Predicting Comprehension from Students' Summaries

Mihai Dascălu, Lucia Larise Stavarache, Philippe Dessus, Stefan Trausan-Matu, Danielle S. Mcnamara, Maryse Bianco

Predicting Comprehension from Students' Summaries

Mihai Dascalu[1(✉)], Larise Lucia Stavarache[1], Philippe Dessus[2], Stefan Trausan-Matu[1], Danielle S. McNamara[3], and Maryse Bianco[2]

[1] Computer Science Department, University Politehnica of Bucharest, Bucharest, Romania
{mihai.dascalu,stefan.trausan}@cs.pub.ro,
larise.stavarache@ro.ibm.com
[2] LSE, University Grenoble Alpes, Grenoble, France
{philippe.dessus,maryse.bianco}@upmf-grenoble.fr
[3] LSI, Arizona State University, Tempe, USA
dsmcnama@asu.edu

Abstract. Comprehension among young students represents a key component of their formation throughout the learning process. Moreover, scaffolding students as they learn to coherently link information, while organically constructing a solid knowledge base, is crucial to students' development, but requires regular assessment and progress tracking. To this end, our aim is to provide an automated solution for analyzing and predicting students' comprehension levels by extracting a combination of reading strategies and textual complexity factors from students' summaries. Building upon previous research and enhancing it by incorporating new heuristics and factors, Support Vector Machine classification models were used to validate our assumptions that automatically identified reading strategies, together with textual complexity indices applied on students' summaries, represent reliable estimators of comprehension.

Keywords: Reading strategies · Textual complexity · Summaries assessment · Comprehension prediction · Support vector machines

1 Introduction

The challenges in helping readers understand discourse and achieving coherent underlying mental representations push educators to devise alternative and novel techniques, beyond focusing on classical cognitive reading processes. Devising instruction on reading comprehension strategies emerged from the need to facilitate continuous learning and enable readers to enhance their understanding levels without eliminating or giving up traditional learning methods. For example, *SERT* (Self-Explanation Reading Training) [1] was designed to support readers in self-monitoring their understanding while engaging in effective comprehension strategies. The principal assumption underlying SERT is that, in order to fully understand a text, readers must be able to provide an answer to the basic question "What does this mean?". iSTART [2], the first automated system that scaffolds self-explanations, has demonstrated that SERT is a successful complementary strategy for learning, particularly for high school stu-

C. Conati et al. (Eds.): AIED 2015, LNAI 9112, pp. 95–104, 2015.
DOI: 10.1007/978-3-319-19773-9_10

dents. Psychological and pedagogical research has demonstrated that individuals better understand challenging text if they attempt to explain to themselves what they have read [3], and students do so more effectively if they have been provided training and practice in using comprehension strategies [4, 5]. In addition, by using self-explanation, readers tend to more effectively structure the content and step away from rote learning (which usually results in more rapid memory loss) and towards more organic learning. This process in turn results in more connections between concepts, helping the reader to construct coherent and long lasting mental representations [6].

Based on our previous work [7, 8], this study is focused on comprehension assessment for elementary school students derived from their summaries, by identifying metacognitive comprehension strategies [9] and by applying specific textual complexity factors on their summaries. In terms of building technologies to assess the use of strategies within summaries, primary school students represent a different category of learners than the ones addressed thus far, as they possess less knowledge compared with adult or experienced readers. Hence, their ability to interconnect information based on previous experience is clearly lower. The current work builds on previous research by using refined mechanisms for identifying reading strategies and a comprehensive set of textual complexity indices incorporating classic surface indices derived from automatic essay grading techniques, morphology and syntax [10], as well as semantics and discourse [7, 11]. In addition, Support Vector Machine classification models [12] use combined subsets of reading strategies and textual complexity factors, which are applied on the analyzed summaries, in order to predict students' comprehension levels.

The primary research question addressed in this study is the following: Are reading strategies identified from students' summaries, combined with textual complexity factors also extracted from their summaries, reliable predictors for evaluating the students' comprehension levels? The following sections include an overview of techniques used to identify reading strategies, textual complexity categories from our multi-layered approach, the proposed classification model used to combine the identified reading strategies and textual complexity factors for predicting the comprehension level from students' summaries, ending with conclusions and future work.

2 Reading Strategies Identification

Readers, although sometimes not fully aware, frequently make use of reading strategies to improve their understanding and to interconnect information out of which four main categories are distinguishable [1]: 1 *paraphrasing*, 2) *text-based inferences* consisting of *causality* and *bridging*, 3) *knowledge-based inferences* or elaboration, and 4) monitoring or *control. Paraphrasing* enables users to express their current understanding on the topic by reusing words and concepts from the initial text, which can be considered a first step in building a coherent representation of discourse. *Text-based inferences* build explicit relationships between two or more textual segments of the initial text. On the other hand, *knowledge-based inferences* connect the information from the presented text to the learner's personal knowledge, this being essential

for building the situation model [13]. Last but not least, *control strategies* grant balance during the actual monitoring process, as readers explicitly express what they have or have not understood.

The use and identification of reading strategies as described above on a large scale can be problematic, considering the disproportion between the number of students and tutors. Moreover, assessing the content of a summary is a demanding and a subjectivity-laden activity, which can benefit by being assisted by automated techniques. These are the main motives behind the idea of using a computer program instead of or as support for a human tutor. Additionally, an automated comprehension assessment tool helps learners by enabling them to better track their progress and develop more rapidly.

Starting from the identification strategies previously proposed and validated [14] that were applied to students' self-explanations given at predefined breakpoints in the narration of the reading material, our aim was to adapt the automated extraction methods to better match the processing of summaries, following the previous categories. Causal and bridging strategies had to be separated due to the underlying computational complexity and their corresponding approaches. However, causal inferences can be considered a particular case of bridging, as well as a reference resolution. Altogether, reading strategies highlight inferences made by learners and the connection between the summaries and the referential material.

Causality is identified by using cue phrases or discourse markers such as "*parce que*" (the experiments were performed in French, the translation is "because"), "*pour*" (for), "*donc*" (thus), "*alors*" (then), "*à cause de*" (because of), whereas for *control* different markers are used such as "*je me souviens*" (I remember), "*je crois*" (I believe that), "*j'ai rien compris*" (I haven't understood anything) and are enforced in the pattern matching process. Subsequently, *paraphrases* are extracted through lexical similarities by identifying identical lemmas, stems, or synonyms from lexicalized ontologies – *WordNet* or *WOLF* [15, 16] – with words from the initial text. Adjacent words from students' summaries are clustered into segments of paraphrasing concepts, highlighting contiguous zones strongly related to the initial text.

Further on, an *inferred concept* is considered to be a non-paraphrased word, not present in the original text, yet maintains a high cohesion value with it. Cohesion plays a central role in our discourse representation and analysis [8] and its corresponding value is determined as an aggregated score of semantic similarity measures [8] applied on lexicalized ontologies [15], more specifically Wu-Palmer distance applied on WOLF [16], cosine similarity from Latent Semantic Analysis (LSA) [17] vector spaces, and Jensen-Shannon dissimilarity applied on Latent Dirichlet Allocation (LDA) [18] topic distributions. Both LSA and LDA semantic models were trained on "Le Monde" corpora (French newspaper, approx. 24M words) after applying stop words elimination and lemmatization.

Finally, the measure for *bridging* considers the connections between different textual segments from the initial text and the summary. Therefore, for each sentence in the summary, the sentence with the highest cohesion with the initial text is identified and marked as being linked to the summary if the corresponding cohesion value exceeds a threshold. The imposed threshold is used to limit the linkage of off-topic sen-

tences from the summary with the initial text. Subsequently, a similar aggregation of contiguous sentences from the initial reading material with the bridging segments is performed, which highlights the sections of the reading material that are actually recalled within the summary.

3 Textual Complexity Assessment

Automated evaluation of textual complexity represents a key focus for the linguistic research field; it emphasizes the evolution of technology's facilitator role in educational processes. *E-Rater* [19] can be considered one of the first systems which automatically measures essay complexity by extracting a set of features representing facets of writing quality. The *E-Rater* analyzer supports a multi-layered textual complexity evaluation based on the centering theory about building a model for assessing the complexity of inferences within the discourse [20]. In addition, various indices are considered for measuring complexity [19] such as spelling errors, content analysis based on vocabulary measures, lexical complexity/diction, proportion of grammar and of style comments, organization, and development scores and features rewarding idiomatic phraseology.

Multiple systems were implemented and were widely adopted in various educational programs [21]: *Lexile* (MetaMetrics), *ATOS* (Renaissance Learning), Degrees of Reading Power: *DRP Analyzer* (Questar Assessment, Inc.), *REAP* (Carnegie Mellon University), *SourceRater* (Educational Testing Service), *Coh-Metrix* (University of Memphis) and *Dmesure* (Université Catholique de Louvain). Our implemented system, *ReaderBench* [7, 8], integrates the most common indices from the previous systems as baseline and is centered on semantics and discourse analysis by including additional indices for evaluating textual cohesion and discourse connectivity, described later on in detail.

As presented in [22], there are three main categories of factors considered in the textual complexity analysis of the French language that also include the most common and frequently used indices from the previous solutions. Firstly, the *surface* category is comprised of quantitative measures and the analysis of individual elements (words, phrases, paragraphs) by extracting simple or combined indices (e.g., Page's grading technique for automated scoring including number of words, sentences or paragraphs, number of commas, average word length or words per sentence) [23], as well as word and character entropy [10]. A particular set of factors from surface analysis handles w*ord complexity,* which consist of the distance between the inflected form, lemma and stem, specificity of a concept reflected in its inverse document frequency from the training corpora (in our case, articles from "Le Monde" corpora), the distance in the hypernym tree from the lexicalized ontology WOLF [16], or the word senses count from the same ontology.

Secondly, the *syntactic* category handles the parsing tree by considering the maximum height and size of the tree, as well as the distributions of specific parts of speech. Balanced CAF (Complexity, Accuracy, and Fluency) [24] techniques also add

their contribution to the analysis of the previous category through the introduction of lexical/syntactic diversity and sophistication.

Thirdly, whereas the first two categories are more representative for writing ability, the *semantics* and *discourse analysis* category is more comprehension centered by identifying the underlying cohesive links [7, 8, 11]. This category makes use of lexical chains, semantic distances, and discourse connectives, all centered on cohesion, a key feature in terms of discourse representation [8] and textual complexity analysis. This category is particularly appealing as it addresses the internal structure of the summary and provides clear insights on whether the learner has achieved a coherent representation of the text or if (s)he is facing problems in terms of cohesion when expressing impressions and thoughts within the summary.

4 Validation of the Comprehension Prediction Model

Model validation of learner's comprehension level has been performed using several scenarios comprising of different combinations between reading strategies, textual complexity factors applied to students' summaries, cohesion between each summary and the initial reading material, as well as external factors (e.g., students' oral fluency). Firstly, comprehension prediction based on reading strategies and cohesion has been computed in order to shape the baseline of our approach. Secondly, multiple textual complexity factors employed on the students' summaries were combined, clearly revealing that using all indices together cannot be an accurate predictor of textual complexity and that surface indices are not reliable for the task at hand. The next scenario only used the best matching factors, from both previous scenarios, which proved successful and increased both the average and the individual agreements. Finally, oral fluency has been added as an external factor –one highly related to comprehension–, which in return provided a significant increase in prediction accuracy.

Our experiments [25] have been conducted with students between the ages of 8 and 11 years old (3^{rd}–5^{th} grade), uniformly distributed in terms of their age, and who produced 149 summaries of the two French stories of approximately 450 words (*The Cloud Swallower* and *Matilda*). After their lecture, students explained what they understood by verbally summarizing their impressions and thoughts about the initial text. These summaries were recorded and later on transcribed. Students were also administered a posttest comprising of 28 questions used to assess their comprehension of the reading materials. Predefined rules and patterns were used to automatically clean the transcribed verbalizations. With regards to the proposed textual complexity factors applied on students' summaries, the same factors were used in [22] to predict the difficulty of the selected French stories. As a result, both texts were classified as being optimal for 3^{rd} graders, making them appropriate in terms of reading ease for all the students participating in our experiments. Because the materials were presented to adjacent elementary classes, their levels were adequate for both 4th and 5th graders who did not consider them to be boring, nor childish.

With regards to comprehension prediction, we opted to create three comprehension classes (noted C1, C2, and C3 in the following tables) with a distribution of 30%, 40% and 30% of student posttest scores sorted in ascending order and to apply 3-fold cross-validations for the SVM training process. This distribution created an equitable split of students per comprehension classes and also marked significant differences in terms of covered scores per class from the [0; 28] scale for all questions from the posttest. Multiclass SVMs have been trained to predict the appropriate comprehension class based on the selected factors applied on students' summaries. We opted to use RBF kernels as the corresponding hyperparameters (the regularization constant C and the kernel hyperparameter γ) were optimized through Grid Search [26]. In addition, we must emphasize that average accuracy is quite low as there are some high discrepancies between summaries with similar comprehension scores in terms of structure and complexity, which ultimately misleads the SVM training process.

As expected, reading strategies, paraphrases, control and causality occurrences were much easier to identify than information coming from students' experience. Nevertheless, if we consider each strategy separately, the prediction rate is low, whereas the combination dramatically increases accuracy (see Table 1). Also, for some strategies it was impossible to differentiate among comprehension classes because rather few occurrences exist in the training dataset, or because students equitably use that specific strategy. We also noticed small prediction rates for the first and second classes due to rather small differentiations between adjacent classes and to conflicting instances. The previous instances consisted of encountered cases in which students with a high number of potentially involuntarily used reading strategies pertained to a low comprehension class based on their posttest, although all textual indices from their summary pointed to a higher degree of comprehension.

In order to increase the strength of the link between the summary and the original reading material, cohesion between the entire texts was introduced. However as a singular effect, the overall agreement decreased.

Table 1. Comprehension prediction agreement based on reading strategies and cohesion

Factors	C1	C2	C3	Average agreement
Paraphrasing	.214	.235	.804	*.418*
Text-based inferences	.524	.431	.647	*.534*
Knowledge-based inferences	.095	.020	1	*.372*
Control	0	0	1	*.333*
All reading strategies	.595	.451	.608	*.551*
All reading strategies plus the cohesion value with the initial document	.571	.451	.549	*.524*

Results presented in Table 1 are encouraging based on the limited number of training instances, the reduced number of classification attributes, and the fact that a lot of noise existed within the transcriptions. Nevertheless, additional factors were introduced in order to increase the accuracy of comprehension prediction.

Table 2. Comprehension prediction accuracy based on textual complexity factors

Factors	C1	C2	C3	Average
All textual complexity indices	.167	.137	.941	*.415*
Surface factors and CAF	0	.294	.863	*.386*
Morphology and semantics	.524	.275	.725	*.508*
Morphology, semantics, all reading strategies and cohesion with initial document	.524	.49	.529	*.514*

Table 3. Average prediction accuracy for the best matching indices

Most relevant factors	M	Most relevant factors	M
(C) Causal relation	*.587*	(B) Syntactic Sophistication - CAF	*.433*
(D) Text-based inferences	*.540*	(B) Avg. tree depth	*.429*
(A) Word entropy	*.509*	(D) Cohesion with initial text	*.429*
(B) Avg. number of adverbs	*.494*	(D) Paraphrasing	*.429*
(A) No. words in summary	*.488*	(B) Avg. no. pronouns	*.427*
(A) Avg. word length	*.486*	(A) Mean word polysemy count	*.427*
(C) All connectives	*.474*	(A) Lexical Diversity	*.426*
(C) Logical relation	*.470*	(C) Overall document score	*.424*
(A) Mean distance between words and corresponding stems	*.463*	(A) Avg. no. sentences per paragraph	*.423*
(C) Avg. intra-paragraph cohesion	*.461*	(A) Total no. sentences	*.423*
(A) Avg. sentence length	*.455*	(C) Avg. sentence-block cohesion	*.421*
(A) Avg. words in sentence	*.452*	(A) Normalized no. sentences	*.420*
(A) Standard deviation for words (letters)	*.447*	(B) Third Person Singular Pronouns Count	*.415*
(C) Avg. paragraph score	*.443*	(B) Avg. no. adjectives	*.412*
(A) Normalized no. of commas	*.441*	(B) Avg. tree size	*.411*
(B) Second Person Singular Pronouns Count	*.441*	(B) First Person Singular Pronouns Count	*.410*
(B) Avg. no. prepositions	*.436*	(B) Lexical Sophistication - CAF	*.409*
(A) Normalized no. words	*.435*	(A) Mean word distance in hypernym tree	*.400*

*(A) - surface factors; (B) - syntactic and morphological factors, including CAF; (C) - semantics, discourse analysis and connectives; (D) - reading strategies.

As it can be observed from Table 2, the integration of surface indices collectively has a low prediction rate. Moreover, the use of too many factors (out of which some proved to be inadequate) is also detrimental to the overall classification: the use of all textual complexity indices or of only the surface factors predicted that all summaries were in the highest comprehension class, clearly a problem in the classification due to the structure similarities between all summaries. Therefore, it turned out to be most appropriate to rely only on complementary and stable factors of textual complexity. Moreover, a slight improvement could be observed after considering the adapted reading strategies extracted from the summary and the cohesion with the initial reading material.

In addition, the best matching individual factors from Table 3 represent a balanced and representative mixture of the previously identified analysis categories and their integration marks a significant improvement in the prediction rate: C1: *.571*; C2: *.608*; C3: *.804*, with an average agreement of .661.

Table 4. Comprehension prediction accuracy after introducing oral fluency

Factors	C1	C2	C3	Average
All reading strategies, cohesion with initial document and oral fluency	.667	.529	.784	*.660*
Morphology, semantics, all reading strategies, cohesion with initial document and oral fluency	.714	.549	.784	*.683*

In the end, the addition of external, non-textual factors (e.g., students' oral fluency determined manually as the number of spoken words per minute) improved the overall results (see Table 4), whereas the problems of using all textual complexity indices remain in the identification of the first two comprehension classes. Overall, the combination of morphology, semantics, reading strategies, cohesion with the initial document, and oral fluency turned out to be one of the most reliable predictors of comprehension for students at the given age. In addition, the semantics category of textual complexity factors, corroborated with the semantic similarity between the summary and the original text, emphasize the importance of cohesion, both internally within the summary, but also between the summary and the initial reading material.

5 Conclusion and Future Research Directions

The integration of the two different approaches applied to summaries resulted in a promising direction for improving comprehension prediction among students. Neither of the two approaches by itself is sufficient to obtain a reliable estimation of comprehension, whereas the combination represents leverage for improving the assessment process. Nevertheless, we can state that reading strategies by themselves are good predictors for assessing comprehension, while morphology and semantics provide a solid ground for evaluations that surpass surface factors commonly used in other automated systems. Moreover, we must emphasize the complementarity of the approaches, as reading strategies and cohesion reflect the link with the initial reading material, whereas textual complexity factors are centered on analyzing the summary's internal structure. Furthermore, the performed measurements and validations indicate that reading strategies, mixed with textual complexity factors [21] and essay scoring techniques [27] increase the accuracy of the predictions related to a student's comprehension level.

As described above, students are a special category of learners who pass through an increasingly difficult process of where they constantly receive more and more information that they must assimilate. This transition along with the inspection of the summaries has emphasized the need for introducing additional techniques to improve understanding and to facilitate both their activity and their tutor's. The main goal of

this paper was to expand the research path of assessing comprehension, while the overall scope our system, *ReaderBench,* remained to support tutors through a regularized and predictable process of prediction as an alternative to the subjectivity-laden task of manual evaluation.

Our future aims consist of expanding the experimental components further, by adding the possibility to automatically assess students' reading fluency and by deploying *ReaderBench* in classroom settings in order to analyze student's comprehension levels on a regular basis and to infer possible comprehension issues, more accurately and in a timely manner.

Acknowledgements. This research was partially supported by the ANR DEVCOMP 10-BLAN-1907-01 and the 2008-212578 LTfLL FP7 projects, by the NSF grants 1417997 and 1418378 to Arizona State University, as well as by the POSDRU/159/1.5/S/132397 and 134398 projects. We would also like to thank Aurélie Nardy and Françoise Toffa who helped us to gather the experimental data, as well as the teachers and students who participated in our experiments.

References

1. McNamara, D.S.: SERT: Self-Explanation Reading Training. Discourse Processes **38**, 1–30 (2004)
2. McNamara, D.S., Levinstein, I., Boonthum, C.: iSTART: Interactive strategy training for active reading and thinking. Behavior Research Methods, Instruments, & Computers **36**(2), 222–233 (2004)
3. Millis, K., Magliano, J.P., Wiemer-Hastings, K., Todaro, S., McNamara, D.S.: Assessing and improving comprehension with latent semantic analysis. In: Landauer, T.K., McNamara, D., Dennis, S., Kintsch, W. (eds.) Handbook of Latent Semantic Analysis, pp. 207–225. Erlbaum, Mahwah (2007)
4. McNamara, D.S., O'Reilly, T.P., Best, R.M., Ozuru, Y.: Improving adolescent students' reading comprehension with iSTART. Journal of Educational Computing Research **34**(2), 147–171 (2006)
5. Jackson, G.T., McNamara, D.S.: Motivation and performance in a game-based intelligent tutoring system. Journal of Educational Psychology **105**, 1036–1049 (2013)
6. McNamara, D.S., Magliano, J.P.: Self-explanation and metacognition. In: Hacher, J.D., Dunlosky, J., Graesser, A.C. (eds.) Handbook of metacognition in education, pp. 60–81. Erlbaum, Mahwah (2009)
7. Dascalu, M., Dessus, P., Bianco, M., Trausan-Matu, S., Nardy, A.: Mining texts, learners productions and strategies with *ReaderBench*. In: Peña-Ayala, A. (ed.) Educational Data Mining: Applications and Trends. SCI, vol. 524, pp. 345–377. Springer, Heidelberg (2014)
8. Dascălu, M.: Analyzing Discourse and Text Complexity for Learning and Collaborating. SCI, vol. 534. Springer, Heidelberg (2014)
9. Nash-Ditzel, S.: Metacognitive Reading Strategies Can Improve Self-Regulation. Journal of College Reading and Learning **40**(2), 45–63 (2010)

10. Dascălu, M., Trausan-Matu, S., Dessus, P.: Towards an integrated approach for evaluating textual complexity for learning purposes. In: Popescu, E., Li, Q., Klamma, R., Leung, H., Specht, M. (eds.) ICWL 2012. LNCS, vol. 7558, pp. 268–278. Springer, Heidelberg (2012)
11. Dascalu, M., Dessus, P., Trausan-Matu, Ş., Bianco, M., Nardy, A.: *ReaderBench*, an environment for analyzing text complexity and reading strategies. In: Lane, H.C., Yacef, K., Mostow, J., Pavlik, P. (eds.) AIED 2013. LNCS, vol. 7926, pp. 379–388. Springer, Heidelberg (2013)
12. Cortes, C., Vapnik, V.N.: Support-Vector Networks. Machine Learning **20**(3), 273–297 (1995)
13. van Dijk, T.A., Kintsch, W.: Strategies of discourse comprehension. Academic Press, New York (1983)
14. Dascalu, M., Dessus, P., Bianco, M., Trausan-Matu, S.: Are automatically identified reading strategies reliable predictors of comprehension? In: Trausan-Matu, S., Boyer, K.E., Crosby, M., Panourgia, K. (eds.) ITS 2014. LNCS, vol. 8474, pp. 456–465. Springer, Heidelberg (2014)
15. Budanitsky, A., Hirst, G.: Evaluating WordNet-based Measures of Lexical Semantic Relatedness. Computational Linguistics **32**(1), 13–47 (2006)
16. Sagot, B.: WordNet Libre du Francais (WOLF) (2008). http://alpage.inria.fr/~sagot/wolf.html
17. Landauer, T.K., Dumais, S.T.: A solution to Plato's problem: the Latent Semantic Analysis theory of acquisition, induction and representation of knowledge. Psychological Review **104**(2), 211–240 (1997)
18. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet Allocation. Journal of Machine Learning Research **3**(4–5), 993–1022 (2003)
19. Powers, D.E., Burstein, J., Chodorow, M., Fowles, M.E., Kukich, K.: Stumping e-rater®: Challenging the validity of automated essay scoring. ETS, Princeton (2001)
20. Grosz, B.J., Weinstein, S., Joshi, A.K.: Centering: a framework for modeling the local coherence of discourse. Computational Linguistics **21**(2), 203–225 (1995)
21. Nelson, J., Perfetti, C., Liben, D., Liben, M.: Measures of text difficulty Council of Chief State School Officers, Washington, DC (2012)
22. Dascalu, M., Stavarache, L.L., Trausan-Matu, S., Dessus, P., Bianco, M.: Reflecting comprehension through french textual complexity factors. In: ICTAI 2014, pp. 615–619. IEEE, Limassol (2014)
23. Page, E.: The imminence of grading essays by computer. Phi Delta Kappan **47**, 238–243 (1966)
24. Housen, A., Kuiken, F.: Complexity, Accuracy, and Fluency in Second Language Acquisition. Applied Linguistics **30**(4), 461–473 (2009)
25. Nardy, A., Bianco, M., Toffa, F., Rémond, M., Dessus, P.: Contrôle et régulation de la compréhension. In: David, J., Royer, C. (eds.) L'apprentissage de la Lecture: Convergences, Innovations, Perspectives, p. 16. Peter Lang, Bern-Paris (in press)
26. Bergstra, J., Bengio, Y.: Random Search for Hyper-Parameter Optimization. The Journal of Machine Learning Research **13**, 281–305 (2012)
27. Todd, R.W., Khongput, S., Darasawang, P.: Coherence, cohesion and comments on students' academic essays. Assessing Writing **12**(1), 10–25 (2007)