Predicting Multi-document Comprehension: Cohesion Network Analysis

Bogdan Nicula, Cecile A. Perret, Mihai Dascalu and Danielle S. McNamara

Published June 21st 2019

Abstract. Theories of discourse argue that comprehension depends on the coherence of the learner's mental representation. Our aim is to create a reliable automated representation to estimate readers' level of comprehension based on different productions, namely self-explanations and answers to open-ended questions. Previous work relied on Cohesion Network Analysis to model a cohesion graph composed of semantic links between multiple reference texts and student productions. From this graph, a set of features was derived and used to build machine learning models to predict student comprehension scores. In this paper, we build on top of the previous study by: a) extending the CNA graph by adding new semantic links targeting specific sentences that should have been captured within the learner's productions, and b) cleaning the self-explanations by eliminating frozen expression, as well as entries which seemed nearly identical to the source text. The results are in line with the conclusions of the previous study regarding the importance of both self-explanations and question answers in predicting the students' reading comprehension level. They also outline the limitations of our feature generation approach, in which no substantial improvements were detected, despite adding more fine-grained features.

Published in S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, & R. Luckin (Eds.), Proceedings of the 20th International Conference of Artificial Intelligence in Education (AIED) in Chicago, IL (pp. 358-369). Cham, Switzerland: Springer.

Predicting Multi-document Comprehension: Cohesion Network Analysis

Bogdan Nicula[1], Cecile A. Perret[3], Mihai Dascalu[1,2](✉), and Danielle S. McNamara[3]

[1] University Politehnica of Bucharest, 313 Splaiul Independentei, 060042 Bucharest, Romania
bogdan.nicula@cti.pub.ro, mihai.dascalu@cs.pub.ro

[2] Cognos Business Consulting S.R.L., 32 Bd. Regina Maria, Bucharest, Romania

[3] Department of Psychology, Arizona State University, PO Box 871104, Tempe, AZ 85287, USA
{cperret,dsmcnama}@asu.edu

Abstract. Theories of discourse comprehension assume that understanding is a process of making connections between new information (e.g., in a text) and prior knowledge, and that the quality of comprehension is a function of the coherence of the mental representation. When readers are exposed to multiple sources of information, they must make connections both within and between the texts. One challenge is how to represent this coherence and in turn how to predict readers' levels of comprehension. In this study, we represent coherence using Cohesion Network Analysis (CNA) in which we model a global cohesion graph that semantically links reference texts to different student verbal productions. Our aim is to create an automated model of comprehension prediction based on features extracted from the CNA graph. We examine the cohesion links between the four texts read by 146 students and their (a) self-explanations generated on target sentences and (b) responses to open-ended questions. We analyze the degree to which features derived from the cohesive links from the extended CNA graph are predictive of students' comprehension scores (on a [0 to 12] scale) using either (a) students' self-explanations, (b) responses to comprehension questions, or (c) both. We compared the use of Linear Regression, Extra Trees Regressor, Support Vector Regression, and Multi-Layer Perceptron. Our best model used Linear Regression, obtaining a 1.29 mean absolute error when predicting comprehension scores using both sources of verbal responses (i.e., self-explanations and question answers).

Keywords: Multi-document comprehension and integration · Natural Language Processing · Cohesion Network Analysis · Comprehension modeling · Machine learning

S. Isotani et al. (Eds.): AIED 2019, LNAI 11625, pp. 358–369, 2019.
https://doi.org/10.1007/978-3-030-23204-7_30

1 Introduction

Comprehension is challenging. The process involves understanding the words and sentences within the text (or discourse), connecting the ideas within the text, and linking the ideas to prior knowledge, in order to generate a coherent mental representation of the content. Comprehension processes are further challenged when faced with multiple sources of information. Multiple document comprehension adds on the need to make connections both within and between texts to generate a coherent mental representation of the disparate sources of information. We are faced with these challenges on a regular basis, when reading separate documents, papers, news, blogs, emails, and so on.

One question is how to simulate the coherence of a reader's mental representation and in turn, the extent to which that coherence predicts comprehension. In this study, we examine that extent to which the semantic connections (i.e., coherence) between a text and a reader's constructed responses while reading and after reading multiple documents predict comprehension. Similar modeling and linguistic techniques have been applied in the context of single text comprehension [1, 2]. Techniques evaluating reading comprehension for multiple document scenarios were previously researched by Hastings, Hughes, Magliano, Goldman and Lawless [3]; however, there is a dearth of research attempting to model how individuals integrate information across texts to form a coherent representation of information from separate sources. Cohesion Network Analysis (CNA) [2] is a technique that combines Social Network Analysis (SNA) [4] and Natural Language Processing (NLP) [5] techniques to identify semantic similarities between various sources of discourse and the levels of semantic cohesion within and between networks. This paper applies CNA to multiple document discourse to predict comprehension as well as to better understand the underlying cognitive processes of integrating information from multiple texts. Students' self-explanations and their responses to open-ended questions after reading multiple documents are analyzed in order to evaluate semantic connections between the documents and the students' productions.

1.1 Comprehension of Multiple Documents

Reading comprehension is a difficult and complex task that requires connecting ideas in a text in order to produce a coherent mental representation of the information [6]. Such a task not only requires understanding the semantic relations between words and sentences, but also necessitates connecting ideas from various sentences throughout a text in order to produce a coherent understanding [7]. Thus, successful comprehension of single texts requires an ability to comprehend textbase content (explicit information derived from a single sentence) as well as develop intra-textual inferences that connect adjacent or distal textbase content from that same text.

This is a dynamic process between the reader and the text requiring the integration of information from the immediate sentence with previous sections of the text as well as the reader's own prior knowledge [6]. This continuous construction of a mental representation of textual materials can be enhanced by a reader's ability to integrate information across texts, thus developing a coherent knowledge base about a specific topic [8]. This can in turn aid in developing mental representations of future texts on related topics. However, comprehension becomes increasingly challenging when readers are expected to combine information from disparate sources.

Each text generally adheres to a consistent style, however, texts from different sources are highly variable in these characteristics and are not typically presented as a set [9]. These features can vary across genres and individual texts potentially creating an additional obstacle for integration. Individual texts contain discourse markers of cohesion that signal relations between ideas, whereas these features are not available between texts thus complicating the integration task for readers [10, 11]. Without these connectors to help guide inferencing, the integration of concepts relies on the reader's prior knowledge. This diversity and lack of clear connections may impose additional challenges for comprehension and integration of multiple texts.

1.2 Assessing and Evaluating Comprehension

Writing tasks during online and offline comprehension have been employed as a means of aiding students in making textual inferences. Both online and offline tasks enhance a reader's ability to process information and potentially integrate ideas across texts.

Offline comprehension tasks, such as essays, recall tasks, and comprehension questions, are often used to assess comprehension. However, they can also be used to support comprehension through the reactivation of relevant concepts. In particular, the recall-cues present in the questions combined with generating responses to convey understanding prompts readers to reactivate concepts, in turn aiding comprehension [12].

Online tasks, such as self-explanations and think-alouds, prompt readers to actively process text information. Self-explanation, the process of explaining information to oneself while employing reading comprehension strategies, is a valuable reading strategy that encourages deeper comprehension throughout the reading process, thus facilitating the construction of a more coherent mental model [13, 14].

Self-explanations also provide insights into a reader's cognitive processing of the text. When students generate responses to sequential text sections as they do in self-explanation tasks, their aggregated responses reveal semantic overlap across sections as well as connectives and other signaling devices that indicate specific connections of causal events. The cohesive devices expressed within reader's self-explanations provide insight into their coherence building processes because they can inform on the reader's depth of comprehension. For example, surface level processing is associated with the overlap of specific words across sentences or the amount of semantic information that can be traced back to previous portions of the text. Deeper comprehension processes also contain semantic overlap, but also have greater lexical diversity of the content relating to the text, suggesting the use of external information such as prior knowledge [1].

This study includes both students' self-explanations during reading and their responses to open-ended questions after reading multiple documents. Our objective here is to examine the semantic connections between the documents and (a) students' self-explanations, and (b) students' responses to questions. These semantic connections are assumed to represent the coherence of students' mental representations of the content. Students' constructed responses provide a glimpse into their processing of text, and thus a potential means of predicting students' comprehension. Here, we represent comprehension via students' score on the comprehension questions (i.e., expert ratings), and we assess the coherence of students' comprehension via semantic links between the documents and students' responses during and after reading. We do so by combining computational linguistics and SNA using CNA.

1.3 Cohesion Network Analysis

Cohesion Network Analysis (CNA) [2] was first introduced to assess participation in Computer Supported Collaborative Learning, but its underlying representation is suitable for any type of discourse. CNA relies on cohesion that is estimated using multiple semantic similarity metrics [15], combines advanced NLP techniques, and integrates SNA measurements applied on the resulting cohesion graph [16, 17]. The cohesion graph can be perceived as a proxy for the underlying semantic content of discourse within a document. It is represented as a multi-layered graph that considers both macro-level and micro-level constituents present at different levels (i.e., sentences, paragraphs, or the entire text). A document is decomposed into its paragraphs and, subsequently, into the underlying sentences and words. Cohesive links are defined between different layers of the hierarchy in order to measure the strength of the inclusion, represented as the relevance of a sentence with regards to the entire document or the impact of a word within each sentence. Cohesive links are also introduced between adjacent sentences and paragraphs in order to model the information flow throughout the discourse; these links are also indicative of cohesion gaps that are often caused by changes in topics. In addition, cohesive links are introduced between highly related discourse constituents in order to better reflect both high local or global text cohesion.

2 Method

We propose a method that extends CNA [2] for performing multi-document evaluations in order to predict students' comprehension of information presented in multiple texts. CNA considers text content and discourse structure in terms of cohesive links that are defined between multiple levels (i.e., sentences, paragraphs and the entire text). CNA can be used to quantify both local and global cohesion while relying on multiple semantic similarity models.

2.1 Dataset

Undergraduate students (n = 146) from a southwestern university in the United States participated in the study. Students first completed a demographics survey followed by a reading task composed of four texts about green living (i.e., lifestyle centered on balancing the usage, as well as preserving Earth's natural resources). As they read, each student wrote 30 self-explanations on specific target sentences distributed throughout the four texts. Target sentences were presented every two to four sentences and were selected on the basis that self-explanations could support inference generation of the content. After reading all of the texts, students answered 12 open-ended comprehension questions covering information from one or multiple texts followed by a reading skill test and a prior knowledge test. The questions are categorized under three types (textbase, intra-textual, and inter-textual) with four questions per category so as to cover the different comprehension and inferencing tasks in which readers engage. Each of the 12 questions were assigned a score of 0 to 1.0 and then summed to provide an assessment of the overall performance on a [0 to 12] scale. The final dataset consists of four independent texts (labeled A, B, C and D), 30 self-explanations, and 12 question responses per student (labeled from 1 to 12).

Table 1. Question identifiers (Questions 1 to 12) as a function of question type

Question type	Number of questions	Question identifiers
Text-base	4	Q4, Q7, Q8, Q10
Intra-textual	4	Q1, Q2, Q5, Q11
Inter-textual	43	Q3, Q6, Q9, Q12

2.2 Multi-document Cohesion Network Analysis

Figure 1 introduces an extension for CNA that considers multiple texts and student responses. Our aim is to build an overarching undirected cohesion graph for each student that semantically links the initial texts as a whole, or specific paragraphs or sequences from them, to individual representations of students' self-explanations or their question responses. This CNA network graph addresses coherence by building a global cohesion map in which we semantically link reference texts to different student constructed responses. Thus, the extended CNA network graph contains as nodes individual cohesion graphs generated for each target text level, as well as for each student response. The cohesive links within the extended graph are established based on the instructional setup and denote semantic relatedness between nodes of interest. For example, textbase and intra-textual questions are related to a specific text, whereas inter-textual questions are related to all four texts. Self-explanations are linked to sequences from the corresponding text (e.g., all prior text, adjacent text). The semantic distances were computed using the *ReaderBench* framework [18], which allowed us to experiment with several semantic models (i.e., LSA, LDA, and word2vec) and semantic distances in WordNet [19].

We extracted features describing the semantic relatedness between the reference texts and students' self-explanations or question responses to provide comparisons on what information most accurately predicts students' comprehension. Our feature extraction approach has slight differences in the way we process the self-explanations and the question answers based on the generated cohesive links, namely the granularity of the reference texts, as well as the consideration of one versus all texts.

In addition, we group together the cohesive links between a question answer/self-explanation and the corresponding paragraphs, and compute aggregate statistical metrics such as mean, median, max, and standard deviation when analyzing the links in the extended CNA graph. In the case of inter-textual questions, we also compute an average of the semantic distances between the question answer and all of the existing texts. For a given student, we obtained 42 sets of features (30 self-explanations and 12 question responses). These features were then grouped into question-related and self-explanation-related features, together with their corresponding aggregated statistical metrics.

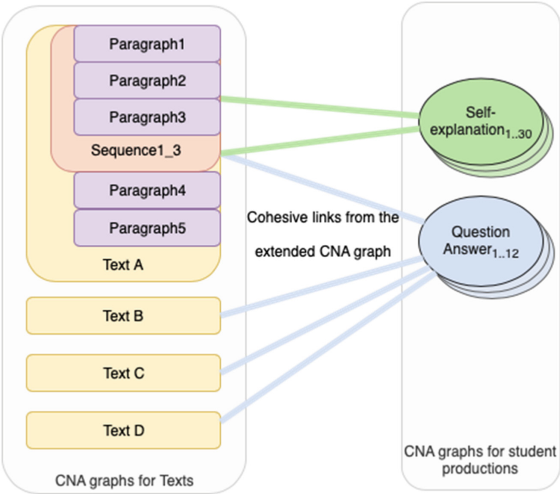

Fig. 1. The CNA multi-document graph.

2.3 Classification Methods

In order to predict the comprehension scores, we used regressor models which are statistical models aimed at making predictions based on a set of features. The models chosen for this experiment are the ones which are known to fare well on a dataset with a small number of examples. We used standard implementations, present in the Python library Scikit-learn [20], for the following models: Linear Regression, Extra Trees Regressor, Support Vector Regression (SVR), and Multi-Layer Perceptron (MLP). The four models were chosen in order to have a varied set of prediction tools, ranging from the least-sophisticated (Linear Regression) to the most complex (Extra Trees Regressor, or SVR). Existing neural network models are unsuited for a regression task with so few data points; thus, from that family of models we opted to solely examine the accuracy of an MLP model.

3 Results

The ReaderBench framework offers several semantic distances, which are related to one another. For each of those, around 300 possible features could be extracted from the CNA graph, meaning that the set of possible features could easily be of the order of thousands. This is why a multiple-step approach was required in order to keep only the most useful features. First, we determined the most suitable semantic distance for our dataset. Afterwards, we filtered the features to retain only the most relevant ones. Third, once we settled on a metric and a restricted group of features; we trained multiple models to observe their predictive accuracy with regards to student comprehension scores.

3.1 Selecting the Best Semantic Measures

In alignment with previous studies on CNA [21], we calculated cohesion using a variety of NLP techniques: vector space models (Latent Semantic Analysis (LSA) [22] and word2vec [23]), topic distributions (Latent Dirichlet Allocation (LDA) [24]), and non-latent word-based semantic distances (i.e., Wu-Palmer ontology-based semantic similarity) [25]. We created CNA graphs limited to using only the question answers/self-explanations and the referred texts, and we computed the semantic distances with each of the metrics, for each user. We then computed mean scores for all self-explanations and question responses. Table 1 presents the correlations between the mean semantic similarity scores and students' comprehension scores. Overall, the most relevant semantic metric for predicting the reading comprehension was provided by word2vec, followed by LSA. Interestingly, LDA performed worst with negative relatedness scores, which means that the topic distributions were considerably different. Moreover, students' responses to the questions provided a better predictor for estimating comprehension score than self-explanations. This was expected, given that the comprehension score was directly based on the responses to the questions. Nonetheless, this indicates that the semantic connections estimated using CNA correlated highly with scores.

Table 2. Pearson correlations between comprehension scores and SE/QA average semantic similarities.

	Score and SE average	Score and QA average
WU-Palmer	.014	.529
LSA	.034	.591
LDA	–.033	.433
word2vec	**.019**	**.675**

3.2 Features Filtering

By employing all the strategies presented in Sect. 2.3, we computed a total of 362 features based on word2vec semantic distances, 272 covering self-explanations, and 90 covering question answers. To reduce multicollinearity, a **baseline filtering** step removed indices with inter-correlations above .9, leaving 126 features (34 for question answers and 92 for the self-explanations). A **second filtering step** consisted in eliminating all the features that had a correlation lower than 0.4 with the comprehension score. The resulting set consisted of 20 features (13 for question answers and 7 for self-explanations). After the second filtering step, the features relating to question answers were almost twice as many as those relating to self-explanation, despite being drawn from a much smaller pool of features. One reason for this is the fact that the questions cover an entire text or a group of texts, while the self-explanations are always centered on a small set of paragraphs.

As displayed in Table 2, six of these features were aggregated features over all exercises in a specific task (question answering or self-explanation), and 13 features were related to the student's performance on a particular task. The notation SE_*X*_P*y* considers the cohesive link between the first *y* paragraphs from text *X* to the self-explanation (denoted as SE), where as SE_*X*_P*y*_*z* reflects the cohesive link between paragraphs *y* to *z* from text X and the SE. The most highly correlated feature score is the mean of the averaged distances between each question and all texts. The best particular task feature is the median over the distances between the answer to question 10 which required intra-textual integration ("Explain how and why these claims might be misleading") and all the paragraphs from the referred text.

Table 3. Correlation between the best features and the comprehension scores.

Aggregated features	*r*
Links between Qs and all texts (M)	.672
Links between Qs and primary text targeted (SD)	.557
Links between SEs and the median of their links to target paragraphs (M)	.527
Links between Qs and the max of their links to target paragraphs (Med)	.515
Links between SEs and target sentence (SD)	.470
Links between SEs for current and prior targeted sentences (Med)	.418
Particular task features	*r*
Links between Q11 and target paragraphs (Med)	.560
Links between Q6 and all texts (M)	.531
Links between Q2 and target paragraphs (Maximum)	.521
Links between Q4 and target paragraphs (Med)	.504
Links between Q6 and target paragraphs (M)	.462
Links between SE_A_P1_3 and target paragraphs (M)	.451
Links between SE_B_P3_4 and target paragraphs (Med)	.448
Links between Q3 and all texts (M)	.432
Link between Q2 and target text	.430
Link between Q7 and target text	.425
Links between Q8 and target paragraphs (Maximum)	.412
Links between Q10 and target paragraphs (M)	.410
Links between SE_B_P4_6 and target paragraphs (Med)	.410
Links between SE_A_P4_7 and target paragraphs (Maximum)	.403

Note: Q = question; SE = self-explanation; M = mean; Med = median; SD = standard deviation.

When analyzing the most important feature for a question in relation to the question type (textbase, intra-textual, inter-textual), we observed that the feature type depends on the question type. The best predicting features for 2 out of 3 inter-textual questions (Q3, Q6) evaluated the semantic similarity between the answer and all the texts. This is in line with how those questions were constructed (as queries for information appearing throughout the 4 texts). In the case of textbase and intra-textual questions which considered information found in a single text, the main features are the aggregating ones (mean, median, or max) applied on the semantic similarity between the answers and all the paragraphs of the text. A second observation is that some question answers are much better predictors for the overall comprehension task than others. The main features for Q11, Q6, Q2, Q4 have a correlation coefficient with the final score above .5, while the main features for Q1, Q5, Q9, and Q12 have a correlation coefficient of around .35 or slightly below. This result is likely due to the complexity of the task, as the four latter questions required inter-textual or intra-textual inferences, which are more complex than textbase questions.

3.3 Predicting Reading Comprehension

We used 5-fold cross-validation as our dataset only has 146 examples. For each model, we trained and tested 5 independent models and report the average and minimum values for mean absolute error (i.e., the measure of difference between the predicted and observed comprehension scores). We examined the models based on the baseline filter (filtering based on multicollinearity) and models using features correlated above .4 with the comprehension score. Table 3 indicates that models using fewer and more highly correlated features were more predictive. This is notably circular given that our ultimate objective is to provide predictions without having the score. Nonetheless, this provides some evidence that the CNA provides good estimates of comprehension scores (Table 4).

Table 4. Prediction performance for the chosen models.

	Classifier	Filtered		Filtered over 0.4	
		MAE average	MAE min	MAE average	MAE min
SE	Linear regression	3.230	2.907	**1.612**	**1.317**
	Extra trees	**1.679**	**1.525**	1.664	1.361
	SVR	1.828	1.497	1.701	1.359
	MLP	1.813	1.401	1.771	1.426
QA	Linear regression	1.551	1.302	**1.434**	**1.096**
	Extra trees	**1.466**	**1.142**	1.508	1.228
	SVR	1.569	1.333	1.435	1.163
	MLP	1.668	1.357	1.600	1.280
Both	Linear regression	5.335	4.372	**1.298**	**0.886**
	Extra trees	**1.480**	**1.221**	1.446	1.133
	SVR	1.721	1.425	1.415	1.097
	MLP	1.853	1.425	1.632	1.259

In addition, results indicate that question-related features are overall more predictive than the self-explanation ones, which was expected, given that the comprehension scores were based on the question answering task. The large discrepancy between question answers and self-explanations that was identified in the first analysis (see Table 2) was considerably lower for comprehension predictions (i.e., a 0.2 MAE difference between the best models for self-explanations and questions answers). This is normal taking into account that self-explanations relate only to one text and they provide a reduced contextualization in contrast to a more detailed question answer. Overall, the best results are obtained using a Linear Regression model on the most highly filtered set of features from both question answers and self-explanations. This shows that even though the question response features are more predictive, self-explanations provide extra information that improves model performance.

Regarding the regressor models, we observed that Extra Trees obtained the best results when trained using a large set of features. However, when switching to the small feature set, the linear regression model narrowly outmatched Extra Trees in all three cases (question answers, self-explanations, and both), despite the fact that its poor performance without filtering based on correlations above .4.

4 Conclusions and Future Work

In this paper, we represent coherence using Cohesion Network Analysis (CNA) in which we model a global cohesion graph that semantically links reference texts to different student constructed responses in order to predict comprehension. We modeled performance using a dataset containing four documents for which students provided self-explanations and answers to open-ended comprehension questions addressing both individual documents as well as aggregated information from multiple sources. Several features were extracted and then filtered by eliminating those that were highly correlated among themselves, or those with weak correlations with the comprehension scores. Four regressor models were trained based on these features, side-by-side comparisons were made in order to highlight which models displayed the lowest MAEs for scores between 1 and 12. The best model without filtering based on correlations with the score was the Extra Trees model, providing between 1.1 and 1.7 MAE. The best model using the added correlation-based filter was Linear Regression, providing between 0.9 and 1.6 MAE. Both outcomes are encouraging - demonstrating that the features extracted from an extended CNA cohesion graph are capable of estimating student's comprehension scores within acceptable margins of error.

Our results showed that answers to some questions may be more suitable predictors than others and question complexity decreased performance. For example, three questions for which the answers were not good predictors of comprehension required inter-textual or intra-textual inferences. Self-explanations also offered valuable insights regarding the students' comprehension. When training a model with self-explanation-related features, the model without filtering provided a close proximation to comprehension scores (i.e., 1.5 MAE). This means that even without having students answer comprehension questions, we can estimate comprehension with relatively good accuracy.

As future developments, this experiment needs to be replicated on various datasets with different text sets and populations. Ultimately, our objective is to twofold: (a) simulate comprehension of multiple documents on line, thus providing the means for feedback, and (b) model the coherence of students' comprehension of multiple documents. The current study is our initial foray toward reaching these objectives.

Acknowledgments. This research was supported by the ReadME project "Interactive and Innovative application for evaluating the readability of texts in Romanian Language and for improving users' writing styles", contract no. 114/15.09.2017, MySMIS 2014 code 119286, the FP7 2008-212578 LTfLL project, the Institute of Education Sciences (R305A180144 and R305A180261), and the Office of Naval Research (N00014-17-1-2300).

References

1. Allen, L.K., Jacovina, M., McNamara, D.S.: Cohesive features of deep text comprehension processes. In: 38th Annual Meeting of the Cognitive Science Society, pp. 2681–2686. Cognitive Science Society, Philadelphia, PA (2016)
2. Dascalu, M., McNamara, D.S., Trausan-Matu, S., Allen, L.K.: Cohesion network analysis of CSCL participation. Behav. Res. Methods **50**(2), 604–619 (2018)
3. Hastings, P., Hughes, S., Magliano, J.P., Goldman, S.R., Lawless, K.: Assessing the use of multiple sources in student essays. Behav. Res. Methods **44**(3), 622–633 (2012)
4. Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications. Cambridge University Press, Cambridge (1994)
5. Jurafsky, D., Martin, J.H.: An Introduction to Natural Language Processing. Computational Linguistics, and Speech Recognition. Pearson Prentice Hall, London (2009)
6. McNamara, D.S., Magliano, J.: Toward a comprehensive model of comprehension. Psychol. Learn. Motiv. **51**, 297–384 (2009)
7. Kintsch, W.: The role of knowledge in discourse comprehension: a construction-integration model. Psychol. Rev. **95**(2), 163 (1988)
8. Wiley, J., Goldman, S.R., Graesser, A.C., Sanchez, C.A., Ash, I.K., Hemmerich, J.A.: Source evaluation, comprehension, and learning in Internet science in-quiry tasks. Am. Educ. Res. J. **46**(4), 1060–1106 (2009)
9. Magliano, J., McCrudden, M.T., Rouet, J.F., Sabatini, J.: The modern reader: should changes to how we read affect research and theory? In: Schober, M.F., Rapp, D.N., Britt, M. A. (eds.) Handbook of Discourse Processes, 2nd Ed. Taylor & Francis, New York (2018)
10. Goldman, S.R., Braasch, J.L., Wiley, J., Graesser, A.C., Brodowinska, K.: Comprehending and learning from Internet sources: processing patterns of better and poorer learners. Read. Res. Q. **47**(4), 356–381 (2012)
11. Goldman, S.R., Rakestraw, J.A.: Structural aspects of constructing meaning from text. In: Kamil, M.L., Mosenthal, P.B., Pearson, P.D., Barr, R. (eds.) Handbook of Reading Research, vol. 3, pp. 311–335. Erlbaum, Mahwah (2000)
12. Ozuru, Y., Briner, S., Kurby, C.A., McNamara, D.S.: Comparing text comprehension measured by multiple-choice and open-ended questions. Can. J. Exp. Psychol. **67**, 215–227 (2013)
13. Chi, M.T., De Leeuw, N., Chiu, M.-H., LaVancher, C.: Eliciting self-explanations improves understanding. Cogn. Sci. **18**(3), 439–477 (1994)

14. McNamara, D.S.: Self-explanation and reading strategy training (SERT) improves low-knowledge students' science course performance. Discourse Processes **54**(7), 479–492 (2017)
15. Dascalu, M., Trausan-Matu, S., McNamara, D.S., Dessus, P.: ReaderBench – automated evaluation of collaboration based on cohesion and dialogism. Int. J. Comput.-Support. Collaborative Learn. **10**(4), 395–423 (2015)
16. Trausan-Matu, S., Dascalu, M., Dessus, P.: Textual complexity and discourse structure in computer-supported collaborative learning. In: Cerri, S.A., Clancey, W.J., Papadourakis, G., Panourgia, K. (eds.) ITS 2012. LNCS, vol. 7315, pp. 352–357. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-30950-2_46
17. Dascalu, M., Trausan-Matu, S., Dessus, P.: Cohesion-based analysis of CSCL conversations: holistic and individual perspectives. In: 10th International Conference on Computer-Supported Collaborative Learning (CSCL 2013), vol. 1, pp. 145–152. ISLS, Madison, USA (2013)
18. Dascalu, M., Dessus, P., Bianco, M., Trausan-Matu, S., Nardy, A.: Mining texts, learner productions and strategies with ReaderBench. In: Peña-Ayala, A. (ed.) Educational Data Mining. SCI, vol. 524, pp. 345–377. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-02738-8_13
19. Budanitsky, A., Hirst, G.: Evaluating WordNet-based measures of lexical semantic relatedness. Comput. Linguist. **32**(1), 13–47 (2006)
20. Pedregosa, F., et al.: Scikit-learn: machine learning in Python. **12**, 2825–2830 (2011)
21. Dascalu, M.: Analyzing Discourse and Text Complexity for Learning and Collaborating. Studies in Computational Intelligence, vol. 534. Springer, Switzerland (2014). https://doi.org/10.1007/978-3-319-03419-5
22. Landauer, T.K., Foltz, P.W., Laham, D.: An introduction to latent semantic analysis. Discourse Processes **25**(2/3), 259–284 (1998)
23. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representation in vector space. In: Workshop at ICLR, Scottsdale, AZ (2013)
24. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. J. Mach. Learn. Res. **3**(4–5), 993–1022 (2003)
25. Wu, Z., Palmer, M.: Verb semantics and lexical selection. In: 32nd Annual Meeting of the Association for Computational Linguistics ACL 1994, pp. 133–138. ACL, New Mexico (1994)