Reference Resolution over a Restricted Domain: References to Documents

Andrei POPESCU-BELIS
ISSCO/TIM/ETI
University of Geneva
Bd. du Pont d'Arve 40
Geneva, CH-1211, Switzerland
andrei.popescu-belis@issco.unige.ch

Denis LALANNE
DIUF
University of Fribourg
Ch. du Musée 3
Fribourg, CH-1700, Switzerland
denis.lalanne@unifr.ch

Abstract

This article studies the resolution of references made by speakers to documents discussed during a meeting. The focus is on transcribed recordings of press review meetings, in French. After an overview of the required framework for reference resolution—specification of the task, data annotation, and evaluation procedure—we propose, analyze and evaluate an algorithm for the resolution of references to documents (ref2doc) based on anaphora tracking and context matching. Applications to speech-to-document alignment and more generally to meeting processing and retrieval are finally discussed.

1 Introduction

The references made by the speakers to the entities that they talk about are one of the keys to the understanding of human dialogs. When speakers discuss one or more documents, as in a press review meeting, the references to these documents constitute a significant proportion of all the occurring references.

A computer representation of the referents is available in this case, unlike references to more abstract objects, since here the documents can be stored in electronic format. Reference resolution amounts thus to the construction of links between each *referring expression (RE)* and the corresponding document element. For example, if someone 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.

We propose here an algorithm for the resolution of references to documents, or ref2doc. Its implementation and evaluation require a computational framework that includes several types of data—documents, transcriptions, and links—and an evaluation measure.

We summarize our view of reference resolution over a restricted domain in Section 2. Then, we situate the present task in the overall speech-to-document alignment process (Section 3). The annotated data and the evaluation metric are described in Section 4, along with empirical results regarding the patterns of the REs. The resolution algorithm is presented in Section 5, and the results obtained in various configurations are analyzed in Section 6, with conclusions about their relevance. Section 7 outlines the applications of the ref2doc algorithm to the exploitation of documents in meeting processing and retrieval applications.

2 Challenges of Reference Resolution over a Restricted Domain

From a cognitive point of view, the role of referring expressions in discourse is to specify the entities about which the speaker talks. It has long been observed that a more accurate view is that REs rather specify representations of entities in the speaker's or hearer's mind, an abstraction called *discourse entities* or *DEs* (Sidner, 1983; Grosz et al., 1995).

Reference resolution can be defined as the construction of the discourse entities specified by referring expressions, or rather, the construction of computational representations of DEs. This difficult but important task in discourse understanding by computers appears to be more tractable when enough knowledge about a domain is available to a system (Gaizauskas and Humphreys, 1997), or when the representations are considerably simplified (Popescu-Belis et al., 1998).

The coreference and anaphoric *links*, that is, links between REs only, are somewhat different aspects of the phenomenon of reference (Devitt and Sterelny, 1999; Lycan, 2000). Coreference is the relation between two REs that specify the same DE. Anaphora is a relation between two REs, called antecedent RE and anaphoric RE, where the DE specified by the latter is determined by knowledge of the DE specified by the former. In other terms, the DE specified by the anaphoric RE cannot be fully determined without knowledge of the antecedent RE.

Depending on how the referent of the second RE is determined by the referent of the first one, the two REs may be coreferent, as in example (1) below, or they can be related by other referring relations, e.g. whole/part, function/value, etc., as in (2).

1. *The first article*$_i$ is particularly relevant to our company. *It*$_i$ discusses ...
2. *The first article*$_i$ is particularly relevant to our company. *The title*$_j$ suggests that we...

In the present case, reference resolution over a restricted domain differs significantly both from anaphora resolution (Mitkov, 2002) and from coreference resolution (Hirschman, 1997; van Deemter and Kibble, 2000). The REs available in the dialog transcript must be matched against the set of potential referents or DEs, which can be derived from the document structure. Therefore a computational representation of the referents is here available to serve as DEs. This advantage results directly from our present research goal and could be later extended to DEs derived computationally from document content, such as the persons mentioned in an article.

Reference resolution in a restricted domain presents similarities with problems in natural language generation (NLG) and in command dialogs, that is, when the sets of referents are known *a priori* to the system. In NLG, the problem is to generate REs from existing computational descriptions of entities—see Paraboni and van Deemter (2002) for an application to intra-document references. In command dialogs, the problem is to match the REs produced by the user against the objects managed by the interface, again known formally to the system (Huls et al., 1995; Skantze, 2002).

3 Components of a Fully Automated Ref2doc System

3.1 Overview

Within the overall goal of a fully automated understanding of references to documents in meeting dialogs, several related sub-tasks can be distinguished, most simply envisaged as separate processes in a computational architecture:

1. Generate a transcript of the utterances produced by each speaker.
2. Detect the REs from the transcripts that make references to the documents of the meeting.
3. Generate a formal representation of the documents: articles, titles, etc.
4. Connect or match each RE to the document element it refers to.

Each of these components can be further subdivided. Our main focus here is task (4). For this task, an evaluation procedure, an algorithm, and its evaluation are provided respectively in Sections 4.3, 5, and 6. Task (3) is discussed below in Section 3.2.1.

Task (1), which amounts more or less to automated speech recognition, is of course a standard one, for which the performance level, as measured by the word error rate (WER), depends on the microphone used, the environment, the type of the meeting, etc. To factor out these problems, which are far beyond the scope of this paper, we use manual transcripts of recorded meetings (see Section 4.2.1).

The present separation between tasks (2) and (4) needs further explanations—see also (van Deemter and Kibble, 2000; Popescu-Belis, 2003) for more details. Our interest here is the construction of reference links between REs and document elements (from which coreference can be inferred), so we do not focus on task (2). Instead, we use a set of REs identified by humans.

Task (2) is not trivial, but could be carried out using a repertoire of pattern matching rules. The patterns of the manually detected REs shown in Table 1 (Section 4.4) are a first step in this direction. The difficulty is that sometimes task (2) proposes candidate REs, for which only task (4) can decide whether they can really be matched to a document element or not. For instance, REs such as pronouns ('it') or deictics ('this') that refer to document elements can only be detected using a combination of (2) and (4). This is one of our future goals.

3.2 Construction of the Logical Structure of Documents

Inferring the structure of a document from its graphical aspect is a task that can be automated with good performances, as explained elsewhere (Hadjar et al., 2004). Here, the documents are front pages of newspapers, in French. We first define the template of document structures, then summarize the construction method.

3.2.1 Targeted Document Structure

Many levels of abstraction are present in the layout and content of a document. They are conveyed by its various structures: thematic, physical, logical, relational or even temporal. The form of a document, i.e. its layout and its logical structure, carries important (and often underestimated) clues about the content, in particular for newspaper pages,

```
Newspaper ->       Date, Name,
                   MasterArticle,
                   Highlight*, Article+,
                   Other*, Filename

MasterArticle ->   Title, Subheading?,
                   Summary*, Author*,
                   Source?, Content?,
                   Reference?, Other*,
                   JournalArticle*

Article ->         Title, Subtitle?,
                   Source?, Content,
                   Author*, Summary*,
                   Reference*, Other?

JournalArticle ->  Title, Source?,
                   Summary*, Content?,
                   Reference+

Highlight ->       Title, Subtitle,
                   Reference+
```

Figure 1: Logical structure of a newspaper front page (in DTD style). Terminal nodes contain text.

where articles are organized by zones, and titles are clearly marked.

We consider that newspaper front pages have a hierarchical structure, which can be expressed using a very simple ontology. This is summarized in Figure 1 using a DTD-like declaration, as the document structure is encoded in XML. For instance, the first rule in Figure 1 states that a `Newspaper` front page bears the newspaper's `Name`, the `Date`, one `Master Article`, zero, one or more `Highlights`, one or more `Articles`, etc. Each content element has an `ID` attribute bearing a unique index.

3.2.2 Document Structure Extraction

The document structure can be extracted automatically from the PDF version of a document, along with a logical representation of the layout. Our approach merges low level extraction methods applied to PDF files with layout analysis of a synthetically generated TIFF image (Hadjar et al., 2004). A segmentation algorithm first extracts from the image the threads, frames and text lines, then separates image and text zones, and finally merges lines into homogeneous blocks. In parallel, the objects contained in the PDF file (text, images, and graphics) are extracted and matched with the result of the layout analysis; for instance, text is associated to physical (graphical) blocks. Finally, the cleaned PDF is parsed into a unique tree, which can be transformed either into SVG or into an XML document, and used for various applications.

```
<dialog>
 <channel id="1">
   ...
   <er id="12">The title</er>reads...
 </channel>
 ...
 <ref2doc>
 ...
 <ref er-id="12"
      doc-file="LeMonde030404.Logic.xml"
      doc-id="//Article[@ID='3']/Author"/>
     ...
 </ref2doc>
</dialog>
```

Figure 2: Sample annotation of a dialog transcription with ref2doc information (`er` stands for RE).

4 Evaluation Method and Data

Two important elements for testing are the available data (4.2), which must be specifically annotated (4.1), and a scoring procedure (4.3), which is quite straightforward, and provides several scores.

4.1 Annotation Model

The annotation model for the references to documents builds upon a shallow dialog analysis model (Popescu-Belis et al., 2004), implemented in XML. The main idea is to add external annotation blocks that do not alter the master resource—here the timed meeting transcription, divided into separate channels. However, REs are annotated on the dialog transcription itself. A more principled solution, but more complex to implement, would be to index the master transcriptions by the number of words, then externalize the annotation of REs as well (Salmon-Alt and Romary, 2004).

As shown in Figure 2, the `ref` pointers from the REs to the document elements are grouped in a `ref2doc` block at the end of the document, using as attributes the index of the RE (`er-id`), the document filename (`doc-file`), and an XPath expression (`doc-id`) that refers to a document element from the XML document representation.

4.2 Annotation Procedure and Results

4.2.1 Data Recording and Transcription

A document-centric meeting room has been set up at the University of Fribourg to record different types of meetings. Several modalities related to documents are recorded, thanks to a dozen cameras and eight microphones. These devices are con-

trolled and synchronized by a master computer running a meeting capture and archiving application, which helps the users organize the numerous data files (Lalanne et al., 2004).

At the time of writing, 22 press-review meetings of ca. 15 minutes each were recorded, between March and November 2003. In such meetings, participants discuss (in French) the front pages of one or more newspapers of the day. Each participant presents a selection of the articles to his/her colleagues, for information purposes. In general, after a monologue of 5-10 utterances that summarize an article, a brief discussion ensues, made of questions, answers and comments. Then, the chair of the meeting shifts the focus of the meeting to another article.

The recordings of the 22 meetings were manually transcribed using Transcriber,[1] then exported as XML files. The structure of the documents was also encoded as XML files using the procedure described above (3.2.1) with manual correction to ensure near 100% accuracy.

4.2.2 Ref2doc Annotation

The annotation of the ground truth references was done directly in the XML format described above (Figure 2). We have annotated 15 meetings with a total of 322 REs. In a first pass, the annotator marked the REs (with <er>...</er> tags), if they referred to an article or to one of its parts, for instance its title or author. 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. Table 1 synthesizes the observed patterns of REs.

The REs were then automatically indexed, and a template for the `ref2doc` block and an HTML view were generated using XSLT. In a second pass, the annotator filled in directly the attributes of the `ref2doc` block in the template. The annotators were instructed to fill in, for each RE (`er-id`), the name of the journal file that the RE referred to (`doc-file`), and the XPath to the respective document element (`doc-id`), using its `ID`. Examples were provided for XPath expressions. The following separate windows are all required for the annotation:

- text/XML editor for the `ref2doc` block of the dialog annotation file;
- HTML browser for the serialized HTML transcript (with REs in boldface);
- XML browser for the document structure representation (one per document);
- PDF viewer for the actual layout of the articles (one per document).

4.2.3 Inter-Annotator Agreement

We tested the reliability of the annotators on the second part of their task, viz., filling in the `ref2doc` blocks. The experiment involved three annotators, for the three meetings that discuss several documents at a time, with a total of 92 REs. In a first stage, annotation was done without any communication between annotators, only using the annotation guidelines. The result was on average 96% agreement for document assignment (that is, 3 errors for 92 REs), and 90% agreement on document elements (that is, 9 errors).[2]

In a second stage, we analyzed and solved some of the disagreements, thus reaching 100% agreement on document assignment, and 97% agreement on document elements, that is only two disagreements. These resulted from different interpretations of utterances—e.g., *they* in "they say..." could denote *the author*, *the newspaper*, etc.—and could not be solved.

This experiment shows that ref2doc annotation is a very reliable task: referents can be clearly identified in most cases. A perfect system would match the human performance at more than 95%.[3]

4.3 Evaluation Metrics

Unlike intra-document coreference resolution, for which evaluation is a complex task (Popescu-Belis, 2003), the evaluation of reference resolution over a specific domain is quite straightforward. One must compare for each RE the referent found by the system with the correct one selected by the annotators. If the two are the same, the system scores 1, otherwise it scores 0. The total score is the number of correctly solved REs out of the total number of REs (100% means perfect). The automatic evaluation measure we implemented using the XML annotation described above provides in fact three scores:

1. The number of times the document an RE refers to is correctly identified. This is informative only when a dialog deals with more than one document.

[1] `www.etca.fr/CTA/gip/Projets/Transcriber`

[2] These numbers were found using the evaluation software described below (Section 4.3). Document element agreement means here that the elements had the same `ID`.

[3] As for the first part of the process, recognizing the REs that refer to documents, we can only hypothesize that inter-annotator agreement is lower than for the second part.

2. The number of times the document element, characterized by its `ID` attribute, is correctly identified. Here, the possible types of document elements are article: `MasterArticle`, `JournalArticle`, `Article` or `Highlight`.

3. The number of times the specific part of an article is correctly identified (e.g., content, title, author, image, as indicated by the XPath annotation in the XML output format).

The third score is necessarily lower than the second one, and the second one is necessarily lower than the first one. The third score is not used for the moment, since our ref2doc algorithms do not target sub-article elements. To help adjust the resolution algorithm, the scoring program also outputs a detailed evaluation report for each meeting, so that a human scorer can compare the system's output and the correct answer explicitly.

4.4 Empirical Analysis of Occurring REs

The patterns of the annotated REs are synthesized in Table 1 according to the type of entity they refer to. This analysis attempts to derive regular expressions that describe the range of variation of the REs that refer to documents, but without generalizing too much. Words in capital letters represent classes of occurring words: NEWSP are newspaper names, SPEC is a specifier (one or more words, e.g., an adjective or a relative sentence), DATE and TITLE are obvious. Items in brackets are optional, and | indicates an exclusive-or. The patterns derived here could be used to recognize automatically such REs, except for two categories—anaphors and (discourse) indexicals—that must be disambiguated.

5 Ref2doc Algorithms

5.1 Preliminary Study

The first resolution method we implemented uses co-occurrences of words in the speech transcript and in the documents. More precisely, for each RE annotated in the transcript as referring to documents, the words it contains and the words surrounding it in the same utterance are matched, using the cosine metric, with the bag of words of each logical block of the document: article, title, author, etc. To increase the importance of the words within the REs, their weight is double the weight of the surrounding words. The most similar logical block is considered to be the referent of the RE, provided the similarity value exceeds a fixed threshold (confidence level).

Referent	#	RE
Journal	6	(le|du) NEWSP
	2	le journal
Front page (*une*)	33	la une NEWSP
	6	la une DATE+NEWSP
	5	(la|une) une
Article	33	(l'|le premier|le dernier) article
	31	cet article
	15	[l'] article suivant
	14	un [petit] article SPEC
	11	[un] autre article [SPEC]
	7	l'article SPEC
	5	[l'article] "TITLE"
Title	10	le [grand] titre [principal]
	4	(premier|second|autre) titre
Other text elements	12	[un] autre (point|sujet|fait) [SPEC]
	10	. . . (rubrique|encart|enquête|page|actualité|highlight|analyse) . . .
	5	(premier|dernier) point
	3	un [petit] point [SPEC]
	3	les grands points de l'actualité
	3	(le|au) point de vue [SPEC]
Graphic elements	11	. . . (dessin|photo|schéma|image|figure) . . .
Authors	6	l'auteur
	5	le journaliste
Anaphors	27	ils
	12	il
	8	l'
	4	(le|au) dernier
	3	autre chose [SPEC]
	2	on
Indexicals	5	là
	4	ça
	4	celui-là
	2	celui-ci
	2	celui SPEC

Table 1: Patterns of REs that refer to documents, in French, ordered by the type of the referent (9 REs out of 322 did not follow these patterns).

5.2 Algorithm based on Anaphora Tracking

A more complex algorithm was designed, which is based on the identification of anaphoric vs. non-anaphoric REs, as well as co-occurrences of words. The algorithm scans each meeting transcript linearly (not by channel/speaker), and stores as variables the 'current document' and the 'current document element' or article. For each RE, the algorithm

assigns first the hypothesized document, 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 simple method does not handle complex references such as 'the other newspaper', but obtains nevertheless a sufficient score (see Section 6 below).

The algorithm then attempts to assign a document element to the current RE. First, it attempts to find out whether the RE is anaphoric or not, by matching it against a list of typical anaphors found in the meetings: 'it', 'the article' (bare definite), 'this article', 'the author' (equivalents in French). If the RE is anaphoric, then it is associated to the current article or document element—a very simple implementation of a focus stack (Grosz et al., 1995)—except if the RE is the first one in the meeting, which is never considered to be anaphoric.

If the RE is not considered to be anaphoric, then the algorithm attempts to link it to a document element by comparing the content words of the RE with those of each article. The words of the RE are considered, as well as those of its left and right contexts. A match with the title of the article, or the author name, is weighted more than one with the content. Finally, the article that scores the most matches is considered to be the referent of the RE, and becomes the current document element.

Several parameters govern the algorithm, in particular the weights of the various matches—the nine pairs generated by {RE_word, left_context_word, right_context_word} $\times$ {title_or_subtitle_word, author_word, contents_word}—and the size of the left and right context—the number of preceding and following utterances, and the number of words retained. Evaluation provides insights about the best values for these parameters.

6 Results and Observations

6.1 Baseline and Best Scores

We provide first some baseline scores on the set of 15 meetings and 322 REs, that is, scores of very simple methods against which our algorithms must be compared (rather than against a 0% score). For RE $\leftrightarrow$ document association, always choosing the most frequent newspaper leads to 82% accuracy (265 REs out of 322). But some meetings deal only with one document; if we look only at meetings that involve more than one newspaper, then the score of this baseline procedure is 50% (46/92), a much lower value. Regarding RE $\leftrightarrow$ document element association, if the referent is always the front page as a whole (`/Newspaper`), then accuracy is 16%. If the referent is always the main article (`/MasterArticle[ID='1']`), then accuracy is 18%—in both cases quite a low value.

The word co-occurrence algorithm (described in Section 5.1) correctly solves more than 50% of the selected REs, in a preliminary evaluation performed on six meetings. This simple algorithm gives interesting results especially when REs belong to an utterance that is thematically close to the content of a document's logical block. However, the method uses only thematic linking and, furthermore, does not take advantage of all the various document structures.[4] The 50% score should thus be considered more as another baseline.

The second algorithm (described in Section 5.2) reaches 98% accuracy for the identification of documents referred to by REs, or 93% if we take into account only the meetings with several documents; remember that baseline was 82%, respectively 50%.

The accuracy for document element identification is 73% (237 REs out of 322). If we score only REs for which the document was correctly identified, the accuracy is 74% (236 REs out of 316), a little higher.

6.2 Score-based Analysis of the Algorithm

The best scores quoted above are obtained when only the right context of the RE is considered for matching (i.e. the words after the RE), not the left one. Also, the optimal number of words to look for in the right context is about ten. If the right context is not considered either, the score drops at 40%.

Regarding the weights, a match between the RE and the title of an article appears to be more important than one between the right context and the title, and much more important than matches with the content of the article: weights are about 15 vs. 10 vs. 1. All these values have been determined empirically, by optimizing the score on the available data. It is possible that they change slightly when more data is available.

If anaphor tracking is disabled, the accuracy of document element identification drops at 65%, i.e. 35% of the REs are linked to the wrong document element. Anaphor tracking is thus useful, though apparently not essential: dropping it leads to an algorithm close to our first attempt (Section 5.1).

Since the automatic scorer provides a detailed evaluation report for each meeting, we are in the

[4] For instance, it cannot solve references related to the document topological information (e.g. 'the figure at the bottom'), or related to the document logical structure (e.g. 'the author of the first article'), which need a semantic analysis of the REs.

process of analyzing the errors to find systematic patterns, which could help us improve the algorithm. Rules depending on the lexical items in the RE seem to be required.

7 Applications

7.1 Speech to Document Alignment

The resolution of references to documents is part of a cross-channel process aimed at detecting links between what was said during a meeting and the documents related to the meeting. The process enhances dialog and document processing, as well as the multi-media rendering of the results. Transcript-to-document alignment allows the generation of an enhanced transcript which is aligned also with the relevant documents, thanks to hyperlinks from transcript to document zones. Such a mechanism is integrated in the query and browsing interfaces that we are building.

Reference-based alignment is not the only way to align documents with the speech transcript. We have proposed two other techniques (Mekhaldi et al., 2003; Lalanne et al., 2004). *Citation-based alignment* is a pure lexicographic match between terms in documents and terms in the speech transcription. *Thematic alignment* is derived from semantic similarity between sections of documents (sentences, paragraphs, logical blocks, etc.) and units of the dialog structure (utterances, turns, and thematic episodes). We have implemented an algorithm that uses various state-of-the-art similarity metrics (cosine, Jaccard, Dice) between bags of weighted words.

For 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 (Lalanne et al., 2004). Thematic units have not been considered yet, for want of reliable automatic segmentation.

Reference-based alignment is complementary to other methods; these could be integrated in a common framework, so that they can be consolidated and compared. Their fusion should allow for more robust document-to-speech alignment.

7.2 Overall Application: Meeting Processing and Retrieval

A promising use of human dialog understanding is for the processing and retrieval of staff or business meetings (Armstrong et al., 2003). When meetings deal with one or several documents, it is important to link in a precise manner each episode or even utterance of the meeting to the sections of the documents that they refer to. Considering users who have missed a meeting or want to review a meeting that they attended, this alignment is required for two types of queries that appear in recent studies of user requirements (Lisowska et al., 2004). First, the users could look for episodes of a meeting in which a particular section of a given document was discussed, so that they can learn what was said about that section. Second, the relevant documents could automatically be displayed when the users browse a given episode of a meeting—so that a rich, multimodal context of the meeting episode is presented.

8 Conclusion

This article described a framework and an algorithm for solving references made to documents in meeting recordings by linking referring expressions to the document elements they denote. The implementation of the algorithm, together with test data (annotated meeting documents and transcripts) and an evaluation metric, show that the best results are obtained when combining anaphora tracking with a weighted lexical matching between RE plus right context, against title plus article contents.

An extension of the present algorithm is under study, in which REs are processed differently according to their type: REs explicitly referring to an article (‘the article’, ‘the section’), REs referring to positions (‘the article at the bottom left’), REs referring to the entities of the contents, etc. These could be matched to various data categories from the document representations.

Since printed documents and spoken interaction are two important modalities in communication, this article is also a step towards cross-modal applications. The reference-based alignment between transcripts and documents generates enriched transcripts, with explicit information about the contents and the timing of document mentions; conversely, it also helps document structuring. These in turn enhance browsing and searching capabilities for multimodal meeting processing and retrieval.

Acknowledgements

This work is part of (IM)2, Interactive Multimodal Information Management, a NCCR

supported by the FNS / Swiss Government (www.im2.ch). The authors are involved in two (IM)2 projects: IM2.MDM, Multimodal Dialogue Management (see http://www.issco.unige.ch/projects/im2/mdm/) and IM2.DI, Document Integration (see http://diuf.unifr.ch/im2/). The data we used is available from http://diuf.unifr.ch/im2/data.html. We thank Emmanuel Palacio, intern at ISSCO, for his contribution to the inter-annotator agreement test. We are also grateful to the reviewers for their helpful suggestions.

References

Susan Armstrong, Alexander Clark, Giovanni Coray, Maria Georgescul, Vincenzo Pallotta, Andrei Popescu-Belis, David Portabella, Martin Rajman, and Marianne Starlander. 2003. Natural language queries on natural language data: a database of meeting dialogues. In *NLDB 2003*, Burg/Cottbus, Germany.

Michael Devitt and Kim Sterelny. 1999. *Language and Reality: an Introduction to the Philosophy of Language*. The MIT Press, Cambridge, MA, USA, 2nd edition.

Robert Gaizauskas and Kevin Humphreys. 1997. Using a semantic network for information extraction. *Natural Languge Engineering*, 3(2-3):147–169.

Barbara J. Grosz, Aravind K. Joshi, and Scott Weinstein. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2):203–225.

Karim Hadjar, Maurizio Rigamonti, Denis Lalanne, and Rolf Ingold. 2004. Xed: a new tool for extracting hidden structures from electronic documents. In *Workshop on Document Image Analysis for Libraries*, Palo Alto, CA, USA.

Lynette Hirschman. 1997. MUC-7 coreference task definition 3.0. Technical report, MITRE Corp., 13 July 1997.

Carla Huls, Wim Claassen, and Edwin Bos. 1995. Automatic referent resolution of deictic and anaphoric expressions. *Computational Linguistics*, 21(1):59–79.

Denis Lalanne, Dalila Mekhaldi, and Rolf Ingold. 2004. 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.

Agnes Lisowska, Andrei Popescu-Belis, and Susan Armstrong. 2004. User query analysis for the specification and evaluation of a dialogue processing and retrieval system. In *LREC 2004*, Lisbon, Portugal.

William G. Lycan. 2000. *Philosophy of Language: a Contemporary Introduction*. Routledge, London, UK.

Dalila Mekhaldi, Denis Lalanne, and Rolf Ingold. 2003. Thematic alignment of recorded speech with documents. In *ACM DocEng 2003*, Grenoble, France.

Ruslan Mitkov. 2002. *Anaphora Resolution*. Longman, London, UK.

Ivandré Paraboni and Kees van Deemter. 2002. Towards the generation of document deictic references. In Kees van Deemter and Rodger Kibble, editors, *Information Sharing: Reference and Presupposition in Language Generation and Interpretation*, pages 329–352. CSLI Publications, Stanford, CA, USA.

Andrei Popescu-Belis, Isabelle Robba, and Gérard Sabah. 1998. Reference resolution beyond coreference: a conceptual frame and its application. In *Coling-ACL '98*, volume II, pages 1046–1052, Montréal, Canada. Université de Montréal.

Andrei Popescu-Belis, Maria Georgescul, Alexander Clark, and Susan Armstrong. 2004. Building and using a corpus of shallow dialogue annotated meetings. In *LREC 2004*, Lisbon, Portugal.

Andrei Popescu-Belis. 2003. Evaluation-driven design of a robust reference resolution system. *Natural Language Engineering*, 9(3):281–306.

Susanne Salmon-Alt and Laurent Romary. 2004. RAF: Towards a reference annotation framework. In *LREC 2004)*, Lisbon, Portugal.

Candace Sidner. 1983. Focusing in the comprehension of definite anaphora. In M. Brady and R. Berwick, editors, *Computational Models of Discourse*, pages 267–330. MIT Press, Cambridge, MA.

Gabriel Skantze. 2002. Coordination of referring expressions in multimodal human-computer dialogue. In *ICSLP 2002*, Denver, CO, USA.

Kees van Deemter and Rodger Kibble. 2000. On coreferring: Coreference in muc and related annotation schemes. *Computational Linguistics*, 26(4):629–637.