Exploiting structural meeting-specific features for topic segmentation

Maria GEORGESCUL[1], Alexander CLARK[2], Susan ARMSTRONG[1]

[1] ISSCO/TIM/ETI, University of Geneva

[2] Department of Computer Science, Royal Holloway University of London

maria.georgescul@eti.unige.ch, alexc@cs.rhul.ac.uk,
susan.armstrong@issco.unige.ch

Résumé. Dans cet article, nous traitons de la segmentation automatique des textes en épisodes thématiques non superposés et ayant une structure linéaire. Notre étude porte sur l'utilisation des traits lexicaux, acoustiques et syntaxiques et sur l'influence de ces traits sur la performance d'un système automatique de segmentation thématique. Nous appliquons notre approche, basée sur des machines à vecteurs support, à des transcriptions des dialogues multi-locuteurs.

Abstract. In this article we address the task of automatic text structuring into linear and non-overlapping thematic episodes. Our investigation reports on the use of various lexical, acoustic and syntactic features, and makes a comparison of how these features influence performance of automatic topic segmentation. Using datasets containing multi-party meeting transcriptions, we base our experiments on a proven state-of-the-art approach using support vector classification.

Mots-clés : segmentation automatique en épisodes thématiques, machines à vecteurs support, dialogues multi-locuteurs.

Keywords: automatic topic segmentation, support vector machines, multi-party dialogues.

1   Introduction

Georgescul et al. (2006b) proposed a support vector machine approach to the task of text segmentation which demonstrates improvements over state-of-the-art techniques, by modeling large scale (merely lexical) features and non-linear relations in an efficient and stable way. Their experimental results showed that word distributions in texts provide relevant information for the detection of boundaries between thematic episodes in data sets covering different domains. In this paper, we put the emphasis on tackling the topic segmentation problem in the context of recorded and transcribed multi-party dialogs. In particular, we extend the work of Georgescul et al. (2006b) by exploring potential information provided by 'surface' cues in multi-party dialogues such as syntactic knowledge, cue-phrases and acoustic cues. We investigate the pertinence of these factors individually and in combination with information provided by word distributions through the intermedium of transductive support vector machines.

In order to identify boundaries, we model the thematic segmentation task as a binary classification problem. The features considered for designing the classifier are described in Section 2. In Section 3 we highlight how the classification model is constructed by using transductive support vector learning. A comparative analysis of the support vector classifier performance by using these cues is provided in Section 4.

2 Input features

As in (Georgescul *et al.*, 2006b), we consider the thematic segmentation task as a binary classification problem, where each utterance should be classified as a topic boundary or not. As explained in Section 3, we employ a support vector machine classifier which is given as input a vectorial representation of the utterance to be classified and its context. Each dimension of the input vector indicates the value of a certain feature characterizing the utterance. For utterance characterization, Georgescul et al. (2006b) only considered features based on observations of patterns in vocabulary use. Here, in addition to these lexical features, we consider meeting-specific features as described in the following.

Note that, similar types of features examined in our study have been previously proposed for analyzing discourse structure in state-of-the-art studies like those described in (Litman & Passonneau, 1995; Hirschberg & Nakatani, 1996; Galley *et al.*, 2003). These include speaker activity, discourse markers, prosodic and syntactic features.

2.1 Speaker activity

According to (Pfau *et al.*, 2001), patterns of *speech activity* are valuable data for discourse analysis. In order to explore this claim, the first pattern we chose to investigate is speaker activity. In particular, we start with the hypothesis that in meeting data the contribution of each participant in the discussion can signal a new topic. For instance, some participants could have a preference for certain subjects of discussion.

We take into account the changes in speaker activity by measuring the number of words each participant uttered before and after each utterance candidate to a thematic boundary. This is formalized in the following manner. Let A_k^s be the number of words that the participant s said in utterance u_k. For each meeting participant s and for each i-th utterance u_i, we take into account the number of words that the participant s uttered before and during u_i in an interval of size $activityWS$, by considering the vector $\vec{fl}_i^s$ as: $\vec{fl}_i^s = \left(A_{i-activityWS+1}^s, A_{i-activityWS+2}^s, ..., A_i^s \right)$. We also store in a vector $\vec{fr}_i^s$ the number of words that the participant s uttered after u_i in an interval of size $activityWS$: $\vec{fr}_i^s = \left(A_{i+1}^s, A_{i+2}^s, ..., A_{i+activityWS}^s \right)$. We then normalize the two vectors $\vec{fl}_i^s$ and $\vec{fr}_i^s$ to form two probability distributions l_i^s and r_i^s, respectively. That is, we perform the normalization by simply dividing each element in the vector by the sum of all entries in the vector.

We measure significant changes in speaker activity by using the *information radius* between the probability distributions given by the speaker activity at the left and right side of the current (u_i)

utterance:

$$IRad(l_i^s, r_i^s) = \frac{1}{2} \left[\sum_{\substack{i \\ l_i^s \neq 0}} l_i^s log \frac{l_i^s}{m_i} + \sum_{\substack{i \\ r_i^s \neq 0}} r_i^s log \frac{r_i^s}{m_i} \right] \tag{1}$$

where $m_i = \frac{l_i^s + r_i^s}{2}$ is the average distribution of the two random variables l_i^s and r_i^s.

Finally, $IRad(l_i^s, r_i^s))$ will constitute the entry for one dimension of the vectorial representation for utterance u_i.

2.2 Discourse markers

Previous studies (Litman & Passonneau, 1995; Marcu, 2000) addressed questions regarding discourse relations and their realization by discourse markers. Here, we are interested in finding those discourse markers that indicate thematic shifts in our data. We started with the following list of discourse markers that has been synthesized from a commonly used list of discourse markers: "accordingly", "actually", "after all", "also", "although", "anyway", "back to", "basically", "but", "fine", "for example", "furthermore", "generally", "however", "like", "moreover", "nevertheless", "nor", "now", "of course", "okay", "really", "similarly", "since", "speaking of", "so", "still", "that's all", "then", "therefore", "well". For each discourse marker in this list, we automatically examine if it occurs in each utterance that is a candidate for marking a thematic boundary. That is, our SVM takes as input binary features indicating whether each discourse marker occurs in the current utterance. We retain as input features to our system only those discourse markers that occur at least once in our corpus.

2.3 Syntactic features

The use of syntax-based features is to a large extent motivated by previous work (Passonneau & Litman, 1993; Litman & Passonneau, 1995) relating discourse structure and noun phrase anaphora. Regarding the pronominal reference, we are mainly following the intuitive assumption that nouns and verbs appear more frequently at the beginning of a new topic, while pronouns appear more frequently in the middle of a thematic episode.

The syntactic features considered in our study are the distributions of different part-of-speech categories before and after a potential thematic boundary. That is, we extracted frequencies of pronouns, (proper) nouns and verbs before and after each utterance candidate to a thematic boundary. For the annotation of part-of-speech information, we used TreeTagger (Schmid, 1994).

This component was formalized as follows. Let P_i, N_i, V_i be the number of pronouns, nouns and verbs, respectively, in utterance u_i. We store in a vector $\vec{fl}_i^p$ the number of pronouns occurring in utterances situated before u_i in an interval of size $synWS$: $\vec{fl}_i^p = (P_{i-synWS+1}, P_{i-synWS+2}, ..., P_i)$. We also store in a vector $\vec{fr}_i^p$ the number of pronouns occurring in utterances situated after u_i in an interval of size $synWS$: $\vec{fr}_i^p = (P_{i+1}, P_{i+2}, ..., P_{i+synWS})$. Similarly, we store in vectors $\vec{fl}_i^n$, $\vec{fl}_i^v$ the number of nouns and verbs, respectively occurring before u_i in an interval of size $synWS$ utterances. Then, the vectors $\vec{fr}_i^n$, $\vec{fr}_i^v$ will contain the number of nouns and verbs, respectively occurring after u_i in an interval of size $synWS$ utterances.

As in Section 2.1, we normalize the resulting vectors of counts $\vec{fl}_i^{\,p}, \vec{fr}_i^{\,p}, \vec{fl}_i^{\,n}, \vec{fr}_i^{\,n}, \vec{fl}_i^{\,v}, \vec{fr}_i^{\,v}$ to obtain probability distributions $l_i^p, r_i^p, l_i^n, r_i^n, l_i^v, r_i^v$, respectively. Finally, we measure changes in the distribution of pronouns, nouns and verbs at the left and right side of the current utterance by using the information radius (see Equation 1). That is, for each utterance u_i, we measure $IRad\left(l_i^p, r_i^p\right), IRad\left(l_i^n, r_i^n\right), IRad\left(l_i^v, r_i^v\right)$, which will constitute entries for three dimensions of the vectorial representation for utterance u_i (taken as input to the SVM classifier).

2.4 Silences and overlaps

Silences and overlaps, as well as other acoustic information can also give evidence whether a major topic shift occurred. In particular, studies on discourse structure (Hirschberg & Nakatani, 1996) exploit various prosodic information such as pitch range (raised at segment-initial phrases and lower at segment-final phrases), speech rate (accelerating at segment-final phrases), amplitude and contour.

We investigated the pertinence of these features with the following formalization. Let S_i be the silence duration between utterance u_{i-1} and u_i. Let O_i be the speaker overlap duration between utterance u_{i-1} and u_i. We normalize the S_i and O_i values by speaker and the resulting values S_i', O_i' are used to compute the following quantities: $sl_i = \sum_{k=i-silenceWSL+1}^{i}\left(S_i'\right)$; $sr_i = \sum_{k=i+1}^{i+silenceWSR}\left(S_i'\right)$; $ol_i = \sum_{k=i-overlapWSL+1}^{i}\left(O_i'\right)$; and $or_i = \sum_{k=i+1}^{i+overlapWSR}\left(O_i'\right)$.

We include silences and overlaps as part of the utterance context representation by considering the sl_i, sr_i, ol_i, or_i quantities as dimensions of the vector characterizing the utterance u_i.

3 Methodology

As introduced in the previous section, we employ a vectorial representation containing lexical, acoustic and syntactic information to characterize each utterance. The topic segmentation task is thus reduced to a binary classification problem: each utterance has to be classified as marking the presence or the absence of a topic shift in the text.

In order to infer eventual dependencies between the binary class label and observations of patterns (provided by the lexical, acoustic and syntactic information), we employ a discriminative approach based on transductive support vector learning. A brief overview on inductive support vector learning for topic segmentation has been described in (Georgescul *et al.*, 2006b). In this section, we give some highlights representing the main elements in using transductive support vector learning for topic segmentation.

The support vector learner $\mathcal{L}$ is given a *training set* $S_{train} = ((\vec{u}_1, y_1), ..., (\vec{u}_n, y_n)) \subseteq (U \times Y)^n$ containing n examples drawn independently and identically distributed (i.i.d.) according to a fixed distribution $Pr(u, y) = Pr(y|u)Pr(u)$. Following the transductive setting proposed by Joachims (1999), the learner is also given an i.i.d. sample, $S_{test} = (\vec{u}_1^*, \vec{u}_2^*, ..., \vec{u}_k^*)$, containing k test examples from the same distribution as $\vec{u}_1, \vec{u}_2, ... , \vec{u}_n$. Each training example from S_{train} consists of a high-dimensional vector $\vec{u}$ describing an utterance and the class label y. The class label y has only two possible values: +1 (corresponding to a 'thematic boundary') or -1 (corresponding to a 'non-thematic boundary'). We represent each utterance instance by a feature vector $\vec{u}$ with attributes containing 'surface' meeting-specific information (as described in Section 2)

plus the attributes given by the bag-of-words representation of word frequencies, as described in (Georgescul *et al.*, 2006b).

Given a hypothesis space $\mathcal{H}$, of functions $h : U \rightarrow \{-1, +1\}$ having the form $h(\vec{u}) = sign(\langle \vec{w}, \vec{u} \rangle + b)$, the transductive learner $\mathcal{L}_{transd}$ seeks a decision function h_{transd} from $\mathcal{H}$, using S_{train} and S_{test} so that the expected number of erroneous predictions on the test examples is minimized. Using the structural risk minimization principle (Vapnik, 1998), the smallest bound on the test error is calculated by minimizing the following cost function $\mathcal{W}^{transd}$:

$$\mathcal{W}^{transd}(y_1^*, ...y_k^*, \vec{w}, b, \xi_1, \xi_2, ..., \xi_n, \xi_1^*, \xi_2^*, \xi_k^*) =$$
$$= \frac{1}{2} \langle \vec{w}, \vec{w} \rangle + C^+ \sum_{i=0, y_i=1}^{n} \xi_i + C^- \sum_{i=0, y_i=-1}^{n} \xi_i + C^* \sum_{j=0}^{k} \xi_j^*, \tag{2}$$

subject to:

$$\begin{cases} y_i \left[\langle \vec{w} \cdot \vec{u}_i \rangle + b \right] \leq 1 - \xi_i \; for \; i = 1, 2, ..., n; \\[2mm] y_j^* \left[\langle \vec{w} \cdot \vec{u}_i^* \rangle + b \right] \leq 1 - \xi_i^* \; for \; j = 1, 2, ..., k; \\[2mm] y_j^* \in \{-1, 1\} \; for \; j = 1, 2, ..., k. \\[2mm] \xi_i \geq 0 \; for \; i = 1, 2, ..., n; \\[2mm] \xi_j^* \geq 0 \; for \; j = 1, 2, ..., k; \end{cases}$$

The so-called *slack variables* ξ_i and ξ_j^* are introduced in order to be able to handle non-separable data. The *regularization parameters* C^- and C^+ are tuned as described in Section 4.1.

4 Experiments and results

4.1 Parameter estimation

We train and evaluate the effectiveness of our technique on the ICSI-MR dataset (Janin *et al.*, 2004) containing transcribed multi-party dialogs.

We divide the ICSI-MR data set into two disjoint parts: a training dataset composed of 80% of the initial data set, while the remaining 20% is held out for testing purposes. That is, the training set is used to determine the best model settings for the SVM classifier, while the test set is used to determine the final topic segmentation error rate.

We select the best model parameters, by running five-fold cross validation for SVM parameter estimation, using the Gaussian RBF kernel. During this preliminary step we estimate the performance of the SVM classifier by using the precision and recall, i.e. the precision/recall-breakeven point (Joachims, 1999). The choice of binary evaluation metrics in this step was motivated by the fact that posing the topic segmentation task as a classification problem involves a loss of the sequential nature of the data, which is an inconvenience in computing the P_k (Beeferman *et al.*, 1999) or Pr_{error} (Georgescul *et al.*, 2006a) measures.

Parameter	Interval for grid search	Best window size
activityWS	$5 \ldots 50$ step 5	35 utterances
synWS	$5 \ldots 50$ step 5	30 utterances
silenceWSL	$2 \ldots 10$ step 1	6 utterances
silenceWSR	$2 \ldots 10$ step 1	3 utterances
overlapWSL	$2 \ldots 10$ step 1	2 utterances
overlapWSR	$2 \ldots 10$ step 1	4 utterances

TAB. 1 – Grid search interval over parameters involved in data representation.

Given that the data used in our experiments contains only about 0.07% utterances marking thematic boundaries relative to the total number of utterances in the corpus, we handle the imbalance between the number of positive and negative examples for the SVM classifier by using an assymetric soft margin optimization, which charges more for false negatives than for false positives. That is, we set the regularization parameter C^+ several times larger than C^-: $C^+ = \left\lceil \frac{n}{n^+ - 1} - 1 \right\rceil \cdot C^-$, where n is the total number of training examples and n^+ is the number of positive training examples.

Model selection is done in two phases, as described below.

The first step in model selection consists of searching for the most appropriate utterance representation by using each individual category of features. That is, we look for appropriate values for the size of the windows (intervals) considered when measuring "speaker activity" and when taking into account "syntactic information" and "silences and overlaps" for the utterance instance (cf. Section 2). This is determined by performing a grid search interval over various values for *activityWS*, *synWS*, *silenceWSL*, *silenceWSR*, *overlapWSL* and *overlapWSR*. For each "window size (WS)" parameter, the range of values we select from is given in the second column of Table 1. Note that for the features based on lexical reiteration, we have used the optimal parameter settings that have been determined in (Georgescul *et al.*, 2006b). In this step, we train the SVMs with fixed values for both the RBF kernel parameter and the regularization parameters C^+ and C^-, i.e. the magnitude of the penalty for violating the soft margin has been set to: $C^- = 1$; while the RBF kernel parameter has been set to: $\gamma = 1$.

Using the entire set of features with the representations selected in the first step (cf. the third column of Table 1), the second step in model selection consists in optimizing the parameters of the classifier, i.e. the regularization parameters C^+ and C^- and the RBF kernel parameter γ. That is, we perform grid search interval over the following values: $C^{-1} \in \{10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3\}$, $\gamma \in \{2^{-6}, 2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 1, 2, 2^2, 2^3, 2^4, 2^5, 2^6\}$.

4.2 Results

The results obtained on the ICSI-MR corpus using only the proposed surface conversational cues, (i.e. excluding the features based on lexical reiteration), in our SVM approach for thematic segmentation are illustrated in Figure 1. The table gives means for the percentage error rates given by P_k metric (Beeferman *et al.*, 1999) and the Pr_{error} metric (Georgescul *et al.*, 2006a) for the systems we have used throughout our work. We provide as baselines the error rates obtained when using *TextTiling* (Hearst, 1997), *C99* (Choi, 2000), *TextSeg* (Utiyama & Isahara, 2001) and *Random*, a naive segmentation algorithm (by which the number of boundaries is randomly selected and boundaries are randomly distributed throughout text).

Exploiting structural meeting-specific features for topic segmentation

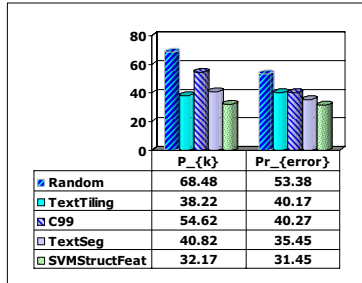|	P_{k}	Pr_{error}
Random	68.48	53.38
TextTiling	38.22	40.17
C99	54.62	40.27
TextSeg	40.82	35.45
SVMStructFeat	32.17	31.45

FIG. 1 – Comparative performance of our SVM approach using only structural features with various topic segmentation systems run on ICSI-MR data.

From Figure 1, we observe that by following the quantitative assessment of both P_k error and the Pr_{error}, our method, labeled as *SVMStructFeat*, using only surface-features outperforms other topic segmentation systems reported on in the literature.

The error values for topic segmentation on the ICSI-MR corpus when using the entire set of features (i.e. lexical, syntactic and prosodic information) are given in the first row of Table 2. The error rates of our method using both lexical and structural features, i.e the error rates of $SVM_{Lexical+StructFeat}$ in the first row of Table 2, as compared to those obtained in (Georgescul *et al.*, 2006b), i.e. the error rates of $SVM_{LexicalFeat}$ in second row of Table 2, show that performance gains can be achieved with the help of surface features in addition to word distribution-based features.

System	P_k error rate	Pr_{error} error rate
$SVM_{Lexical+StructFeat}$	20.94 %	20.17%
$SVM_{LexicalFeat}$	21.68%	21.83%

TAB. 2 – Comparative performance of our SVM approach when using only lexical features (second row) and when using both lexical and structural features (first row).
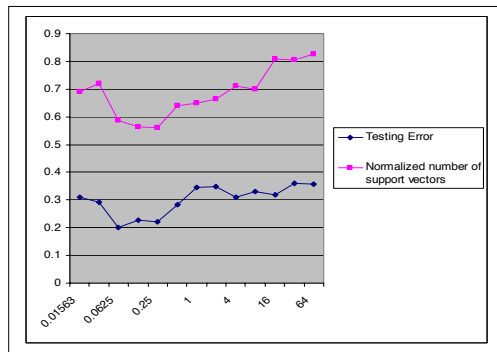

FIG. 2 – Plotting the error rates on testing data and the normalized number of support vectors when tuning γ, the RBF kernel parameter.

Figure 2 shows the influence of the kernel width both on the testing error curves and on the number of support vectors when $C^- = 10^{-2}$ (the optimal value selected through the procedure described in Section 4.1). We observe that in the optimality region the curve representing the error rates has a similar behavior as the curve corresponding to the normalized number of support vectors. That is, the minimum area in the number of support vectors corresponds to minimum error values of SVM-based topic segmentation on testing data. Therefore the number of support vectors is a good indicator of the optimality region.

We also observe from Figure 2 that the number of support vectors is rather large for all tuning values of γ. This reflects the fact that the positive samples (corresponding to the 'topic boundary' class) are not easily separable from the negative examples (corresponding to the 'non-topic boundary' class) due to noise. Moreover, our SVM approach has the critical property of differentiating between positive and negative class members by effectively removing the existing uninformative patterns from the data.

5 Comparison to other work

Comparing the performance of our model to other similar existing studies is not straightforward due to differences in corpora, in experimental design, and/or different input assumptions. Nevertheless, in the following we discuss some related work, by exemplifying some common aspects of the work and the experimental results.

Kauchak and Chen (2005) examined how the boundaries of thematic episodes can be detected in encyclopedia articles and in two books. They employ a supervised technique based on support vector machines using a variety of information including, for instance, features based on the presence of paragraph breaks, pronouns and named entities. When evaluating their topic segmentation model on encyclopedia articles, they obtained a P_k error rate of 39.8%.

Note that, in the context of spontaneous multiparty dialogue, the lack of paragraphs makes the topic segmentation task more difficult than the topic segmentation of narrative written text. For instance the chance of each paragraph break being a topic boundary is about 39.1% in expository texts (Hearst, 1997), while in the ICSI-MR corpus, the chance of each utterance to be a subtopic segment boundary is approximately 0.07% for top-level boundaries. Moreover, meeting dialogues provide particular challenges since topic changes are not always clearly delimited in contrast to e.g. broadcast news or written texts.

The model proposed in (Galley *et al.*, 2003) is the most similar to our model in terms of incorporating multi-party meeting specific features such as cue phrases, silences and conversation overlaps. Using such strucural features in addition to lexical chains, Galley et al. (2003) trained a decision tree which achieved a P_k error rate of 23% on a subset of the ICSI-MR corpus.

6 Conclusions and future work

In this article, we have presented an approach to learn the thematic structure of texts in the context of recorded and transcribed multi-party dialogs. Each utterance is represented as a collection of features obtained from lexical, syntactic and prosodic information. A SVM-based classifier has been trained to discriminate between utterances marking thematic and non-

thematic boundaries in meeting transcriptions.

Our contribution is fivefold. First, we introduce a series of different linguistic and acoustic cues to represent each utterance and we evaluate whether the proposed surface (meeting-specific) cues are useful for thematic segmentation. Second, we check the suitability of our SVM approach combining meeting-specific surface features with large-scale lexical features. Third, we evaluate the compatibility of SVM classification for various thresholds. Fourth, we study the influence of the kernel width on the testing error rate and on the (normalized) number of support vectors. Fifth, we compare the results with existing state-of-the-art methods for topic segmentation. We demonstrate that using 'surface' meeting specific features, our SVM approach generates competitive results on meeting data sets.

As a continuation of this work, it would be interesting to replicate our experiments on larger training sets. The proposed method can potentially be improved by exploiting additional sources of information, including for instance other prosodic information such as speech pitch range and speech rate. It would be also interesting to evaluate whether our topic segmentation approach can be further improved via other kernel methods.

Aknowledgments

References

BEEFERMAN D., BERGER A. & LAFFERTY J. (1999). Statistical Models for Text Segmentation. *Machine Learning*, **34**(Special Issue on Natural Language Learning), 177–210.

CHOI F. (2000). Advances in Domain Independent Linear Text Segmentation. In *Proceedings of the 1st Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, Seattle, USA.

GALLEY M., MCKEOWN K., FOSLER-LUISSIER E. & JING H. (2003). Discourse Segmentation of Multi-Party Conversation. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics (ACL)*, p. 562–569, Sapporo, Japan.

GEORGESCUL M., CLARK A. & ARMSTRONG S. (2006a). An Analysis of Quantitative Aspects in the Evaluation of Thematic Segmentation Algorithms. In *Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue*, p. 144–151, Sydney, Australia: Association for Computational Linguistics.

GEORGESCUL M., CLARK A. & ARMSTRONG S. (2006b). Word Distributions for Thematic Segmentation in a Support Vector Machine Approach. In *Proceedings of the 10th Conference on Computational Natural Language Learning (CoNLL)*, p. 101–108, New York City, USA.

HEARST M. (1997). TextTiling: Segmenting Text into Multi-Paragraph Subtopic Passages. *Computational Linguistics*, **23**(1), 33–64.

HIRSCHBERG J. & NAKATANI C. (1996). A Prosodic Analysis of Discourse Segments in Direction-Giving Monologues. In *Proceedings of the 34th Annual Meeting on Association for Computational Linguistics (ACL)*, p. 286–293, Santa Cruz, California, USA.

JANIN A., ANG J., BHAGAT S., DHILLON R., EDWARDS J., MACIAS-GUARASA J., MORGAN N., PESKIN B., SHRIBERG E., STOLCKE A., WOOTERS C. & WREDE B. (2004). The ICSI Meeting Project: Resources and Research. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP), Meeting Recognition Workshop*, Montreal, Quebec, Canada.

JOACHIMS T. (1999). Making Large-Scale Support Vector Machine Learning Practical. In B. SCHÖLKOPF, C. BURGES & A. SMOLA, Eds., *Advances in Kernel Methods - Support Vector Learning*. Cambridge, MA: MIT Press.

KAUCHAK D. & CHEN F. (2005). Feature-Based Segmentation of Narrative Documents. In *Proceedings of the ACL Workshop on Feature Engineering for Machine Learning in Natural Language Processing*, p. 32–39, Ann Arbor, Michigan, USA.

LITMAN D. J. & PASSONNEAU R. J. (1995). Combining Multiple Knowledge Sources for Discourse Segmentation. In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics (ACL)*, p. 108–115, Cambridge, Massachusetts, USA.

MARCU D. (2000). *The Theory and Practice of Discourse Parsing and Summarization*. MIT Press Cambridge, MA, USA.

PASSONNEAU R. J. & LITMAN D. J. (1993). Intention-based Segmentation: Human Reliability and Correlation with Linguistic Cues. In *Proceedings of the 31st Conference on Association for Computational Linguistics (ACL)*, p. 148 – 155, Columbus, Ohio, USA.

PFAU T., ELLIS D. P. & STOLCKE A. (2001). Multispeaker Speech Activity Setection for the ICSI Meeting Recorder. In *Proceedings of the IEEE Workshop on Automatic Speech Recognition and Understanding*, p. 107–110.

SCHMID H. (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees. In *Proceedings of the International Conference on New Methods in Language Processing*, Stuttgart, Germany.

UTIYAMA M. & ISAHARA H. (2001). A Statistical Model for Domain-Independent Text Segmentation. In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics and the 10th Conference of the European Chapter of the Association for Computational Linguistics (ACL/EACL)*, p. 491–498, Toulouse, France.

VAPNIK V. N. (1998). *Statistical Learning Theory*. A Volume in the Wiley Series on Adaptive and Learning Systems for Signal Processing, Communications, and Control. Berlin: Springer-Verlag.