

# Topic Segmentation : A First Stage to Dialog-Based Information Extraction

Narjes Boufaden Guy Lapalme Yoshua Bengio  
{boufaden, lapalme, bengioy}@iro.umontreal.ca  
Departement of Computer Science and Operations Research  
University of Montreal, Quebec Canada

## Abstract

We study the problem of topic segmentation of manually transcribed speech in order to facilitate information extraction from dialogs. Our approach is based on a combination of multi-source knowledge modeled by hidden Markov models. We experiment with different combinations of linguistic-level cues on dialogs dealing with search and rescue missions. Results show the effectiveness of multi-source knowledge.

## 1 Introduction

The experiments presented here are part of the Search And Rescue project originally conducted by the Canadian Defense Research Establishment Valcartier. It is aiming to develop a decision support tool that help producing search and rescue plans given information from telephone sources (Figure 1). It is a three part process which include a system that extract relevant information from the transcribed telephone conversations. A system that uses the extracted data to induce hypothesis on probable location of the searched object. And a third system that uses the hypothesis and generates search and rescue plans. In this paper, we are concerned with the first system called EXITTS (EXtraction of Information from Telephone and Text sources). In particular, we highlight a necessary stage in dialog-based information extraction : topic segmentation.

Information extraction (IE) consists in collecting relevant events from a document in

a particular domain of application. Previous work on IE has generally been concerned with well-structured texts (Hobbs et al., 1996; Appelt et al., 1993; Cardie, 1997; Lenhert et al., 1992).

### Speaker Utterance

1	O	<i>RCC.</i>
2	C	<i>How are you doing ?</i>
3	O	<i>Good.</i>
4	C	<i>I'm wondering if, Unit1, it might be an oppor, a possibility.</i>
5	C	<i>Ther're, they're good for surface search as well and in fact their radar is perhaps, even a little more sensitive than, than the other one, than Unit2.</i>
6	C	<i>Ha, do, is there, is there a Unit1 available?</i>
7	C	<i>Ha, would you guys make one available before I go out make the request ?</i>
8	O	<i>Yes, let me see what we got on the line here.</i>
9	C	<i>Ok.</i>
10	O	<i>I'll make one available (INAUDIBLE).</i>
11	C	<i>yeah.</i>
12	C	<i>Thanks.</i>
13	O	<i>All right.</i>
14	O	<i>Bye</i>

Figure 1: Excerpt of a conversation involving two speakers: **Caller (C)** and **Operator (O)**. For the sake of confidentiality, named entities have been replaced here with generic terms.

Unlike those systems, dialog-based IE systems have to deal with scattered information and disfluencies embedded in dialogs. In many cases, relevant information is com-

pleted through several consecutive utterances belonging to one or more speakers. Question-answer pairs, widely used in dialogs, are examples where bits of information are conveyed through consecutive utterances. In this paper we investigate the usefulness of topic segmentation of dialogs as a preprocess to information extraction tasks. We noticed that segmenting dialogs into topics facilitates the localization and extraction of scattered relevant information. By identifying topical information, it is possible to ensure the extraction of coherent and complete key answers. Besides, topic segmentation is a valuable preprocessing for coreference resolution, which is a difficult task in IE.

We have therefore developed a topic segmentation system based on a multi-knowledge source modeled by a hidden Markov model. We investigate a fine-grained topic segmentation approach at the level of template fields that fit information extraction needs. We evaluate the system on transcripts of telephone conversations dealing with search and rescue missions.

## 2 Work context

EXITTS is the first part of the SAR project. It has to extract typical informations from transcribed telephone conversations dealing with completed search and rescue missions and fill in different templates. Conversations mostly report incidents such as missing persons, overdue boats and ongoing missions. Typical informations are, for example, time, location, descriptions of the missed object and descriptions of the unit search used during the mission. They are extracted to fill in four types of templates:

- mission related templates “**Search mission**” which encompass fields for the date, time of the beginning and completion of the mission, region where the mission took place, weather conditions.
- templates related to the resources used for the search “**Search unit**” (Figure 2)
- incident related templates “**Incident**”

### Search Unit Template :

**Id:** identifier of the unit instance used to search  
**Category:** aircraft, helicopter, boat, etc.  
**Type :** the brand of the unit instance  
**Detection equipment :** radar, etc.  
**Unit availability:** yes or no.

Figure 2: Excerpt of search unit template

with fields for the cause of the incident, the initial alert, the time and location of the incident.

- missing object templates “**Search object**” with fields for the category of the missed object, the type, status and the call sign referring to the object.

Both transcripts and templates were provided by the Canadian Defense Research Establishment Valcartier (DREV). The remaining sections present the first module of the system EXITTS: topic segmentation.

## 3 Topic Segmentation for Information Extraction

Topic segmentation of dialogs is concerned with automatically dividing dialogs into blocks of utterances that convey the same communicative intention; where the communicative intention is the purpose of a speaker move (Levelt, 1989). However, unlike projects related to the Topic Detection and Tracking (TDT) initiative (Allan et al., 1998) that deal with topic segmentation of transcribed speech and broadcast news, we are targeting a fine-grained segmentation of dialogs at the level of template fields. This implies that the key answer for one template field is obtained at most from a single segment. For example, utterances 6 to 11 (Figure 3) represent an ongoing negotiation about availability of a “*Unit1*” for a mission. From these utterances, it is possible to extract the key-answer of the slot “**Unit availability**” of the template **Search unit** (Figure 2). In particular, since they are part of the same topic segment, it is easier to map

utterance 10 (the answer) to the utterance 6 (the question).

Hence, by using topic segmentation at the level of template fields, we aim at improving information extraction in two ways. First, by gathering topical information in the same segment, the system can locate the set of utterances related to the same relevant piece of information. This allows the extraction of coherent and complete key answers.

Second, because of the limited length of segment topics ( $\simeq 3.4$  utterances per topic segment), ambiguity problems related to coreference resolution can be reduced by providing valuable restrictions for determining the referent.

#### 4 Approach

Our goal is to automatically locate topic boundaries at the template fields level and identify. For this purpose, we manually located the topic boundaries and identify what knowledge could help doing automatic segmentation. Figure 3 is an example of segmented conversation where dotted lines are topic boundaries and tags indicate the class of each utterance.

We annotated the corpus with cues, mostly discourse markers and used a learning approach to predict topic boundaries using the linguistic cues. Basically, the training corpus is composed of contiguous utterances. Since our inputs are transcripts, we did not have to choose a definition for the utterance to use. Otherwise, identifying utterance boundaries in continuous speech is in itself problematic (Traum and Heeman, 1997).

Utterances were classified into topic boundaries and non topic boundaries. We distinguished utterances introducing a new topic (topical change: TC) from those developing it (no topical change: No-TC) and those ending it (end topic: ET). Also two other tags were added to capture utterances beginning a conversation (begin conversation: BC) and those ending it (end conversation: EC).

In fact, topic segmentation became a problem of utterance classification. Various techniques have been used for such classification

problems; such as decision trees and probabilistic models (Tür et al., 2001; Litman and Passonneau, 95). We use a probabilistic model because of the prominence of temporal dependencies of utterances within a topic segment. Question-answer pairs are good examples of such temporal dependencies.

	Speaker	Tag	Utterance
1	O	BC	<i>RCC.</i>
2	C	BC	<i>How are you doing ?</i>
3	O	BC	<i>Good.</i>
			.....
4	C	TC	<i>I'm wondering if, Unit1, it might be an oppor, a possibility.</i>
5	C	No-TC	<i>Ther're, they're good for surface search as well and in fact their radar is perhaps,even a little more sensitive than, than the other one, than Unit2.</i>
			.....
6	C	TC	<i>Ha, do, is there, is there a Unit1 available?</i>
7	C	No-TC	<i>Ha, would you guys make one available before I go out make the request ?</i>
8	O	No-TC	<i>Yes, let me see what we got on the line here.</i>
9	C	No-TC	<i>Ok.</i>
10	O	No-TC	<i>I'll make one available (INAUDIBLE).</i>
11	C	ET	<i>yeah.</i>
			.....
12	C	EC	<i>Thanks.</i>
13	O	EC	<i>All right.</i>
14	O	EC	<i>Bye</i>

Figure 3: Except of Figure 1 is segmented into topics with dotted lines and each utterance is tagged with topic boundary tags.

In the following sections, we explain our choice of linguistic cues and describe our language model.

## 4.1 Linguistic Cues

Many studies on discourse analysis and experiments on topic segmentation of monologues (Halliday and Hassan, 1976; Renkema, 1993; Tür et al., 2001) emphasize the usefulness of lexical and syntactic cues in detecting coherence between utterances. On the other hand, Maynard and Sacks explain that topicality is observable in a patterned way. In a two-party conversation, each speaker usually demonstrates interest or understanding through typical answers such as *ok*, *yeah*, *right*. Depending on the speaker ship, lexical cues can entice the other party to continue his topic development or not (Maynard, 1980; Sacks et al., 1974). Maynard defines the **topical speaker** as the speaker who has the responsibility of developing the topic, whereas, the **recipient** may keep the line of talk or interrupt it. We will see that speaker ship helps in the detection of topical transitions. Finally, our experiments shows that extra-linguistic cues such as interruptions can help identifying topic boundaries. Hence, we retained the following cues :

- **Discursive cues** which are typically topical-speaker and recipient. For these experiments they were manually extracted. However, we are running some experiments in order to automatically predict the speaker ship using syntactic and lexical cues combined with the speaker identity.
- **Syntactic cues** are conjunctions (*and*, *or*, *but ...*), questions marks and temporal adverbs (*then*, *before...*). These cues are automatically extracted.
- **Lexical cues** are essentially repetitions, and acknowledgment words (*ok*, *yeah*, *thank you*, *sure...*), continuers (*hum*, *I mean...*). These lexical cues are automatically extracted.
- **Interruptions** transcribed by successive dots (...) are automatically extracted.

## 4.2 Language Model

We built a hidden Markov model with five states, each representing a topic boundary class. Two states are used to capture formulaic speech patterns used in opening (**BC**) and closing (**EC**) conversations. Three other classes classify utterances into initiating a topic (**TC**) developing it (**No-TC**) and ending it (**ET**). The observation likelihoods for the HMM state represent the probability of generating a sentence belonging to one of these classes.

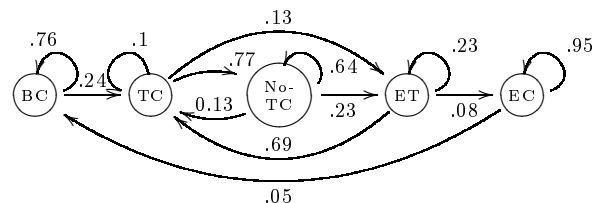


Figure 4: A first order HMM for topic segmentation

Two states are for the utterances beginning (**BC**) and those ending (**EC**) the conversation and three states (**TC**, **No-TC**, **ET**) are for utterances beginning a new topic, those developing it and those ending it. Parameters of the model are computed from relative frequencies. For testing, we run the model as a hidden topic segment model, hypothesizing topic boundaries between cues vectors. Our goal therefore, is to find the most likely topic boundary distribution using different linguistic level cues. Formally, we denote the string of boundary classification as  $Q$  and the string of cues extracted from a conversation as  $Y$ . The most likely sequence  $q^*$  of topic boundaries for a string of linguistic cues  $Y$  is given by :

$$q^* = \operatorname{argmax}_q P(q|Y)$$

We use a Viterbi algorithm to infer the best topic boundary distribution for each conversation.

## 5 Experiments and Results

The initial corpus is composed of 300 transcripts of telephone conversation totaling 12,688 utterances. The length of the conversations ranges from 20 to 65 utterances per conversation. Our experiments were conducted on 52 conversations (2,000 utterances) which were annotated with linguistic cues (section 4.1). About 2.2% of the utterances have too many inaudible segments and are tagged with a special tag for noisy utterances. Highest rates of utterances with only the speaker ship tag are in TC boundary class with 40% of the utterance and in No-TC boundary class with 34.8% of utterance.

In fact, the purpose is to find out what combination of linguistic level cues best classify utterances and the language model that performs well on utterances having only the speaker ship tag. The next two sections present out-of-sample cross validation results of the experiments.

### 5.1 Combination of Linguistic Cues

To measure effectiveness of multi-knowledge sources on topic segmentation, we trained a first order HMM with combinations of different linguistic level cues. We retained the following combinations :

- A combination of lexical and discursive cues (LD)
- Lexical, syntactic and discursive cues combination (LSD)
- Lexical, syntactic and interruption cues combination (LSI)
- All the cues (SLDI)

Table 1 shows error rates per boundary class and the weighted error rate for each combined model. To facilitate interpretation, we recalculate error rates considering ET and NO-TC as one class.

The best weighted error rate is achieved by the SLDI combined model. It also appears that discursive cues are effective in topic segmentation since the SLDI combined model

Models	LD	SLD	SLI	SLDI
BC	18.8	17.7	17.8	17.6
EC	4.0	5.0	1.6	5.0
TC	46.2	34.0	38.8	32.7
No-TC	16.2	32.5	22.6	19.7
Weighted Error.	22.1	30.3	24.7	21.6

Table 1: Error rate per boundary class for all the linguistic models

Models	SLDI-1	SLDI-2
BC	17.6	15.9
EC	5.0	5.9
TC	32.7	29.3
No-TC	19.7	21.4
Weighted Error	21.6	21.9

Table 2: Error rate per boundary class for the first and second order HMM. SLDI-1 is the SLDI combined model with a first order HMM and the SLDI-2 is with a second order HMM.

outperforms the SLI combined model. However, the best score in detecting end conversations (EC) is achieved with the SLI combined model, which does not use discursive cues. This result can be explained by the fact that most utterances closing a conversation are rather formulaic.

Utterances beginning new topics (TC) are the most problematic as shown in table 1. Since 40% of those utterances have only speaker ship tags it is difficult to detect topical transitions.

An alternative solution to this problem is to add more contextual informations and take more advantage of temporal dependencies.

### 5.2 Increasing Contextual Informations

To decrease TC error rate, we performed experiments with the SLDI combined model with a second order HMM. This allowed us to measure the effect of more contextual information in managing lack of cues in utterances beginning new topics.

Surprisingly, the weighted error rate is almost the same for both models and No-TC

Topic boundary class	Recall	Prec.
BC	88.2	87
TC	76.6	67.4
No-TC	76.8	80.5
EC	96.9	93.9
Weighted score	78.6	78.9

Table 3: Recall and precision of topic detection for the model SLDI-2

error rate increased (Table 2). We suspect the size of the training corpus to be the cause of this result. However, as we anticipated, TC error rate decreased as does BC error rate. Despite the increase of No-TC error rate, it appears that more contextual information can improve topic segmentation.

The weighted Recall of the SLDI-2 combined model (Table 3) indicates that 78.6% of the topic segments are correctly detected and 78.9% of the proposed topic segment are correct. Although the results are encouraging, further calculations need to be done to determine how many relevant topic segments were correctly detected since the final purpose is information extraction.

As far as we know, no previous published work on automatic topic segmentation of dialogs has been presented, hence, it is difficult to compare our results. However, (Litman and Passonneau, 95) achieved a 46% recall and 67% precision on topic segmentation of transcribed speech monologues.

## 6 Conclusions

The purpose of the work presented here is to facilitate the development of the dialog-based information extraction system. The baseline tasks in IE are locating and extracting relevant information. However, for dialogs it is difficult to efficiently fulfill these tasks because of scattered information in the text. We based our IE approach on the assumption that topic segmentation can help resolve the problem of scattered information in dialogs. Our topic segmentation system is based on a multi-knowledge source modeled by a hidden Markov model. Experiments show that

the SLDI combined model outperforms all the other combined models. Although error rates of the first and second order SLDI combined model are almost the same, we observed improvement in topical transition detection. This confirmed our intuition that more contextual information could compensate for the lack of cues in utterances. Topic detection precision is about 78.9% and recall is about 78.6%. Both results are encouraging, although further calculations need to be done to measure effectiveness in gathering relevant utterances.

The results reported here should be regarded as a baseline for further improvement. In particular, more lexical cues such as referent could improve the model. Combining several models is also another direction for future improvements.

## References

- J. Allan, J. Carbonnel, G. Doddington, J. Yamron, and Y. Yang. 1998. Topic detection and tracking pilot study final report. In *Proc. of the DARPA Broadcast News Transcription and Understanding Workshop*.
- E.D. Appelt, J.R Hobbs, J. Bear, D. Israel, and M. Tyson. 1993. Fastus: A finite-state processor for information extraction from real-world Text. In *Proceedings of IJCAI*, pages 1172–1178.
- C Cardie. 1997. Empirical methods in information extraction. In *AI Magazine*, volume 18 of 4, pages 65–79. American Association for Artificial Intelligence.
- M.A.K Halliday and R. Hassan. 1976. *Cohesion in English*. Longman, London.
- J. Hobbs, D Appelt, J. Bear, D. Israel, and Kame, 1996. *FASTUS: A Cascaded Finite-State Transducer for Extracting Information from Natural-Language Text*. MIT Press.
- W. Lenhert, C. Cardie, D. Fisher, J. McCarthy, E. Riloff, and E. Soderland. 1992. Description of the CIRCUS system as used for MUC-4. In Morgan Kuffmann, editor, *Proceedings of the fourth Message Understanding Conference (MUC-4)*, pages 282–288, San Mateo, CA, June.
- W.J.M. Levelt. 1989. *Speaking From Intention to Articulation*. MIT Press.

- D.J. Litman and R.J. Passonneau. 95. Combining multiple knowledge sources for discourse segmentation. In *Proc. of ACL '95*, pages 108–115.
- D.W. Maynard. 1980. Placement of topic changes in conversation. In *Semiotica*, volume 30, pages 263–290. Mouton Publishers.
- Jan Renkema. 1993. *Discourse Studies : An Introductory Textbook*. John Benjamins, Philadelphia.
- H. Sacks, E.A. Schegloff, and G. Jefferson. 1974. A simplest systematics for the organization of turn-taking. In *Language*, volume 50, pages 696–735.
- D. Traum and P. Heeman. 1997. Utterance units in spoken dialogue. In E. Maier, M. Mast, and S. LuperFoy, editors, *Dialogue Processing in Spoken Language Systems*, LNAI. Springer-Verlag.
- G. Tür, D. Hakkani-Tur, A. Stolcke, and E. Shriberg. 2001. Integrating prosodic and lexical cues for automatic topic segmentation. *Computational linguistics*, 1(27):31–57.