Detection and Resolution of References to Meeting Documents

Andrei Popescu-Belis[1] and Denis Lalanne[2]

[1] University of Geneva,
School of Translation and Interpreting (ETI), TIM/ISSCO,
40, bd. du Pont d'Arve,
CH-1211 Geneva 4, Switzerland
andrei.popescu-belis@issco.unige.ch
[2] University of Fribourg,
Faculty of Science,
DIUF/DIVA,
3, ch. du Musie,
CH-1700 Fribourg, Switzerland
denis.lalanne@unifr.ch

Abstract. This article proposes a method for document/speech alignment based on explicit references made in speech to documents and parts of documents, in the context of multimodal meetings. The motivation and the main components of the method are first described briefly. Then, the article focuses on the two main stages of dialogue processing: the detection of the expression referring to documents in transcribed speech, and the recognition of the documents and document elements that they refer to. The detailed evaluation of the implemented modules, first separately and then in a pipeline, shows that results are well above the baseline, and that the various features of the proposed algorithms are all relevant. The integration of this document/speech alignment technique with other ones is finally discussed.

1 Introduction

Documents are often the main support for presentations and discussions that take place during meetings. For instance, slides are used for talks, and are generally presented in sequence, being thus naturally aligned with the presenter's utterances. This is not the case, however, when the supporting documents are not so obviously set into focus, for instance when reports or articles are discussed during a meeting. In this case, it becomes necessary for proper meeting browsing to detect the temporal alignment of speech and of documents, or, more specifically, parts of documents such as sections, figures or pages. This kind of alignment has to be derived from the linguistic content of speech and documents, and from clues in other modalities.

We study in this paper the alignment of (transcribed) speech and (electronic) documents based on the references made explicitly in speech to document elements, such as "the title of our latest report" or "the article about X". To

implement and evaluate this method, a number of processing modules are described in Section 2, in particular a brief outline of techniques for document structuring (2.2). Section 3 defines reference-based document/speech alignment (3.1) and describes the methods we propose for the two stages, namely the detection of expressions referring to documents (3.2) and the recognition of the document parts they refer to (3.3). The press-review meetings used in this experiment, and the novel evaluation methods, are described in Section 4. Results appear in Section 5. Finally, the place of reference-based alignment among other techniques is discussed in Section 6.

2 Document/Speech Alignment for Meeting Browsing

A meeting processing and retrieval application helps a professional who has missed a meeting to browse through its content directly to the most relevant points, without, say, viewing or listening the entire recording. The application is also relevant to someone who attended a meeting but would like to review some points (e.g., the decisions), or to someone who would like to track the progress of issues over several meetings. The meeting browser must enhance the retrieval function. For instance, once an episode of interest has been spotted in a meeting, the user should be able to explore the transcript, or watch/listen the recording of the episode, or check the documents that were discussed.

2.1 Importance of References to Documents for Meeting Browsing

When meetings deal with one or several documents, it becomes important to align in a precise manner each episode of the meeting to the sections of the documents that are discussed in it, and vice-versa. Thus, a meeting browser can retrieve the episodes of a meeting in which a particular section of a document was discussed, to find out what was said about it. Conversely, the application can also display in real-time the documents relevant to a given episode of a meeting, while the user browses through the episode. A user requirement study has shown that many queries require information related to meeting documents [1].

The references made in speech to the meeting documents are a fined-grained type of information that allows document/speech alignment. Using these references, the multi-media rendering of the meeting can be enhanced as shown in Figure 1. The transcript is color-coded for the different speakers, and retains timing information. The expressions that refer to documents are coded, in our implementation, as hyperlinks towards the right part of the window: clicking on such a link highlights the article referred to by that expression. The resolution of references to documents is thus a cross-channel process that enhances dialogue and document browsing.

The automated understanding of references to documents in meeting dialogs requires several preliminary operations, as outlined in Figure 2. The most significant ones are: (1) the generation of a transcript of the utterances produced by each speaker (since the performance level of automated speech recognition

Fig. 1. Aligned browsing of meeting transcript and documents. Clicking on a referring expression (underlined) highlights the corresponding document element.

was low on our data, we use manual transcripts here); (2) the generation of an abstract representation of each document structure (see next sub-section); (3) the detection of the expressions from the transcripts that refer to meeting documents; and (4) the identification of the document element each of these expressions refers to. The third and fourth tasks are the main object of the remaining sections.

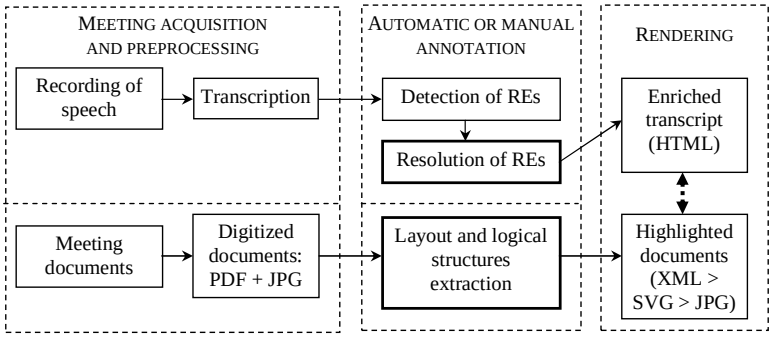

Fig. 2. Components of an application for the resolution of references to documents.

2.2 Construction of the Logical Structure of Documents

Since PDF can be easily generated from almost every document format, it has become very common for disseminating nearly any kind of printable documents. However, because its use is limited to displaying and printing, its use for retrieval and extraction is considerably reduced. Our experience has clearly shown that the reading order of a text is often not preserved, especially in documents having a complex multi-column layout, such as newspapers. Even recent tools that

extract the textual content of PDF documents do not fulfill our requirements, because they do not reveal the physical and logical structures of documents. To overcome this limitation, we designed and implemented Xed, a tool that reverse engineers electronic documents and extracts their layout structure. The approach merges low-level text extraction methods with layout analysis performed on synthetically generated TIFF images [2]. Xed has been tested with success on various document classes with complex layouts, including newspapers.

In the present study, we considered that newspaper front pages have a hierarchical structure, i.e. a `Newspaper` front page bears the newspaper's `Name`, the `Date`, one `Master Article`, zero, one or more `Highlight`s, one or more `Article`s, etc. Each content element has an `ID` attribute bearing a unique index. An `Article` is composed of a `Title`, a `Subtitle`, a `Source`, the `Content` (mandatory), and one or more `Author`s and `Reference`s.

To obtain ground truth data for the application to reference resolution—i.e. 100% correct document structure—the correct document segmentations have been validated manually according to the structure mentioned above, encoded in an XML DTD. Information about the layout structure, i.e. the bounding boxes of each logical block, topological positions, fonts, etc., was stored in separate annotation files, using pointers to the ID attributes of the logical blocks.

3 Reference-based Document/Speech Alignment

3.1 What Are References to Documents?

From a cognitive point of view, speakers use *referring expressions (REs)* to specify the entities about which they talk, or more accurately the representations of entities in the speaker's mind [3]. When speakers discuss one or more documents, as in press-review meetings, they refer quite often explicitly to documents or various parts of documents (e.g. "the title", "the next article", etc.).

Reference resolution amounts to the construction of links between each RE and the corresponding document element. For example, if a speaker says: "I do not agree with the title of our latest report", then 'our latest report' refers to a document available as a computer file, and 'the title of our latest report' refers precisely to its title, an element that can be retrieved from the file.

Two important notions are *coreference* and *anaphora*. RE_1 is coreferent to RE_2 if they refer to the same entity, here a document element. RE_2 is an anaphor with respect to RE_1 if the element it refers to cannot be identified without making use of RE_1, called its antecedent. In the following example, *the first article* is the antecedent and the pronoun *it* is the anaphor: "*The first article* is particularly relevant to our company. *It* discusses ..." The resolution of references to documents offers the advantage of a restricted set of entities, when compared to anaphora resolution [4] or coreference resolution [5, 6].

The process has two main stages: (1) identify the REs that refer to documents; (2) identify to which document and document element each RE refers. In a preliminary study [7], only the second stage could be automated: no results where therefore available for the entire process.

3.2 The Detection of REs

We designed a grammar-based component that spots the REs referring to documents in the transcript of meeting dialogues (in French). We chose to consider a manual transcript because an automatic one would contain too many errors, which would make the evaluation of our alignment impossible.

The CLaRK XML environment[3] [8] was used to write a tokenizer and a grammar for the detection of REs. Each channel is segmented into utterances following the SDA.XML format used in our project [9].

In order to detect REs that refer to documents, we created a set of pattern matching rules that apply to the words of the utterances, with sometimes a left or a right context. The rules were derived from a prior analysis of the REs annotated by humans in the corpus [7]. Most of the references are made to entire articles, using REs such as (translated) "the article", "the [first|last] article", "a short article about *X*", or "the front page of Le Monde" ('Le Monde' being the name of a French newspaper). The challenge in writing the detection grammar was to combine *a priori* knowledge about the form of REs with the empirical observations on our corpus, and to tune the coverage of the grammar to avoid too many false positives or true negatives (corresponding respectively tp precision and recall errors).

The resulting grammar has about 25 pattern matching rules—but since most of them contain one or more logical 'or' and optional tokens, they are equivalent to several hundred pattern matches. The application of the grammar by the CLaRK tool results in the recognized REs being tagged as `<re>...</re>` in the XML corpus. The accuracy of the tagging is commented in Section 5.1.

The main remaining problem of this method—apart from increasing the coverage and accuracy of the grammar—is the intrinsic ambiguity of certain REs, which may or may not refer to documents depending on their context. A typical example are the pronouns (such as 'it') and the indexicals (such as 'this' or 'this one'), which cannot be tagged as referring to documents without being resolved at the same time, i.e. attached to their antecedent. We currently study a classifier for this task, which could use information about the surrounding REs to decide whether a pronoun refers to a document or not. In the meanwhile, we tested several pattern matching rules, and kept the ones that increased recall without reducing precision too much. The failure to detect the pronouns is, however, quite penalizing for the document/speech alignment task.

3.3 The Recognition of References to Documents

Once the REs are detected, another component must recognize to which document and document element each of them refers, from the set of potential referents that is derived from the document structure. A first idea is to consider co-occurrences of words in the speech transcript and in the documents. For each RE, its words and the words surrounding it in the same utterance are

[3] Available at: `http://www.bultreebank.org/clark/`.

matched, using the cosine metric, with the bag of words of each logical block of the document: article, title, author, etc. Words within the RE count double the ones surrounding it. The most similar document element is considered to be the referent of the RE, provided the similarity value exceeds a fixed threshold (confidence level).

The theories of reference resolution acknowledge, however, the importance of keeping track of the referents that were mentioned, in particular of the current referent [10]. We integrated this features into a more complex algorithm, which attempts to identify anaphoric vs. non-anaphoric REs by matching them against a list of typical anaphors, then to processes each RE accordingly.

The algorithm determines first the document referred to, from the list of documents associated to the meeting. REs that make use of a newspaper's name are considered to refer to the respective newspaper; the other ones are supposed to refer to the current newspaper, i.e. they are anaphors (this method does not handle complex references such as 'the other newspaper').

The algorithm then attempts to determine the document element that the current RE refers to. If the RE is anaphoric (e.g. 'the article' or 'it', provided they are not the first RE in the meeting), then its referent is the current document element—a very simple implementation of a focus stack [3, 10]. If the RE is not anaphoric, then co-occurrences of words are used as above to find the document element it refers to: the words of the RE and the surrounding ones are matched with document elements; the one that scores the most matches is considered to be the referent of the RE. Then, the 'current document' and the 'current document element' variables are updated, before processing the next RE.

Several parameters govern the algorithm, in particular the weights of the various matches between words of the RE, of its left/right context, and the words from document elements (title, author, body of article). There are thus nine weights, generated by the pairings of {RE_word, left_context_word, right_context_word} with {title_or_subtitle_word, author_word, contents_word}. Other parameters are the sizes of the left and right contexts, that is, the number of preceding and following words and utterances considered for matching. Evaluation results in Section 5.2 provide insights about the best values of these parameters.

4 Data Annotation and Evaluation

The data was recorded in the document-centric meeting room set up at the University of Fribourg. Several modalities related to documents were recorded, thanks to a dozen cameras and eight microphones. These devices are controlled and synchronized by a meeting capture and archiving application, which also helps the users organize the numerous data files [11].

In this study, we use 22 press-review meetings of ca. 15 minutes each, recorded between March and November 2003[4]. In these meetings, participants discuss (in French) the front pages of one or more newspapers of the day. Each participant

[4] Available at: `http://diuf.unifr.ch/im2/data.html`.

presents a selection of the articles; after a short monologue, a brief discussion ensues, then, the chair of the meeting shifts the focus to another article. The meetings were manually transcribed using the Transcriber tool[5] and exported as XML files. The structure of the documents was also encoded as XML files.

4.1 Annotation of Ground Truth REs and References

The annotation model for the references to documents was described in an earlier paper [7]. The main idea is to separate the annotation of REs, marked on the XML transcript as <re ID='...'>...</re>, from the annotation of the references to documents. These are encoded in a separate block at the end of the transcript as links between the index of the RE, a document filename, and an XPath designation of the document element that the RE refers to, in the XML representation of the document structure.

In a first pass, the annotators marked the REs, following simple guidelines. The most litigious cases were the impersonal references to the creator of an article, such as (in English) "and *they* say that..." We considered this as a reference to the author of the article, or at least to the entire article (the actual scoring procedure allows this ambiguity). However, REs that corresponded only to quotations of an article's sentences were not annotated, since they refer to entities mentioned in the documents, rather than to the document elements. A total of 437 REs were annotated in the 22 meetings.

The REs were then automatically indexed, and in a second pass, the annotator filled in directly the attributes of the references-to-documents links at the end of the transcript (these templates were also generated automatically for each RE). The annotators were instructed to fill in, for each RE, the name of the document and the XPath to the respective document element, using its ID. Examples were provided for XPath expressions. When in doubt, annotators were instructed to link the RE to the most general element, i.e. the article or even the entire front page.

Inter-annotator agreement for the second stage [7], with three annotators on 25% of the data, is 96% for document assignment and 90% for document element assignment (using the evaluation metric below). After discussion among annotators, we reached 100% agreement on document assignment, and 97% agreement on document elements.

4.2 Evaluation of RE Detection

The evaluation of the first processing stage is done by comparing the correct REs with those found automatically, in terms of precision and recall. Two difficulties must however be solved: to what extent some variability on the RE boundaries is tolerated? And how are embedded REs processed?[6]

[5] Available at: http://www.etca.fr/CTA/gip/Projets/Transcriber.

[6] Brackets indicate embedded REs in this example: "[the title of [the next article]]".

Given our intended applications, recognizing only a fragment of an RE is sufficient for the alignment, if it is correctly linked to a document element[7]. Therefore, if the <re>...</re> found by the RE detector are comprised within the correct ones, then the RE counts as a correct hit. This decision is somewhat similar to the MUC-7 guidelines [5], except that they considered also a minimal substring for each RE, below which an RE detector would not score a hit. For our application, this is quite unnecessary, and is long to annotate.

Regarding embedded REs, if only exact matches would count as correct, then they would pose no problem, as there would be no risk of confusion between REs. However, since in the present case "included" matches count as correct too, one should avoid counting them several times in the case of embedded REs. We propose therefore the following scoring method:

- Loop through all the correct RE following their chronological order in the meeting; for embedded REs, start with the deepest (hence smallest) ones.
  - For each correct RE, test whether the detector has found an RE included in, or equal to, the current RE:
    - if it has, remove this RE from the set of found REs (this ensures it will not be counted again for REs that embed the current one);
    - if it has not, then count one recall error.
  - When all correct REs have been tested:
    - recall error is the number of recall errors counted above, divided by the total number of correct REs;
    - precision error is the number of detected REs remaining in the list (unmatched by correct ones), divided by the total number of detected REs.

4.3 Evaluation of RE Resolution

If the resolution of REs is attempted on the correct set of REs, then the evaluation of the resolution is done simply in terms of correctness or accuracy [7]. For each RE the referent found by the system is compared with the correct one. Three scores are computed, then normalized by the total number of REs:

1. The number of times the document is correctly identified.
2. The number of times the document element, characterized by its ID attribute, is correctly identified.
3. The number of times the specific part of an article is correctly identified by its XPath (e.g., content, title, author, image).

The third score is necessarily lower than the second one, which is lower than the first one. In what follows, we will use only the first two, since our algorithms do not target sub-article elements yet.

[7] The correct linking is the responsibility of the second component. Our component is still operational on fragments of REs, if enough surrounding words are considered for matching with document elements.

When the resolution of REs is combined with their recognition, the evaluation method must be adapted to rule out wrong REs are linked to document elements, or to count missing REs (not linked to an element), or to count REs which are included into correct REs are still acceptable. We modified the RE scoring algorithm proposed at 4.2 in order to *synchronize* the detected REs with the correct ones. Indeed, when a correct RE is matched by the scorer to a response RE, the latter receives the same index as the former. The three scores above can now be computed by looking only at the REs that were correctly detected, and using the standard annotation file to find out whether they are correctly linked to document elements or not. REs wrongly detected, or REs not detected, do not contribute to the score, and the score is finally normalized by the number of correct REs.

5 Results

5.1 Scores for the Detection of REs

The grammar for the detection of REs (3.2) is evaluated in terms of precision, recall, and f-measure (4.2). The initial grammar based on prior knowledge and on corpus observation reaches $R = 0.65$, $P = 0.85$ and $f = 0.74$. Experiments with the grammar and subsequent tuning slightly improve these values. Naturally, when adding a rule that marks all third person pronouns as referring to documents, precision decreases dramatically, with insufficient increase in recall: $R = 0.71$, $P = 0.52$ and $f = 0.60$. Adding a rule that marks all indexicals as referring to documents produces an even lower performance: $R = 0.70$, $P = 0.46$ and $f = 0.56$.

It appears however that in the present context of press-review meetings, the indexicals 'celui-ci' and 'celui-là' ('this one' and 'that one', without their feminine counterparts) are almost always used to refer to articles. Therefore, the best scores are obtained after tuning and adding this rule: $R = 0.675$, $P = 0.875$ and $f = 0.762$. Without this particular rule, that is without rules for pronouns and indexicals at all, f-measure is only slightly lower, at 0.756. We aim in the future at using the left and right contexts of pronouns and indexicals to find which of them refer to documents, either by extending the above grammar, or by using a trainable classifier.

5.2 Scores for the Resolution of REs

Baseline scores are scores of very simple methods against which our algorithms can be compared (rather than against a 0% score). For RE↔document association, always choosing the most frequent newspaper leads to ca. 80% baseline accuracy. However, when considering meetings with at least two newspapers, the score of this random procedure is 50%, a much more realistic, and lower, baseline. Regarding RE↔document element association, if the referent is always the front page as a whole, then accuracy is 16%. If the referent is always the main article, then accuracy is 18%—in both cases quite a low baseline.

The RE resolution algorithm applied on the set of *correct* REs reaches 97% accuracy for the identification of documents referred to by REs, that is, 428 REs out of 437 are correctly resolved—or 93% if only meetings with several documents are considered. This is a very high score which proves the relevance of the word co-occurrence and anaphora tracking techniques.

The accuracy for document element identification is 67%, that is, 303 REs out of 437 are correctly resolved at the level of document elements. If we consider only REs for which the correct document was previously identified, the accuracy is 68% (301 REs out of 428). The figure does not change much because most of the REs are already correctly resolved with respect to the document.

The best scores are obtained when only the right context of the RE is considered for matching, i.e. the words after the RE, and not the ones before it. Empirically, the optimal number of words to look for in the right context is about ten. Regarding the other optimal parameters, a match between the RE and the title of a document element appears to be more important than one involving the right context, and much more important than matches with the content of the article: optimal weights are about 15 vs. 10 vs. 1. If anaphor tracking is disabled, the accuracy of document element identification drops at ca. 60%. We are now in the process of analyzing the errors to find systematic patterns, which could help us improve the algorithm.

5.3 Combination of RE Detection and Resolution

When the two modules are combined in a pipeline, their errors cumulate in an unforeseeable manner, which can only be assessed empirically, as in this study. The best configurations were selected for the two modules and, on a perfect transcript, the obtained result was only 60% document accuracy (265 REs out of 427) and 32% document element accuracy (141 REs out of 437). If we compute document element accuracy only on the REs which have the correct document attached, the score is 46% (123 REs out of 265). The main observation is thus that error rates do not combine linearly. If they did, the scores would have been respectively ca. 73% and ca. 50%.

The reason for the lower scores obtained lies quite certainly in the context-based algorithm used for RE resolution, in which each RE depends on the correct resolution of the previous one, thanks to the monitoring of the "current document element". When too many REs are missing (since recall is only 67%), and especially when all the pronouns are missing, this algorithm loses track of the current document element. Therefore, an improvement of RE detection should considerably increase the overall score.

6 Other Document/Speech Alignment Techniques

The resolution of references to documents is only one method for the cross-channel alignment of what was said during a meeting with the meeting documents. Document/speech alignment enhances the multi-media rendering of the

meetings by query and browsing interfaces. Other methods for document/speech alignment are *citation-based alignment*, a pure lexical match between terms in documents and in speech transcription, and *thematic alignment*, a semantic match between sections of documents (sentences, paragraphs, logical blocks, etc.) and units of dialogue structure (utterances, turns, and thematic episodes).

We have implemented a robust thematic alignment method, using various state-of-the-art metrics (cosine, Jaccard, Dice) and considering document and speech units as bags of weighted words [11]. After suppression of stop-words and proper stemming, and after calculation of terms frequency in their section relative to their frequency in the whole document (TF.IDF), the content of the document elements is compared with the content of the speech transcript units.

When matching spoken *utterances* with document *logical blocks*, using cosine metric, recall is 0.84, and precision is 0.77, which are encouraging results. And when matching speech *turns* with logical blocks, recall stays at 0.84 and precision rises to 0.85. On the other hand, alignment of spoken *utterances* to document *sentences* is less precise but is more promising since it relies on less processing. Using Jaccard metric, recall is 0.83, and precision is 0.76 [11]. Thematic alignment of spoken *utterances* to document *sentences* has recently been used for joint thematic segmentation of both documents and speech transcripts. A recent evaluation of this method has shown that this bi-modal thematic segmentation outperforms standard mono-modal segmentation methods, which tends to prove that combining modalities improves considerably segmentation scores [12].

In another recent evaluation, we evaluated the effect of combining the various document/speech alignments—citation, references, and various levels of thematic alignments—on the general document/speech alignment performances [13]. Eight meetings have been tested, with a total of 927 utterances, and 116 document logical blocks. After combination of the three methods, the values of recall, precision, and f-measure were respectively 67%, 72% and 68%, whereas their independent use does not lead to scores above, respectively, 55%, 75% and 63%. These results tend to prove the benefit of merging the various methods of document/speech alignment.

7 Conclusion

Printed documents and spoken interaction are two important modalities in communication. This article presented an attempt to align these modalities based on their semantic content, in the context of a meeting browser that makes use of the mentions of documents in the dialogue.

The results presented here demonstrate the feasibility of a reference-based alignment technique using word co-occurrence and anaphora tracking, on the REs found by a grammar-based module. The two modules were evaluated separately, then in sequence: the scores obtained at this level are still above the baseline for the second module. Future feasibility studies will evaluate the degradation induced in the pipelined process by other automated modules, such as speech recognition or automatic document structuring. Together with other alignment

techniques, we believe that our approach will contribute to the design of a robust multi-modal meeting browser.

References

1. Lisowska, A., Popescu-Belis, A., Armstrong, S.: User query analysis for the specification and evaluation of a dialogue processing and retrieval system. In: LREC 2004, Lisbon, Portugal (2004)
2. Hadjar, K., Rigamonti, M., Lalanne, D., Ingold, R.: Xed: a new tool for extracting hidden structures from electronic documents. In: Workshop on Document Image Analysis for Libraries, Palo Alto, CA, USA (2004)
3. Sidner, C.: Focusing in the comprehension of definite anaphora. In Brady, M., Berwick, R., eds.: Computational Models of Discourse. MIT Press, Cambridge, MA (1983) 267–330
4. Mitkov, R.: Anaphora Resolution. Longman, London, UK (2002)
5. Hirschman, L.: MUC-7 coreference task definition 3.0. Technical report, MITRE Corp. (1997)
6. van Deemter, K., Kibble, R.: On coreferring: Coreference in muc and related annotation schemes. Computational Linguistics **26** (2000) 629–637
7. Popescu-Belis, A., Lalanne, D.: Reference resolution over a restricted domain: References to documents. In: ACL 2004 Workshop on Reference Resolution and its Applications, Barcelona, Spain (2004) 71–78
8. Simov, K., Simov, A., Ganev, H., Ivanova, K., Grigorov, I.: The CLaRK system: Xml-based corpora development system for rapid prototyping. In: LREC 2004, Lisbon, Portugal (2004) 235–238
9. Popescu-Belis, A., Georgescul, M., Clark, A., Armstrong, S.: Building and using a corpus of shallow dialogue annotated meetings. In: LREC 2004, Lisbon, Portugal (2004) 1451–1454
10. Grosz, B.J., Joshi, A.K., Weinstein, S.: Centering: A framework for modeling the local coherence of discourse. Computational Linguistics **21** (1995) 203–225
11. Lalanne, D., Mekhaldi, D., Ingold, R.: Talking about documents: revealing a missing link to multimedia meeting archives. In: Document Recognition and Retrieval XI - IS&T/SPIE's Annual Symposium on Electronic Imaging, San Jose, CA, USA (2004)
12. Mekhaldi, D., Lalanne, D., Ingold, R.: Using bi-modal alignment and clustering techniques for documents and speech thematic segmentations. In: Thirteenth Conference on Information and Knowledge Management (CIKM 2004), Washington D.C., U.S.A (2004)
13. Mekhaldi, D., Lalanne, D., Ingold, R.: From searching to browsing through multimodal documents linking. In: 8th International Conference on Document Analysis and Recognition (ICDAR2005), Seoul, Korea (2005)