trying to differentiate between interruptions of A. As a starting point for the second round of parsimonious estimation, we take the maximum likelihood estimates of all words that are not removed in the first round. Then again we estimate parsimonious probabilities and remove words with probabilities below the threshold. The resulting word clouds now contain more informative words.

4. EVALUATION

We have executed a user study to evaluate our system. The test persons are 20 experts familiar with the Dutch political landscape. First of all we asked our test persons what information they get from the graph that visualizes the structure of the debate. The most frequent answers are in general: Who interrupts who, and how often, and who is actively involved in the debate. More specifically, test persons see that coalition partners do not interrupt each other often and that opposition parties interrupt governing parties. On the question if the graph gives a good overview of the structure of the debate, we get a score of 3.7 on a 5-point Likert scale, where 1 means strongly disagree, and 5 means strongly agree.

Secondly, we take a look at the word clouds. We asked the test persons whether they think the word clouds are useful summaries of the speeches and the interruptions. For the interruptions we get an average score of 3.1, for the speeches 3.0, so the test persons do not agree or disagree with this statement. Furthermore, the test persons were asked to judge a number of word clouds of speeches as well as interruptions. Each test person gets 3 speeches and 3 to 5 interruptions. For each word in the clouds they mark whether they think the word is informative or not. We have defined informative as 'a word that gives a good impression of the contents of a speech or interrupt.' It should be both a word 'relevant' to the debate, as well as 'discriminative' for the speaker or party. Of the word clouds of the speeches, 47% of the words are considered informative. Of the interruptions, less words are considered informative, i.e. 41%.

An example of a word cloud produced by our parsimonious model and a cloud produced using a log-likelihood model [3] can be found in Table 1. The translated word clouds are from the speech of the party leader of the *Animal Rights Party*. Comparing the two clouds, we see that our parsimonious model correctly removes non-informative stopwords that still remain in the log-likelihood cloud. Although this example parsimonious cloud does not contain any (corpus specific) stopwords, the test persons consider only 58% of the words in this cloud informative.

Our final question is whether looking at the graph and the word clouds gives a good first impression of the debate. Most of our test persons agree with this statement, the average score is 3.8. We can conclude that our tools adequately capture the structure and content of the debate at an aggregated level.

5. CONCLUSIONS AND FUTURE WORK

We have exploited the annotations of meeting notes of the parliament to visualize the structure of the debate in a graph. The speeches and interruptions can be summarized into word clouds using a parsimonious model. While the word clouds still contain a considerable number of words that are regarded as non-informative, the graph together with the word clouds does give a good first impression of the debate. From the comments that persons gave in our user study we can see some useful extensions on our system. Instead of words without any context, we could select multiword phrases consisting of sentences or parts of sentences to summarize speeches and interruptions in a cloud. Another direction for future work is to use the graph structure to visualize aggregated keyword search results instead of a debate.

REFERENCES

[1] T. Gielissen and M. Marx. Exemelification of parliamentary debates. In *Proceedings of the 9th Dutch-Belgian Workshop on Information Retrieval (DIR 2009)*, 2009.

[2] D. Hiemstra, S. Robertson, and H. Zaragoza. Parsimonious language models for information retrieval. In *Proceedings SIGIR 2004*, pages 178–185. ACM Press, New York NY, 2004.

[3] P. Rayson and R. Garside. Comparing corpora using frequency profiling. In *Proceedings of the workshop on Comparing Corpora*, 2000.