Adapting Lexical Chaining to Summarize Conversational Dialogues

Iryna Gurevych
EML Research gGmbH
Schloss-Wolfsbrunnenweg 33
D-69118 Heidelberg, Germany
www.eml-research.de/∼gurevych

Thade Nahnsen
International University in Germany
School of Information Technology
D-76646 Bruchsal, Germany
thade.nahnsen@gmx.de

Abstract

We present a system for summarizing transcripts of conversational dialogues based on lexical chaining. The experiments were carried out with twenty Switchboard dialogues (LDC, 1993). We designed and implemented four summarization methods employing lexical chains as their source representation. The summarization task is defined as extracting the most relevant utterances conveying the meaning of the dialogue. We evaluate the methods against *lead* and *random* baseline systems and show that lexical chaining outperforms them in terms of precision and recall.
Keywords: summarization, dialogue, lexical chains.

1 Introduction

The paper addresses the challenge of summarizing transciprts of spoken dialogues in unrestricted domains. Previous work on summarization focused on such genres as news articles (McKeown et al., 1995), web pages (Berger and Mittal, 2000), scientific texts (Teufel and Moens, 2002). In dialogue summarization, the motivation is the automatic transcription and summarization of multi-party dialogues, e.g. meetings (Alexandersson and Poller, 1998; Reithinger et al., 2000; Zechner, 2002). Therefore, it needs to deal with the whole range of dialogue and speech phenomena. Alexandersson and Poller (1998) present a system for generating meeting minutes in multiple languages. The approach is domain-sensitive as it relies on a database of handcrafted knowledge. The summary is produced using natural language generation techniques. The employment of this methodology in unrestricted domains is not feasible, as deep understanding of unrestricted spoken discourse is still an unsolved problem. Going beyond restricted domains requires domain-independent processing. The system presented by Zechner (2002) is designed for summarizing conversational dialogues in unrestricted domains. He uses pre-processing techniques to "normalize" the dialogue input, i.e. remove speech disfluencies, false-starts, detect question-answer pairs, etc. Statistical techniques are used to create the summaries. The output of the system is based on words in the input.

Gurevych and Strube (2004) employ a set of WordNet-based semantic similarity metrics to perform dialogue summarization. The methods evaluate the noun portion of WordNet in order to determine semantic similarity between utterances and a whole dialogue. The approach operates on manually disambiguated nouns. Bellare et al. (2004) determine subgraphs of WordNet, which are most relevant with respect to the semantics of the document. The sentence selection is performed based on the synsets that are most relevant to the text. Erkan and Radev (2004) approach text summarization from a graph-theoretical point of view. Their approach assigns weights to connections based on the number of occurrence and on the type of elements a specific element is connected to.

Our approach attempts to perform dialogue summarization with the help of lexical semantics, thus bridging the gap between domain-dependent deep analysis and domain-independent statistical processing. The system is based on the intuition that if lexical chains are used as intermediate representation in dialogue summarization, then "strong" lexical chains will be represented by the most relevant utterances. We designed and implemented four different methods to summarize dialogues based on representations constituted by lexical chains.

2 Research on Lexical Chains

Lexical chains are defined as sets of lexical items, which are either identical or related to each other by conceptual similarity. Conceptual similarity is determined on the basis of a certain lexical-semantic resource, e.g. WordNet (Fellbaum, 1998) and lexical-semantic relations between individual lexemes. Work on lexical cohesion dates back to Halliday and Hasan (1976) and even earlier. Morris and Hirst (1991) suggest lexical chains to determine the discourse structure of the text. The criterion for the inclusion of the word in a chain is a cohesive relation, which is figured out with the help of a thesaurus. Hirst and St-Onge (1998) propose to employ WordNet as a knowledge source for building lexical chains. Their definition of semantic relatedness is, hence, based on WordNet and synsets. Three kinds of relations can be distinguished: 1) extra-strong (holds between a word and its repetition); 2) strong (a synset is common to two lexemes, or there is a horizontal link, such as ANTONYMY, SIMILARITY, SEE-ALSO, or there is any kind of link between a synset associated with each word if one word is a compound phrase that includes the other); 3) medium-strong (there is a legal path connecting the synsets associated with each word).

Barzilay and Elhadad (1999) describe an algorithm for text summarization employing lexical chains as its intermediate representation. The algorithm includes three steps: 1) constructing lexical chains; 2) identifying strong chains; 3) extracting significant sentences from the text. The authors evaluate their algorithm on 30 texts. However, their evaluation is informal and does not provide an empirical proof whether the lexical chains model outperforms alternative summarization techniques. Also, there is no intrinsic evaluation, i.e. whether lexical chains constitute an appropriate representation of the discourse to be summarized. Silber and McCoy (2002) extend the work by Barzilay and Elhadad (1999). Two main contributions of their work are the following:

- an algorithm for computing lexical chains that is linear in time and space, thus eliminating one of the disadvantages in the earlier work, i.e. an exponential inefficiency for computing the chains. This makes it computationally feasible to compute lexical chains for large documents in real time;
- a new method for the evaluation of lexical chains as an intermediate representation in the summarization process. Their evaluation is based on a corpus of 10 scientific articles and 14 chapters from university textbooks.

Galley and McKeown (2003) focus on the lexical chaining algorithm in the context of work on word sense disambiguation (WSD). Along with the computational inefficiency mentioned earlier, a lack of accuracy in WSD is known to be a drawback of lexical chaining based algorithms. Galley and McKeown employ a different algorithm for computing lexical chains based on the *"one sense per discourse"* assumption. Their algorithm: 1) builds a representation of all possible interpretations of the text; 2) disambiguates all words; 3) finally constructs the lexical chains. The authors evaluate their algorithm with respect to the task of word sense disambiguation on the SEMCOR corpus. Their algorithm outperforms both Barzilay and Elhadad's and Silber and McCoy's algorithm (accuracies of 62.09%, 56.56% and 54.48% WSD respectively). No attempt is made to evaluate any further aspects of lexical chains.

The discourse type underlying our research, i.e. conversational dialogues, does not conform with the *one sense per discourse* constraint. In our corpus, topical changes occur rather frequently. Thus, one word may have different meanings within a single discourse. Therefore, our algorithm for building lexical chains follows other previous work (cf. e.g. Silber and McCoy (2002). Though slightly inferior in terms of WSD, it is both computationally efficient and imposes no constraints on the number of meanings that a single lexeme may have within a discourse.

The goals of this paper are the following: design summarization techniques based on lexical chaining for a new genre, i.e. conversational dialogue and carry out an extrinsic evaluation of lexical chains in dialogue summarization.

3 Experiments on Dialogue Summarization

3.1 Corpus

The experiments were carried out with twenty Switchboard dialogues on various topics, e.g. child care, dressing code. Data on our corpus is given in Table 1. The dialogue transcripts were manually annotated by three humans by selecting about 10% of utterances as being *relevant*, s. Table 7 for an excerpt from one of the dialogues. The reconciled version of the annotations, i.e. the *gold standard* was produced by selecting utterances labeled *relevant* by at least two annotators. It includes 9.47% of all utterances. When calculated for the whole corpus, the Kappa coefficient yielded .43. While this is not a high agreement rate on a general scale, it is comparable to what has been reported concerning the task of summarization in general (cf. (Mitra et al., 1997; Radev et al., 2003)).

3.2 Computing Lexical Chains

Lexical chains are computed on the basis of the noun portion of WordNet1.7. In the first step, the dialogue is processed and noun instances are selected. Thus, the dialogue D is represented as a set of nouns $D = \{N_1, ..., N_n\}$, each of them having a set of possible interpretations (synsets) $I_N = \{s_1, ..., s_m\}$ in WordNet. Then, the algorithm by Silber and McCoy (2002)

dialogue	words	utt./ markables	*relevant* utterances	lex. chains	*strong* lex. chains
1	2350	267	24	80	3
2	1069	79	15	50	2
3	1180	110	15	52	3
4	969	60	12	37	2
5	1428	133	15	55	1
6	1417	160	17	34	3
7	1159	131	15	28	2
8	2092	254	20	56	1
9	1284	162	12	43	2
10	1316	149	14	43	3
11	1521	138	16	37	3
12	1225	110	18	41	2
13	4046	416	22	83	2
14	2604	229	16	62	2
15	1542	53	9	49	3
16	1576	144	14	38	1
17	1966	159	11	54	3
18	1799	157	14	55	2
19	2751	210	15	66	2
20	1536	154	16	42	2
Total:	**34830**	**3275**	**310**		

Table 1: Descriptive corpus statistics

	1 utt.	**3 utt.**	**5 utt.**	**Default**
Identical word	1	1	1	1
Synonym	1	1	1	1
Hypernym	1	0.5	0.5	0.5
Sibling	1	0.3	0.2	0

Table 2: Computing word contributions to chains

is employed to automatically perform word sense disambiguation of the nouns.

We adapted the scheme for computing the contribution of a word to the chain as compared to that employed by Silber and McCoy due to a different discourse type, i.e. dialogues. Table 2 summarizes the values which are used to compute contributions of words to lexical chains in our system. It is similar to the original scheme in that it is based on two essential parameters: the nature of semantic relations between synsets and the distance between noun instances in the discourse. However, due to a different genre, i.e. dialogue versus text, the distance is defined in terms of utterances rather than paragraphs. Following Silber and McCoy, we allow different types of relations existing within the chain to contribute differently to that chain. The disambiguated sense of the noun is related to other synsets, see Table 3.

We store the corresponding interpretation s (synset) for each N (noun), resulting in the dialogue D being interpreted as a set of synsets $D = \{s_1, ..., s_m\}$. In Table 4, the "head" synsets of lexical chains that a given noun is related to are presented. On the other side, for each synset a corresponding lexical chain is stored, see Table 5. When the chains have been computed, they are ranked according to the scoring function defined by Barzilay and Elhadad (1999): $Score(Chain) = Length * Homogeneity$, where $Length$ is the total number of synset occurrences in the chain, while $Homogeneity$ is (1 – the number of distinct synset occurrences divided by $Length$). Strong chains are then defined as follows: $Score(Chain) > Average(Scores) + 2 * StandardDeviation(Scores)$. Table 1 gives an overview over the distribution of strong lexical chains in our data, and Table 6 gives examples of some initial synsets of chains ranked according to their strength.

3.3 Creating summaries

We designed and implemented four dialogue summarization methods operating on lexical chains. The set of lexical chains in D is represented as a two-dimensional matrix LC with the dimensions $(\#c \times \#s)$, where $\#c$ and $\#s$ denote the overall numbers of lexical chains and synsets in the dialogue, respectively. This can be formalized as: $LC = (b_{cs})_{1,...,\#c, s=1,...,\#s}$, where the matrix elements b_{cs} are the boolean values denoting whether the chain contains the corresponding synset or not. The chains are sorted numerically in a descending order according to their strength, i.e. the dialogue is also represented by the vector of lexical chains $(c_1, ..., c_{\#c})$. The knowledge represented by the lexical chains can be utilized in two ways by the summarization algorithm: from chains to utterances and from utterances to chains.

3.3.1 From chains to utterances

Utterances in the dialogue are ranked according to the strength of the strongest chain crossing them and their discourse position. The heuristics presented by Barzilay and Elhadad (1999) extract one sentence for each *strong* lexical chain. Method 1, called *one utterance per chain* method is similar to this heuristic, as we extract exactly one utterance per chain. However, it is also different from the original heuristics – we consider all lexical chains instead of only the strong ones, as the number of strong chains in our dialogues is small. The rest of the utterances are appended at the end in the order of their occurrence in the dialogue. This is done in order to fit a given compression rate when a summary is generated.

```
Step 1
  For each chain beginning with the strongest one
    Find the 1st utterance containing at least 1
    element belonging to the chain
    Insert the utterance into summary
```

noun	synset offset	gloss
child care	922884	a service involving care for other people's children
	922515	an act of help or assistance; "he did them a service"
	923360	childcare during the day while parents work

Table 3: Synsets related to a given sense of the noun

noun	synset offset	gloss
subject	5303	a human being; "there was too much for one person to do"
child care	922884	a service involving care for other people's children
children	5303	a human being; "there was too much for one person to do"
facility	15787	a man-made object taken as a whole
opinions	5079811	any cognitive content held as true
thoughts	5079811	any cognitive content held as true

Table 4: Disambiguated nouns

```
Step 2
 For each utterance
  If the utterance is not in the summary
   Append the utterance to the summary
```

Method 3, called *many utterances per chain* is similar to the previously introduced one. However, instead of extracting exactly one utterance per chain, we extract all utterances per chain (in the order of their dialogue occurrence), and process all chains in a descending order. At the end, we attach the utterances which are not represented by any chains in the order of their dialogue occurrence.

```
Step 1
  For each chain beginning with the strongest one
   Find all utterances containing at least 1
   element belonging to the chain
   Insert the utterances into summary
Step 2
 For each utterance
  If the utterance is not in the summary
   Append the utterance to the summary
```

3.3.2 From utterances to chains

The overall utterance score is a function of the number and type of chains crossing a particular utterance. In Methods 2 & 4, we find all noun instances in the utterance represented by synsets and assign a score to the noun based on the synset's chain membership. For Method 2, if a particular synset belongs to a *strong* chain, the contribution of the noun to the overall utterance score is 2, otherwise the contribution is 1. The utterance score is defined as a sum of all noun contributions. Then, the utterances are sorted numerically in a descending order according to their ranks.

```
  For each utterance
   For each synset
   If synset belongs to a strong chain
    Add 2 to the utterance score
   Else
    Add 1 to the utterance score
 Sort utterances numerically by overall score
```

For Method 4, the only difference is the scoring heuristic: instead of using binary weights (2 corresponding to a "strong" chain and 1 to any other chain), we employ the absolute weights of the respective lexical chains as scores for the synsets belonging to them.

```
  For each utterance
   For each synset
    Add the strength score of the chain to the
    utterance score
 Sort utterances numerically by overall score
```

4 Evaluation

Evaluating summaries produced on the basis of lexical chains is not straight-forward. We define dialogue summarization as the extraction of *relevant* utterances from the dialogue transcript. *Relevant* utterances are defined as those carrying the essential content of the dialogue. As it is desirable to support varying lengths of the resulting summaries, the *compression rate* is adjustable. Therefore, the summarization method supports ranking of all utterances in the dialogue, rather than a selection of individual utterances. We reformulate the problem in terms of standard information retrieval evaluation metrics: Precision, Recall and F-measure.

synset offset	gloss	nouns in the chain
5303	a human being; "there was too much for one person to do"	subject, children, child, children, child, children, person children, child, child, case, child, child, person
922884	a service involving care for other people's children	child care, child care, child care, child care, day care child care, child care
15787	a man-made object taken as a whole	facility, facilities, stuff, facility
5079811	any cognitive content held as true	opinions, thoughts, thought

Table 5: Synsets and lexical chains they belong to

synset offset	words and gloss	strength
5303	person, individual, someone, somebody, mortal, human, soul – (a human being; "there was too much for one person to do")	11.0
22634	group, grouping – (any number of entities (members) considered as a unit)	9.0
11745254	condition, status – (a state at a particular time; "a condition (or state) of disrepair"; "the current status of the arms negotiations")	6.0
12814143	time-of-life – (a period of time during which a person is normally in a particular life state)	6.0
922884	childcare, child-care – (a service involving care for other people's children)	5.0
8522773	parent – (a father or mother; one who begets or one who gives birth to or nurtures and raises a child; a relative who plays the role of guardian)	3.0

Table 6: Chains represented by initial synsets and their strengths

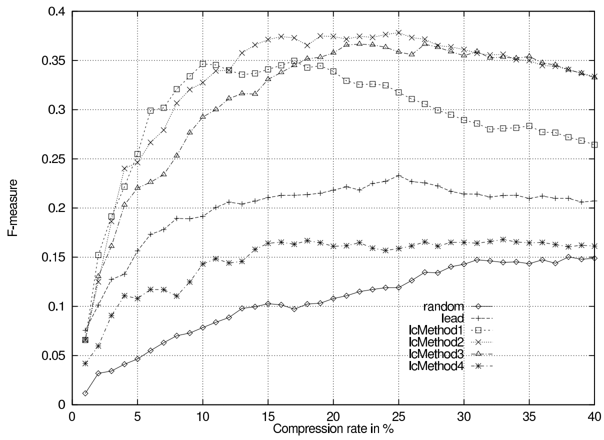

Figure 1: F-measure versus compression rate [1;40]

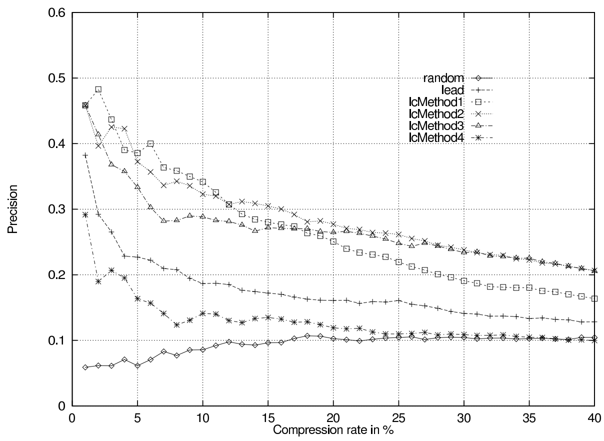

Figure 2: Precision versus compression rate [1;40]

Two baseline systems are employed in the evaluation. The first system is a *random* baseline, where relevant utterances (depending on the compression rate) were selected by chance. The second baseline, *lead*, is based on the intuition that the most important utterances tend to occur at the beginning of the discourse.

Figures 1 and 2 show that all lexical chaining based summarization methods, except for Method 4, outperform the baselines. Method 4 computes a score for each utterance by summing up the weights of nouns defined as the strength values of their respective chains. This strongly favours the utterances containing nouns belonging to the strongest chains, while the importance of other chains is minimized. Apparently this assumption is not true. Method 2 performs better than Methods 1 and 3, but this difference is not significant. The precision of all methods is rather low, e.g. about 23% for the compression rate 20%. Nevertheless the utterances selected by them differ (see Table 7), which suggests that an algorithm integrating multiple knowledge types is needed.

Our results are comparable to the results reported by Gurevych and Strube (2004) for the same dataset, e.g. at *compression rate* 25%, F-measure improves from .35 to .37. Both approaches employ WordNet as a knowledge source to determine the most relevant utterances. However, our algorithm disambiguates word senses automatically, whereas the results by Gurevych and Strube (2004) are based on manually disambiguated word senses. A comparison to the work by Zechner (2002) which is also based on Switchboard, i.e. domain-independent conversa-

utterance	gold standard	Method1	Method2	Method3	Method4
Go ahead.	none	21	39	39	39
oh, okay.	none	22	56	40	56
Yeah	none	23	52	41	52
the, uh, subject is child care and how to determine child care,	relevant	1	17	1	29
and that's, uh, an interesting one for me to talk about since I have no children,	none	24	33	2	11
but I did run a child care facility for a while.	none	9	19	27	1
Um .	none	25	46	42	46
And, uh, have some,	none	26	49	43	49
Well, you should, you should have some opinions on that, then.	none	5	36	29	20
I do have some thoughts on that,	none	27	35	30	15
yeah.	none	28	40	44	40
Uh, it's, uh, an interesting experience to be a surrogate parent for, or parent for a lot of people there,	none	2	3	12	6
and, uh, it's also very interesting in terms of how people choose the child care facilities	relevant	29	16	13	28
Well, I guess if I were going to choose, I mean, my first consideration would be safety.	relevant	3	23	19	25
My second consideration would be, uh, uh, health.	relevant	31	29	20	38
And, uh, I guess my third consideration would be, uh, warm environment, warm personal environment.	relevant	10	15	28	26
Well, right.	none	33	54	47	54
Uh, in Texas, we have to meet certain state standards in order to operate on a, at an institutional level and at a, like a small home level	none	8	4	21	10
so you meet the standards,	none	34	31	36	4
but then after that there's, there's a lot more.	none	35	34	34	14
I think it's important as the safety and health and that kind of stuff, is qualification of people who work there,	relevant	7	6	14	16
...					

Table 7: Utterances, their ranks and *gold standard*

tional dialogues, is not directly possible. He adopts a different view of the task, where summarization is performed by summarizing topical segments of dialogues (determined manually in his evaluation). In our approach, topic segmentation is performed implicitly through lexical chains. Additionally, his evaluation scheme is broken down to the word level. We redefine dialogue summarization as selecting higher-level *relevant* units, i.e. utterances, yielding much better inter-annotator agreement as originally reported by Zechner (.126), see Section 3.1.

5 Conclusions

We presented a system which adapts lexical chaining to summarize a new discourse type, i.e. conversational dialogues. Our research extends previous work on dialogue summarization by incorporating a broad coverage domain independent knowledge source and automatic word sense disambiguation. It is domain independent as opposed to approaches which aim at the deep semantic analysis and summary generation. Nevertheless, it is based on the semantic meaning of a dialogue as opposed to statistical approaches.

Additionally, we extend previous work on lexical chains by providing an extrinsic evaluation of the method against the human *gold standard* for the task of extracting the most relevant utterances. This relates the performance of the summarization model based on lexical chains to alternative models, e.g. *lead* and *random* baselines. Currently, our approach has been confined to the noun portion of WordNet, no predicates are considered and no anaphora resolution (about 10% of relevant utterances do not contain any nouns due to e.g. referential expressions) is performed.

Future research will, thus, aim at evaluating an extension to capture synsets of verbs and adjectives, as well. To achieve this goal, these will need to be conceptually integrated into the lexical chains algorithm, which currently is optimized to consider noun relationships. Furthermore, the impact of using anaphora resolution, which frequently occurs in dialogues, on selection performance should be evaluated. Using the above mentioned additional computational steps, it will be possible to evaluate utterances such as "He lived there", which were annotated as relevant, but could not be captured by lexical chains because the utterance does not contain any noun.

Some other interesting points concern the definition of the summarization task used in this study as summarizing dialogue transcripts by selecting relevant utterances. So far, we did not address the issues of speech recognition errors and automatic utterance boundary detection. Those will entail imperfect input to the lexical chains algorithm, with which respect its robustness to errors has to be further investigated. Also, the unit of analysis has been defined as *utterance*. Replacing *utterance* with *adjacency pairs* (Galley et al., 2003) capturing information about the speaker interaction, such as question – answer, offer – acceptance can be considered in a new annotation study.

Topical changes and the dialogue structure represent further interesting challenges. While topics of the dialogue are reflected in *strong* lexical chains, the interplay with the resulting summary has to be analysed. Finally, this will provide important implications for summary presentation. E.g., the summary can be generated by selecting adjacency pairs referring to specific topics and converting those to reported speech complemented by a high-level description of the original dialogue.

Acknowledgments

This work has been funded by the Klaus Tschira Foundation. We thank the reviewers for their valuable comments concerning this work.

References

Jan Alexandersson and Peter Poller. 1998. Towards multilingual protocol generation. In *Proceedings of the International Workshop on Natural Language Generation 1998,* Niagara-On-The-Lake, Canada, August 1998, pages 198–207.

Regina Barzilay and Michael Elhadad. 1999. Using lexical chains for text summarization. In Inderjeet Mani and Mark T. Maybury, editors, *Advances in Automatic Text Summarization*, pages 111–121. Cambridge/MA, London/England: MIT Press.

Kedar Bellare, Anish Das Sharma, Atish Das Sharma, Navneet Loiwal, and Pushpak Bhattacharyya. 2004. Generic text summarization using WordNet. In *Language Resources Engineering Conference (LREC'2004),* Barcelona, May.

Adam L. Berger and Vibhu O. Mittal. 2000. OCELOT: a system for summarizing web pages. In *Research and Development in Information Retrieval*, pages 144–151.

Günes Erkan and Dragomir R. Radev. 2004. LexRank: Graph-based centrality as salience in text summarization. *Journal of Artificial Intelligence Research*, 22.

Christiane Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, Mass.

Michel Galley and Kathleen McKeown. 2003. Improving word sense disambiguation in lexical chaining. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI-03),* Mexico.

Michel Galley, Kathleen R. McKeown, Eric Fosler-Lussier, and Hongyan Jing. 2003. Discourse segmentation of multi-party conversation. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics,* Sapporo, Japan, 7–12 July 2003, pages 562–569.

Iryna Gurevych and Michael Strube. 2004. Semantic similarity applied to spoken dialogue summarization. In *Proceedings of the 20th International Conference on Computational Linguistics,* Geneva, Switzerland, 23 – 27 August 2004, pages 764–770.

M. A. K. Halliday and Ruqaiya Hasan. 1976. *Cohesion in English*. London: Longman.

Graeme Hirst and David St-Onge. 1998. Lexical chains as representations of context for the detection and correction of malapropisms. In Christiane Fellbaum, editor, *WordNet: An electronic lexical database and some of its applications*, pages 305–332. Cambridge, MA: The MIT Press.

LDC. 1993. Switchboard. Linguistic Data Consortium. University of Pennsylvania, Philadelphia, Penn.

Kathleen R. McKeown, Jaques Robin, and Karen Kuckich. 1995. Generating concise natural language summaries. *Information Processing and Management*, 31(5):703–733.

Mandar Mitra, Amit Singhal, and Chris Buckley. 1997. Automatic text summarization by paragraph extraction. In *Proceedings of ACL Workshop on Intelligent Scalable Text Summarization,* Madrid, Spain, July 1997, pages 39–46.

Jane Morris and Graeme Hirst. 1991. Lexical cohesion computed by thesaural relations as an indicator of the structure of text. *Computational Linguistics*, 17(1):21–48.

Dragomir R. Radev, Simone Teufel, Horacio Saggion, Wai Lam, John Blitzer, Hong Qi, Arda Celebi, Danyu Liu, and Elliott Drabek. 2003. Evaluation challenges in large-scale document summarization. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics,* Sapporo, Japan, 7–12 July 2003, pages 375–382.

Norbert Reithinger, Michael Kipp, Ralf Engel, and Jan Alexandersson. 2000. Summarizing multilingual spoken negotiation dialogues. In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics,* Hong Kong, 1–8 August 2000, pages 310–317.

H. Gregory Silber and Kathleen F. McCoy. 2002. Efficiently computed lexical chains as an intermediate representation for automatic text summarization. *Computational Linguistics*, 28(4):487–496.

Simone Teufel and Marc Moens. 2002. Summarizing scientific articles: Experiments with relevance and rhetorical status. *Computational Linguistics*, 28(4):409–445.

Klaus Zechner. 2002. Automatic summarization of open-domain multiparty dialogues in diverse genres. *Computational Linguistics*, 28(4):447–485.