

Joining Meeting Documents to Strengthen Multimodal Thematic Alignment

Dalila Mekhaldi

Computational Linguistics Research Group
University of Wolverhampton
Wolverhampton, UK
dalila.mekhaldi@gmail.com

Denis Lalanne

Departement d'Informatique
Université de Fribourg
Fribourg, Switzerland
denis.lalanne@unifr.ch

Abstract: A study carried out on multimodal thematic alignment in the context of multi-document meetings is presented in the current paper. First, a classification of documents discussed in meetings is described. The alignment of dialog transcript with discussed documents is then presented based on two different approaches, either combining all documents within one (jointed), or align each document individually with the dialog transcript (disjointed). Furthermore, an entailment process that aims to improve the thematic alignment results is presented. The efficiency of this entailment process is assessed based on the two document arrangement strategies (disjointed/jointed), in addition to other features of discussed document (total/partial discussion and homogeneous/heterogeneous content). An evaluation of the presented techniques over different meeting corpora is then presented. The results tend to prove that the jointed strategy generates better results than the disjointed strategy, for both thematic alignment and entailment process. Moreover, the entailment process might have different effect on the thematic alignment results, depending on the features of the discussed documents.

Keywords- thematic alignment; thematic links entailment; multi-document meetings; jointed alignment, disjointed alignment

I. INTRODUCTION

Meetings are an indispensable event in the life of any organization nowadays, in which several input/output resources from various modalities are used, exchanged or generated. Written documents represent one of the most important resources during meetings, which are used before, during or after the meeting. Written documents are used as support documents by meeting participants, such as printed budget documents, displayed slideshows, etc. They are also present in the form of an agenda in order to monitor the meeting, or in the form of emails that are exchanged between participants before meetings.

Current researches in image and video analysis are willing to automatically create indexes and pictorial video summaries to help users browse through multimedia meeting corpora. However, those methods are often based on low-level visual features and lack semantic information. Other research projects use language understanding techniques or

text caption derived from OCR, in order to create more powerful indexes and search mechanisms. Our assumption is that in a large proportion of multimedia applications (e.g. lectures, meetings, news, etc.), classical printable documents play a central role in the thematic structure of discussions. Further, we believe printable documents could provide a natural and thematic mean for browsing and searching through multimedia meeting repository. In our previous work, we developed a multimodal document alignment framework [6] that follows the idea of linking documents with other multimedia meeting data in order to enhance browsing capabilities. FaericWorld [12] and FriDoc [4] are multi-modal browsers that allow navigating through a large multimedia corpus of meetings, and in a meeting respectively. For instance, FriDoc allows replaying a meeting using documents as structure vectors to multimedia content (see Fig. 1). Thanks to the alignment technique [6], all the representations are synchronized, meaning they all have the same time reference, and clicking on one of them causes all the components to visualize their content at the same time. Clicking on a document part positions audio/video clips at the time when it was discussed, positions the speech transcription at the same time, and displays the document that was projected. Several user evaluations performed on our document centric meeting browsers have shown the usefulness of document alignment as a way to improve question answering while replaying a meeting. In addition to FaericWorld and FriDoc, there have been also other multimodal browsers of meetings data, such as Archivus [5] for accessing processed and stored recorded multimodal meetings, including video and audio, electronic copies of all documents used or referred to in the meetings, handwritten notes made by participants during the meeting, as well as the text transcript of the meeting itself.

In this paper, a study is carried out on meetings where more than one written document is discussed, which highlights their features and characteristics. This study is an extension of our previous work about multimodal document alignment [6, 7]. The main aim of this study is to decide whether it is better to combine all discussed meeting documents within one resource that will be aligned with the dialog transcript (called jointed strategy), or to keep them as independent resources (disjointed strategy).

This paper is organized as follow; in section II our previous work on multimodal document alignment is briefly described. In section III, features that are related to multi-document meetings are discussed. In sections IV and V, the thematic alignment and an entailment process are studied and evaluated for the two strategies of arranging documents (jointed/disjointed). Finally, section VI presents some related work where documents or information sources are combined.

II. MULTIMODAL DOCUMENTS ALIGNMENT

In our previous studies, an alignment framework was developed in order to align written documents that are discussed during meetings with dialogs transcript [6, 7]. The aim of this alignment framework was to link the various parts of meeting dialogs to the related parts of discussed documents. Several alignment dimensions have been studied, mainly quotations, references and thematic. The three alignment dimensions aim at linking document to speech and vice versa, so we can easily answer questions such as “what was said about a particular part of a document?”, “when was a particular part of document being discussed?”, “which document part was discussed at time T?”, “in which order the various document / document parts are discussed during a meeting?”, etc. Quotation alignment links the parts of documents that are quoted by speakers. Reference alignment corresponds to reference, anaphora or cataphora, made by speakers to the various document parts [11]. Thematic alignment, studied further in this paper, is based on a thematic similarity between the document and speech transcript parts.

Detecting thematic links between these documents required first their decomposition into segments. The static documents are segmented into sentences or into logical blocks. The speech transcript is segmented into speaker utterances or turns. After the segmentation step, the generated segments are cleaned from stop words, and then remaining tokens are reduced into stems. Later on, the thematic similarity is computed between pre-processed segments using Cosine metric, which is based on terms co-occurrences in compared segments.

In order to highlight the more frequent terms in each segment S_i , TF.IDF coefficient (term frequency-inverse document frequency) is associated to each term t , and is computed according to the following equation:

$$TF.IDF_{t,S_i} = occ_{t,S_i} / \sum_{j=0}^N occ_{t,S_j} \quad (1)$$

This TF.IDF formula was chosen since it was more efficient than other formulae, especially the logarithmic based one. In addition to Cosine, other similarity metrics were used in our previous work (with and without TF.IDF) mainly Dice and Jaccard. However, Cosine+TF.IDF has generated the best results amongst several corpora [6], and therefore it was used in the current work.

The evaluation of the detected links is based on a comparison to a manual ground truth, which was performed

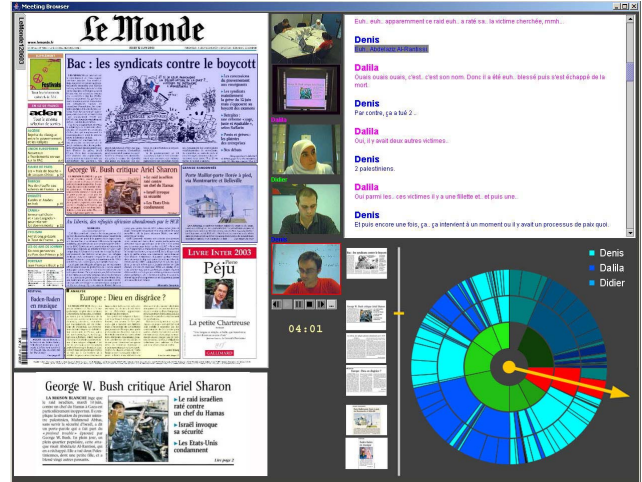


Figure 1. FriDoc multimodal browser for a press review meeting.

by a human. The manual annotation of a source segment involves detecting the id(s) of the best segment(s) from the target file that could be aligned with it (i.e. the best match according to a human annotator). The comparison of the generated alignment links with the manual ground truth consists in associating a binary score to each detected alignment link, “1” if it matches with the manual ground truth, and “0” otherwise. In order to measure the performance of the thematic alignment process, recall, precision and efficiency measure F are used.

Several meeting domains were considered in the evaluation of our multimodal document alignment framework, mainly press reviews and decision-making meetings (furniture proposals and movie club). Among these corpora, two types of meetings are distinguished, mono-document meetings, and multi-document meetings where several documents are discussed.

III. DISCUSSING DOCUMENTS IN MEETINGS

Depending on their types and on their contents, documents discussed during meetings could be classified into various categories. They might be classified into primary or secondary according to their role during the meeting. They are considered as totally discussed or partially discussed according to the usage of their content by speakers. Finally, meeting documents are regarded as having homogeneous or heterogeneous content depending on the similarity of their respective contents, or parts of their contents. From another side, the scenario of the meeting varies according to the way of presenting and discussing documents, if it is made in parallel or in sequence. These various classifications of discussed documents and meeting scenarios are detailed in the following sections.

A. From the document perspective

A document is considered as *primary* if it is discussed or referenced along the meeting. This might happen if many documents are present during the meeting but only one of them is discussed, which reduces the meeting from a multi-document into a mono-document one. Another use case of this documents category is due to the nature of their contents which leads to use them in parallel with other documents. Meeting agenda is an example of this category of documents whose elements that constitute the outline of the meeting are usually followed by speakers during the whole meeting. Primary documents might be very useful in order to detect the topics of the meeting, mainly by projecting their topics segmentation on the dialog transcript. From another hand, a document is considered as *secondary* if it is discussed only during a specific period of the meeting, and is preceded or followed by other documents. Examples of this category are documents discussed during press reviews, where several speakers present several newspapers one after the other.

A document might be categorized as *totally* discussed if almost all its content is mentioned, either in one speaker turn (i.e. the speaker discussing this document is not interrupted by any other speaker), or in several turns. This categorization of documents is based on the importance of their content during the meeting, and also on the proportion of its content being discussed. Totally discussed documents could be very functional in order to increase the automatic speech recognition accuracy, especially for spontaneous speech such as in meetings and presentations. If only a part of the document is used during the meeting, then it is regarded as *partially* discussed.

Depending on the meeting nature, speakers might discuss and present support documents that have *heterogeneous* content, i.e. they have completely distinct thematic content. In other cases, discussed documents might share a *homogeneous* content. If we take the example of a movie club meeting (described in section 5.1), then many documents, including posters and slideshows, show information about same movies. In press review meetings, the current event happening in the date of recording is usually shared between the various newspapers discussed in the same meeting. This similarity between documents contents leads to align them with the same speech transcript part. This aspect of homogeneity of documents has an impact on the precision of the alignment framework results, which generates ambiguous alignment links between the speech and the documents (e.g. links *x* and *y* in Fig. 2.a to documents D2 and D3 respectively, having the same speech source segment).

B. From the dialog perspective

Regarding the time interval during which documents are consulted, two meetings scenarios are distinguished, either consulting documents in *parallel* or in *sequence*. In *parallel*

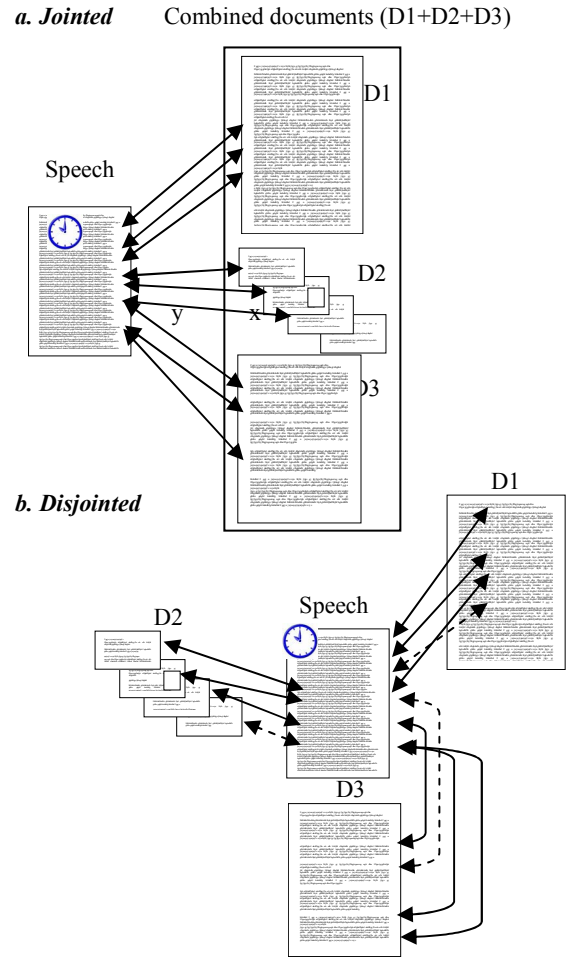


Figure 2. The two strategies (jointed/disjointed) for arranging documents for the thematic alignment in a multi-document meeting. Dashed links represent new links that appear in disjointed alignment strategy only.

meeting scenario, speakers might want to compare the content of several documents regarding one topic. For instance, compare the various newspapers reports about the same event (e.g. war in Iraq in press review meetings). Speakers might also want to complement missing information about the discussed event, by referring to a document D2 whilst discussing a document D1. In the *sequential* meeting scenario, documents are presented independently each one after the other. The variety of scenarios according to which documents are discussed in a meeting, either in parallel or in sequence, means that the thematic alignment of these documents with the dialog transcript should be performed according to the same scenarios. Thus if documents are discussed in sequence one after the other, then they need to be jointed within one global document to be aligned with the dialog transcript (Fig. 2.a). If documents are discussed in parallel, then a disjointed alignment process is required, where the dialog transcript is aligned with each individual document

independently from other documents (Fig. 2.b). In the following sections the two terms, disjointed and jointed, will be associated to the thematic alignment and the entailment process of the generated thematic links.

IV. ALIGNMENT IN MULTI-DOCUMENT MEETINGS

Our previous strategy to deal with documents discussed in multi-document meetings consisted in extracting and then merging their content in one resource [6, 7]. Our assumption was that all meeting documents are related to the same speech transcript, and thus they should be merged in one global document. This means that the alignment process is performed according to the jointed strategy, as presented in Fig. 2.a. In this figure, the three available documents (D1, D2 and D3) in the meeting were combined together, and then aligned with the dialog transcript as being one document. However, in our recorded multi-document meetings there was a similarity of the content of some documents which has affected our alignment precision. Our aim in the current section is to put the focus on these multi-document meetings having homogeneous document contents, by performing the thematic alignment process between documents and dialogs transcript according to the disjointed strategy, i.e. align the dialog transcript with each document individually, as if these documents were discussed by speakers in parallel. The two strategies of alignment (disjointed/jointed) might generate similar or dissimilar alignment results, depending on how much does the matching metric, which is used to detect these alignment links, rely on the interaction between the various document contents. In the following section, the data used to measure the effect of the two alignment strategies on the thematic alignment are described.

A. Test data

Among three meeting corpora available in our previous work, 6 meetings that contain more than one document have been selected for the current study, respectively four French press reviews, one English movie club and one English furniture proposal meeting (Table 1).

The four press review meetings, which were recorded at the smart meeting room in Fribourg [14], include respectively two, three, four and four documents. The discussed documents represent the front page of newspapers containing several topics. Within each of these press review meetings, the presented documents were almost totally discussed. Two among the four meetings were recorded on successive days in April 2003, which led to have partially homogeneous documents (war in Iraq). The two other meetings were recorded few months later on (in July and October 2003 respectively), which generated heterogeneous documents. The general meeting scenario for this corpus was mainly sequential, i.e. discussing one document after the other. The movie club meeting, recorded at the Idiap Smart Meeting Room [9], contains three written

documents including the agenda of the meeting and two other documents presenting a selection of movies, three posters for a selected movie, and two slideshows. The first three documents in this corpus were partially homogeneous, totally discussed and presented in sequential way. Whereas the three posters were totally homogeneous, partially discussed and presented in parallel. The last corpus, the furniture proposal meeting, has been recorded by ISSCO Research group at the Idiap Smart Meeting Room [9, 10]. In this meeting, the furniture that should be chosen for a reading room is discussed. In this meeting six documents are discussed, among them three slideshows are presented, one of them corresponds to the agenda of the meeting, and the two others represent some ideas for choosing the furniture (furniture item snapshots, prices, etc.). The main characteristic of one of the slideshows is the similarity of textual content of some of its parts (furniture items names), even if the accompanied snapshots of these similar items were different. In addition to the slideshows, three written documents were considered in this meeting. These three documents were almost not used during the meeting.

For each of the evaluated meeting corpora, appropriate segmentation methods are used in order to prepare the documents and the speech transcript for the thematic alignment process. For press reviews, speech transcript was segmented into speaker turns. The written documents were segmented into logical blocks, where a logical block corresponds to a newspaper article. For the furniture proposal and the movie club meetings the speech transcript was segmented into speaker utterances (smallest homogeneous part of a turn), and the documents/slideshows were segmented into logical blocks, where a logical block corresponds to a document section and to a slide respectively. The results of the segmentation of all meetings are presented in Table 1.

B. Effect of documents arranging strategies on alignment

The thematic alignment is based on a thematic similarity between segments, computed using similarity metrics. For each segment from the source file (speech transcript or document), all similar segments from the target file (document or speech transcript) are detected. In our evaluation, only thematic links having a similarity that overcomes 10% (chosen after observation of significant links) are considered as relevant and thus preserved. The used similarity metric was the Cosine which generated the best scores compared to other metrics mainly Dice and Jaccard [6]. Two varieties of Cosine are used, without then with term frequency TF.IDF. The obtained scores based on the disjointed /jointed alignment strategies are described in the following paragraphs.

In order to perform the disjointed alignment strategy, the speech transcript was first aligned with each discussed document individually. Later on, the generated alignment links with the respective documents were combined within

one alignment file and then evaluated in comparison to manual ground truth.

1) Alignment with Cosine

When the thematic links are detected using the Cosine metric, the two alignment strategies (disjointed/jointed) have generated the same scores (Table 2), where the obtained F value was 48% for press reviews corpus, 7.5% for movie club corpus and 7.6% for the furniture proposal corpus. This means that the combination of documents does not affect the alignment scores. This is mainly due to the non-consideration of the TF.IDF that is influenced by the interaction between the various documents contents.

2) Alignment with Cosine+TF.IDF

The integration of the TF.IDF in Cosine metric has generated different scores for the two alignment strategies (see Table 3). The best results were obtained using the jointed strategy for all meetings. When using the disjointed strategy, the F value for press reviews, movie club and furniture proposal respectively decreased from 65.7% to 58%, from 36% to 31%, and from 29% to 24%. The difference in performance between the two strategies is due to the impact of the combination of documents. By combining all documents within one resource in the jointed strategy, the most important terms in the meeting have more chance to appear, which decreases the number of noisy links and thus generates a better precision.

TABLE 1: STATISTICS OF DOCUMENTS AND DIALOGS IN THE EVALUATED MULTI-DOCUMENT MEETINGS

Corpus	Press reviews	Movie club	Furniture proposal	Total (min)
Discussed document	13	8	6	29
Document segment	114	48	44	206
Dialog duration	43	48	37	128
Speech segment	582	1348	935	2865

TABLE 2: THEMATIC ALIGNMENT RESULTS USING JOINTED/DISJOINTED STRATEGIES (COSINE)

Corpus	Press reviews		Movie club		Furniture proposal	
	Join.	Disj.	Join.	Disj.	Join.	Disj.
Recall	67	67	71	71	76	76
Precision	47	47	4	4	4	4
F	48	48	7.5	7.5	7.6	7.6

TABLE 3: THEMATIC ALIGNMENT RESULTS USING JOINTED/DISJOINTED STRATEGIES (COSINE+TF.IDF)

Corpus	Press reviews		Movie club		Furniture proposal	
	Join.	Disj.	Join.	Disj.	Join.	Disj.
Recall	58.7	57	51	54	42	59
Precision	81	63	28	22	23	20
F	65.7	58	36	31	29	24

V. ENTAILING THEMATIC LINKS

In the current section, an entailment process of the generated thematic alignment links between speech transcript and documents is studied. This entailment process aims to prune alignment links especially by adding missing links, which should help to enhance the alignment scores. The entailment of thematic links is based on various features of terms within aligned segments, linguistic, structural, etc. In this paper we focus on the linguistic features, especially the matching of named entities (NE) between aligned segments. Our assumption is that the existence of common NE between two compared segments means that they are thematically similar, and thus a thematic alignment link is established between them if it was not yet generated by the thematic alignment process. The NE matching between two segments is defined on the basis of a NE similarity score that is computed according to equation Eq. (2), where NE_{S_i} , NE_{S_j} are the named entities detected within segments S_i , S_j respectively. It should be noted that other NE similarity formulas were tried out, which consisted in considering the denominator either as $|NE_{S_i}|$ or $|NE_{S_i} \cup NE_{S_j}|$. However, all formulas have generated similar scores.

$$NE \text{ similarity } (S_i, S_j) = |NE_{S_i} \cap NE_{S_j}| / |NE_{S_j}| \quad (2)$$

In the following example, the speaker utterance and the document sentence share one movie name “The Big Lebowski”, thus the NE similarity score is 1:

- <utterance id="9" StartTime="28.6" EndTime="32.8"> Okay, I'll go with <NE>**The Big Lebowski**</NE>, since posters are already ready </utterance>
<sentence id="15" doc-id="D1">Friday 29th April, 20h, <NE>**The Big Lebowski**</NE> by Joel Coen</sentence>

The detection of named entities in respective segments is based on predefined gazetteers. Three distinct gazetteers containing named entities for the three respective meeting corpora were manually created. The press reviews gazetteer contains names of famous personalities in politic, sport, etc., as well as countries, cities. The movie club gazetteer contains names of movies, actors, etc. The furniture proposal gazetteer contains names of furniture items.

After detecting the thematic links and computing the NE similarity between all speech transcript segments and documents segments, the generated thematic alignment links are compared to the NE links, in order to add missing thematic links. Several thresholds were defined in order to decide whether to consider a NE link as thematic link or not (see the program code at Fig. 3). First, a threshold Th_E that aims to filter candidate NE links according to their similarity values is defined. If a NE similarity score overcomes Th_E that is dynamically varied from 10% to 100%, then a new thematic link might be established between the corresponding segments (line (4) in Fig. 3).

```

Program Entailing thematic links
1. For each segment  $S_i$  from the speech transcript
2. For each segment  $S_j$  from the documents
3. Compute  $NE\_similarity(S_i, S_j)$ 
4. If  $NE\_similarity(S_i, S_j) > Th_E$ 
5. If  $S_i$  is not yet thematically aligned with  $S_j$ 
6. If  $\exists$  one thematic link  $T(k_1, l_1)$  where  $k_1 \in [S_i - Th_M, S_i]$ 
7. If  $\exists$  one thematic link  $T(k_2, l_2)$  where  $k_2 \in [S_i, S_i + Th_M]$ 
8. Establish a thematic link between  $S_i$  and  $S_j$ 
End

```

Figure 3. Program code for our entailment process.

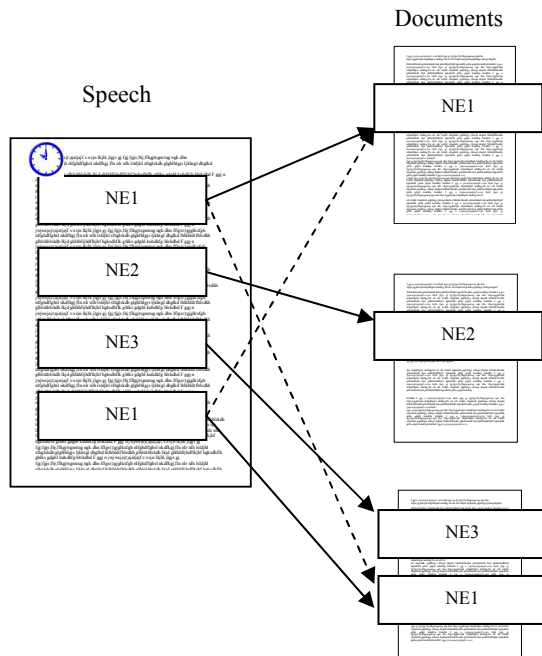


Figure 4. NE matching between speech transcript and discussed documents. Black lines correspond to relevant NE links, dashed lines correspond to ambiguous NE links.

Later on, for each source speech segment S_i for a given NE_{S_i, S_j} link with a document segment S_j , a window of fixed size Th_M (varied dynamically from 0 to 10 segments) is defined along the speech transcript segments for segments preceding and following S_i (lines (6) and (7) in Fig. 3). The window size corresponds to the interval (in term of number of speech segments) in which the speech segment S_i has neighbours speech segments that are sources for thematic links. The Th_M threshold aims to filter the ambiguous NE links which are mainly due to the similar content of some documents or document logical blocks (e.g. dashed links in Fig. 4). At the same time Th_M threshold aims to strengthen

the relevant NE links that are surrounded by neighbours' thematic links. That means a specific NE link is not considered as relevant (to create a thematic link) if it is not surrounded by existing thematic alignment links within a scope of $\pm Th_M$ speech transcript segments. If a candidate NE_{S_i, S_j} link between two segments S_i and S_j satisfies the two conditions related to the thresholds Th_E and Th_M , then a new thematic link is established between S_i and S_j (if it does not exist yet).

From another side, and in order to resolve the NE ambiguity that affects the thematic alignment (Fig. 4), our attempt was based on the exploitation of the disjointed strategy whilst detecting the NE based links between speech transcript and documents, instead of the jointed strategy. Our assumption is that detecting NE links between the speech transcript and each of the documents individually, then filtering these links according to the thresholds defined previously (Th_E and Th_M), should help to preserve only relevant NE links, and thus to create relevant thematic links. In order to validate this assumption, experiments were carried out using the disjointed and jointed strategies for both thematic alignment and NE links detection. These experiments are described in the following sections.

A. Impact of alignment strategies on the entailment

In order to measure the impact of the alignment strategies (disjointed/jointed) on resolving ambiguous NE links, our entailment process has been performed on the three corpora described previously, according to the two strategies. In the first set of experiments, the thematic links were detected according to the jointed strategy, and then entailed according to the same strategy. In the second set of experiments, the latter experiments are performed using the disjointed strategy, and then the individual thematic/NE links are grouped and evaluated.

In the following paragraphs, the obtained entailment scores for the three corpora, based on the jointed/disjointed strategies, are presented. For each strategy in each corpus, the methods considered are:

- M1: Cosine+TF.IDF method (as initial thematic alignment method);
- M2: NE based matching method;
- M3: Entailment method (Cosine+TF.IDF+NE), using only the Th_E threshold to filter non relevant NE links;
- M4: Entailment method (Cosine+TF.IDF+NE), using Th_E threshold. In addition Th_M threshold is used from one neighbour side only, either before or after the current segment (i.e. either condition (6) or (7) in Fig. 3);
- M5: Entailment method (Cosine+TF.IDF+NE) using Th_E , and Th_M from both sides (conditions (6), (7) in Fig. 3).

B. Entailment results

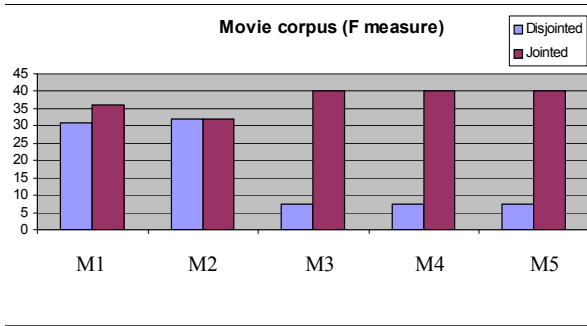


Figure 5. Evaluation results for the movie corpus.

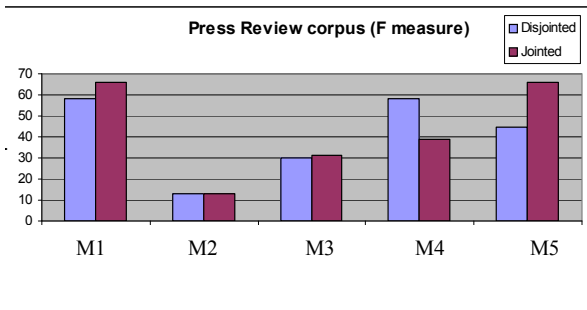


Figure 6. Evaluation results for the press review corpus.

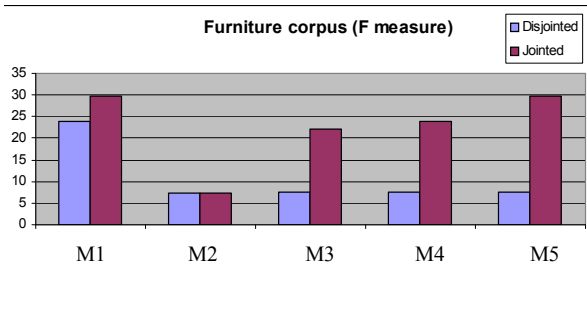


Figure 7. Evaluation results for the furniture proposal corpus.

The entailment results for the three corpora are presented in Fig. 5, 6 and 7. A common aspect that was remarked among all these corpora is that the same scores are obtained for NE matching (method M2), using either the jointed or disjointed strategy. This is explained by the fact that the TF.IDF is not considered in the formula that computes the NE similarity, which keeps the scores independent from the way of arranging documents.

The obtained entailment scores for the movie club corpus (Fig. 5) show the interaction between the various defined parameters (Th_E and Th_M) and the NE matching process to generate relevant NE links, which might depend on the nature of the corpus. In this corpus, a NE link is

relevant if it has a neighbour thematic link in a scope of $Th_M = \pm 7$ speech segments, and if its NE score $\geq 60\%$ (M5).

If only one side of neighbour segments is considered (M4), then the best margin threshold Th_M decreases to 5 segments. If the margin threshold Th_M is not considered (M3), then the NE similarity should be at least 70% to be a relevant NE link. Under these conditions, the initial F value (M1) increased from 36% to 40.2%. This positive effect of the entailment, when thresholds Th_E and Th_M are considered, shows its relevance to resolve NE ambiguity that exist between speech transcript and homogeneous documents. For the press review corpus (Fig. 6), the entailment process with its variations M3 and M4 did not improve the alignment scores. Many added thematic links were noisy which decreased the precision and thus the F value. The best F value was obtained under the jointed strategy using M5 method when $Th_E = 0$ and $Th_M = \pm 0$, which is similar to the initial value using method M1 (66%). This is due to the nature of this corpus, where documents are totally discussed and only few of them are partially homogeneous, which led to a non-necessity of the entailment. For the furniture proposal corpus (Fig. 7), and similarly to the previous corpus, the entailment process did not improve the alignment scores. The best F value was obtained under the jointed strategy using method M5 when $Th_E = 0$ and $Th_M = \pm 0$, which is similar to the initial score using M1 (29.7%). This negative effect of the entailment process is explained by the nature of this corpus, where 50% of documents were not discussed, and those discussed had a poor textual content (only names of furniture items). All these factors led to the non-effectiveness of the entailment process to add relevant thematic links.

After these various experiments, we conclude that the jointed arrangement of documents is more significant than the disjointed one to generate relevant thematic alignment links, for all evaluated corpora. Regarding the entailment process, the obtained results for the movie corpus showed that the jointed arrangement of documents is still the most relevant where the F value has increased from 36% to 40.2%. However, the entailment of thematic links in the other corpora was influenced by the features of documents being discussed during meetings, mainly total discussion of documents with heterogeneous content in press reviews, or partial discussion of documents with poor textual content in furniture proposal. The two latter scenarios of discussing documents made the entailment non-necessary and non-efficient to add pertinent thematic links.

VI. INFORMATION FUSION: RELATED WORK

The strategy of combining the various sources of information, known also as information fusion or integration, was considered in several research areas such as in computer security, image analysis and in other data mining fields. This integration of information refers to different representations of knowledge, either in textual form or in rich media form.

In a topic segmentation study [15], the accuracy of generated thematic segments for multiple documents was improved when the weighted mutual information, which is

based on shared information between similar documents, was used. Similar improvement was obtained in multimedia information retrieval [3] and multimedia distributed digital libraries [2] fields, when multiple sources of information were combined. In multi-document summarization [13], Sakai and Masuyama have proposed a system that presents to the user a list of keywords extracted from the overall documents to summarize. The user will choose the relevant terms for his summary from this list. The selected keywords are then used by the system to personalize and control the summary of multiple documents, according to the user needs. Finally, in a bi-modal method for thematic segmentation of meetings [8, 6], the shared information between documents discussed during the meeting and the dialog transcripts was exploited in order to narrow and then generate the meeting topics.

VII. CONCLUSION

In this paper various features of multi-document meetings have been studied, where several document categories are distinguished based on their use, their roles in the meeting and the similarity of their contents. Later on, two distinct document arrangement strategies (jointed/disjointed), related to the thematic alignment and the entailment the generated thematic links have been described and evaluated. The experiments carried out using the two strategies on several meeting corpora showed that combining discussed documents is more significant not only to generate relevant thematic alignment links between speech transcript and documents, but also to entail these thematic links. From another side, experiments showed that the used entailment process (based on NE matching), has different effect on the various meeting corpora, either adding relevant missing thematic links, or adding noisy links. This effect depends on the scenario of the meeting and the nature of its discussed documents, i.e. if they are totally or partially discussed, if they have heterogeneous or homogeneous textual content, etc. In the future, the effect of the features of discussed documents on other alignment strategies between dialog transcript and documents, principally the so-called references [11] and quotation [6] alignments, will be studied. From another side, other entailment techniques will be considered to detect and delete non-relevant thematic links. Finally, advanced techniques will be developed in order to resolve the ambiguous NE matching between dialog transcript and discussed documents.

REFERENCES

- [1] M. Ailomaa, M. Melichar, A. Lisowska, M. Rajman, and S. Armstrong, "Archivus: A Multimodal System for Multimedia Meeting Browsing and Retrieval," Proc. Conf. Association for Computational Linguistics (ACL 06), Australia, 2006.
- [2] J. Callan, F. Crestani, H. Nottelmann, P. Pala, and X. M. Shou, "Resource Selection and Data Fusion in Multimedia Distributed Digital Libraries," Proc. ACM SIGIR Conf. 26th International ACM Conference on Research and Development in Information Retrieval, Canada, 2003.
- [3] J. Kludas, E. Bruno, and S. Marchand-Maillet, "Information Fusion in Multimedia Information Retrieval," Proc. Conf. 5th International Workshop on Adaptive Multimedia Retrieval, France, 2007.
- [4] D. Lalanne, R. Ingold, D. von Rotz, A. Behera, D. Mekhaldi, and A. Popescu-Belis, "Using Static Documents as Structured and Thematic Interfaces to Multimedia Meeting Archives," Proc. Machine Learning for Multimodal Interaction (MLMI 05), LNCS 3361, 2005, pp. 87-100.
- [5] A. Lisowska, "Multimodal Interface Design for Multimedia Meeting Content Retrieval". PhD thesis, Genève, Switzerland, 2007.
- [6] D. Mekhaldi, "A Study on Multimodal Document Alignment: Bridging the Gap between Textual Documents and Spoken Language". PhD thesis, Fribourg, Switzerland, 2006.
- [7] D. Mekhaldi, "Multimodal Document Alignment: towards a Fully-indexed Multimedia Archive," Proc. Multimedia Information Retrieval Workshop, SIGIR, the Netherlands, 2007.
- [8] D. Mekhaldi, D. Lalanne, and R. Ingold, "Using Bi-modal Alignment and Clustering Techniques for Documents and Speech Thematic Segmentations", Proc. 13th Conference on Information and Knowledge Management (CIKM 04), Washington D.C., U.S.A, November 8-13, 2004, pp. 69-77
- [9] D. Moore, "The Idiap Smart Meeting Room," Technical Report, Idiap, 2002.
- [10] A. Popescu-Belis, M. Georgescu, A. Clark, and S. Armstrong, "Building and Using a Corpus of Shallow Dialogue Annotated Meetings," Proc. International Conference on Language Resources and Evaluation (LREC 04), Portugal, 2004.
- [11] A. Popescu-Belis, and D. Lalanne, "Detection and Resolution of References to Meeting Documents," Proc. Machine Learning for Multimodal Interaction (MLMI 06), LNCS 3869, 2006, pp. 64-75.
- [12] M. Rigamonti, D. Lalanne, and R. Ingold, "FaericWorld: Browsing Multimedia Events through Static Documents and Links," Proc. International Conference on Human-Computer Interaction (Interact 07), Brazil, 2007, pp. 102-115.
- [13] H. Sakai, and S. Masuyama, "Multiple-Document Summarization System with User Interaction," Proc. International Conference on Computational Linguistics (Coling 04), Switzerland, 2004, pp. 1001-1007.
- [14] The Smart meeting room recorded data. Available from <http://diuf.unifr.ch/im2/>
- [15] B. Sun, P. Mitra, C. L. Giles, J. Yen, and H. Zha, "Topic Segmentation with Shared Topic Detection and Alignment of Multiple Documents," Proc. SIGIR, the Netherlands, 2007, pp. 199-206.