INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600



Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/502

UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

ELICITATION OF KNOWLEDGE DIFFERENCES IN READING COMPREHENSION USING LATENT SEMANTIC ANALYSIS WITH MULTIPLE SEMANTIC SPACES

A Dissertation

SUBMITTED TO THE GRADUATE FACULTY

In partial fulfillment of the requirements for the degree of

Doctor of Philosophy

By

PETER M. MOERTL

Norman, Oklahoma

2002

UMI Number: 3059902



UMI Microform 3059902

Copyright 2002 by ProQuest Information and Learning Company.
All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

© Copyright by Peter M. Moertl 2002

All Rights Reserved.

ELICITATION OF KNOWLEDGE DIFFERENCES IN READING COMPREHENSION USING LATENT SEMANTIC ANALYSIS WITH MULTIPLE SEMANTIC SPACES

A Dissertation APPROVED FOR THE

DEPARTMENT OF PSYCHOLOGY

BY

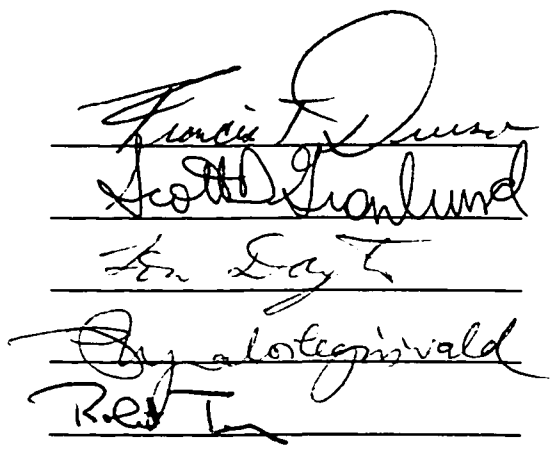


Table of Contents

Abstract v

Introduction 1

Latent Semantic Analysis 4

Description of LSA 4

Empirical evaluations and applications of LSA 8

LSA representations in multiple semantic spaces 11

Experiment 1 13

Experiment 2 24

Conclusions 25

References 31

Tables 37

Figure captions 47

Figures 48

Appendix A 55

Appendix B 56

Appendix C 74

Abstract

Previous research has proposed Latent Semantic Analysis (LSA) as a model and technique of knowledge representation that represents knowledge differences in single semantic spaces (e.g. Grolier's Academic American Encyclopedia, Landauer & Dumais 1997). In this project, LSA knowledge representations were constructed in multiple semantic spaces to represent user knowledge differences for adaptive information retrieval. Semantic spaces with varying degrees of background knowledge were constructed for two versions of a story that participants had read. The two versions induced either complete or incomplete story comprehension. The results indicated that optimal LSA representations depended on the level of story comprehension: LSA representations that were derived from semantic spaces of any size resembled participants' complete story comprehension but matched incomplete story comprehension only if semantic spaces included sufficient information. Larger semantic spaces captured more background knowledge than smaller spaces (Experiment 2). This led to the conclusion that participants with incomplete comprehension relied more on background knowledge to rate word pair relatedness than in the Solved condition where they relied more on story knowledge. Comparing LSA representations in multiple semantic spaces was found to be a viable mean for representing knowledge dependent on a reader's background. Implications of these findings for the representation of user knowledge for automated adaptive information retrieval are discussed.

Latent semantic analysis (LSA) is a model and technique for computerized knowledge representation (Landauer & Dumais, 1997). Knowledge is represented in semantic spaces that are derived from large corpora of text. A semantic space consists of the principal components that are extracted from the frequencies of word co-occurrences in text (see Appendix A). The texts that people have read (Foltz, Kintsch, Landauer, 1998; Foltz & Wells, 1999) or written (for example, Foltz, 1996) can be automatically mapped into a semantic space. This process permits the automated extraction of knowledge from text without manual intervention of human domain experts (e.g. Kintsch, 1998). This represents a significant advantage compared to traditional knowledge elicitation methods such as interviews, observations, or conceptual card sorting techniques that usually require human intermediaries (see Cooke, 1999). Previous research has found strong agreement between LSA knowledge representations and human knowledge for single semantic spaces.

The current project investigates LSA representations for user adaptive information retrieval. Human information retrieval intermediaries (e.g. librarians) actively explore clients' background knowledge (e.g. Taylor, 1968) in order to adapt to particular information needs. Traditional automated information retrieval systems however, are usually independent of user knowledge because information is only matched to a given search query but not related to user knowledge. Automated information retrieval systems that adapt to information needs therefore require the incorporation of user knowledge representations. This project investigates if LSA representations can be used to capture user knowledge differences for such purpose.

Semantic spaces for LSA knowledge representations have been constructed for domains such as history and introductory psychology (Foltz & Wells, 1999), medicine (Derwester, Dumais, Furnas, Landauer, & Harshman 1990; Wolfe, Schreiner, Rehder, Laham, Foltz, Kintsch, & Landauer 1998) technical reports (Foltz, 1990), newsgroups (Foltz, 1992), and general knowledge (Landauer & Dumais, 1997). LSA knowledge representations were applied to automatic essay grading (Foltz, 1996; Foltz, Britt, Perfetti, 1996), assessment of textual coherence (Foltz, Kintsch, & Landauer, 1998), knowledge assessment (Foltz, Britt, & Perfetti, 1996; Foltz, & Wells, 1999), automatic tutoring (Graesser et al., 2000; Kintsch, Steinhart, Stahl, & LSA research group, 2000), text summarization (Kintsch, 2002) and for human computer interaction (Soto, 1999), such as automatic indexing (Derwester, et al. 1990; Dumais, 1990, 1994), information exploration (Gordon & Dumais, 1998), personalized information filtering (Foltz, 1990; Foltz & Dumais, 1992) and cross-language information retrieval (Landauer & Littman, 1990).

LSA knowledge representations are constructed by mapping texts with to-be-assessed-knowledge into semantic spaces. Semantic spaces are constructed from larger collections of text with relevant background information in a specific knowledge domain. In this way, LSA knowledge representations are comparable because they are built in the same semantic space and knowledge similarities can be assessed. For example, the similarity between LSA representations of student essays with prototypical essays of domain experts has been shown to be a good predictor for essay quality (Foltz 1996; Wolfe, Schreiner, Rehder, Laham, Foltz, Kintsch & Landauer, 1998). Thereby, the essays

of students and experts are mapped into the same semantic space which is constructed for specific knowledge domains.

The requirement of building LSA knowledge representations in specific semantic spaces currently limits the wider applicability of LSA. Semantic spaces that are built for general knowledge (e.g. based on Grolier's Academic American Encyclopedia, Landauer & Dumais, 1997) usually do not contain sufficient information about more specific domains. Foltz and Wells (1999), for example, chose to create different domain specific semantic spaces for their LSA representations of historical or psychological knowledge.

For a wide variety of applications, knowledge in different domains needs to be assessed. In libraries, for example, librarians retrieve information for information needs of clients with highly diverse backgrounds. Thereby, librarians initiate their assistance by a negotiation process that attempts to assess clients' knowledge, experience, and context in order to facilitate the formulation of the information need (see, for example, Belkin, Brooks & Oddy, 1982; Kuhltau, Turock, George & Belvin, 1990; Taylor, 1968). If clients' backgrounds differ, the same question can indicate different information needs. Information retrieval experts attempt to account for these differences by making information retrieval dependent on a client's background. Automated user-adaptive information retrieval systems therefore require the automated assessment of clients' background knowledge.

Though LSA seems applicable to the automation of knowledge assessment for such purposes, the technique is currently limited to knowledge representations in predefined domains. Differences in knowledge are thereby represented by mapping different texts into the same semantic space that is chosen to be representative of the

knowledge in a specific domain. This poses a limitation if information (e.g. the retrieval document) must be evaluated dependent on a client's background knowledge. Seen from the perspective of clients with different background knowledge, a given retrieval document might contain different relevant information. Only when semantic spaces capture clients background knowledge accurately, retrieval information could be appropriately related to clients information needs.

This project proposes a solution to this limitation. Instead of building one semantic space for all LSA knowledge representations, multiple spaces are built to capture differences in background knowledge. LSA representations are then compared across multiple semantic spaces that reflect differences in background knowledge. In particular, this project investigates if LSA knowledge representations in multiple semantic spaces can be used to elicitate knowledge differences at different levels of comprehension. Thereby, LSA knowledge representations that are based on multiple semantic spaces of different texts are compared with knowledge that participants have acquired by reading those texts.

The next section briefly describes the LSA technique of representing knowledge. For a full description of the method see Landauer and Dumais (1997) and Deerwester et al. (1990).

Latent Semantic Analysis

Description of LSA

LSA knowledge representations are created from word co-occurrence information in text. A schematic depiction of this process is printed in appendix A. The first step in the process of creating an LSA knowledge representation is similar to Salton's vector

space model (SMART, Salton, 1971; Salton, Wong & Young, 1975/1997; Salton & McGill, 1983) that was developed for the automatic indexing of documents. Syntactic text information is disregarded and frequencies of words that co-occur in the same context are input in a word co-occurrence matrix. The word co-occurrence matrix represents words in rows and paragraphs in columns. Cells of the matrix contain the frequencies of words within one paragraph context. Dependent on the application, contexts are defined as documents, paragraphs, sentences, sets of words, or combinations of these. No clear preference for either context definition exists in current literature. However, Moertl and Durso (2000) found an advantage of LSA knowledge representations with contexts based on paragraphs over those based on fixed numbers of words. Accordingly, paragraphs are used as contexts in this project.

Varieties of texts and sizes have been used for the construction of LSA knowledge representations. Landauer and Dumais (1997) based their semantic space on 30,473 articles from Grolier's Academic American Encyclopedia. Howard and Kahana (2002) used an LSA semantic space that was based on 37,651 documents with 92,409 unique terms (TASA-ALL: see http://lsa.colorado.edu/spaces.html) and Foltz and Wells (1999) created a semantic space for introductory psychology (4903 paragraphs). Smaller semantic spaces were created by Foltz and Wells (1999) in the domains of history (607 paragraphs), by Foltz, Kintsch, and Landauer (1998) using texts about heart disease (830 sentences) and Freeman, Thompson and Cohen (1999) for air combat (377 paragraphs).

The second step in the creation of an LSA knowledge representation is weighing the entries of the co-occurrence matrix that was derived in step 1. Different weighing mechanisms have been proposed such as multiplying term frequencies by a factor that is

inversely related to the number of documents in which that term occurs (Salton, Wong & Young, 1975/1997). Landauer and Dumais (1997) reported that the quality of LSA knowledge representations was improved by transforming cell frequencies to [ln(1 + F)/E], where F = cell frequency for a given word, and E = entropy of words over all contexts = $-\Sigma$ (p) log(p). Thus the influence of frequently occurring words is dampened in favor of less frequently occurring words. This transformation presumably corresponds to the effect of human learning that tends to slow down as the amount of learning increases (Landauer & Dumais, 1997; Rescorla & Wagner 1972; Wagner & Rescorla 1972).

The benefit of log-entropy term weighing cannot necessarily be expected to transfer to smaller semantic spaces that are based on smaller texts. In large corpora of text, words that occur few times might carry more significant information than words that occur more often. In smaller texts, however, most words can be expected to occur only a few times and weighing them more might overestimate their informational value. Therefore, the impact of weighing on the quality of LSA representations is considered as an empirical question in the present study.

The third step of creating LSA knowledge representations involves the extraction of principal components from the word co-occurrence matrix by using a mathematical procedure called singular value decomposition (SVD, for larger matrices see Berry, 1992). SVD is a commonly used mathematical procedure that is behind frequently used methods for data reduction and analysis like principal component analysis and factor analysis. SVD allows the extraction of linear components that are optimized to reproduce the analyzed matrix. Whereas principal component analysis is intended to extract

components from square matrices that usually contain correlations or covariances, SVD can be used to decompose any rectangular, non-singular matrix into three matrices (see Landauer & Dumais, 1997). One of the three extracted matrices is usually used and represents word meanings in rows of extracted components. Retaining only a subset of the strongest components results in a compressed representation of the initial matrix. This compressed representation eliminates redundant information and exposes underlying semantic patterns. Words that co-occur with the same words tend to become more similar to each other in this process. For example, if words X and Y never co-occur in the same context but both co-occured with word Z, the process of dimensional reduction captures this similarity between words X and Y (see Landauer & Dumais, 1997).

A crucial decision is the selection of the optimal dimensionality to extract semantic structure. The optimal dimensionality of these LSA word-vectors is usually determined empirically by comparing LSA representations with human knowledge and selecting the vector dimensionality that yields optimal match with participants' performance (see Landauer & Dumais, 1997). Once the optimal dimensionality has been determined for a given semantic space, it is used for subsequent LSA-knowledge representations in this space. The maximal dimensionality of a semantic space without semantic extraction reflects pure word-coocurrences.

Optimal dimensionalities of LSA word vectors vary somewhat across different studies but have been previously found to be generally less than 20 % of maximal dimensionality (see, e.g. Foltz & Wells 1999, Landauer & Dumais 1997). Larger text corpora seem to require semantic spaces with relatively smaller optimal dimensionalities. Large semantic spaces that were built from reading material with more than 37,000

documents contain about 1.1 % of maximal dimensionality whereas smaller space built from texts with about 7,000 documents contain 6.2 % of maximal dimensionality (see http://lsa.colorado.edu/spaces.html). Moertl and Durso (2000) used optimal LSA spaces with 7.1 % dimensions of maximal dimensionality for 3167 documents. Smaller LSA representations were used by Foltz and Wells (1999) who used LSA spaces that were based on 607 documents with 16.5 % dimensions of maximal dimensionality (experiment 1) and Freeman, Thompson, and Cohen (1999) who used an LSA-space based on 377 documents with 19.89 % of maximal dimensionality. Accordingly, optimal LSA spaces for even smaller texts might be expected to contain somewhat higher relative dimensionalities.

Once the optimal dimensionality, M, has been determined for a given semantic space, words are represented by M-dimensional vectors. The semantic similarity between words is calculated by the geometric cosine between two word vectors in this M-dimensional space. Other measures such as dot-product, Euclidean distance, or vector length have been considered by Rehder, Schreiner, Wolfe, Laham, Landauer and Kintsch (1998). The meaning of multiple words such as sentences (Foltz, Kintsch, Landauer, 1998) or essays (Foltz, 1996; Foltz, Britt, & Perfetti, 1996; Landauer, Foltz, Laham, 1998; Wolfe et al. 1998) is represented by averages of LSA word vectors.

Empirical evaluations and applications of LSA

LSA knowledge representations have been created for various domains and have been compared with human performance. Landauer and Dumais (1997) created a semantic space of the Grolier's Academic American Encyclopedia and mapped synonym questions of a language test (Test of English as a Foreign Language, Educational Testing

Service) into that semantic space. Using a multiple choice test to select correct synonyms among several alternatives, they found that LSA knowledge representations led to synonym selections that were highly similar to those of students. Also, Foltz and Wells (1999) built LSA representations for essays that novices and experts wrote in the domain of introductory psychology and mapped both types of texts onto the semantic space that was derived from an introductory psychology textbook. They reported significant differences in LSA knowledge representations between novices and experts. Similarly, Foltz (1996) used LSA to predict the source from which students had acquired knowledge and Foltz, Kintsch, and Landauer (1998) used LSA to predict the coherence of texts (see also Foltz, Britt, & Perfetti 1996). In the same publication, Wolfe et al. (1998) predicted the amount of learning for participants with different background knowledge. More recently, Howard and Kahana (2002) used LSA to explain the influence of word similarities on episodic retrieval effects.

Prior to the emergence of psychological interest into LSA knowledge representations, LSA had been successfully applied to various aspects of human performance. First applications of LSA were in the field of information retrieval. Derwester et al. (1990) used LSA for automatic indexing of documents and demonstrated retrieval advantages compared to traditional methods (see also Dumais, 1991). In a similar effort, Foltz (1990) and Foltz and Dumais (1992) applied LSA to the task of information filtering for matching user interest. More recently Gordon and Dumais (1998) applied LSA to literature based discoveries by supporting the detection of conceptual relationships in literature and Kuralenok and Nekrest'yanov (2000) used LSA to automatically classify text. These successful applications of LSA to information

retrieval led Moertl (2001) to suggest the use of LSA for the automatic summarization of text (see also Kintsch, 2002).

Beside information retrieval, LSA has also been applied to knowledge assessment (Chen 1997; Shapiro & McNamara, 2000) and other domains such as computer interface design (Soto, 1999; Sutcliffe & Ennis, 2000). LSA knowledge assessment was used by Kintsch et al. (in press) to aid the development of summarization skills of students by giving adaptive feedback (see also Butcher & Kintsch 2001). Similarly, Graesser et al. (2000) developed an automatic tutoring device based on LSA-created feedback.

Recently criticism arose about the epistemological status of LSA knowledge. Glenberg and Robertson (2000), for example, pointed out that LSA knowledge representations are different from human knowledge because they are not connected to a perceptual system and are therefore perceptually ungrounded. This is, they argue, an essential difference from human knowledge which is tied to perceptual experience and the affordances of real objects. Consequently, human knowledge is directly tied to the world. Knowledge representations such as LSA that do not contain this essential link must necessarily fail as model for human knowledge. Landauer and Dumais (1997) deflect some of this criticism by suggesting a way to incorporate perceptual systems into LSA. Additionally, Landauer (1999) stated that LSA spaces are based on text created by humans who have perceptually experienced the world and developed a system to organize these experiences into language. Therefore, the structure of the language symbols on which LSA is based is an expression of human experiences in the real world. An LSA representation might not require a separate grounding mechanism because it is based on grounded symbols.

LSA representations in multiple semantic spaces

Previous research created and compared LSA knowledge representations within single semantic spaces. Though different semantic spaces are frequently built, the actual comparison between LSA knowledge representations is performed within a single semantic space (e.g. the comparison of students' essays with experts' essays, Wolfe, et al., 1997). In this way, the LSA representation of a text that participants wrote is created in a semantic space that is assumed to reflect the participants' background knowledge. This causes a problem if writers' background knowledge is not appropriately reflected by the semantic space. Two identical texts might reveal different information about a domain expert or novice. Individual differences in background knowledge are important for automated information retrieval systems that attempt to adapt to human users. A solution to this problem is to create individual semantic spaces that reflect individual background knowledge. This would make it possible to create different LSA knowledge representations of the same text that are sensitive to a person's background knowledge. This project proposes and investigates LSA knowledge representations in multiple semantic spaces and compares the representations between different semantic spaces.

One important question is if differences in semantic spaces lead to LSA knowledge representations that are comparable to differences in human knowledge. For example, are the knowledge differences of two groups of readers who read two slightly different stories detectable in two different semantic spaces that incorporate these stories? LSA knowledge representations within one semantic space should match knowledge of readers of the same text while they should be different between readers of different texts. In other words, when participants acquire different levels of comprehension by reading

two versions of a story, are differences in their knowledge structure detectable in the LSA knowledge representations that are based on different semantic spaces of the version of the story?

Different levels of comprehension were induced by letting participants read two different versions of a story: one version facilitating complete story comprehension, the other version incomplete comprehension. Participants' knowledge structures were assessed by their ratings of the relatedness of key concepts in the stories. Conceptual relatedness ratings have been frequently used for the assessment of knowledge structures. For example, Durso, Rea, and Dayton (1994) used relatedness ratings to assess different levels of insight and Schvaneveldt, Durso, Goldsmith, Breen, and Cooke (1985) used them to assess differences in expertise.

Differences in comprehension were induced by providing readers with two different versions of a story that induce measurable knowledge differences in readers. These texts are then incorporated into semantic spaces so that LSA knowledge representations of key-concepts can be derived from them.

Two factors are expected to influence the quality of such LSA knowledge representations. One factor is the size of the texts from which semantic spaces are constructed. In order to measure this influence, semantic spaces were created from texts of different sizes. A second factor is the dimensionality of derived LSA word vectors. Semantic extraction is caused by the dimensional reduction of word vectors (see discussion above). Therefore, LSA knowledge representations with semantic extraction (defined as reduced dimensionalities) are compared with those without semantic extraction.

LSA knowledge representations are then compared to the participants' concept relatedness ratings (participant knowledge) and two measures of representational quality are assessed. The first measure is the match between LSA and participants' knowledge. Match is measured by the correlation between concept relatedness ratings of participants in each condition and the corresponding LSA-representation. The second measure of representational quality is the differentiation of knowledge structures between the complete versus the incomplete comprehension conditions. Differentiation is measured by the percentage of correct classification of participants based on their similarity to LSA-knowledge representation. Comprehension differences between participants should be reflected in differences of the LSA knowledge representations. Both measures of knowledge quality are essential for determining the cognitive veracity of LSA-knowledge representations in different semantic spaces.

Experiment 1

Method

Participants

Forty-four psychology undergraduate students participated for partial fulfillment of an experimental requirement. Participants were tested in groups of up to 4 individuals. All participants were native English speakers.

Materials

Two versions of a story by Arthur Conan Doyle (The Red-headed League, Doyle, 1892) were created, one for each experimental condition. One version contained the solution to the story (thereafter referred to as Solved), the other version did not (referred

to as Unsolved). The length of the two versions was made approximately the same by excluding purely embellishing and story-irrelevant details from the Solved version. Table 1 indicates the length of the original story and the two versions.

A panel of three judges independently selected key concepts from the original story and then agreed on 10 key concepts for the original version of the story. The key concepts were the most important elements of the story and included main characters, objects, and places. In order for these key concepts to occur at the same frequency in both versions the two versions were slightly altered. The words and their frequencies in both texts are printed in Table 2. All modifications are listed in Table 3. The story is printed in Appendix B.

In order to assess the participants' comprehension of the story, a comprehension questionnaire with 18 multiple choice questions with four alternatives each was created. Additionally, one question assessed the participant's familiarity with Arthur Conan Doyle stories in general and asked specifically if they had previously read "The Red-headed League". The questionnaire is printed in Appendix C.

Design and Procedure

Participants were randomly assigned to one of two experimental conditions. Participants were asked to read carefully one version of the Red-headed League. They were told that they would later have to answer questions about the story. Then the story was presented to them in a booklet. In one condition participants read the Solved version whereas in the other condition they read the Unsolved version of the Red-headed League. They were informed that other participants might read a different text and therefore might

be done sooner or later than them. The times when participants started and finished reading were recorded.

After reading the Red-headed league, participants were asked to indicate their general impression of the relatedness of all possible 45 pairs of ten key concepts in the story on a scale from 0 to 9. The word pairs are listed in Table 4. They were informed that they could consider a number of factors including similarity, co-occurrence, or dependency and were asked not to ponder about their judgments. Word pairs were printed on a separate sheet in randomized order. Word position was also randomized.

After the relatedness ratings, participants filled out the comprehension questionnaire. At the conclusion of the experiment, they were given information about the experiment.

Results and Discussion

Semantic spaces of the two versions of the "The Red-headed League" were constructed. In order to include background knowledge, additional, larger semantic spaces were constructed by incrementally adding sets of Sherlock Holmes stories to each version of "The Red-headed League". Adding 5 sets of stories resulted in 12 semantic spaces, 6 for each version of "The Red-headed League". The smallest semantic spaces, LSA 1(S) and LSA 1(U) consisted of the solved and unsolved versions of "The Red-headed League", respectively. For the semantic spaces LSA 3(S,U), LSA 6(S,U), LSA 9(S,U), LSA 12(S,U), and LSA 15(S,U), additional Sherlock Holmes stories were randomly selected from a pool of 15 stories and incrementally added to each version of "The Red-headed League". LSA 3(S) for example, consisted of the Solved version of the Red-headed League plus the same two other stories that were added to LSA 3 (U). Each

of those stories were of similar length (see Table 5) and were added in the same order. For the largest semantic space, LSA 15(S,U), three of Conan Doyle's considerably longer stories were added. Note, that any larger semantic space contained all the stories of the smaller spaces; LSA 15, for example, contained the same stories as LSA 12 plus 3 additional stories. Table 5 lists the stories and their word counts (for simplification, the condition indicators S and U are left off).

Simple word co-occurrence frequencies were used to construct the semantic spaces. An initial evaluation of the impact of weighing the word co-occurrence matrix by logarithmic entropy (see also Landauer & Dumais, 1997) yielded knowledge representations that were inferior to using simple word co-occurrence frequencies. Therefore, the unweighted word co-occurrence matrix was used.

After LSA representations were constructed for all possible dimensions, optimal LSA representations were determined with and without semantic extraction. Optimal LSA representations were thereby defined as those that are most similar to human performance (see Landauer & Dumais, 1997). For this purpose, the participants' 45 relatedness ratings in each condition were averaged and correlated with the cosines between the corresponding word vectors for each of the 12 LSA representations (Solved and Unsolved versions of LSA 1 to LSA 15) at each dimensionality of word vectors. For each of the 12 LSA representations, only word vectors with the dimensionality that correlated maximal with human ratings were considered optimal and were retained for the following analysis. For example, Figure 1 shows correlations between LSA and human ratings with all possible word vector dimensionalities for LSA 12 in the Solved condition.

Three participants were excluded from the analysis because they had previously read the stories. One randomly selected participant was eliminated in order to balance the design. Overall, the data of 40 participants were included, 20 in each condition.

The responses on the comprehension questionnaires indicated that participants' comprehension differed significantly between the conditions. Participants in the Solved condition answered a significantly larger proportion of questions correctly than participants in the Unsolved condition, $t(38)=3.1$, $p < .05$, (α was set to 0.05 except where otherwise noted). This difference was expected and confirmed that participants acquired different knowledge in the two conditions.

The agreement among participants' knowledge structures was assessed by correlating each participant's 45 word relatedness ratings with the average of all other participants' ratings in each condition. The resulting correlations were normalized using Fisher's r to z transformation and then averaged for the Solved and Unsolved condition. The averages were then transformed back to correlation coefficients. The 45 word similarity ratings correlated with $r(44) = 0.633$, $p < .01$ within the Solved condition and $r(44) = 0.450$, $p < .01$ within the Unsolved condition. This indicates significant agreement of participants' knowledge for both conditions. The correlation is somewhat lower in the Unsolved condition (but not significantly lower, $p > .20$). This could be due to the fact that the Unsolved version did not allow participants to integrate the key concepts of the story to the same degree as in the Solved condition. Because they did not read the end of the story, participants in the Unsolved condition probably relied on background knowledge to rate word similarities, which is consistent with the greater variability of the ratings.

Table 6 displays the correlations between optimal LSA representations and participants' judgments for both conditions. All correlations were found to be significant. Importantly, Table 6 shows that the dimensionalities that yielded a maximal match between LSA and participants' ratings were different between the conditions. Optimal dimensionalities in the Unsolved condition seemed higher than in the Solved condition.

Figure 3 displays the optimal dimensionalities of LSA representations in percentages of maximal dimensionalities (word vectors include all principal components). This indicates that semantic extraction had a different effect on the quality of LSA knowledge representations in the Solved and Unsolved condition. Optimal LSA-representations in the Unsolved condition had slightly reduced dimensionalities between 91 % and 96 % of all dimensions except for optimal LSA 15 with 0.6 % of all dimensions (see dashed line in Figure 3 and percentages in Table 7). Compare this to the Solved condition where optimal LSA-representations contained all less than a third of all dimensions (see solid line in Figure 3 and percentages in Table 7). Semantic extraction was crucial to match knowledge in the Solved condition but not in the Unsolved condition except for LSA 15 where semantic extraction led to match in both conditions.

This indicates that the effect of semantic extraction on the match of the LSA knowledge with participants' knowledge depends on the level of readers' comprehension and the amount of background knowledge in the LSA representations. Only if enough stories are included in the Unsolved LSA representations, low vector dimensionalities resulted in an optimal match with human participants. Compare this to the Solved condition where all optimal LSA representations had low vector dimensionalities.

Using semantic extraction, the correlations of word similarity ratings between LSA and participants in the Solved condition were all significant and increased with the size of the semantic space from LSA 1 to LSA 12 (see solid line in Figure 4 and correlations in column 4 in Table 8), indicating that the quality of LSA representations increased as semantic spaces incorporated more texts. Note that LSA 15 resulted in slightly lower representational quality compared with LSA 12. At LSA 12, word similarity ratings reached maximal match and correlated about as high with averaged participant ratings as participants agreed among themselves, indicated by the solid line in Figure 4 (participants' agreement, $r(44) = 0.634$, $p < .01$, straight solid line in Figure 4, and LSA – participants $r(44) = 0.638$, $p < .01$, curved, solid line in Figure 4). At LSA 15, the match with participants' ratings was again lower than at LSA 12, $r(44) = 0.564$, $p = .05$ (not significantly lower than at LSA 12, $p > 0.5$) which suggests an upper boundary for optimal semantic spaces. Inclusion of more information into the semantic space might result in less optimal LSA knowledge representations because too much non-relevant information distorts the associations between the key concepts.

In the Unsolved condition, the correlations of word similarity ratings between LSA and participants did not correlate significantly for any size of semantic space below LSA 15. At LSA 15, the correlation was significant ($r(44) = 0.537$, $p < .01$, see curved dotted line in Figure 4) and was higher than the participants agreement among themselves (participants' agreement: $r(44) = 0.45$, $p < .01$, LSA – participants: $r(44) = 0.537$, $p < .01$, see straight, dotted line in Figure 4). LSA 15 was more similar to each participant's knowledge than participants averaged knowledge structures. This suggests that participants in the Unsolved condition relied more on their background knowledge when

rating word similarities than in the Solved condition. Because small semantic spaces did not contain enough information to capture participants' background knowledge, those semantic spaces did not facilitate LSA knowledge presentations that matched participants' responses.

The findings are in concordance with previous research that found that optimal LSA knowledge representations match human knowledge and that they do so better for consistent knowledge structures such as of experts. Foltz and Wells (1998) reported that LSA knowledge representations correlated with expert knowledge within the range of $r = 0.31$ to $r = 0.37$ but for novice knowledge at around $r = 0.19$. Like experts, readers in the Solved condition were better able to extract information from their text and did so more consistently. Interestingly, the correlations in this experiment were higher than in Foltz and Wells (1998) which might be due to differences of the used texts (fiction versus non-fiction text) and the experimental procedure (e.g. rating 45 word pairs in this experiment versus 120 word pairs in their experiment).

The effect of semantic extraction was assessed by comparing LSA knowledge with and without semantic extraction to participants' knowledge. LSA knowledge without semantic extraction was constructed by including all dimensions into LSA word vectors. Such LSA knowledge correlated in the Solved condition lower with participants' judgments than when semantic extraction was used (see Figure 5 and column 5 in Table 8). The correlations decreased with larger semantic spaces. This agrees with the expectation that the quality of LSA knowledge representations is caused by the dimensional reduction of word vectors. However, in the Unsolved condition the knowledge match between LSA and participant ratings was higher for LSA 1 to LSA 12

without semantic extraction. Without semantic extraction, word similarity ratings between LSA and participants correlated significantly for LSA 1 to LSA 9 but not for larger LSA representations (LSA 12 and LSA 15, dashed line in Figure 5 and column 5 in Table 8).

These are unexpected findings as they indicate that word relatedness ratings in the Unsolved condition are better predicted by pure word co-occurrence information than by the semantic structure of the text. It seems less surprising that neither LSA nor participants could extract enough semantic information from the Unsolved version of the Red-headed League that did not allow complete story comprehension. It is interesting however, that in that case word co-occurrences predicted word similarities better.

A second measure of LSA knowledge quality is the difference in LSA knowledge representations between the two conditions. Discriminative analysis was used to test if participants' knowledge in one condition was consistently more similar to the LSA knowledge representation in this condition and different from knowledge in the other condition. Again, the similarity between LSA word vector similarities and participants' word relatedness ratings were correlated and used to measure knowledge agreement. The results of a multivariate analysis (see Table 9) indicated that participants could be correctly categorized at a rate greater than chance. The best discrimination was evidenced in LSA 15 where 85 % of participants' level of comprehension could be correctly categorized (see Table 9). Overall, LSA representations were more similar to participants within each condition and were dissimilar to the participants in the other condition. The connection between the size of the semantic space and quality of differentiation appeared to be somewhat curvilinear, yielding best differentiation at the smallest (LSA 1: 80 %

correct categorizations) and largest semantic space (LSA 15: 85 % correct categorizations). This corresponds to percentages of correct classification shown in Table 9 and that are plotted as solid line in Figure 6.

This indicates that differentiation among participants was accomplished in two different ways. First, small semantic spaces seemed to allow differentiation among participants because of the relatively strong impact of the differences in the included texts. For LSA 1 this difference was greatest as it contained only the two different texts that participants had read. Secondly, for large semantic spaces the impact of the different experimental texts was relatively smaller because large amounts of the identical text were added to semantic spaces in both conditions. Instead, differentiation was caused by differences in the underlying semantic structure that was revealed by the dimensional reduction of the LSA representations.

LSA knowledge representations seemed to discriminate about as well among participants as the measurement quality of word similarity ratings permitted. A way to estimate the quality of this measurement was to see how many participants would be grouped correctly into their condition by the similarity of their ratings with the ratings in the same condition and the dissimilarity from ratings in the other condition. A discriminative analysis revealed significant differences between conditions (multivariate analysis, $F(2,37) = 27.93$, $p<.01$) with 85 % correct categorizations (dotted line in Figure 6). This is the same percentage of correct classifications that was obtained with LSA 15.

Overall, the results of Experiment 1 indicated that the factors for building optimal LSA knowledge representations strongly depend on the amount of knowledge that the experimental texts revealed. The story in the Solved condition revealed enough

knowledge to allow for complete text comprehension and LSA knowledge representations at any size correlated significantly with participants' ratings. However, if stories did not allow complete comprehension, only LSA knowledge representation in the largest semantic space (LSA 15) matched participants' ratings in the Unsolved condition.

A similar pattern was found in the analysis of the effect of dimensional reduction of LSA word vectors. Dimensional reduction led to optimal knowledge representations in the Solved condition at any size of semantic spaces. In the Unsolved condition, dimensional reduction had an effect only for the largest semantic space (LSA 15). None of the other semantic spaces in the Unsolved condition facilitated optimal LSA knowledge representations using dimensional reduction (see Table 8).

Participants appear to use different knowledge to rate word pair relatedness, dependent on their level of story comprehension. With full story comprehension they relied on story knowledge to rate word pair similarities but did less so with incomplete comprehension in the Unsolved condition. Participants could not acquire enough information from the story and therefore relied on their background knowledge to rate the relatedness of word pairs. Therefore, only if enough background information was included in the semantic spaces LSA representations would correlate significantly with participants' word relatedness ratings in the Unsolved condition. However, it remains to be seen if LSA knowledge representations derived from larger semantic spaces capture background knowledge better than smaller semantic spaces. Experiment 2 was intended to answer this question. In Experiment 2 participants rated the same key-concepts that were rated in Experiment 1 but never read the "Red-headed League".

Experiment 2

Method

Participants

Thirty-eight undergraduate students participated for partial fulfillment of an experimental requirement. None of them had previously read the Red-headed League and all participants were native English speakers.

Design and Procedure

The same 10 concepts from experiment 1 were used in this experiment. The participants were asked to indicate their general impression of the relatedness of all pairs of concepts on a scale from 0 to 9. They were informed that they could consider a number of factors including similarity, co-occurrence, or dependency and were asked not to ponder about their judgments. For cases when they would not be able to decide about word relatedness because of unknown words, they were instructed to give minimal ratings. Word pairs were printed on a separate sheet in randomized order. Also the word positions were randomized and word order and positions were different for every participant. After completing the relatedness ratings, participants indicated their familiarity with Sherlock Holmes stories and whether they had previously read the Red-headed League.

LSA knowledge representations

The similarity ratings of all participants were averaged. This average of word similarity ratings is referred to as participants' background knowledge because it reflects the participants' associations among key concepts in the Red-headed League without participants' knowledge of the story. Participants' background knowledge was then

correlated with the word similarity ratings that were calculated from the semantic spaces LSA 1 to LSA 15 for the Solved and Unsolved versions of the story. The semantic spaces were the same as used in Experiment 1. The correlations were averaged for each size LSA 1 to LSA 15 across the Solved and Unsolved spaces.

Results and Discussion

Figure 7 shows the correlations between participants' word relatedness ratings and the ratings derived from LSA-representations. The match between averaged LSA and participants' ratings increased with the size of the LSA text corpus and a significant correlation was found for the largest semantic space, LSA 15, $r(44) = 0.513$ $p < .01$ (the correlation is not significantly greater than LSA 12, $p > 0.2$, see Table 10). This indicates that LSA knowledge representations agree more with the background knowledge of participants in larger semantic spaces than smaller ones.

Conclusions

The present research evaluated LSA representations in multiple semantic spaces to represent knowledge differences between participants. Differences in text comprehension were successfully represented in different semantic spaces. LSA representations matched participants' comprehension of a story but matched incomplete story comprehension only if semantic spaces included sufficient background knowledge. Larger semantic spaces do match participants' knowledge better than smaller spaces. LSA representations with sufficient background knowledge also allowed the differentiation among participants with different levels of comprehension.

Implications for Knowledge Elicitation

These findings have important implications for the application of LSA as an automated tool of knowledge elicitation. One of the previous limitations of LSA has been its restriction to predefined domains for which semantic spaces had to be constructed. Knowledge could be only assessed and compared within this domain. The current research indicates that LSA knowledge representations can be compared across multiple semantic spaces.

Automated knowledge assessment for user-adaptive information retrieval could benefit from the proposed technique. Research in information retrieval from large information sources has for a long time considered ways to represent user knowledge for automated user-adaptive retrieval. For example, Oddy (1977) developed an automated information retrieval system that attempted the construction of user models by means of an interactive human-computer dialog. Similarly, Belkin (1980) developed a system to support identification of incomplete knowledge states of users and to match those states with retrieval documents. Other automated information retrieval algorithms incorporated the knowledge of user intermediaries and expert systems such as CONIT, CANSEARCH, PLEXUS, and MONSTRAT (see Davis, 1996) who used production rules that reflected the knowledge of information retrieval experts. One limitation of such systems has been described as their application to rather well-defined and restricted domains (Davis, 1996).

The current research led to the expectation that user knowledge differences can be represented by LSA knowledge representations in multiple semantic spaces. User background knowledge would be captured in semantic spaces and a user's search query could be compared with the LSA representations of retrieval documents. Semantic spaces

could be derived from text samples that clients previously had read. It is important that this process formulates search queries dependent on user background knowledge because the queries' LSA representations are built in the clients' semantic space. It is expected that the influence of user background knowledge for adaptive information retrieval should yield higher ratios of relevant search results. Empirical investigations of implementations of such retrieval systems will have to test this expectation.

Texts are needed for this process to represent users' background knowledge for the construction of relevant semantic spaces. For example, texts that users browse on the web could be collected and be used to categorize users into user profiles for which semantic spaces are constructed. Methods for clustering user sessions are currently investigated by Heer and Chi (2002).

Dimensional Reduction and Comprehension

The current research links the effect of dimensional reduction in LSA knowledge representations to the amount of information and level of text comprehension. The findings confirm the relationship between dimensional reduction and semantic extraction (Foltz & Wells, 1999; Landauer & Dumais, 1997). Previous research has found that semantic extraction is essential for LSA representations to agree with human performance. The current research replicates these findings for LSA knowledge representations in large semantic spaces. In smaller semantic spaces, however, dimensional reduction worked only if the texts contained enough relevant information (Solved condition). If texts did not include enough relevant information, participants' word pair relatedness ratings correlated higher with word co-occurrences than with information that was extracted using dimensional reduction.

Accordingly, participants who did not acquire full story comprehension seemed to use at least two different strategies to rate the relatedness of concepts. First, participants used information about how words co-occurred to rate concept relatedness. Participants tended to rate concepts that occurred in the same paragraphs as more related. The proximity of those concepts might have led to concurrent activation in memory during reading which influenced relatedness judgments when concepts could not be linked through comprehension. Secondly, participants relied on their background knowledge to infer the relatedness between concepts. This was revealed by the match of LSA knowledge representations that included the most background knowledge (LSA 15). For this largest LSA knowledge representation (LSA 15), the same amount of dimensional reduction was needed in the Solved as in the Unsolved condition to match participants' ratings. Therefore, the dimensional reduction of LSA word vectors seems to reflect participants' reliance on the story's semantic structure in the Solved condition and background knowledge in the Unsolved condition for their concept relatedness ratings.

Text comprehension between the Solved and Unsolved condition differed by their influence of background knowledge. In the Solved condition participants rated concept relatedness based on actual story comprehension and used less background knowledge than in the Unsolved condition (decreased match of LSA 15 compared to LSA 12 see solid line in Figure 4). In the Unsolved condition however, concept relatedness ratings were strongly affected by participants background knowledge (increased match of LSA 15 compared to LSA 12, see dashed line in Figure 4). This indicates that readers used flexible strategies to achieve text comprehension that depended on the amount of information contained in the reading material. If not enough information was contained in

the reading material, readers filled the gaps in the semantic structure with their background knowledge and tended to relate concepts according to their co-occurrence in paragraphs.

Size of Semantic Space and Level of Comprehension

The results in Experiment 2 showed a direct relation between the size of semantic spaces and their agreement with participants' story relevant background knowledge. Note that any of the semantic spaces beside LSA 1 were derived from texts that participants had not previously read. Still, those semantic spaces facilitated knowledge representations that were similar to participants' background knowledge prior to reading the story. LSA representations in very similar semantic spaces that included more than 98 % of the same texts (e.g. LSA 15) still yielded significantly different LSA knowledge representations. The possibility of such clear differences indicates that LSA knowledge representations are highly sensitive to relatively small text differences in underlying semantic spaces.

Future directions

Further investigations should explore the influence of specific text information on the quality of LSA representations. In the current study corpora of text were incrementally added to semantic spaces. Investigating the sequence of adding texts into the semantic spaces might reveal an important connection between specific information in a semantic space and the information that participants require for text comprehension. Finally, studies using different types of text and knowledge could be used to generalize the current findings to a broader range of applications.

A practical implementation of LSA knowledge representations for automated information retrieval involves several steps, ranging from the development of a methodology to collect texts for user knowledge representation, to the construction of an appropriate information retrieval system, and the testing of retrieval performance of such a system. The current research proposed the modification of LSA for such tasks and demonstrated how computerized knowledge representations correspond to human knowledge for different levels of story comprehension. Further investigations and applications are expected to build on the gained knowledge.

References

Berry, M. W. (1992). Large scale singular value computations. International Journal of Supercomputer Applications, 6, 13-49.

Belkin, N. J. (1980). Anomalous states of knowledge as the basis of information retrieval, Canadian journal of information science, 5, 133-143.

Belkin, N., Brooks, H., & Oddy R. (1982). ASK for information retrieval, Part 1. Journal of Documentation, 38, 61-71.

Chen, C. (1997). Tracking latent domain structures: an integration of pathfinder and latent semantic analysis. AI & Society, 11, 48-62.

Conan Doyle, A. (1892). The Red-headed League. In: The Adventures of Sherlock Holmes. London: George Newnes. [On-line]. Available: ftp://sailor.gutenberg.org/pub/gutenberg/etext99/advsh10.txt

Cooke, N., J. (1999). Knowledge elicitation. In: F. T. Durso (Ed.), Handbook of applied cognition (pp. 479-510). Chichester, England: John Wiley & Sons Ltd.

Davis, E. (1996). Progress and Problems in Information Retrieval. Library Association Publishing: London, UK.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T., K., & Harschman, R. (1990). Indexing by Latent Semantic Analysis. Journal of the American Society for Information Science, 41(6), 391 – 407.

Dumais, S. T. (1991). Improving the retrieval of information from external sources. Behavior Research Methods, Instruments and Computers, 23, 229-236.

Dumais, S. T. (1994). Latent semantic indexing (LSI) and TREC-2. In D. Harman (Ed.), National Institute of Standards and Technology Text Retrieval Conference. NIST special publication.

Durso, F. T., Rea, C., & Dayton, T. (1994). Graph-theoretic confirmation of restructuring during insight. Psychological Science, 5(2), 94-98.

Foltz, P. W. (1990). Using latent semantic indexing for information filtering. In R. B. Allen (Ed.), Proceedings of the Conference on Office Information Systems, Cambridge, MA, 40-47.

Foltz, P. W. (1992). Personalized information delivery: an analysis of information filtering methods. Communications of the ACM, 35(12), 51-60.

Foltz, P. W. (1996). Latent semantic analysis for text-based research. Behavior Research Methods, Instruments and Computers, 28(2), 197-202.

Foltz, P. W., Britt M. A., & Perfetti, C. A. (1996). Reasoning from multiple texts: An automatic analysis of readers' situation models. In G. W. Cottrell (Ed.), Proceedings of the 18th Annual Cognitive Science Conference (pp. 110-115). Lawrence Erlbaum, NJ.

Foltz, P. W., Kintsch, W., Landauer, T. K. (1998). The measurement of textual coherence with latent semantic analysis. Discourse Processes, 25(2-3), 285-307.

Foltz, P. W., & Wells, A. D. (1999). Automatically deriving reader's knowledge structures from texts. Behavior Research Methods, Instruments, and Cognition, 31(2), 208-214.

Freeman, J. T., Thompson, B. T., Cohen, M. S. (1999). Modeling and Diagnosing Domain Knowledge using Latent Semantic Indexing. Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting, 233-236.

Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: a comparison of high-dimensional and embodied theories of meaning. Journal of Memory and Language, 43, 379-401.

Gordon, M. D., & Dumais, S. (1998). Using latent semantic indexing for literature based discovery. Journal of the American Society for Information Science, 49(8), 674-685.

Graesser, A., Wiemer-Hastings, P., Wiemer-Hastings, K., Harter, D., Person, N., Tutoring Research Group, & Person N. (2000). Using Latent Semantic Analysis to Evaluate the Contributions of Students in AutoTutor. Interactive Learning Environments, 8(2), 129-147.

Heer J., & Chi, E. H. (2002). Separating the swarm: categorization methods for user sessions on the web. Proceedings of the Conference on human factors in computing systems, 4, 243-250.

Howard, M. W., & Kahana, M. J. (2002) When does semantic similarity help episodic retrieval? Journal of Memory and Language, 46, 85-96.

Kintsch, W. (1998). The representation of knowledge in mind and machines. International Journal of Psychology, 33(6), 411-420.

Kintsch, W. (2002). On the notions of theme and topic in psychological process models of text comprehension. In M. Louwerse & van Peer (Eds.), Thematics: Interdisciplinary Studies (pp. 157-170). Amsterdam: Benjamins.

Kintsch, E., Steinhart, D., Stahl, G. & LSA research group (2000). Developing Summarization Skills through the Use of LSA-Based Feedback. Interactive learning environments,8(2) , 87-109.

Kuhltau, C. C., Turock, B. J., George, M. W., & Belvin, R. J. (1990). Validating a model of the search process: A comparison of academic, public and school library users. Library Information Science Research, 12, 5-31.

Kuralenok, I., & Nekrest'yanov (2000). Automatic document classification based on latent semantic analysis. Programming and Computer Software, 26(4), 199-206.

Landauer, T., K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. Psychological Review, 104(2), 211-240.

Landauer, T. K., & Littman, M. L. (1990). Fully automatic cross-language document retrieval using latent semantic indexing. Proceedings of the Sixth Annual Conference of the UW Centre for the New Oxford English Dictionary and Text Research (pp. 31-38). Waterloo, Ontario: UW Centre for the New OED.

Landauer, T. K. (1999). Latent semantic analysis (LSA), a disembodied learning machine, acquires human word meaning vicariously from language alone. Behavioral and Brain Sciences, 22(4), 624-625.

Moertl, P. M. (2001). Modeling User Knowledge and Semantic Structure for Information Extraction from Text. Proceedings of the Fourth International Conference for Cognitive Modeling, Fairfax, Virginia, 283-284.

Moertl, P. M., & Durso, F. T. (2000). Representing Knowledge in Air Traffic Control Using Latent Semantic Analysis. Paper presented at the meeting of the Society for Applied Research in Memory and Cognition, Miami, Florida, June 7 - 8.

Oddy, R. N. (1977). Information retrieval through man-machine dialogue. Journal of documentation, 33, 1-14.

Rehder, B., Schreiner, M. E.; Wolfe, M. B. W., Laham, D., Landauer, T. K., Kintsch, W. (1998). Using latent semantic analysis to assess knowledge: Some technical considerations. Discourse Processes, 25(2-3), 337-354.

Rescorla, R. A. & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokasy (Eds.), Classical conditioning: Vol. 2. Current research and theory. New York: Appleton-Century-Crofts.

Salton, G. (1971). The SMART retrieval system – experiments in automatic document processing, Englewood Cliffs, Prentice Hall, 1971.

Salton, G., & McGill, M. J. (1983). Introduction to modern information retrieval, New York, McGraw Hill.

Salton, G., Wong, A., & Yang (1997). A vector space model for automatic indexing. In: K. Sparck Jones & P. Willett (Eds.), Readings in Information Retrieval (pp. 273-280) Morgan Kaufman Publishers: San Francisco, CA. (Reprinted from Communications of the ACM, 18, 1975, 613-620).

Schvaneveldt, R. W., Durso, F. T., Goldsmith, T. E., Breen, T. J., Cook, N, Tucker, R. G., & De Maio, J. (1985). Measuring the structure of expertise. International Journal of Man-Machine Studies, 22, 699-738.

Soto, R. (1999). Learning and performing by exploration: label quality measured by latent semantic analysis. Proceedings of the Computer Human Interaction Conference on Human Factors and Computing Systems, 418-425.

Taylor, R. (1989). Question-Negotiation and Information Seeking in Libraries. College and Research Libraries, 29(3), 178-194.

Wagner, A. R. & Rescorla, R. A. (1972). Inhibition in Pavlovian conditioning: Application of a theory. In R. A. Boakes & M. S. Halliday (Eds.), Inhibition and learning. New York: Academic Press.

Wolfe, M. B. W, Schreiner, M. E., Rehder, B., Laham, D., Foltz, P. W., Kintsch, W., & Landauer, T. K. (1998). Learning from text: matching readers and texts by latent semantic analysis. Discourse Processes, 25(2-3), 309-336.

Table 1. Word and paragraph count of the experimental texts that were used in experiments 1 and 2.

	Original Story "The Red-headed League"	Version "Solved"	Version "Unsolved"
Words	9115	6840	6910
Paragraphs	218	174	177

Table 2. Key concepts and their frequency in both versions of the story.

Frequency of keywords (major elements of story)	Keyword occurrence in each experimental text "mystery solved" and " mystery solved"
Assistant	13
Bank	6
Hair	12
Sherlock Holmes [a]	40
Jabez Wilson [b]	8
Vincent Spaulding [a]	8
Pawnbroker	5
Red	26
Tunnel	3
Cellar	11

Note. [a] only the last name was presented to participants for their word pair relatedness ratings, [b] only the first name was presented.

Table 3. Changes to the story "The Red-headed League" that resulted in two version of the story.

Line	Action	Description	Version
3	Add	Because I had run late from an appointment with my bank	unsolved
6	change	helper -> assistant	unsolved
15- 29	Delete		solved
44-87	Delete		solved
68	Change	as as ship's carpenter -> my apprenticeship in a tunnel construction company	unsolved
87	Add	said Holmes	unsolved
142	change	he -> my assistant	unsolved
158	add	as founding member of a large bank and tunnel construction company	unsolved
182	add	in their hair	unsolved
186	add	that was in the cellar	unsolved
186	change	up -> down	unsolved
210	add	cellar	unsolved
212	change	up from below -> down from above	unsolved
243	add	in this cellar	unsolved
277	change	in -> to the cellar	unsolved
278	add	cellar	unsolved
281	change	hoped with diligence that I might get on to the B's before very long -> just started to write about Bank	unsolved
287	add	went down into the cellar	unsolved
307	add	bank	unsolved
311	add	in the cellar	unsolved
312	change	red-headed man -> man with red hair	unsolved
315	add	cellar	unsolved
318	add	cellar	unsolved
369-383	delete		solved
395	add	pawnbroker	unsolved
437-453	delete		solved
468-479	delete		solved
490-502	delete		solved
515-518	delete		solved
527-693	delete	(conclusion of mystery)	unsolved
536-548	delete		solved
568-587	delete		solved
623-641	delete		solved

Note. The line numbers refer to the line numbering of the story in Appendix B.

Table 4. List of 45 word pairs used in Experiments 1 and 2.

assistant	bank	Holmes	Jabez
assistant	cellar	Holmes	pawnbroker
assistant	hair	Holmes	red
assistant	Holmes	Holmes	Spaulding
assistant	Jabez	Holmes	tunnel
assistant	pawnbroker	Jabez	pawnbroker
assistant	red	Jabez	red
assistant	Spaulding	Jabez	Spaulding
assistant	tunnel	Jabez	tunnel
bank	cellar	pawnbroker	red
bank	hair	pawnbroker	Spaulding
bank	Holmes	pawnbroker	tunnel
bank	Jabez	red	Spaulding
bank	pawnbroker	red	tunnel
bank	red	Spaulding	tunnel
bank	Spaulding		
bank	tunnel		
cellar	hair		
cellar	Holmes		
cellar	Jabez		
cellar	pawnbroker		
cellar	red		
cellar	Spaulding		
cellar	tunnel		
hair	Holmes		
hair	Jabez		
hair	pawnbroker		
hair	red		
hair	Spaulding		
hair	tunnel		

Table 5. Selections of stories of Sir Arthur Conan Doyle from which semantic spaces were constructed.

Semantic space of:	Included stories:	Word count	Paragraph count
1 story	The Red-headed League	6840	174
3 stories	+ The Man with the Twisted Lip	25792	1966
	+ The Adventure of the Beryl Coronet		
6 stories	+ A Case of Identity	50726	4330
	+ The Boscombe Valley Mystery		
	+ The Adventure of the Engineer's Thumb		
9 stories	+ The Adventure of the Copper Beeches	75895	6744
	+ The Five Orange Pips		
	+ The Adventure of the Blue Carbuncle		
12 stories	+ The Adventure of the Noble Bachelor	102412	9361
	+ The Adventure of the Speckled Band		
	+ A Scandal in Bohemia		
15 stories	+ The Hound of the Baskervilles	263041	24890
	+ The Sign of the Four		
	+ The Valley Of Fear		

Note. Numbers refer to the Solved version of the story. The Unsolved version is 90 words longer and has 3 more paragraphs.

Table 6. Correlations between optimal LSA representations and participants' ratings

Semantic space	Condition	LSA-Size in Words	Correlation with participant ratings	LSA dimen-sionality	Maximal LSA Dimen-sionality
LSA 1	Solved	6,840	0.3377*	44	154
	Unsolved	6,910	0.475**	145	152
LSA 3	Solved	25,792	0.4851**	22	565
	Unsolved	25,861	0.3561*	538	566
LSA 6	Solved	50,726	0.5476**	15	1063
	Unsolved	50,795	0.3843**	1003	1066
LSA 9	Solved	75,895	0.5810**	20	1593
	Unsolved	75,964	0.3349*	1469	1600
LSA 12	Solved	102,412	0.6381**	27	2211
	Unsolved	102,482	0.2977*	2109	2217
LSA 15	Solved	263,041	0.5642**	31	5310
	Unsolved	263,066	0.5370**	31	5319

Note: * $p<0.05$, ** $p < 0.01$

Table 7. Optimal dimensionality of word vectors as expressed by percentage of maximal dimensionality

Semantic space	Percentage of maximal dimensionality	
	Solved	Unsolved
LSA 1	28.6 %	95.4 %
LSA 3	3.9 %	95.1 %
LSA 6	1.4 %	94.1 %
LSA 9	1.3 %	91.8 %
LSA 12	1.2 %	95.1 %
LSA 15	0.6 %	0.6 %

Table 8. Experiment 1: Results of Comparison between LSA and participant's knowledge and details of LSA

Semantic space	Condition	LSA-Size in Words	Correlation with participant ratings using semantic extraction	Correlation with Human ratings at maximal dimensionality	LSA dimen-sionality for semantic extraction	Maximal LSA Dimen-sionality
LSA 1	Solved	6,840	0.3377*	0.2641	44	154
	Unsolved	6,910	0.0360	0.4350**	45	152
LSA 3	Solved	25,792	0.4851**	0.2793	22	565
	Unsolved	25,861	0.0250	0.3004*	173	566
LSA 6	Solved	50,726	0.5476**	0.2213	15	1063
	Unsolved	50,795	0.0198	0.3413*	327	1066
LSA 9	Solved	75,895	0.5810**	0.0482	20	1593
	Unsolved	75,964	0.0563	0.3181*	19	1600
LSA 12	Solved	102,412	0.6381**	0.1110	27	2211
	Unsolved	102,482	0.0505	0.2670	18	2217
LSA 15	Solved	263,041	0.5642**	-0.1292	31	5310
	Unsolved	263,066	0.5370**	0.0294	31	5319

Note: * $p<0.05$, ** $p < 0.01$. LSA dimensions below one third of the maximal dimensionality reflect semantic extraction.

Table 9. Experiment 1: Results of discriminant analysis.

Semantic space	F (2, 37) for multivariate analysis of variance	Percentage of correct categorizations
LSA 1	27.14**	80 %
LSA 3	18.52**	82.5 %
LSA 6	5.83**	67.5 %
LSA 9	5.34**	67.5 %
LSA 12	9.56**	72.5 %
LSA 15	23.27**	85 %

Note. Baseline of participants' agreement with average judgment was 85 %

Table 10. Experiment 1: Correlations between LSA and participants' word similarity ratings without reading the story (story relevant background knowledge).

Semantic space	Correlation coefficient
LSA 1	0.090
LSA 3	0.145
LSA 6	0.186
LSA 9	0.263
LSA 12	0.297*
LSA 15	0.513**

Note. * $p < .05$, ** $p < .01$

Figure Captions

Figure 1. The effect of number of dimension retained in LSA representations on the match with participants' word similarity ratings.

Figure 2. Experiment 1: Match between LSA and Participants Knowledge per Experimental Condition and LSA-Size at Optimal Dimensionality

per LSA-Size at Optimal Dimensionality

Figure 3. Experiment 1: Percentage Optimal of Maximal LSA-dimensionality per Experimental Condition, Lower Values Indicating Higher Levels of Semantic Extraction

Figure 4. Experiment 1: Match between LSA and Participant Knowledge per Experimental Condition and LSA-Size Using Semantic Extraction.

Figure 5. Experiment 1: Comparison of the Effect of Semantic Extraction on the Match between LSA and Participant Knowledge.

Figure 6. Experiment 1: Percentage of Correct Categorizations of Reader Type from Match between LSA and Participant Knowledge

Figure 7. Experiment 2: Influence of LSA-Size on Match with Participants' Story Relevant Background Knowledge Without Reading the Story

Figure 1

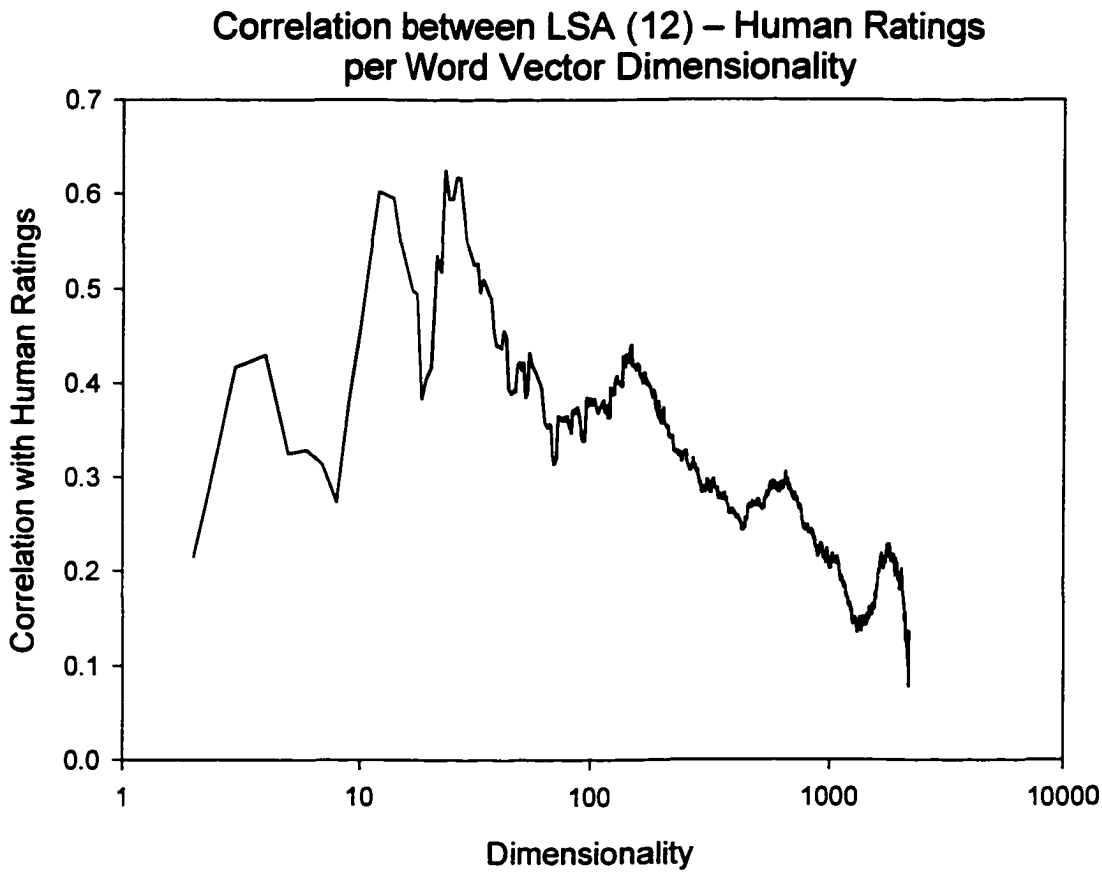

Correlation between LSA (12) – Human Ratings
per Word Vector Dimensionality
Correlation with Human Ratings
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
1
10
100
1000
10000
Dimensionality

Figure 2

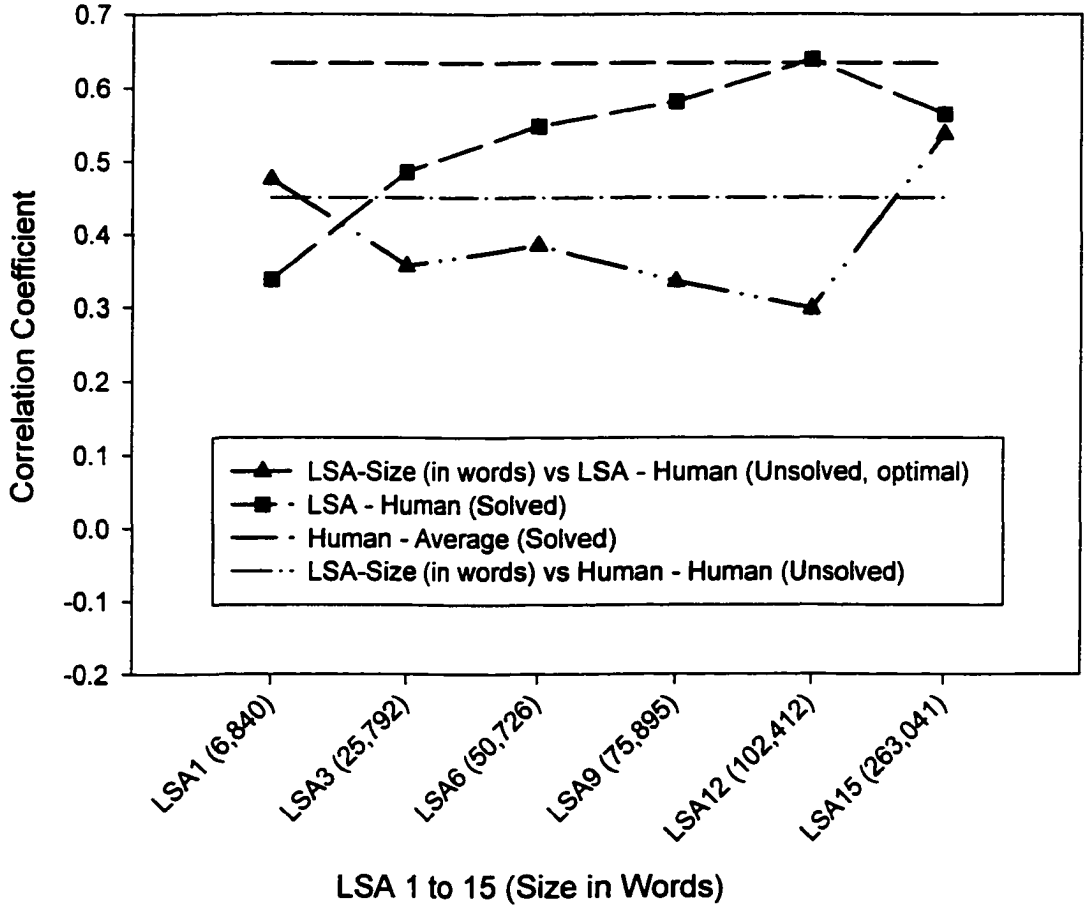

0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
Correlation Coefficient
LSA-Size (in words) vs LSA - Human (Unsolved, optimal)
LSA - Human (Solved)
Human - Average (Solved)
LSA-Size (in words) vs Human - Human (Unsolved)
LSA1 (6,840)
LSA3 (25,792)
LSA6 (50,726)
LSA9 (75,895)
LSA12 (102,412)
LSA15 (263,041)
LSA 1 to 15 (Size in Words)

Figure 3

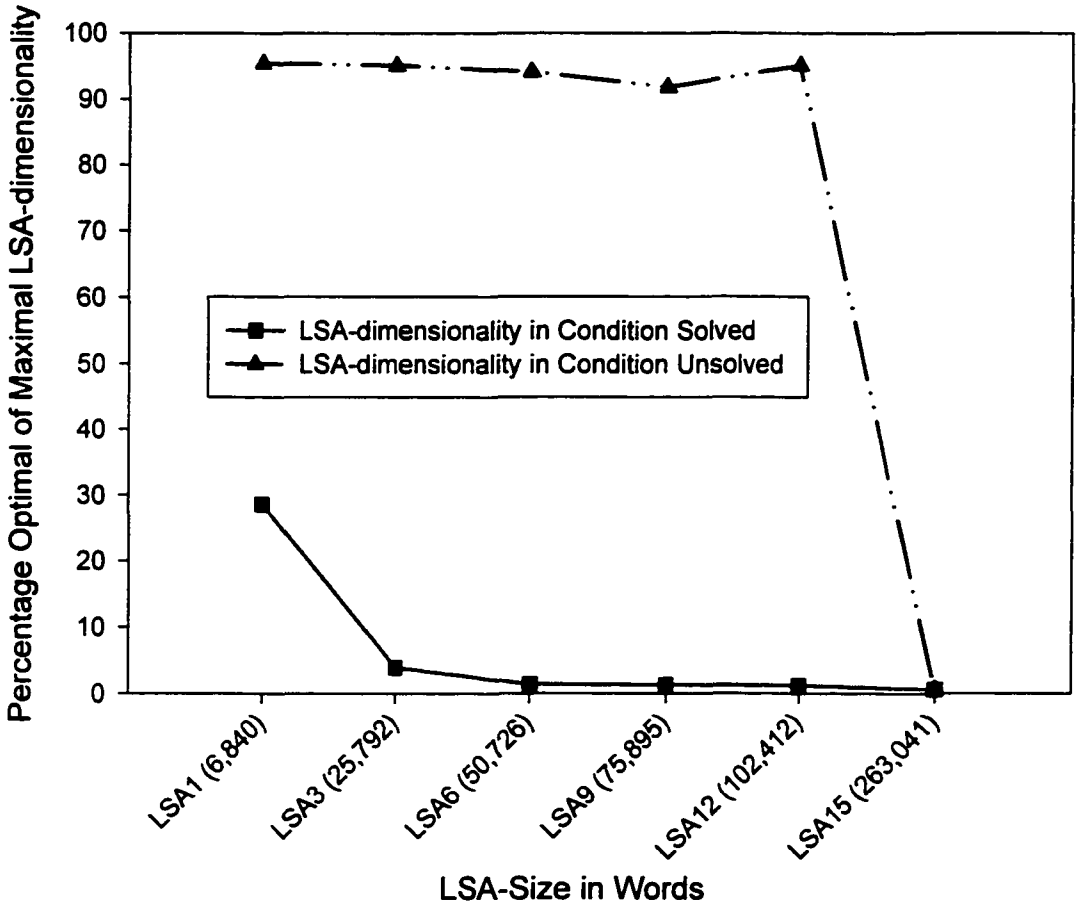

Percentage Optimal of Maximal LSA-dimensionality
100
90
80
70
60
50
40
30
20
10
0
LSA-dimensionality in Condition Solved
LSA-dimensionality in Condition Unsolved
LSA1 (6,840)
LSA3 (25,792)
LSA6 (50,726)
LSA9 (75,895)
LSA12 (102,412)
LSA15 (263,041)
LSA-Size in Words

Figure 4

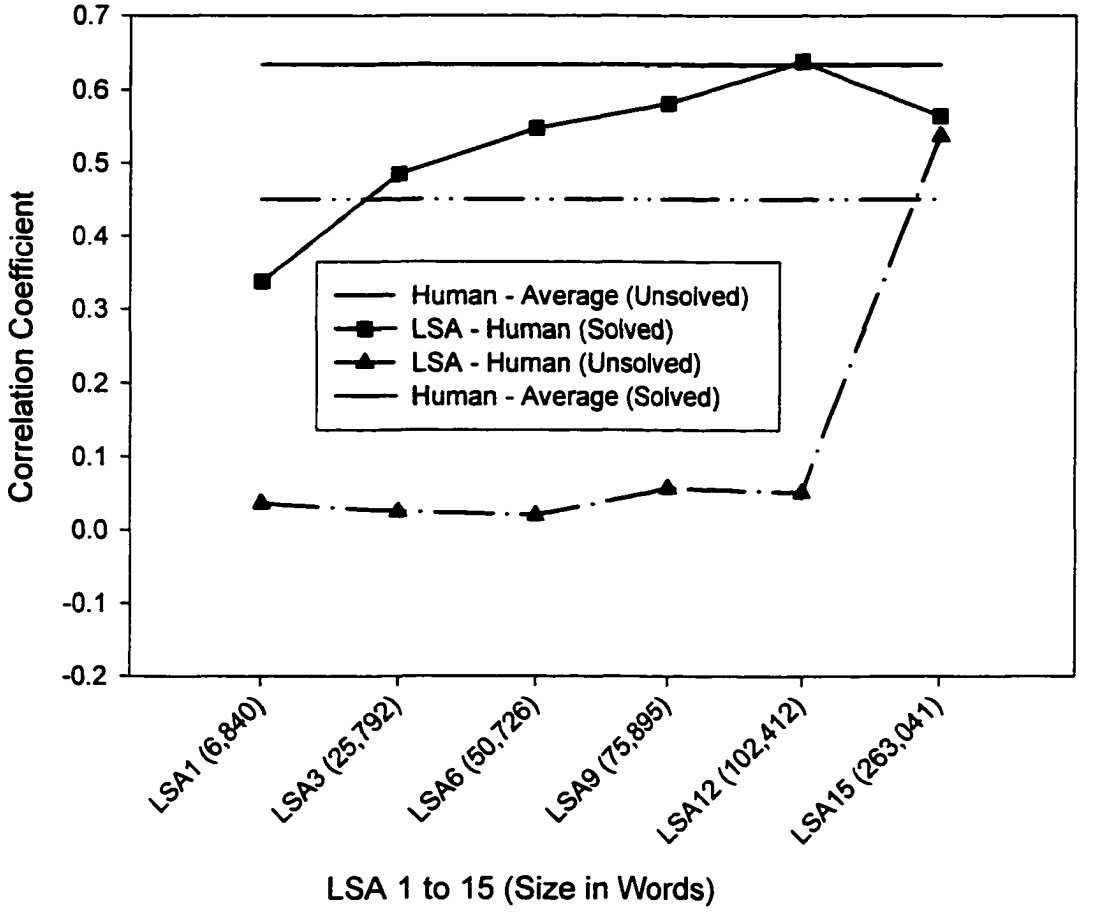

0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
Correlation Coefficient
Human - Average (Unsolved)
LSA - Human (Solved)
LSA - Human (Unsolved)
Human - Average (Solved)
LSA1 (6,840)
LSA3 (25,792)
LSA6 (50,726)
LSA9 (75,895)
LSA12 (102,412)
LSA15 (263,041)
LSA 1 to 15 (Size in Words)

Figure 5

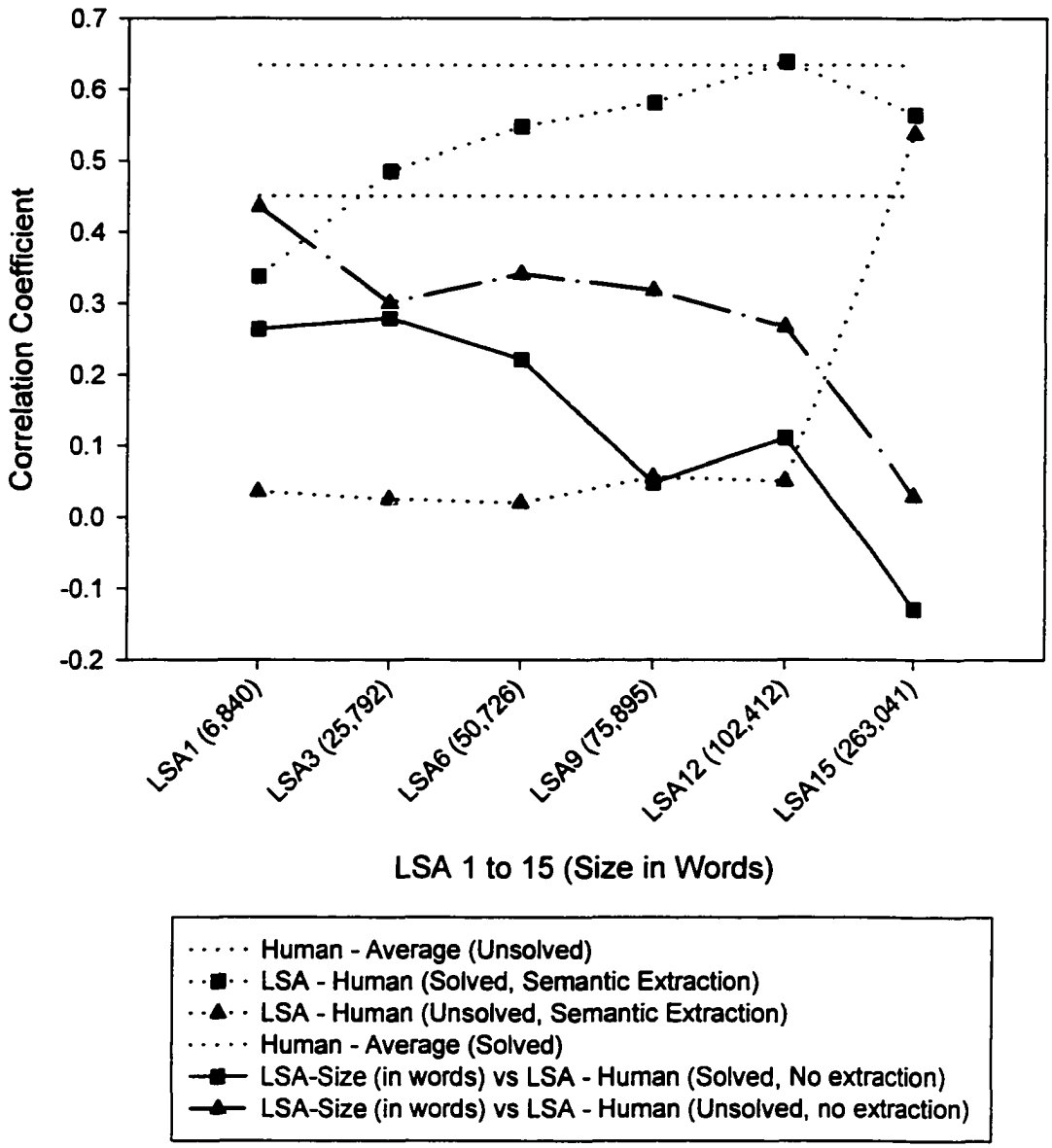

0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
Correlation Coefficient
LSA1 (6,840)
LSA3 (25,792)
LSA6 (50,726)
LSA9 (75,895)
LSA12 (102,412)
LSA15 (263,041)
LSA 1 to 15 (Size in Words)
Human - Average (Unsolved)
LSA - Human (Solved, Semantic Extraction)
LSA - Human (Unsolved, Semantic Extraction)
Human - Average (Solved)
LSA-Size (in words) vs LSA - Human (Solved, No extraction)
LSA-Size (in words) vs LSA - Human (Unsolved, no extraction)

Figure 6

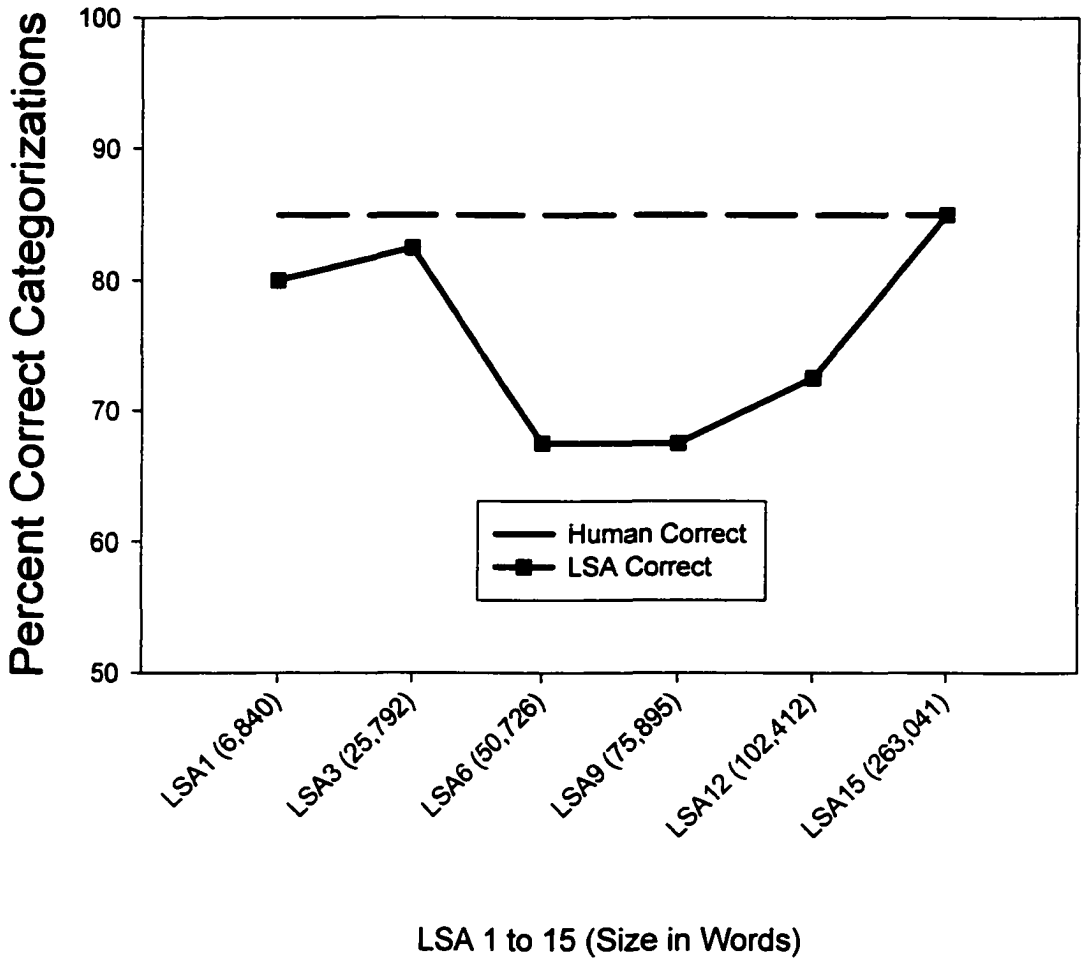
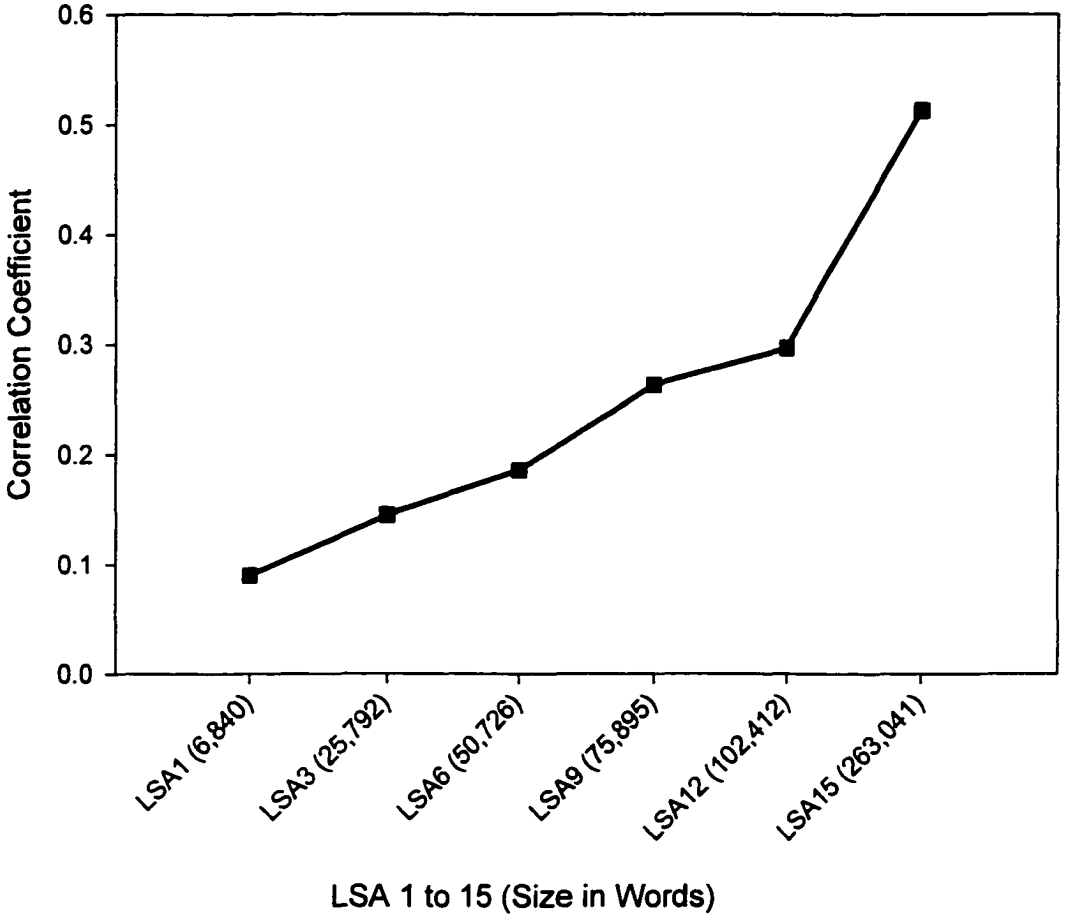
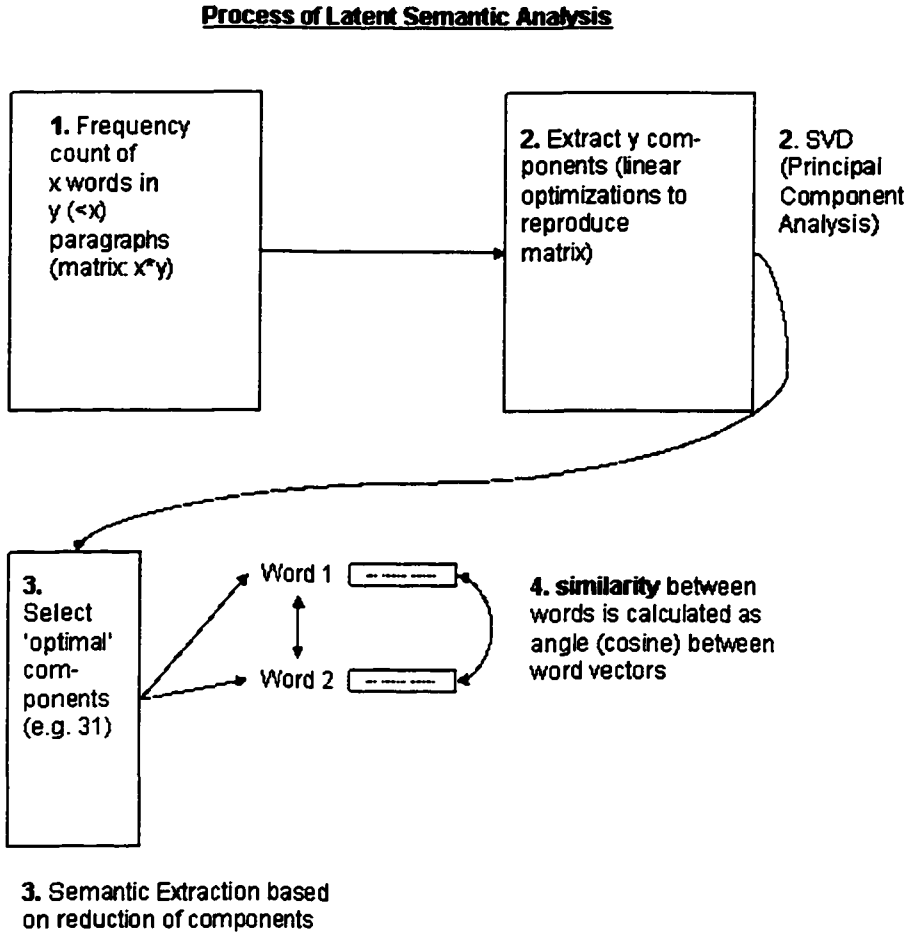

Percent Correct Categorizations
100
90
80
70
60
50
Human Correct
LSA Correct
LSA1 (6,840)
LSA3 (25,792)
LSA6 (50,726)
LSA9 (75,895)
LSA12 (102,412)
LSA15 (263,041)
LSA 1 to 15 (Size in Words)

Figure 7


0.6
0.5
0.4
0.3
0.2
0.1
0.0
Correlation Coefficient
LSA1 (6,840)
LSA3 (25,792)
LSA6 (50,726)
LSA9 (75,895)
LSA12 (102,412)
LSA15 (263,041)
LSA 1 to 15 (Size in Words)

Appendix A


Process of Latent Semantic Analysis
1. Frequency count of x words in y (<x) paragraphs (matrix x*y)
2. Extract y components (linear optimizations to reproduce matrix)
2. SVD (Principal Component Analysis)
3. Select 'optimal' components (e.g. 31)
Word 1
Word 2
4. similarity between words is calculated as angle (cosine) between word vectors
3. Semantic Extraction based on reduction of components

Appendix B

The Red-headed League by Sir Arthur Conan Doyle

I had called upon my friend, Mr. Sherlock Holmes, one day in the autumn of last year and found him in deep conversation with a very stout, florid-faced, elderly gentleman with fiery red hair. With an apology for my intrusion, I was about to withdraw when Holmes pulled me abruptly into the room and closed the door behind me.

"You could not possibly have come at a better time, my dear Watson," he said cordially.

"I was afraid that you were engaged."

"So I am. Very much so."

"Then I can wait in the next room."

"Not at all. This gentleman, Mr. Wilson, has been my partner and helper in many of my most successful cases, and I have no doubt that he will be of the utmost use to me in yours also."

The stout gentleman half rose from his chair and gave a bob of greeting, with a quick little questioning glance from his small fat-encircled eyes.

"Try the settee," said Holmes, relapsing into his armchair and putting his fingertips together, as was his custom when in judicial moods. "I know, my dear Watson, that you share my love of all that is bizarre and outside the conventions and humdrum routine of everyday life. You have shown your relish for it by the enthusiasm which has prompted you to chronicle, and, if you will excuse my saying so, somewhat to embellish so many of my own little adventures."

"Your cases have indeed been of the greatest interest to me," I observed.

"You will remember that I remarked the other day, just before we went into the very simple problem presented by Miss Mary Sutherland, that for strange effects and extraordinary combinations we must go to life itself, which is always far more daring than any effort of the imagination."

"A proposition which I took the liberty of doubting."

"You did, Doctor, but none the less you must come round to my view, for otherwise I shall keep on piling fact upon fact on you until your reason breaks down under them and acknowledges me to be right. Now, Mr. Jabez Wilson here has been good enough to call upon me this morning, and to begin a narrative which promises to be one of the most singular which I have listened to for some time. You have heard me remark that the strangest and most unique things are very often connected not with the larger but with the smaller crimes, and occasionally, indeed, where there is room for doubt whether any positive crime has been committed. As far as I have heard it is impossible for me to say whether the present case is an instance of crime or not, but the course of events is certainly among the most singular that I have ever listened to. Perhaps, Mr. Wilson, you would have the great kindness to recommence your narrative. I ask you

not merely because my friend Dr. Watson has not heard the opening part but also because the peculiar nature of the story makes me anxious to have every possible detail from your lips. As a rule, when I have heard some slight indication of the course of events, I am able to guide myself by the thousands of other similar cases which occur to my memory. In the present instance I am forced to admit that the facts are, to the best of my belief, unique."

The portly client puffed out his chest with an appearance of some little pride and pulled a dirty and wrinkled newspaper from the inside pocket of his greatcoat. As he glanced down the advertisement column, with his head thrust forward and the paper flattened out upon his knee, I took a good look at the man and endeavored, after the fashion of my companion, to read the indications which might be presented by his dress or appearance.

I did not gain very much, however, by my inspection. Our visitor bore every mark of being an average commonplace British tradesman, obese, pompous, and slow. He wore rather baggy gray shepherd's check trousers, a not over-clean black frock-coat, unbuttoned in the front, and a drab waistcoat with a heavy brassy Albert chain, and a square pierced bit of metal dangling down as an ornament. A frayed top-hat and a faded brown overcoat with a wrinkled velvet collar lay upon a chair beside him. Altogether, look as I would, there was nothing remarkable about the man save his blazing red head, and the expression of extreme chagrin and discontent upon his features.

Sherlock Holmes's quick eye took in my occupation, and he shook his head with a smile as he noticed my questioning glances. "Beyond the obvious facts that he has at some time done manual labour, that he takes snuff, that he is a Freemason, that he has been in China, and that he has done a considerable amount of writing lately, I can deduce nothing else."

Mr. Jabez Wilson started up in his chair, with his forefinger upon the paper, but his eyes upon my companion.

"How, in the name of good-fortune, did you know all that, Mr. Holmes?" he asked. "How did you know, for example, that I did manual labour. It's as true as gospel, for I began as a ship's carpenter."

"Your hands, my dear sir. Your right hand is quite a size larger than your left. You have worked with it, and the muscles are more developed."

"Well, the snuff, then, and the Freemasonry?"

"I won't insult your intelligence by telling you how I read that, especially as, rather against the strict rules of your order, you use an arc-and-compass breastpin."

"Ah, of course, I forgot that. But the writing?"

"What else can be indicated by that right cuff so very shiny for five inches, and the left one with the smooth patch near the elbow where you rest it upon the desk?"

"Well, but China?"

"The fish that you have tattooed immediately above your right wrist could only have been done in China. I have made a small study of tattoo marks and have even contributed to the literature of the subject. That trick of staining the fishes' scales of a delicate pink is quite peculiar to China. When, in addition, I see a Chinese coin hanging from your watch-chain, the matter becomes even more simple."

Mr. Jabez Wilson laughed heavily. "Well, I never!" said he. "I thought at first that you had done something clever, but I see that there was nothing in it, after all."

"I begin to think, Watson," said Holmes, "that I make a mistake in explaining. 'Omne ignotum pro magnifico,' you know, and my poor little reputation, such as it is, will suffer shipwreck if I am so candid. Can you not find the advertisement, Mr. Wilson?"

"Yes, I have got it now," he answered with his thick red finger planted halfway down the column. "Here it is. This is what began it all. You just read it for yourself, sir."

I took the paper from him and read as follows.

TO THE RED-HEADED LEAGUE: On account of the bequest of the late Ezekiah Hopkins, of Lebanon, Pennsylvania, U. S. A., there is now another vacancy open which entitles a member of the League to a salary of 4 pounds a week for purely nominal services. All red-headed men who are sound in body and mind and above the age of twenty-one years, are eligible. Appiy in person on Monday, at eleven o'clock, to Duncan Ross, at the offices of the League, 7 Pope's Court, Fleet Street.

"What on earth does this mean?" I ejaculated after I had twice read over the extraordinary announcement.

Holmes chuckled and wriggled in his chair, as was his habit when in high spirits. "It is a little off the beaten track, isn't it?" said he. "And now, Mr. Wilson, off you go at scratch and tell us all about yourself, your household, and the effect which this advertisement had upon your fortunes. You will first make a note, Doctor, of the paper and the date."

"It is The Morning Chronicle of April 27, 1890. Just two months ago."

"Very good. Now, Mr. Wilson?"

"Well, it is just as I have been telling you, Mr. Sherlock Holmes," said Jabez Wilson, mopping his forehead; "I have a small pawnbroker's business at Coburg Square, near the City. It's not a very large affair, and of late years it has not done more than just give me a living. I used to be able to keep two assistants, but now I only keep one; and I would have a job to pay him but that he is willing to come for half wages so as to learn the business."

"What is the name of this obliging youth?" asked Sherlock Holmes.

"His name is Vincent Spaulding, and he's not such a youth, either. It's hard to say his age. I should not wish a smarter assistant, Mr. Holmes; and I know very well that he could better himself and earn twice what I am able to give him. But, after all, if he is satisfied, why should I put ideas in his head?"

"Why, indeed? You seem most fortunate in having an employee who comes under the full market price. It is not a common experience among employers in this age. I don't know that your assistant is not as remarkable as your advertisement."

"Oh, he has his faults, too," said Mr. Wilson. "Never was such a fellow for photography. Snapping away with a camera when he ought to be improving his mind, and then diving down into the cellar like a rabbit into its hole to develop his pictures. That is his main fault, but on the whole he's a good worker. There's no vice in him."

"He is still with you, I presume?"

"Yes, sir. He and a girl of fourteen, who does a bit of simple cooking and keeps the place clean--that's all I have in the house, for I am a widower and never had any family. We live very quietly, sir, the three of us; and we keep a roof over our heads and pay our debts, if we do nothing more.

"The first thing that put us out was that advertisement. Spaulding, he came down into the office just this day eight weeks, with this very paper in his hand, and he says:

"'I wish to the Lord, Mr. Wilson, that I was a red-headed man.'

"'Why that?' I asks.

"'Why,' says he, 'here's another vacancy on the League of the Red-headed Men. It's worth quite a little fortune to any man who gets it, and I understand that there are more vacancies than there are men, so that the trustees are at their wits' end what to do with the money. If my hair would only change color, here's a nice little crib all ready for me to step into.'

"'Why, what is it, then?' I asked. You see. Mr. Holmes, I am a very stay-at-home man, and as my business came to me instead of my having to go to it, I was often weeks on end without putting my foot over the door-mat. In that way I didn't know much of what was going on outside, and I was always glad of a bit of news.

"'Have you never heard of the League of the Red-headed Men?' he asked with his eyes open.

"'Never.'

"'Why, I wonder at that, for you are eligibile yourself for one of the vacancies.'

"'And what are they worth?' I asked.

"'Oh, merely a couple of hundred a year, but the work is slight, and it need not interfere very much with one's other occupations.'

"Well, you can easily think that that made me prick up my ears, for the business has not been over-good for some years, and an extra couple of hundred would have been very handy.

"'Tell me all about it,' said I.

"'Well ' said he, showing me the advertisement, 'you can see for yourself that the League has a vacancy, and there is the address where you should apply for particulars.

As far as I can make out, the League was founded by an American millionaire, Ezekiah Hopkins, who was very peculiar in his ways. He was himself red-headed, and he had a great sympathy for all red-headed men; so when he died it was found that he had left his enormous fortune in the hands of trustees, with instructions to apply the interest to the providing of easy berths to men whose hair is of that color. From all I hear it is splendid pay and very little to do.'

"'But,' said I, 'there would be millions of red-headed men who would apply.'

"'Not so many as you might think,' he answered. 'You see it is really confined to Londoners, and to grown men. This American had started from London when he was young, and he wanted to do the old town a good turn. Then, again, I have heard it is no use your applying if your hair is light red, or dark red, or anything but real bright, blazing, fiery red. Now, if you cared to apply, Mr. Wilson, you would just walk in; but perhaps it would hardly be worth your while to put yourself out of the way for the sake of a few hundred pounds.'

"Now, it is a fact, gentlemen, as you may see for yourselves, that my hair is of a very full and rich tint, so that it seemed to me that if there was to be any competition in the matter I stood as good a chance as any man that I had ever met. Vincent Spaulding seemed to know so much about it that I thought he might prove useful, so I just ordered him to put up the shutters for the day and to come right away with me. He was very willing to have a holiday, so we shut the business up and started off for the address that was given us in the advertisement.

"I never hope to see such a sight as that again, Mr. Holmes. From north, south, east, and west every man who had a shade of red in his hair had tramped into the city to answer the advertisement. Fleet Street was choked with red-headed folk, and Pope's Court looked like a coster's orange barrow. I should not have thought there were so many in the whole country as were brought together by that single advertisement. Every shade of color they were--straw, lemon, orange, brick, Irish-setter, liver, clay; but, as Spaulding said, there were not many who had the real vivid flame-colored tint. When I saw how many were waiting, I would have given it up in despair; but Spaulding would not hear of it. How he did it I could not imagine, but he pushed and pulled and butted until he got me through the crowd, and right up to the steps which led to the office. There was a double stream upon the stair, some going up in hope, and some coming back dejected; but we wedged in as well as we could and soon found ourselves in the office."

"Your experience has been a most entertaining one," remarked Holmes as his client paused and refreshed his memory with a huge pinch of snuff. "Pray continue your very interesting statement."

"There was nothing in the office but a couple of wooden chairs and a deal table, behind which sat a small man with a head that was even redder than mine. He said a few words to each candidate as he came up, and then he always managed to find some fault in them which would disqualify them. Getting a vacancy did not seem to be such a very easy matter, after all. However, when our turn came the little man was much

more favorable to me than to any of the others, and he closed the door as we entered, so that he might have a private word with us.

"'This is Mr. Jabez Wilson,' said my assistant, 'and he is willing to fill a vacancy in the League.'

"'And he is admirably suited for it,' the other answered. 'He has every requirement. I cannot recall when I have seen anything so fine.' He took a step backward, cocked his head on one side, and gazed at my hair until I felt quite bashful. Then suddenly he plunged forward, wrung my hand, and congratulated me warmly on my success.

"'It would be injustice to hesitate,' said he. 'You will, however, I am sure, excuse me for taking an obvious precaution.' With that he seized my hair in both his hands, and tugged until I yelled with the pain. 'There is water in your eyes,' said he as he released me. 'I perceive that all is as it should be. But we have to be careful, for we have twice been deceived by wigs and once by paint. I could tell you tales of cobbler's wax which would disgust you with human nature.' He stepped over to the window and shouted through it at the top of his voice that the vacancy was filled. A groan of disappointment came up from below, and the folk all trooped away in different directions until there was not a red-head to be seen except my own and that of the manager.

"'My name,' said he, 'is Mr. Duncan Ross, and I am myself one of the pensioners upon the fund left by our noble benefactor. Are you a married man, Mr. Wilson? Have you a family?'

"I answered that I had not.

"His face fell immediately.

"'Dear me!' he said gravely, 'that is very serious indeed! I am sorry to hear you say that. The fund was, of course, for the propagation and spread of the red-heads as well as for their maintenance. It is exceedingly unfortunate that you should be a bachelor.'

"My face lengthened at this, Mr. Holmes, for I thought that I was not to have the vacancy after all; but after thinking it over for a few minutes he said that it would be all right.

"'In the case of another,' said he, 'the objection might be fatal, but we must stretch a point in favor of a man with such a head of hair as yours. When shall you be able to enter upon your new duties?'

"'Well, it is a little awkward, for I have a business already,' said I.

"'Oh, never mind about that, Mr. Wilson!' said Vincent Spaulding. 'I should be able to look after that for you.'

"'What would be the hours?' I asked.

"'Ten to two.'

"Now a pawnbroker's business is mostly done of an evening, Mr. Holmes, especially Thursday and Friday evening, which is just before pay-day; so it would suit me very

well to earn a little in the mornings. Besides, I knew that my assistant was a good man, and that he would see to anything that turned up.

"'That would suit me very well,' said I. 'And the pay?'

"'Is 4 pounds a week.'

"'And the work?'

"'Is purely nominal.'

"'What do you call purely nominal?'

"'Well, you have to be in the office, or at least in the building, the whole time. If you leave, you forfeit your whole position forever. The will is very clear upon that point. You don't comply with the conditions if you budge from the office during that time.'

"'It's only four hours a day, and I should not think of leaving,' said I.

"'No excuse will avail,' said Mr. Duncan Ross; 'neither sickness nor business nor anything else. There you must stay, or you lose your billet.'

"'And the work?'

"'Is to copy out the Encyclopaedia Britannica. There is the first volume of it in that press. You must find your own ink, pens, and blotting-paper, but we provide this table and chair. Will you be ready to-morrow?'

"'Certainly,' I answered.

"'Then, good-bye, Mr. Jabez Wilson, and let me congratulate you once more on the important position which you have been fortunate enough to gain.' He bowed me out of the room and I went home with my assistant, hardly knowing what to say or do, I was so pleased at my own good fortune.

"Well, I thought over the matter all day, and by evening I was in low spirits again; for I had quite persuaded myself that the whole affair must be some great hoax or fraud, though what its object might be I could not imagine. It seemed altogether past belief that anyone could make such a will, or that they would pay such a sum for doing anything so simple as copying out the Encyclopaedia Britannica. Vincent Spaulding did what he could to cheer me up, but by bedtime I had reasoned myself out of the whole thing. However, in the morning I determined to have a look at it anyhow, so I bought a penny bottle of ink, and with a quill-pen, and seven sheets of foolscap paper, I started off for Pope's Court.

"Well, to my surprise and delight, everything was as right as possible. The table was set out ready for me, and Mr. Duncan Ross was there to see that I got fairly to work. He started me off upon the letter A, and then he left me; but he would drop in from time to time to see that all was right with me. At two o'clock he bade me good-day, complimented me upon the amount that I had written, and locked the door of the office after me.

"This went on day after day, Mr. Holmes, and on Saturday the manager came in and planked down four golden sovereigns for my week's work. It was the same next week, and the same the week after. Every morning I was there at ten, and every afternoon I left at two. By degrees Mr. Duncan Ross took to coming in only once of a morning, and then, after a time, he did not come in at all. Still, of course, I never dared to leave the room for an instant, for I was not sure when he might come, and the billet was such a good one, and suited me so well, that I would not risk the loss of it.

"Eight weeks passed away like this, and I had written about Abbots and Archery and Armour and Architecture and Attica, and hoped with diligence that I might get on to the B's before very long. It cost me something in foolscap, and I had pretty nearly filled a shelf with my writings. And then suddenly the whole business came to an end."

"To an end?"

"Yes, sir. And no later than this morning. I went to my work as usual at ten o'clock, but the door was shut and locked, with a little square of card-board hammered on to the middle of the panel with a tack. Here it is, and you can read for yourself."

He held up a piece of white card-board about the size of a sheet of note-paper. It read in this fashion:

THE RED-HEADED LEAGUE

IS

DISSOLVED.

October 9, 1890.

Sherlock Holmes and I surveyed this curt announcement and the rueful face behind it, until the comical side of the affair so completely overtopped every other consideration that we both burst out into a roar of laughter.

"I cannot see that there is anything very funny," cried our client, flushing up to the roots of his flaming head. "If you can do nothing better than laugh at me, I can go elsewhere."

"No, no," cried Holmes, shoving him back into the chair from which he had half risen. "I really wouldn't miss your case for the world. It is most refreshingly unusual. But there is, if you will excuse my saying so, something just a little funny about it. Pray what steps did you take when you found the card upon the door?"

"I was staggered, sir. I did not know what to do. Then I called at the offices round, but none of them seemed to know anything about it. Finally, I went to the landlord, who is an accountant living on the ground-floor, and I asked him if he could tell me what had become of the Red-headed League. He said that he had never heard of any such body. Then I asked him who Mr. Duncan Ross was. He answered that the name was new to him.

"'Well,' said I, 'the gentleman at No. 4.'

"'What, the red-headed man?'

"'Yes.'

"'Oh,' said he, 'his name was William Morris. He was a solicitor and was using my room as a temporary convenience until his new premises were ready. He moved out yesterday.'

"'Where could I find him?'

"'Oh, at his new offices. He did tell me the address. Yes, 17 King Edward Street, near St. Paul's.'

"I started off, Mr. Holmes, but when I got to that address it was a manufactory of artificial knee-caps, and no one in it had ever heard of either Mr. William Morris or Mr. Duncan Ross."

"And what did you do then?" asked Holmes.

"I went home to Saxe-Coburg Square, and I took the advice of my assistant. But he could not help me in any way. He could only say that if I waited I should hear by post. But that was not quite good enough, Mr. Holmes. I did not wish to lose such a place without a struggle, so, as I had heard that you were good enough to give advice to poor folk who were in need of it, I came right away to you."

"And you did very wisely," said Holmes. "Your case is an exceedingly remarkable one, and I shall be happy to look into it. From what you have told me I think that it is possible that graver issues hang from it than might at first sight appear."

"Grave enough!" said Mr. Jabez Wilson. "Why, I have lost four pound a week."

"As far as you are personally concerned," remarked Holmes, "I do not see that you have any grievance against this extraordinary league. On the contrary, you are, as I understand, richer by some 30 pounds, to say nothing of the minute knowledge which you have gained on every subject which comes under the letter A. You have lost nothing by them."

"No, sir. But I want to find out about them, and who they are, and what their object was in playing this prank--if it was a prank--upon me. It was a pretty expensive joke for them, for it cost them two and thirty pounds."

"We shall endeavor to clear up these points for you. And, first, one or two questions, Mr. Wilson. This assistant of yours who first called your attention to the advertisement--how long had he been with you?"

"About a month then."

"How did he come?"

"In answer to an advertisement."

"Was he the only applicant?"

"No, I had a dozen."

"Why did you pick him?"

"Because he was handy and would come cheap."

"At half-wages, in fact."

"Yes."

"What is he like, this Vincent Spaulding?"

"Small, stout-built, very quick in his ways, no hair on his face, though he's not short of thirty. Has a white splash of acid upon his forehead."

Holmes sat up in his chair in considerable excitement. "I thought as much," said he. "Have you ever observed that his ears are pierced for earrings?"

"Yes, sir. He told me that a gypsy had done it for him when he was a lad."

"Hum!" said Holmes, sinking back in deep thought. "He is still with you?"

"Oh, yes, sir; I have only just left him."

"And has your business been attended to in your absence?"

"Nothing to complain of, sir. There's never very much to do of a morning."

"That will do, Mr. Wilson. I shall be happy to give you an opinion upon the subject in the course of a day or two. To-day is Saturday, and I hope that by Monday we may come to a conclusion."

"Well, Watson," said Holmes when our visitor had left us, "what do you make of it all?"

"I make nothing of it," I answered frankly. "It is a most mysterious business."

"As a rule," said Holmes, "the more bizarre a thing is the less mysterious it proves to be. It is your commonplace, featureless crimes which are really puzzling, just as a commonplace face is the most difficult to identify. But I must be prompt over this matter."

"What are you going to do, then?" I asked.

"To smoke," he answered. "It is quite a three pipe problem, and I beg that you won't speak to me for fifty minutes." He curled himself up in his chair, with his thin knees drawn up to his hawk-like nose, and there he sat with his eyes closed and his black clay pipe thrusting out like the bill of some strange bird. I had come to the conclusion that he had dropped asleep, and indeed was nodding myself, when he suddenly sprang out of his chair with the gesture of a man who has made up his mind and put his pipe down upon the mantelpiece.

"Sarasate plays at the St. James's Hall this afternoon," he remarked. "What do you think, Watson? Could your patients spare you for a few hours?"

"I have nothing to do to-day. My practice is never very absorbing."

"Then put on your hat and come. I am going through the City first, and we can have some lunch on the way. I observe that there is a good deal of German music on the programme, which is rather more to my taste than Italian or French. It is introspective, and I want to introspect. Come along!"

We travelled by the Underground as far as Aldersgate; and a short walk took us to Saxe-Coburg Square, the scene of the singular story which we had listened to in the morning. It was a poky, little, shabby-genteel place, where four lines of dingy two-storied brick houses looked out into a small railed-in enclosure, where a lawn of weedy grass and a few clumps of faded laurel-bushes made a hard fight against a smoke-laden and uncongenial atmosphere. Three gilt balls and a brown board with "JABEZ WILSON" in white letters, upon a corner house, announced the place where our red-headed client carried on his business. Sherlock Holmes stopped in front of it with his head on one side and looked it all over, with his eyes shining brightly between puckered lids. Then he walked slowly up the street, and then down again to the corner, still looking keenly at the houses. Finally he returned to the pawnbroker's, and, having thumped vigorously upon the pavement with his stick two or three times, he went up to the door and knocked. It was instantly opened by a bright-looking, clean-shaven young fellow, who asked him to step in.

"Thank you," said Holmes, "I only wished to ask you how you would go from here to the Strand."

"Third right, fourth left," answered the assistant promptly, closing the door.

"Smart fellow, that," observed Holmes as we walked away. "He is, in my judgment. the fourth smartest man in London, and for daring I am not sure that he has not a claim to be third. I have known something of him before."

"Evidently," said I, "Mr. Wilson's assistant counts for a good deal in this mystery of the Red-headed League. I am sure that you inquired your way merely in order that you might see him."

"Not him."

"What then?"

"The knees of his trousers."

"And what did you see?"

"What I expected to see."

"Why did you beat the pavement?"

"My dear doctor, this is a time for observation, not for talk. We are spies in an enemy's country. We know something of Saxe-Coburg Square. Let us now explore the parts which lie behind it."

The road in which we found ourselves as we turned round the corner from the retired Saxe-Coburg Square presented as great a contrast to it as the front of a picture does to the back. It was one of the main arteries which conveyed the traffic of the City to the

north and west. The roadway was blocked with the immense stream of commerce flowing in a double tide inward and outward, while the footpaths were black with the hurrying swarm of pedestrians. It was difficult to realize as we looked at the line of fine shops and stately business premises that they really abutted on the other side upon the faded and stagnant square which we had just quitted.

"Let me see," said Holmes, standing at the corner and glancing along the line, "I should like just to remember the order of the houses here. It is a hobby of mine to have an exact knowledge of London. There is Mortimer's, the tobacconist, the little newspaper shop, the Coburg branch of the City and Suburban Bank, the Vegetarian Restaurant, and McFarlane's carriage-building depot. That carries us right on to the other block. And now, Doctor, we've done our work, so it's time we had some play. A sandwich and a cup of coffee, and then off to violin-land, where all is sweetness and delicacy and harmony, and there are no red-headed clients to vex us with their conundrums."

My friend was an enthusiastic musician, being himself not only a very capable perfomer but a composer of no ordinary merit. All the afternoon he sat in the stalls wrapped in the most perfect happiness, gently waving his long, thin fingers in time to the music, while his gently smiling face and his languid, dreamy eyes were as unlike those of Holmes, the sleuth-hound, Holmes the relentless, keen-witted, ready-handed criminal agent, as it was possible to conceive. In his singular character the dual nature alternately asserted itself, and his extreme exactness and astuteness represented, as I have often thought, the reaction against the poetic and contemplative mood which occasionally predominated in him. The swing of his nature took him from extreme languor to devouring energy; and, as I knew well, he was never so truly formidable as when, for days on end, he had been lounging in his armchair amid his improvisations and his black-letter editions. Then it was that the lust of the chase would suddenly come upon him, and that his brilliant reasoning power would rise to the level of intuition, until those who were unacquainted with his methods would look askance at him as on a man whose knowledge was not that of other mortals. When I saw him that afternoon so enwrapped in the music at St. James's Hall I felt that an evil time might be coming upon those whom he had set himself to hunt down.

"You want to go home, no doubt, Doctor," he remarked as we emerged.

"Yes, it would be as well."

"And I have some business to do which will take some hours. This business at Coburg Square is serious."

"Why serious?"

"A considerable crime is in contemplation. I have every reason to believe that we shall be in time to stop it. But to-day being Saturday rather complicates matters. I shall want your help to-night."

"At what time?"

"Ten will be early enough."

"I shall be at Baker Street at ten."

"Very well. And, I say, Doctor, there may be some little danger, so kindly put your army revolver in your pocket." He waved his hand, turned on his heel, and disappeared in an instant among the crowd.

I trust that I am not more dense than my neighbors, but I was always oppressed with a sense of my own stupidity in my dealings with Sherlock Holmes. Here I had heard what he had heard, I had seen what he had seen, and yet from his words it was evident that he saw clearly not only what had happened but what was about to happen, while to me the whole business was still confused and grotesque. As I drove home to my house in Kensington I thought over it all, from the extraordinary story of the red-headed copier of the Encyclopaedia down to the visit to Saxe-Coburg Square, and the ominous words with which he had parted from me. What was this nocturnal expedition, and why should I go armed? Where were we going, and what were we to do? I had the hint from Holmes that this smooth-faced pawnbroker's assistant was a formidable man--a man who might play a deep game. I tried to puzzle it out, but gave it up in despair and set the matter aside until night should bring an explanation.

It was a quarter-past nine when I started from home and made my way across the Park, and so through Oxford Street to Baker Street. Two hansoms were standing at the door, and as I entered the passage I heard the sound of voices from above. On entering his room I found Holmes in animated conversation with two men, one of whom I recognized as Peter Jones, the official police agent, while the other was a long, thin, sad-faced man, with a very shiny hat and oppressively respectable frock-coat.

"Ha! Our party is complete," said Holmes, buttoning up his peajacket and taking his heavy hunting crop from the rack. "Watson, I think you know Mr. Jones, of Scotland Yard? Let me introduce you to Mr. Merryweather, who is to be our companion in to-night's adventure."

"We're hunting in couples again, Doctor, you see," said Jones in his consequential way. "Our friend here is a wonderful man for starting a chase. All he wants is an old dog to help him to do the running down."

"I hope a wild goose may not prove to be the end of our chase," observed Mr. Merryweather gloomily.

"You may place considerable confidence in Mr. Holmes, sir," said the police agent loftily. "He has his own little methods, which are, if he won't mind my saying so, just a little too theoretical and fantastic, but he has the makings of a detective in him. It is not too much to say that once or twice, as in that business of the Sholto murder and the Agra treasure, he has been more nearly correct than the official force."

"Oh, if you say so, Mr. Jones, it is all right," said the stranger with deference. "Still, I confess that I miss my rubber. It is the first Saturday night for seven-and-twenty years that I have not had my rubber."

"I think you will find," said Sherlock Holmes, "that you will play for a higher stake to-night than you have ever done yet, and that the play will be more exciting. For you,

Mr. Merryweather, the stake will be some 30,000 pounds; and for you, Jones, it will be the man upon whom you wish to lay your hands."

"John Clay, the murderer, thief, smasher, and forger. He's a young man, Mr. Merryweather, but he is at the head of his profession, and I would rather have my bracelets on him than on any criminal in London. He's a remarkable man, is young John Clay. His grandfather was a royal duke, and he himself has been to Eton and Oxford. His brain is as cunning.as his fingers, and though we meet signs of him at every turn, we never know where to find the man himself. He'll crack a crib in Scotland one week, and be raising money to build an orphanage in Cornwall the next. I've been on his track for years and have never set eyes on him yet."

"I hope that I may have the pleasure of introducing you to-night. I've had one or two little turns also with Mr. John Clay, and I agree with you that he is at the head of his profession. It is past ten, however, and quite time that we started. If you two will take the first hansom, Watson and I will follow in the second."

Sherlock Holmes was not very communicative during the long drive and lay back in the cab humming the tunes which he had heard in the afternoon. We rattled through an endless labyrinth of gas-lit streets until we emerged into Farrington Street.

"We are close there now," my friend remarked. "This fellow Merryweather is a bank director, and personally interested in the matter. I thought it as well to have Jones with us also. He is not a bad fellow, though an absolute imbecile in his profession. He has one positive virtue. He is as brave as a bulldog and as tenacious as a lobster if he gets his claws upon anyone. Here we are, and they are waiting for us."

We had reached the same crowded thoroughfare in which we had found ourselves in the morning. Our cabs were dismissed, and, following the guidance of Mr. Merryweather, we passed down a narrow passage and through a side door, which he opened for us. Within there was a small corridor, which ended in a very massive iron gate. This also was opened, and led down a flight of winding stone steps, which terminated at another formidable gate. Mr. Merryweather stopped to light a lantern, and then conducted us down a dark, earth-smelling passage, and so, after opening a third door, into a huge vault or cellar, which was piled all round with crates and massive boxes.

"You are not very vulnerable from above," Holmes remarked as he held up the lantern and gazed about him.

"Nor from below," said Mr. Merryweather, striking his stick upon the flags which lined the floor. "Why, dear me, it sounds quite hollow!" he remarked, looking up in surprise.

"I must really ask you to be a little more quiet!" said Holmes severely. "You have already imperilled the whole success of our expedition. Might I beg that you would have the goodness to sit down upon one of those boxes, and not to interfere?"

The solemn Mr. Merryweather perched himself upon a crate, with a very injured expression upon his face, while Holmes fell upon his knees upon the floor and, with

the lantern and a magnifying lens, began to examine minutely the cracks between the stones. A few seconds sufficed to satisfy him, for he sprang to his feet again and put his glass in his pocket.

"We have at least an hour before us," he remarked, "for they can hardly take any steps until the good pawnbroker is safely in bed. Then they will not lose a minute, for the sooner they do their work the longer time they will have for their escape. We are at present, Doctor--as no doubt you have divined--in the cellar of the City branch of one of the principal London banks. Mr. Merryweather is the chairman of directors, and he will explain to you that there are reasons why the more daring criminals of London should take a considerable interest in this cellar at present."

"It is our French gold," whispered the director. "We have had several warnings that an attempt might be made upon it."

"Your French gold?"

"Yes. We had occasion some months ago to strengthen our resources and borrowed for that purpose 30,000 napoleons from the Bank of France. It has become known that we have never had occasion to unpack the money, and that it is still lying in our cellar. The crate upon which I sit contains 2,000 napoleons packed between layers of lead foil. Our reserve of bullion is much larger at present than is usually kept in a single branch office, and the directors have had misgivings upon the subject."

"Which were very well justified," observed Holmes. "And now it is time that we arranged our little plans. I expect that within an hour matters will come to a head. In the meantime Mr. Merryweather, we must put the screen over that dark lantern."

"And sit in the dark?"

"I am afraid so. I had brought a pack of cards in my pocket, and I thought that, as we were a partie carree, you might have your rubber after all. But I see that the enemy's preparations have gone so far that we cannot risk the presence of a light. And, first of all, we must choose our positions. These are daring men, and though we shall take them at a disadvantage, they may do us some harm unless we are careful. I shall stand behind this crate, and do you conceal yourselves behind those. Then, when I flash a light upon them, close in swiftly. If they fire, Watson, have no compunction about shooting them down."

I placed my revolver, cocked, upon the top of the wooden case behind which I crouched. Holmes shot the slide across the front of his lantern and left us in pitch darkness--such an absolute darkness as I have never before experienced. The smell of hot metal remained to assure us that the light was still there, ready to flash out at a moment's notice. To me, with my nerves worked up to a pitch of expectancy, there was something depressing and subduing in the sudden gloom, and in the cold dank air of the vault.

"They have but one retreat," whispered Holmes. "That is back through the house into Saxe-Coburg Square. I hope that you have done what I asked you, Jones?"

"I have an inspector and two officers waiting at the front door."

"Then we have stopped all the holes. And now we must be silent and wait."

What a time it seemed! From comparing notes afterwards it was but an hour and a quarter, yet it appeared to me that the night must have almost gone. and the dawn be breaking above us. My limbs were weary and stiff, for I feared to change my position; yet my nerves were worked up to the highest pitch of tension, and my hearing was so acute that I could not only hear the gentle breathing of my companions, but I could distinguish the deeper, heavier in-breath of the bulky Jones from the thin, sighing note of the bank director. From my position I could look over the case in the direction of the floor. Suddenly my eyes caught the glint of a light.

At first it was but a lurid spark upon the stone pavement. Then it lengthened out until it became a yellow line, and then, without any warning or sound, a gash seemed to open and a hand appeared; a white, almost womanly hand, which felt about in the centre of the little area of light. For a minute or more the hand, with its writhing fingers, protruded out of the floor. Then it was withdrawn as suddenly as it appeared, and all was dark again save the single lurid spark which marked a chink between the stones.

Its disappearance, however, was but momentary. With a rending, tearing sound, one of the broad, white stones turned over upon its side and left a square, gaping hole, through which streamed the light of a lantern. Over the edge there peeped a clean-cut, boyish face, which looked keenly about it, and then, with a hand on either side of the aperture, drew itself shoulder-high and waist-high, until one knee rested upon the edge. In another instant he stood at the side of the hole and was hauling after him a companion, lithe and small like himself, with a pale face and a shock of very red hair.

"It's all clear," he whispered. "Have you the chisel and the bags? Great Scott! Jump, Archie, jump, and I'll swing for it!"

Sherlock Holmes had sprung out and seized the intruder by the collar. The other dived down the hole, and I heard the sound of rending cloth as Jones clutched at his skirts. The light flashed upon the barrel of a revolver, but Holmes's hunting crop came down on the man's wrist, and the pistol clinked upon the stone floor.

"It's no use, John Clay," said Holmes blandly. "You have no chance at all."

"So I see," the other answered with the utmost coolness. "I fancy that my pal is all right, though I see you have got his coat-tails."

"There are three men waiting for him at the door," said Holmes.

"Oh, indeed! You seem to have done the thing very completely. I must compliment you."

"And I you," Holmes answered. "Your red-headed idea was very new and effective."

"You'll see your pal again presently," said Jones. "He's quicker at climbing down holes than I am. Just hold out while I fix the derbies."

"I beg that you will not touch me with your filthy hands," remarked our prisoner as the handcuffs clattered upon his wrists. "You may not be aware that I have royal blood in

my veins. Have the goodness, also, when you address me always to say 'sir' and 'please.'"

"All right," said Jones with a stare and a snigger. "Well, would you please, sir, march upstairs, where we can get a cab to carry your Highness to the police-station?"

"That is better," said John Clay serenely. He made a sweeping bow to the three of us and walked quietly off in the custody of the detective.

"Really, Mr. Holmes," said Mr. Merryweather as we followed them from the cellar, "I do not know how the bank can thank you or repay you. There is no doubt that you have detected and defeated in the most complete manner one of the most determined attempts at bank robbery that have ever come within my experience."

"I have had one or two little scores of my own to settle with Mr. John Clay," said Holmes. "I have been at some small expense over this matter, which I shall expect the bank to refund, but beyond that I am amply repaid by having had an experience which is in many ways unique, and by hearing the very remarkable narrative of the Red-headed League."

"You see, Watson," he explained in the early hours of the morning as we sat over a glass of whisky and soda in Baker Street, "it was perfectly obvious from the first that the only possible object of this rather fantastic business of the advertisement of the League, and the copying of the Encyclopaedia, must be to get this not over-bright pawnbroker out of the way for a number of hours every day. It was a curious way of managing it, but, really, it would be difficult to suggest a better. The method was no doubt suggested to Clay's ingenious mind by the color of his accomplice's hair. The 4 pounds a week was a lure which must draw him, and what was it to them, who were playing for thousands? They put in the advertisement, one rogue has the temporary office, the other rogue incites the man to apply for it. and together they manage to secure his absence every morning in the week. From the time that I heard of the assistant having come for half wages, it was obvious to me that he had some strong motive for securing the situation."

"But how could you guess what the motive was?"

"Had there been women in the house, I should have suspected a mere vulgar intrigue. That, however, was out of the question. The man's business was a small one, and there was nothing in his house which could account for such elaborate preparations, and such an expenditure as they were at. It must, then, be something out of the house. What could it be? I thought of the assistant's fondness for photography, and his trick of vanishing into the cellar. The cellar! There was the end of this tangled clew. Then I made inquiries as to this mysterious assistant and found that I had to deal with one of the coolest and most daring criminals in London. He was doing something in the cellar--something which took many hours a day for months on end. What could it be, once more? I could think of nothing save that he was running a tunnel to some other building.

"So far I had got when we went to visit the scene of action. I surprised you by beating upon the pavement with my stick. I was ascertaining whether the cellar stretched out in front or behind. It was not in front. Then I rang the bell, and, as I hoped, the assistant answered it. We have had some skirmishes, but we had never set eyes upon each other before. I hardly looked at his face. His knees were what I wished to see. You must yourself have remarked how worn, wrinkled, and stained they were. They spoke of those hours of burrowing. The only remaining point was what they were burrowing for. I walked round the corner, saw the City and Suburban Bank abutted on our friend's premises, and felt that I had solved my problem. When you drove home after the concert I called upon Scotland Yard and upon the chairman of the bank directors, with the result that you have seen."

"And how could you tell that they would make their attempt to-night?" I asked.

"Well, when they closed their League offices that was a sign that they cared no longer about Mr. Jabez Wilson's presence--in other words, that they had completed their tunnel. But it was essential that they should use it soon, as it might be discovered, or the bullion might be removed. Saturday would suit them better than any other day, as it would give them two days for their escape. For all these reasons I expected them to come to-night."

"You reasoned it out beautifully," I exclaimed in unfeigned admiration "It is so long a chain, and yet every link rings true."

"It saved me from ennui," he answered, yawning. "Alas! I already feel it closing in upon me. My life is spent in one long effort to escape from the commonplaces of existence. These little problems help me to do so."

"And you are a benefactor of the race," said I.

He shrugged his shoulders. "Well, perhaps, after all, it is of some little use," he remarked. " 'L'homme c'est rien--l'oeuvre c'est tout,' as Gustave Flaubert wrote to George Sand."

Appendix C

Following questionnaire assesses your understanding of the story that you have read. Each question has four response alternatives with one correct answer (A – D).

Circle the letter of the response that you think is correct. Make sure to answer every question. If you are not sure about the correct response, try guessing: A few questions might ask about something that you have not previously read in the story.

Please never go back to change an answer that you have previously given.

1. What was Jabez Wilson's first occupation as an apprentice?
A. Tunnel worker
B. Secretary
C. Fisherman
D. Pawnbroker

2. What is the real name of Vincent Spaulding?
A. Duncan Ross
B. John Clay
C. Vincent Spaulding is his real name
D. John Watson

3. Why did Sherlock Holmes beat the pavement?
A. to verify the existence of a tunnel
B. to see how the assistant would react
C. as a way to knock on the door
D. to let the nearby waiting Scotland Yard agent know that he thought this man should be arrested

4. Who was the actual founder of the red-headed league?
A. Jabez Wilson
B. Vincent Spaulding
C. Ezekiah Hopkins
D. Mr. Merryweather

5. Compared to his previous assistants: for how much did Jabez Wilson hire his newest assistant?
A. The regular, minimum salary
B. The double salary
C. Third of the salary
D. Half the salary

6. Why did Sherlock Holmes look at the assistant's knees?
A. to see if the assistant was digging a tunnel
B. to find out if the assistant was very poor

C. to find out if the assistant was actually working as a pawnbroker
D. to see if his trousers were dirty, inferring that he was not home since yesterday.

7. What is the Red-headed league?
A. A communistic party
B. A criminal organization
C. A fictional organization that does not exist
D. An organization that supports red-haired people

8. Who is John Clay?
A. A criminal
B. The director of a bank
C. The name was never mentioned in the story
D. An agent from Scotland Yard

9. What attribute did Sherlock Holmes assign to Vincent Spaulding?
A. smart
B. stout
C. happy
D. funny

10. To which country had Jabez Wilson traveled?
A. America
B. India
C. China
D. never mentioned

11. Which business neighbored to Jabez Wilson's business?
A. A bank
B. The red-headed league
C. St. James Music Hall
D. A knee cap manufacturer

12. What is the current occupation of Jabez Wilson?
A. Tunnel worker
B. Secretary
C. Porter
D. Pawnbroker

13. What is the presumed hair color of Ezekiah Hopkins?
A. cannot be inferred from the story
B. black
C. brown
D. red

14. What is the occupation of Jabez Wilson at the red headed league?
A. building a tunnel
B. copy the Encyclopedia Britannica
C. write the biography for the founder of the red headed league
D. nothing because Jabez Wilson actually invented the whole story

15. What is the most likely reason for the dissolution of the red-headed league?
A. The league went bankrupt.
B. The league was in fact a criminal organization that tried to avoid Sherlock Holmes finding out about them.
C. The league never really existed but was the trick of a criminal and was not needed anymore.
D. One of the trustees took all the money and disappeared in London.

16. What was the presumed position of Duncan Ross in this story?
A. an agent from Scotland Yard
B. Director of a bank
C. Pensioner of the read headed league
D. a pawnbroker

17. What was the most likely intention of Vincent Spaulding?
A. To slowly take over the pawnbroker business and kill its owner
B. To secretely build a criminal organization while the owner of the business was out of the house
C. To rob a bank
D. To escape Scottland Yard while they were in search for him

18. What is the presumed name of Jabez Wilson's secretary?
A. Vincent Spaulding
B. He did not have a secretary
C. John Watson
D. Ezekiah Hopkins

Final question:

19. Please indicate your familiarity with stories from Sir Arthur Conan Doyle (more than one response alternative is possible):
A. I have never heard of Sherlock Holmes before.
B. I have never read a Sherlock Holmes mystery or seen a Sherlock Holmes movie before but I have heard the name.

C. I have seen at least one Sherlock Holmes movie before but have not read any.
D. I have read at least one Sherlock Holmes mystery (regardless of having seen movies).

If you indicate yes on D., please indicate also:

1. Estimate how many Sherlock Holmes mysteries you have read:

_____

2. You have read "The red-headed league" before: yes no

Thank you for your participation !!