Jane Morris
Faculty of Information Studies, University of Toronto

Readers' Perceptions of Lexical Cohesion in Text

Abstract: Preliminary results from an experimental study of readers' perceptions of lexical cohesion and lexical semantic relations in text are presented. Readers agree on a common "core" of groups of related words and exhibit individual differences. The majority of relations reported are "non-classical" (not hyponymy, meronymy, synonymy, or antonymy). A group of commonly used relations is presented. These preliminary results indicate potential for improving both relations existing in lexical resources, and methods dependent on lexical cohesion analysis.

1. Introduction and Overview

This work reports on preliminary results from an experimental study of readers' perceptions of lexical cohesion and of the associated lexical semantic relations in text. Lexical cohesion (Halliday and Hasan 1976) is the contribution to a text's meaning that is made by groups of related words found within it. Lexical semantic relations are the individual relations that exist between pairs of words, and are therefore involved in the creation of lexical cohesion.

In linguistics, lexical cohesion is used to explain one aspect of how a text's meaning is created, through "continuity of lexical meaning" (Halliday and Hasan 1976, 320). Hasan (1984; Halliday and Hasan 1989) extended the theory to include the concept of *cohesive harmony* (discussed below in sections 2.2 and 4), and demonstrated that it correlates with users' judgements of the coherence of text. Lexical cohesion has also been used in determining the structure of text (Morris and Hirst 1991). Computationally, lexical cohesion has been used in text summarization (Barzilay and Elhadad, 1999; Silber and McCoy 2002) and spelling correction (Hirst and Budanitsky 2004), and lexical semantic relations have been used in query expansion in IR (Information Retrieval) (Voorhees 1998).

The goal of the experimental study is to investigate the nature of the groups of related words (i.e., the lexical cohesion) and the individual relations between these words, as they are perceived by readers. The identification of lexical cohesion in text has been done so far by linguists with particular points of view (Hasan 1984; Martin 1992), and research on types of lexical semantic relations has proceeded largely out of the context of text, with the exception of the original work of Halliday and Hasan (1976), where any semantically "recognizable" relation between words in text was used in forming word groups.

A practical reason for analyzing reader-identified relations is the potential for improving the relations that exist in current lexical resources such as WordNet (Fellbaum 1998) and IR thesauri. Similarly, the potential exists for improving methods dependent on lexical cohesion analysis that are used in NLP (Natural Language Processing) and IR applications.

As part of a larger study of three different texts, reader data has been analyzed for both the first text, which will be referred to as Text 1, and a pilot study. In all cases, the text was the first 1.5 pages of a *Reader's Digest* article. There are two parts to the analysis: the overall lexical cohesion, and the individual lexical semantic relations. In the analysis of lexical cohesion, agreement was calculated for the word groups (sometimes referred to in the literature as lexical chains) marked by the readers, to give an indication of their subjectivity, as measured by individual differences. The overall nature of the word groups has not yet been formally analyzed, but aspects of interest with examples are discussed in sections 2.2 and 4 below.

In the analysis lexical semantic relations are classified as either *classical* (hyponymy, meronymy, synonymy, and antonymy) or *non-classical* (details below in section 2.1), and for all cases in which at least two readers marked the same word pair from within a group of related words, their agreement on the actual relation given between the words was determined. The data were also analyzed qualitatively to determine whether a "smallish" common set of relation types was being used.

2. Theoretical Background

2.1 Lexical Cohesion and Lexical Semantic Relations

The two fundamental areas of research pertaining to this work are lexical cohesion and lexical semantic relations, and they will each be discussed in turn. The linguistic theory of lexical cohesion was first presented by Halliday and Hasan in their 1976 classic *Cohesion in English*. Lexical cohesion is one of the five types of cohesion detailed therein, and as stated earlier, its contribution to the meaning of text is provided by the continuity of lexical meaning created by the different groups of related words that run through a text.

Here is an example showing four word groups identified by a reader from the first paragraph of Text 1:

> I attended a *funeral service* recently. **Kind** words, *Communion*, *chapel* overflowing, speeches by <u>lawyers</u>, <u>government workers</u>, friends, all speaking of the *deceased's* **kindness**, his **brilliance** in mathematics, his **love** of SCRABBLE and CHESS, his great **humility** and **compassion**, his **sense of humour**.

Each different word group is shown here with a different emphasis. The italicized word group is about funerals, the bold word group about positive human characteristics, the underlined group about types of people/jobs, and the capitalized group about board games. These inter-sentence word groups have no size or location restrictions, and can therefore span any portion of the text, possibly all of it.

Hasan extended this theory in 1984 to include reference cohesion, but the most significant addition was to combine the inter-sentence relations already a part of the original theory with intra-sentence relations similar to those of Fillmore's (1968) case relations. The result consists of structured and more tightly knit units of lexical cohesion within a text. Related ideas are found in Cruse's (1986) concept of patterns of lexical affinities, and in Barsalou's (1989) concept of ad hoc categories. Details of these three

related ideas and their similarities are given by Morris, Beghtol and Hirst (2003). Examples from the experimental study are also given below in sections 2.2 and 4.

The other fundamental area of research pertaining to this work is on lexical semantic relations. Within the work on lexical cohesion, the issue of which lexical semantic relations particular researchers include in their analysis has obviously been addressed. There has also been a considerable amount of work on the nature and behaviour of lexical semantic relations outside the context of lexical cohesion. Both of these two approaches are now reviewed.

In the original work of Halliday and Hasan (1976) on lexical cohesion, any recognizable relation was considered relevant. In subsequent work (Hasan 1984; Halliday and Hasan 1989), the only relations used in analysis were classical ones, that is, those that are characterized by a sharing of the same individual defining properties between the words, and a requirement that the words be in the same syntactic class. Hasan (1984, 213) notes that these relations are "ready made" types from linguistics, and furthermore, that the rest are too inter-subjective to use. Indeed, in work on lexical semantic relations outside the context of lexical cohesion, the focus has also been on the classical relations (Cruse 1986). This focus is also reflected in WordNet, which is the major lexical resource used in computational linguistics, and also in psychology; however, McRae and Boisvert (1998) advocate expanding beyond the use of classical relations.

On the other hand, in LIS (Library and Information Science), non-classical relations are included and referred to as "associative" relations or "related terms" in thesaurus-based research (Neelameghan 2001; Milstead 2001). This is also true of *Roget's Thesaurus,* which was used as the source for relation identification in work on lexical cohesion identification (Morris and Hirst 1991). However, in both *Roget's Thesaurus* and the LIS thesauri, these non-classical relations are not classified by type, which is clearly a loss of valuable semantic information.

As an example of the type of data collected on lexical semantic relations in this study, consider again the example above. From within the italicized word group, which the reader described as "processes involved when someone dies", the reader identified *funeral service* and *chapel* as a related word pair, and gave the lexical semantic relation between them as "a funeral service is most often held at a place of worship e.g. chapel". Notice that this relation is non-classical, as are many found by the readers (details below in section 3), which is why the distinction between classical and non-classical relations is relevant. If non-classical relations are heavily used, then there is a potential benefit to incorporating them into applications dependent on lexical semantic relation identification.

2.2 Context

Lexical cohesion and lexical semantic relations will now be considered with respect to the following two aspects of context: the text, and the reader. Other aspects of context such as social, cultural, and environmental factors are not specifically addressed here, except insofar as they are reflected by the group of readers used in the study (details below in section 3.1).

Lexical cohesion is, by definition, studied within the context of text; however, as stated above, the lexical semantic relations used to create it are largely not. In Hasan's (1984) work on lexical cohesion, the broader view of lexical semantic relations became restricted to classical relations, which were taken from the existing linguistic types that were developed out of the context of text. Her goal was an objective analysis of the relation between cohesive harmony (i.e., her expanded concept of lexical cohesion) and subjects' judgements of coherence, and she found a correlation. Hoey (1991) and Hearst (1997) have shown that lexical cohesion analysis that involves only repetition can be useful in determining text structure, and repetition clearly does not depend on the context of text.

The goal in this paper is different in that it is precisely the inter-subjectivity that is under investigation. Individual differences in readers' perceptions of the membership of words in the word groups and the lexical semantic relations between these words are used as a measure of subjectivity. Rather than being viewed as a potential problem, individual differences are viewed as an important aspect of research on what is "in the text" versus what is "in the reader". Recent work that focuses on automatic corpus analysis in both linguistics and computational linguistics is, by definition, only able to analyze or discover what is "in the text", although in some cases human-annotated corpora are used as starting points, and results can be judged by humans. A well-known earlier approach to discourse analysis (Grosz and Sidner 1986) emphasized the text, in what was referred to as the linguistic structure, the writer, in what was called the intentional structure, and the reader's focus, in the attentional structure. Clearly their theory attempts to account for more than what is "in the text".

In part because of the difficulty of understanding and analyzing writers' intentions and readers' subjective interpretations, as well as because of advances in statistical and machine learning techniques for automatic corpus analysis, research in general has shifted away from including the reader (or writer) as context. However, as aptly stated by Olson (2004):

> This I now believe, is one of the fateful illusions of modernism, the idea that knowledge can be embodied in a text or a computer program or other artifact. Texts are more accurately seen as artifacts, notational devices for representation and thought. Knowledge remains the possession of the knower not of the artifact.

The work reported in this paper attempts to analyze lexical cohesion and lexical semantic relations in the context of both text and reader. The view taken is that knowledge exists in both the text and the reader, and that we need research on the limits of each approach.

Hasan (1984, 201) comments that the classical relations she now uses are "dissociated from a real context of utterance" and therefore become "objective" or immune to reader interpretation. However, some of the word pairs Hasan saw as examples are not so clearly dissociated from the context of their text. As an example, *stroke* and *feed* are considered as related, presumably by co-hyponymy or co-meronymy, but this is a much fuzzier case than that of many other examples of co-hyponymy such as the pair *Scrabble* and *chess* from the example given above in section 2. The parent category would presumable be something like "things to do to pets" or "common things done to dogs", but the category now seems closer in type to those referred to as ad hoc categories by Barsalou (1989).

Ad hoc categories are those that are "made up on the fly" (Lakoff 1987, 45) for some text-specific purpose. Morris, Beghtol and Hirst (2003) suggest that rather than *stroke* and *feed* being related in a pre-existing classical relation (hyponymy or meronymy), it is the mechanism of cohesive harmony, through the linking of inter-sentence classical and intra-sentence case relations, that *creates* the relation. It is in this way that cohesive harmony seems related to the ad hoc categories. Rather than being "dissociated" from the text, the categories are postulated as a direct result of a type of text-specific analysis.

In a word-based hyponymy hierarchy like WordNet, or an LIS thesaurus, the categories are relatively straightforward and classical (Lakoff 1987), and as suggested by Morris, Beghtol and Hirst (2003), giving one word of a word pair related by hyponymy is all the context required to bring to mind the child, parent, or siblings. For example, *bird* will imply types of birds such as *robin*, or *Scrabble* will imply other types of board games such as *chess*, but given *stroke*, does *feed* come to mind in a way that is "dissociated from any real context of utterance" (Hasan 1984, 201)?

Similarly, in the example given above in section 2.1, would *humour* imply *humility*? The bolded word group consisting of *kind*, *kindness*, *brilliance*, *love*, *humility*, *compassion*, and *sense of humour* is described by many readers as positive human characteristics. The question arises as to whether this category is non-classical or classical (Lakoff 1987). A clue seems to reside in the category name, in that categories with names that are not lexicalized as a single word, are, in the very least, fuzzy, and most likely non-classical. Another example is given in section 4 below. The "positive human characteristics" category is considered here to be non-classical.

3. Experimental Study

3.1 Objectives

This study is designed to investigate readers' perceptions of lexical cohesion and lexical semantic relations in text. One goal is to investigate the subjectivity of these perceptions. For both lexical cohesion and lexical semantic relations, the readers' data is analyzed to determine how much agreement there is among them. Another objective is to investigate the nature of the lexical semantic relations and word groups perceived by the readers. The relation types are classified as classical or non-classical to determine the usage of non-classical relations, which are under-represented both in the research and in lexical resources. The relation types are also qualitatively analyzed to determine whether a smallish set of common types is being used. The nature of the word groups has not yet been analyzed in detail and this issue is discussed further below in section 4.

3.2 Method

The results reported here are for the first of three texts in a larger study, referred to as Text 1, and also for a pilot study. In all cases, the first 1.5 pages of a general-interest *Reader's Digest* article were used as text. For the pilot there were five readers: four Ph.D. students and one Masters student from FIS (Faculty of Information Studies,

University of Toronto), and for Text 1 there were nine readers: all Masters students from FIS. All of the readers were self-reported as fluent readers and writers of English.

The subjects (readers) were given a set of 30 coloured pencils with which to mark the word groups. They were instructed to read through the article naturally, and to mark each word group that they perceived with a different colour. The word groups were explained as being words related by meaning in the text. An example was given of the first paragraph from a different *Reader's Digest* article where this had been done, which was then taken away before they started to limit any influence on their choices. They were told that while they should mark the word groups as they read, it was also okay to go back and change their underlining.

Once they were finished marking the word groups in the text, there were four types of data sheets to be filled in. The first was used to record a copy of all of the words in each word group, using one data sheet per word group. The second was used to record the most significant word pairs from each word group, along with the actual relations between each of the word pairs, and as many sheets as necessary were used. The subjects were told that if a set of words within a group were all clearly related by the same relation, they could group them together on the data sheet, and give the one relation holding between them all, but this was not a common occurrence. In the pilot study, they were asked to give all word pairs that they perceived as related, but this caused time and understanding difficulty, so for Text 1, the subjects were instructed to give only the most important related word pairs. The third type of data sheet was used to record the meaning of the word group in the text, and one was filled in for every word group. Lastly, the fourth type was used to record any existing relations between the word groups in the text, and as many as were necessary were used. Data from this last type of data sheet has not yet been analyzed.

3.3 Data Analysis and Results

In order to summarize the degree of agreement between the subjects, three statistics are presented for both the pilot study and Text 1. For the pilot study, the results include data from the four word groups that were marked by all subjects and one of three word groups marked by three of the five subjects. For Text 1, data is used from the eleven word groups that were marked by at least four of the nine subjects. For each of these groups, the subjects' agreement on membership of the group was computed in the following manner: For all possible pairs of subjects, for each pair, the number of words on which they agreed was computed as a percentage of the total number of words they marked. Averaged over all possible pairs of subjects, the agreements were 63% for the pilot, and 63% for Text 1.

Next, the agreement on the word pairs that were identified as directly related within the word groups was computed. This analysis was restricted to *core* words, which are defined as those marked by a majority of subjects. All distinct instances of word pairs that were marked by at least 50% of the subjects were counted, and this number was divided by the total number of distinct word pairs marked. It was found that 13% and 25% of the word pairs were marked by at least 50% of the subjects, in the pilot study and Text 1, respectively.

For the set of word pairs that were identified by more than one subject, the agreement on what the relation between the pair was deemed to be was computed. Because the individual wordings of the relation descriptions varied by subject as expected, analysis and interpretation was done so that relations that were clearly intended to be the same were grouped together. This process is sometimes referred to as conceptual content analysis. It was found that the subjects agreed on 70% of the cases in the pilot study, and 86% of the cases in Text 1.

The data show that the subjects are identifying a common core of groups of related words in the text, but that individual differences do occur. Agreement on which word pairs within word groups are related is much lower. This is most likely due to the fact that it is not as direct or intuitive a task as word group identification, and since all words in a group are related in some way, it is not easy to pick out the most important ones. Also, the word groups could be comprehended as wholes or gestalts, meaning that when a word is added to the group, it could be perceived as related to the whole of the group so far detected, rather than being related to any one particular word. Note that in the relatively rare instances when all words in a group are related by the same relation, the subjects did as instructed, and gave one relation for the whole group. An example of this occurs in the positive human characteristics group shown in bold in the example in section 2, where the reader grouped all of the words together as being positive human characteristics.

For word pairs that are marked by more than one subject, the agreement on relation type seems high. Cruse (1986) notes that subjects are generally not good at explaining meaning, but in his examples of doing this, subjects are asked to explain meaning out of the context of a naturally occurring text. Morris, Beghtol and Hirst (2003) expected that it would be easier for subjects to identify related word pairs than to explain the relation between them. The results obtained so far indicate the opposite, suggesting that subjects do not find it hard to explain meaning in the context of surrounding text.

Regarding the usage of classical and non-classical relations, in Text 1, the overwhelming majority of the relations were non-classical, and there were a few cases of hyponymy, synonymy, and antonymy. In the pilot, the majority of the relations reported were non-classical; however, more classical relations were reported than in Text 1. This is to be expected, since some texts will contain large classical categories, and in the article used in the pilot, there was a large set of lobby groups such as the Gun Control Lobby, Mothers Against Drunk Driving, and the American Lung Association.

Qualitative analysis of the relation types has been done for Text 1. In this analysis, all relations given by the subjects are included, regardless of how many of them used the particular word pairs involved. The goal is to investigate any commonality that might exist among the entire set of relations given by the subjects. Although investigating commonality is the goal, care must be taken not to impose any preconceived notions of what the nature of the commonality should be, and also, not to impose commonality where it does not exist. In the preliminary analysis of Text 1, there seems to be a set of 13 commonly used relation types, and these are listed below. Note that outlier relation types (used by only one subject) are not listed here:

1. Positive qualities (*brilliant* / *kind*).
2. Negative qualities (*homeless* / *alcoholic*).

3. Qualities in opposition (*drunk* / *drying out*).
4. The same relation existing between all members of large categories such as positive human characteristics (*humility* / *humour*), typical major life events (*funeral* / *born* / *married*), and jobs / types of people (*lawyer* / *volunteer*).
5. Words that are each related to a third concept; for example "caring" (*kind* / *gentlemanly*), "remember" (*speeches* / *deceased*), and "education" (*people* / *professors*).
6. Descriptive noun / adjective pairs (*born* / *young*; *professors* / *brilliant*).
7. Commonly co-occurring words (*drunk* / *bottle*). In many cases the readers used subgroups of this category:
    a. Location (*homeless* / *shelter*; *funeral* / *chapel*; *kitchen* / *home*).
    b. Problem and solution / cause / one word leads to the other (*homeless* / *drunk*; *date* / *love*; *date* / *relationship*; *alcoholic* / *rehab program*).
    c. Case relations (*volunteer* / *service*; *people* / *living*; *speeches* / *friends*).
    d. Aspects of an institution: "married" (*son* / *married*), "funeral" (*speeches* / *communion*), and "education" (*college* / *jobs*).
8. Stereotypical relations (*homeless* / *drunk*; *people* / *home*).
9. One word related to a large group of words, in a contrasting role, seemingly with a lot of import: (*homeless* / <the group of positive human characteristics such as *brilliant*, *kind* , *humility*>).
10. Definitional: (*alcoholic* / *drunk*).
11. Quasi-hyponymy relations (*friend* / *relationship*).
12. Synonymy (*relaxed* / *at ease*).
13. Antonymy (*died* / *born*).

It must be emphasized that this set of common relation types is preliminary. Some of the relation types seem similar, for example relations 3 and 13. Perhaps relation 5 should become a sub-type of relation 7. Relation 7, apart from the specific sub-types found, seems very general, but it is used to group a set of relations whose descriptions seem to specifically imply co-occurrence. Examples of such descriptions, with the words that are related given in italics, are: "a *drunk* has a close relation to a *bottle*", " a *bottle* is a fact of life for a *drunk*", and "a *drunk*'s most familiar item is a *bottle*". Some of the relation types are specific, and more like relation types found in the literature, such as 7a, 7b, and 7c, or 12 and 13. Others are more general, such as 5, and 7d.

4. Discussion and Future Work

The preliminary results indicate that the subjects in the pilot study and for Text 1 identified a common core of groups of related words in the text, as well as exhibiting individual differences. Preliminary analysis of the subjectivity of readers' responses reflecting reader attitude in the pilot study is given by Morris and Hirst (2004).

The qualitative relation type analysis done for Text 1 indicates the emergence of some common relation types, and integration of the results from the other two texts will give a better indication of whether a smallish set of common types is emerging. However, even if the relation types in the three general-interest *Reader's Digest* articles exhibit commonality, differences may be found in different types or genres of texts. As well, further analysis of more texts will give an indication of whether topic (or something else) is a major factor in the relation types perceived. Many of the relations reported are non-

classical, which supports their integration into methods or theories that depend on lexical semantic relation analysis in text.

Further research is required on the nature of the word groups found in the text. As mentioned above, the positive human characteristics word group appears to be a non-classical category, but not much is known about how such categories are formed and used in the creation of meaning in text. The overwhelming historical and current research is on hierarchical classical categories and relations.

One non-classical category found in the pilot study was "bad behaviours", which consisted of the following core words (i.e., those used by a majority of subjects): *sex*, *drinking*, *drag racing*, *shooting*, *dangerous*, and *irresponsible behaviours*. These words, (with the exception of *dangerous)* have the same case (Fillmore 1968) relation of agent/action to the movie stars who were mentioned in the article. This is similar to the mechanism of cohesive harmony, since the movie stars are in a "movies" word group and there are at least two instances of the same intra-sentence case relation (a technical requirement of Hasan's 1984 concept of cohesive harmony) linking together two inter-sentence word groups (movies and bad behaviours). It would be useful to determine if there are similarities between cohesive harmony and the creation of Barsalou's (1989) ad hoc categories in the text. Perhaps, as suggested above, some of the word groups can be considered and analyzed as such.

One obvious area for future research is the effect of different types of texts and readers. Demographic data such as age, gender, level of university education, and years of speaking English was collected from the subjects and could be analyzed as the potential source of any differences in responses.

Some areas for further research were suggested by Morris, Beghtol and Hirst (2003), and remain as future goals. For example, the agreement found on identifying specific word pairs within a group was low. In this study, subjects were asked about relations between individual word pairs. This does not directly address the possibility of words going into the word group because of a relation to this unity, rather than to individual words. Subjects could be asked for the relation information at the time of adding a word to a group, but it would have to be determined if this would affect which words go into the word groups, or any other significant aspect of word group formation. Also, insight could be gained from qualitative analysis involving working with subjects closely as they create word groups, eliciting from them how they are doing it, and asking them questions about the process.

In summary, it appears from the preliminary results reported here that the analysis of readers' perceptions of both the word groups and the lexical semantic relations in text will provide results that have potential to improve theories and methods that use lexical cohesion and lexical semantic relations in the analysis of text.

Acknowledgements

This research was supported by the Natural Sciences and Engineering Research Council of Canada. The author is grateful for the many detailed comments and suggestions made by Graeme Hirst and Clare Beghtol on earlier drafts of the paper.

References

Barsalou, L. 1989. Intra-concept similarity and its implications for inter-concept similarity. In *Similarity and analogical reasoning* , eds. S. Vosniadou and A. Ortony. Cambridge: Cambridge University Press.

Barzilay, R., and M. Elhadad. 1999. Using lexical chains for text summarization. In *Advances in text summarization* , eds. I. Mani and M. Maybury. Cambridge, Mass.: The MIT Press.

Cruse, D. 1986. *Lexical semantics.* Cambridge: Cambridge University Press.

Fellbaum, C. 1998. *WordNet: An electronic lexical database.* Cambridge, Mass.: The MIT Press.

Fillmore, C. 1968. The Case for Case. In *Universals in linguistic theory* , eds. E. Bach and R. Harms. New York: Holt, Rinehart and Winston.

Grosz, B., and C. Sidner. 1986. Attention, Intentions, and the Structure of Discourse. *Computational Linguistics* 12(3): 175–204.

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

Halliday, M.A.K., and R. Hasan. 1989. *Language, Context and Text: Aspects of language in a social-semiotic perspective*. Geelong, Victoria: Deakin University Press. (republished by Oxford University Press 1989).

Hasan, R. 1984. Coherence and Cohesive Harmony. In *Understanding reading comprehension: Cognition, language and the structure of prose* , ed. J. Flood. Newark, Delaware: International Reading Association.

Hearst, M. 1997. TextTiling: Segmenting text into multi-paragraph subtopic passages. *Computational Linguistics* 23(1): 33–64.

Hirst, G., and A. Budanitsky. 2004. Correcting read-word spelling errors by restoring lexical cohesion. Submitted for publication.

Hoey, M. 1991. *Patterns of lexis in text.* Oxford: Oxford University Press.

Lakoff, G. 1987. *Women, fire and dangerous things.* Chicago: University of Chicago Press.

Martin, J. 1992. *English text: System and structure.* Amsterdam: John Benjamins.

McRae, K., and S. Boisvert. 1998. Automatic semantic similarity priming. *Journal of Experimental Psychology: Learning, Memory and Cognition* 24(3): 558–572.

Milstead, J.L. 2001. Standards for relationships between subject indexing terms. In

Relationships in the organization of knowledge , eds. C.A. Bean and R. Green. Norwell, Mass.: Kluwer Academic Publishers.

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

Morris, J., C. Beghtol., and G. Hirst. 2003. Term Relationships and their Contribution to Text Semantics and Information Literacy through Lexical Cohesion. In *Proceedings of the 31st Annual Conference of the Canadian Association for Information Science* Halifax, Nova Scotia, June 1-June 4: 153–168.

Morris, J., and G. Hirst. 2004. The subjectivity of lexical cohesion in text. *AAAI Spring Symposium on Exploring Affect and Attitude in Text* Stanford University. To appear.

Neelameghan, A. 2001. Lateral relationships in multicultural, multilingual databases in the spiritual and religious domains: The OM Information Service. In *Relationships in the organization of knowledge* , eds. C.A. Bean and R. Green. Norwell, Mass.: Kluwer Academic Publishers.

Olson, D. R. 2004. Knowledge and its artifacts. In *History of science, history of text* , ed. K. Chemla. Dordrecht, NL: Kluwer.

Roget, P. 1977. *Roget's international thesaurus* (2nd ed.). Harper and Row Publishers Inc.

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

Voorhees, E. (1998). Using WordNet for Text Retrieval. In *WordNet: An Electronic Lexical Database* , ed. C. Fellbaum. Cambridge, Mass.: The MIT Press.