Reflecting Comprehension through French Textual Complexity Factors

Mihai Dascalu, Lucia Larise Stavarache, Stefan Trausan-Matu, Philippe Dessus, Maryse Bianco

Reflecting Comprehension through French Textual Complexity Factors

Mihai Dascalu, Larise Lucia Stavarache, Stefan Trausan-Matu
Computer Science Department
University Politehnica of Bucharest
Bucharest, Romania
mihai.dascalu@cs.pub.ro, larise.stavarache@ro.ibm.com, stefan.trausan@cs.pub.ro

Philippe Dessus, Maryse Bianco
Laboratory of Educational Sciences
University Grenoble Alpes
Grenoble, France
philippe.dessus@upmf-grenoble.fr, maryse.bianco@upmf-grenoble.fr

***Abstract*—Research efforts in terms of automatic textual complexity analysis are mainly focused on English vocabulary and few adaptations exist for other languages. Starting from a solid base in terms of discourse analysis and existing textual complexity assessment model for English, we introduce a French model trained on 200 documents extracted from school manuals pre-classified into five complexity classes. The underlying textual complexity metrics include surface, syntactic, morphological, semantic and discourse specific factors that are afterwards combined through the use of Support Vector Machines. In the end, each factor is correlated to pupil comprehension metrics scores, spanning throughout multiple classes, therefore creating a clearer perspective in terms of measurements impacting the perceived difficulty of a given text. In addition to purely quantitative surface factors, specific parts of speech and cohesion have proven to be reliable predictors of learners' comprehension level, creating nevertheless a strong background for building dependable French textual complexity models.**

Keywords—French textual complexity assessment; Readability; Support Vector Machines; Textual cohesion

I. INTRODUCTION

Automated essay scoring, identification of textual patterns or exploration of term correlation markers, all stand as appealing subjects related to textual analysis debated in both academic and commercial environments. While research groups focus on handling and connecting large amounts of information that the social web produced in the last decade, the commercial market is focused on building instruments that could improve the presentation and induce a better control of the information flow from source to targeted audience.

As a result, there is a lot of unexplored ground regarding other languages that could actually reveal surprising results by also taking into consideration the following variables: the complexity of the vocabulary, morphology, discourse structure and overall aim to enhance the comprehension level among students. The English language has a flexible and continuously expanding vocabulary, and can be considered of medium difficulty with regards to its learning time for non-native speakers. Starting from this assumption and correlating with the English language origins, in this paper we opted to perform French textual complexity analysis from a novel perspective. Our aim was to confirm already proven theories, but also to point out language specificities, relevant factors in terms of comprehension assessment and specific textual traits.

Based on the classic educational scenario, in which learners are introduced to specific learning materials scaffolded by a tutor, the latter's aim should consist of sharing knowledge in the most productive manner, adapted to the learners' corresponding comprehension level. Nevertheless, evaluating comprehension can be achieved in sundry ways like constant feedback gathering, tests, class interactivity or homework. Moreover, evaluating a large number of texts in a relatively short amount of time in order to find an adequate selection is prone to human errors. Therefore, our goal is to support tutors assessing textual materials in a controlled, equitable, reproducible and traceable manner, applicable to a large number of materials in a short timespan. In addition, the constant measurement of students' comprehension can be a good indicator if the presented materials and their underlying structure are actually the best fit.

The following section presents an overview of the textual complexity assessments, their categorization, and other similar automated systems that have been developed to evaluate textual complexity. The third section is centered on the description of the proposed French textual complexity automatic model, while the fourth section introduces the interpretation and extraction of relevant markers for textual complexity assessment. Afterwards, the fifth section emphasizes the refinement strategies for obtaining more accurate results, while the last section is focused on conclusions and future improvements.

II. Overview of Textual Complexity Assessment

Measuring textual complexity in an automatic and adaptive manner has been a goal and a challenge for tutors and researchers in the past years. Education methodologies, essay scoring or class teaching, all need to adapt to a continuously growing number of pupils with a broader access to information. Software tools focusing on automatic textual complexity scoring need to be primarily adaptive, in the sense that, for a given target audience, the estimated levels of textual complexity measured for specific texts should be adequate and relevant. Tutors are constantly facing a difficult task of assessing textual complexity and of presenting adequate materials to learners. Additionally, time plays an important role and tutors should focus on interaction and on passing on their knowledge, not on frequent changes of materials due to students' poor results. In this case, software tools can be perceived as calibrators providing support by pinpointing out the peaks and the lows of the textual materials.

The first dimension encompassing quantitative factors (e.g., word frequency or sentence length) is the most straightforward and computationally feasible. Qualitative factors are centered on the levels of meaning, structure, language conventionality, clarity and knowledge demands. Eventually, reader and task orientation can be considered the most difficult dimension to be taken into account as it considers learners' knowledge, motivation and interests. Anyway, the pool of considered factors represents a key element for properly assessing complexity, as there is no single factor that can be used to have a reliable prediction.

Based on the previous dimensions, the vast majority of existing systems rely solely on simple quantitative factors. As reference, there is a wide variety of solutions used in various English education programs [1] — e.g., *Lexile* (MetaMetrics) [2], *ATOS* (Renaissance Learning) [3], Degrees of Reading Power: *DRP Analyzer* (Questar Assessment, Inc.) [4], the *Pearson Reading Maturity Metric* (Pearson Knowledge Technologies) [5], *Coh-Metrix* (University of Memphis) [6] —, but only one system focusing on French language assessment — *Dmesure* (Catholic University of Louvain-La-Neuve) [7]. The latter addresses lexical and syntactic complexity factors applied on French as a foreign language (FFL) texts in which sets of metrics were aggregated using different classifiers (e.g., multinomial logistic regression, decision trees, bagging and boosting, support vector machines) [8] in order to automatically generate language exercises fit to learners' level.

Overall, providing valid measurements for automatic textual complexity scoring should come not as a tutor replacement, but as support to an evolving education system that reacts to external factors and that should constantly adapt to a developing society. Thus, this paper is centered on finding factors and patterns that precisely measure textual complexity in tight correlation to learner comprehension and textual cohesion derived from the cohesion graph used as underlying discourse structure [9, 10].

III. The Proposed French Textual Complexity Model

A. General Presentation

Automatic textual complexity evaluation is relatively assessed against existing factors and known measurements, but becomes pointless without a validated and fine-tuned complexity model pre-trained on specific corpora. Starting from a textual complexity model previously applied on English texts [10, 11], we propose a multi-dimensional analysis of adapted factors for French language, integrating classic surface metrics derived from automatic essay grading techniques, morphology and syntax factors [11], as well as semantics and discourse factors [10, 12]. In the end, subsets of factors are aggregated through Support Vector Machines [13, 14] which are proven to be the most efficient method [7, 15].

Firstly, the *surface* category is centered on the individual analysis elements (words, phrases, paragraphs) and makes use of simple statistics. The textual analysis factors from this category are based on Page's grading technique for automated scoring [16], simple readability formulas [17, 18], fluency (e.g., number of words, number of commas), structure complexity (e.g., number of sentences and of paragraphs), diction (e.g., word length, average number of syllables per word, or of words per sentence) and word/character entropy [11]. A specific subcategory is focused on w*ord complexity* which consists of several different factors: syllable count, 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 the newspaper "*Le Monde*" comprising approximately 24M words), the distance in the hypernym tree or the word polysemy count from WOLF [19].

Afterwards, the *syntactic and morphological* category changes the focus to the parsing tree by considering the maximum depth and the size of the underlying parsing structure [20], as well as an adaptation of the Balanced CAF (Complexity, Accuracy, Fluency) measure [21] for French. The word information category is centered on specific pronoun forms (singular, plural, first, second or third form) identified as cue phrases. In the end, the *semantics* and *discourse analysis* category is based on the cohesion graph and its underlying cohesive links [10, 12, 22], lexical chains and discourse connectives that are considered the central elements in terms of discourse representation [22]. From a computational perspective, cohesion is determined as a mixture of semantic similarity measures [22] applied on lexicalized ontologies [23], cosine similarity applied on Latent Semantic Analysis (LSA) [24] vector space, and Jensen-Shannon dissimilarity [25] computed on Latent Dirichlet Allocation (LDA) [26] topic distributions.

B. Model Validation

The previously described textual complexity model comprising of 54 individual factors grouped within the predefined categories was trained on 200 documents extracted from French primary school manuals that were manually categorized into five complexity classes, used to categorize and encircle the content within the documents. The five classes are directly mapped onto five primary grade classes for pupils ranging from 6 to 11 years old, or CP – "*Cours préparatoire*" (1st grade), CE1, CE2 – "*Cours élementaire*" (2nd and 3rd grades), CM1 and CM2 – "*Cours moyen*" (4th and 5th grades) of the French national education system. Each document has been manually validated, ranked and scored by expert linguists, whereas the selected training corpora proved sufficient in terms of volume for training a reliable SVM based classifier.

In the end, *3*-fold cross validation [27] was applied for determining precision or exact agreement (*EA*), and adjacent agreement (*AA*) [7], the percent to which the SVM was close at predicting the correct classification [10, 12] (see Table 1). As specific parameters for the SVM, an RBF kernel with degree 3 was selected and a Grid Search method [28] was enforced to increase the effectiveness of the SVM through the parameter selection process for the Gaussian kernel.

TABLE I. TEXTUAL COMPLEXITY DIMENSIONS AND EXACT/ADJACENT AGREEMENT (EA/AA) SCORES

Depth of metrics	**Category of textual complexity factors**	**Avg. *EA***	**Avg. *AA***
Surface Analysis	Fluency factors	.67	.93
	Structure complexity factors	.73	.80
	Diction factors	.33	.73
	Entropy factors (words, characters)	.53	.87
	Word complexity factors	.60	.87
Morphology & Syntax	Balanced CAF (Complexity, Accuracy, Fluency)	.67	.93
	Specific parts of speech factors	.60	.93
	Word information factors	.60	.80
	Parsing tree complexity factors	.73	.93
Semantics & Discourse	Discourse factors	.53	.80
	Connectives factors	.40	.67
	Lexical chains	.60	.80

The integration of the previous factors from all textual complexity dimensions proved that SVMs are reliable predictors and that our model is viable (*EA* = .733 and *AA* = .933). With regards to the results reported for English language (*EA* = .779 and *AA* = .997) [22], we found current evaluations more than encouraging as the first assessments considered an alignment of 1,000 documents with automatic scores, more precisely Degree of Reading Power (DRP) scores divided into six complexity classes [29], whereas the experiments performed on French used manually-classified learning materials.

IV. PREDICTING COMPREHENSION BASED ON TEXTUAL COMPLEXITY FACTORS

Starting from the previously trained textual complexity model, a specific corpus comprising 16 documents (see Table 2) was used in order to determine the alignment of each complexity factor to human comprehension scores. Therefore, in parallel to the automated scoring performed with our system *ReaderBench* [12, 22], a study was performed in French primary schools in which pupils' comprehension was assessed through post-tests related to the presented learning materials. The experiments were performed with pupils from three adjacent classes (CE2, CM1, CM2), but the materials were specifically designed for the lowest one (CE2) in order to ensure comprehension even at this level. As expected, the comprehension scores in the [1, 20] range increased with the pupil's level, but significant differences among the used documents could be observed.

TABLE II. DOCUMENT STATISTICS

ID	**Document**	**CE2**		**CM1**		**CM2**	
		No. pupils	***Avg. score [1; 20]***	***No. pupils***	***Avg. score [1; 20]***	***No. pupils***	***Avg. score [1; 20]***
1	L'avaleur de nuages	138	8.69	136	9.97	131	11.40

ID	Document	CE2		CM1		CM2	
2	Bibamboulor			135	9.63	127	11.68
3	Boudha	393	8.54	410	9.30	377	10.07
4	Brésil	396	4.20	413	5.18	377	5.96
5	Cordophones	398	5.76	414	6.45	378	7.37
6	Destins croisés	383	6.70	406	7.95	376	9.10
7	Et la terre	245	6.16	213	6.50	213	7.10
8	Henri Vallée					378	8.23
9	La puce	141	6.24				
10	Le petit garçon	138	7.92	134	9.08	126	10.25
11	Le roi Crapaud	140	8.05				
12	Lion					381	8.90
13	Matilda	133	9.26	132	10.67	127	12.36
14	Le monde d'en haut	397	6.19	412	7.19	382	8.08
15	Sept Corbeaux	406	7.12	402	7.80		
16	Tom et Léa	135	6.85	136	8.17	127	9.68

Based on the previous corpora we were able to identify the factors and categories that have the highest correlations (both positive and negative) to pupils' comprehension scores (see Tables 3 and 4). High individual correlations for most factors in Table 3 are expectable, ranging from the simplest one (e.g., surface factors) to morphology and all the way to semantics and discourse. The latter prove that cohesion plays an important role in comprehension, as McNamara, et al. [29] also showed. Moreover, these analogies demonstrate the consistency and adaptability of our approach that supports both English and French languages.

TABLE III. CORRELATIONS BETWEEN AGGREGATED FRENCH TEXTUAL COMPLEXITY FACTORS PER CATEGORY AND PUPIL COMPREHENSION SCORES

Aggregated category	CE2	CM1	CM2
Average surface	.76	.78	.75
Average CAF measures	.50	.67	.60
Average morphology	.69	.79	.78
Average discourse (all significant factors)	.60	.59	.67

As a general tendency, an increasing correlation with higher educational levels was observed, showing nevertheless how individual or aggregated textual complexity factors strongly influence comprehension at lower levels. Moreover, the purpose of aggregating factors (see Table 3) was to observe the evolution of correlations by considering multiple factors condensed through the use of the average function applied on their normalized values (or their inverse, in case of negatively correlating factors).

The results highlighted a major improvement with regards to the individual factors from Table 4 and emphasize the premise that textual complexity cannot be reflected in a single factor, but through multiple categories, each capturing specific facets of the discourse. In the end, the correlations between the textual complexity factors and the pupils' understanding enabled us to determine the potential impact of each factor and to estimate the reliability for predicting the overall textual complexity level of the reading material.

TABLE IV. CORRELATIONS BETWEEN INDIVIDUAL FRENCH TEXTUAL COMPLEXITY FACTORS AND PUPIL COMPREHENSION SCORES.

Category	Factors for assessing textual complexity	CE2	CM1	CM2
Structure complexity	Normalized number of words	.38	.56	.59
	Normalized number of sentences	.65	.70	.65

Category	Factors for assessing textual complexity	CE2	CM1	CM2
Diction	Average word length	-.48	-.62	-.52
	Average word in sentence	-.64	-.64	-.51
	Average sentence length	-.47	-.60	-.48
	Standard deviation for words (letters)	-.60	-.72	-.67
Entropy	Word entropy	.36	.46	.46
Word complexity	Mean word distance in hypernym tree	-.38	-.36	-.42
Balanced CAF	Lexical Sophistication	-.66	-.70	-.63
	Syntactic Sophistication	-.51	-.66	-.57
	Balanced CAF	-.37	-.49	-.47
Morphology & Syntax	Average number of nouns	-.56	-.64	-.59
Specific parts of speech	Average number of pronouns	.55	.58	.52
	Average number of adjectives	-.54	-.60	-.53
	Average number of prepositions	-.49	-.63	-.55
Word information	First Person Singular Pronouns Count	.44	.44	.54
	Second Person Singular Pronouns Count	.57	.52	.61
	Third Person Singular Pronouns Count	.50	.50	.56
Parsing tree complexity	Average tree depth	-.66	-.73	-.65
	Average tree size	-.51	-.45	-.33
Discourse	Average sentence-block cohesion	.52	.55	.59
	Average intra-block cohesion	.60	.57	.58
Connectives	Temporal relation	.60	.72	.80

V. Refining the Textual Complexity Predictions

Besides the previous experiments, an additional refinement had to be implemented as the pre-trained textual complexity model (presented in detail in Section 3) applied on the 16 selected documents from Table 2 returned only the third class (CE2) as automatic prediction. This result is consistent with the initial manual selection of the learning materials, but further investigations were required in order to better grasp discrepancies among the documents. Therefore, incremental runs of the *k*-Means++ clustering algorithm [8] were performed to highlight affinities or differences among documents.

Table 5 highlights the resulted clusters (the presented identifiers are consistent with the ones assigned to each document in Table 2), while the corresponding clustroids (centers of each cluster pertaining to the document space) are marked between brackets. Table 5 also presents the overall compactness (a measure of intra-cluster cohesion) and isolation (a measure of inter-cluster separation) between clusters that demonstrate a better separation and definition of clusters with an increasing value of *k*, the global number of clusters. Moreover, besides finding similar documents in terms of their structure reflected in textual complexity factors (e.g., similar vocabulary in some cases or comparable syntactic structures at sentence level in others), clustering also revealed specificities of some texts (e.g., "*Cordophone*" – id 5, an explanative text, not a narrative one, is separated).

TABLE V. Clusters obtained after applying k-Means++

k	Clusters	Compactness	Isolation
2	C1: 4; 5; (11); 15 C2: 1; 2; 3; 6; 7; 8; 9; 10; 12; 13; (14); 16	1.15	.06
3	C1: (5) C2: 1; 2; 3; 6; 7; 8; 9; 12; (14); 16	1.24	.18

k	Clusters	Compactness	Isolation
	C3: 4; 10; 11; 13; (15)		
4	C1: 4; (15) C2: (5) C3: 1; 2; 3; 6; 7; 10; 11; (13); 16 C4: 8; 9; (12); 14	1.34	.37
5	C1: 1; 2; 3; 6; 7; 10; 11; (13); 16 C2: (5) C3: 8; (9); 14 C4: (12) C5: 4; (15)	1.46	.61

Additionally, we've obtained compatible results by applying an agglomerative clustering algorithm that considers the average group distance. After detecting a drop-off in terms of inter-group similarity between textual complexity traits, Tables 6 presents the obtained clusters.

TABLE VI. CLUSTERS OBTAINED AFTER APPLYING AN AGGLOMERATIVE CLUSTERING ALGORITHM

Cluster	Assigned document IDs per group
C1	1, 2, 3, 6, 7, 8, 10, 11, 13, 14, 16
C2	4, 15
C3	5
C4	9, 12

VI. CONCLUSIONS AND FUTURE WORK

As main conclusions for this undergone research on French textual complexity, we argue that the obtained results and measurements are not as precise as those performed on English, but are still very relevant for an initial approach towards multi-lingual analyses. The proposed textual complexity model was adequate in predicting the corresponding difficulty level and the complexity estimations for all the 16 analyzed documents (all CE2) were an additional confirmation of its validity. Moreover, surface, morphology and discourse factors centered on cohesion correlated well with pupils' comprehension, while the overall results improved when several factors were aggregated.

In a nutshell, *ReaderBench* offers flexibility, adaptability to multiple educational scenarios and extensibility reflected in the ease of adding additional textual complexity factors, thus making the textual assessment more reliable and accurate. To conclude, this research comes as a new direction to the authors' previous work, resulting in an accurate and validated interpretation of comprehension reflected in French textual complexity factors.

ACKNOWLEDGMENT

This research was partially supported by an *Agence Nationale de la Recherche* (DEVCOMP) grant, by the 264207 ERRIC–Empowering Romanian Research on Intelligent Information Technologies/FP7-REGPOT-2010-1 project and by the Sectoral Operational Programme Human Resources Development 2007-2013 of the Ministry of European Funds through the Financial Agreements POSDRU/159/1.5/S/ 134398 and 132397. We would also like to thank Hatier Publishing for providing the French corpora, Aurélie Nardy and Françoise Toffa who helped us to gather experimental data, as well as the teachers and pupils who participated in our experiments.

REFERENCES

[1] J. Nelson, C. Perfetti, D. Liben, and M. Liben, "Measures of text difficulty: Testing their predictive value for grade levels and student performance," Council of Chief State School Officers, Washington, DC2012.

[2] A. J. Stenner, "Measuring reading comprehension with the Lexile Framework," MetaMetrics, Inc., Durham, NC1996.

[3] G. D. Borman and N. M. Dowling, "Testing the Reading Renaissance Program Theory: A Multilevel Analysis of Student and Classroom Effects on Reading Achievement," University of Wisconsin-Madison, Madison, WI2004.

[4] B. L. Koslin, S. M. Zeno, S. Koslin, H. Wainer, and S. H. Ivens, *The DRP: An effectiveness measure in reading*. New York, NY: College Entrance Examination Board, 1987.

[5] T. K. Landauer, K. Kireyev, and C. Panaccione, "Word maturity: A new metric for word knowledge," *Scientific Studies of Reading,* vol. 15, pp. 92–108, 2011.

[6] D. S. McNamara, M. M. Louwerse, P. M. McCarthy, and A. C. Graesser, "Coh-Metrix: Capturing linguistic features of cohesion," *Discourse Processes,* vol. 47, pp. 292–330, 2010.

[7] T. François and E. Miltsakaki, "Do NLP and machine learning improve traditional readability formulas?," in *First Workshop on Predicting and improving text readability for target reader populations (PITR2012)*, Montreal, Canada, 2012, pp. 49–57.

[8] X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda*, et al.*, "Top 10 algorithms in data mining," *Knowledge and Information Systems,* vol. 14, pp. 1–37, 2008.

[9] S. Trausan-Matu, M. Dascalu, and P. Dessus, "Textual complexity and discourse structure in Computer-Supported Collaborative Learning," in *ITS 2012*, Chania, Grece, 2012, pp. 352–357.

[10] M. Dascalu, P. Dessus, S. Trausan-Matu, M. Bianco, and A. Nardy, "ReaderBench, an environment for analyzing text complexity and reading strategies," in *AIED 2013*, Memphis, USA, 2013, pp. 379–388.

[11] M. Dascalu, S. Trausan-Matu, and P. Dessus, "Towards an integrated approach for evaluating textual complexity for learning purposes," in *ICWL 2012*, Sinaia, Romania, 2012, pp. 268–278.

[12] M. Dascalu, P. Dessus, M. Bianco, S. Trausan-Matu, and A. Nardy, "Mining texts, learners productions and strategies with ReaderBench," in *Educational Data Mining: Applications and Trends*, A. Peña-Ayala, Ed., ed Switzerland: Springer, 2014, pp. 335–377.

[13] C. Cortes and V. N. Vapnik, "Support-Vector Networks," *Machine Learning,* vol. 20, pp. 273–297, 1995.

[14] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Transactions on Neural Networks,* vol. 13, pp. 415–425, 2002.

[15] S. E. Petersen and M. Ostendorf, "A machine learning approach to reading level assessment," *Computer Speech and Language,* vol. 23, pp. 89–106, 2009.

[16] E. Page, "The imminence of grading essays by computer," *Phi Delta Kappan,* vol. 47, pp. 238–243, 1966.

[17] J. P. Kincaid, R. P. Fishburne, R. L. Rogers, and B. S. Chissom, *Derivation of New Readability Formulas: (automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel*, Naval Air Station Memphis, 1975.

[18] R. Flesch, "A new readability yardstick," *Journal of Applied Psychology,* vol. 32, pp. 221–233, 1948.

[19] B. Sagot, "WordNet Libre du Francais (WOLF)," ed. Paris: INRIA, 2008.

[20] V. Gervasi and V. Ambriola, "Quantitative assessment of textual complexity," in *Complexity in language and text*, M. L. Barbaresi, Ed., ed Pisa, Italy: Plus, 2002, pp. 197–228.

[21] M. Schulze, "Measuring textual complexity in student writing," in *American Association of Applied Linguistics (AAAL 2010)*, Atlanta, GA, 2010, pp. 590–619.

[22] M. Dascalu, *Analyzing discourse and text complexity for learning and collaborating, Studies in Computational Intelligence* vol. 534. Switzerland: Springer, 2014.

[23] A. Budanitsky and G. Hirst, "Evaluating WordNet-based Measures of Lexical Semantic Relatedness," *Computational Linguistics,* vol. 32, pp. 13–47, 2006.

[24] T. K. Landauer and S. T. Dumais, "A solution to Plato's problem: the Latent Semantic Analysis theory of acquisition, induction and representation of knowledge," *Psychological Review,* vol. 104, pp. 211–240, 1997.

[25] C. D. Manning and H. Schütze, *Foundations of statistical Natural Language Processing*. Cambridge, MA: MIT Press, 1999.

[26] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research,* vol. 3, pp. 993–1022, 2003.

[27] S. Geisser, *Predictive inference: an introduction*. New York, NY: Chapman and Hall, 1993.

[28] J. Bergstra and Y. Bengio, "Random Search for Hyper-Parameter Optimization," *The Journal of Machine Learning Research,* vol. 13, pp. 281–305, 2012.

[29] D. S. McNamara, A. C. Graesser, and M. M. Louwerse, "Sources of text difficulty: Across the ages and genres," in *Assessing reading in the 21st century*, J. P. Sabatini and E. Albro, Eds., ed Lanham, MD: R&L Education, in press, p. 27.