Towards an Integrated Approach for Evaluating Textual Complexity for Learning Purposes

Mihai Dascălu, Ștefan Trăușan-Matu, Philippe Dessus

Towards an Integrated Approach for Evaluating Textual Complexity for Learning Purposes

Mihai Dascălu[1,2], Ștefan Trăușan-Matu[1], Philippe Dessus[2]

[1] Politehnica University of Bucharest, Computer Science Department, Romania
{mihai.dascalu, stefan.trausan}@cs.pub.ro
[2] LSE, UPMF Grenoble-2 & IUFM–UJF Grenoble-1, France
philippe.dessus@upmf-grenoble.fr

Abstract. Understanding a text in order to learn is subject to modeling and is partly dependent to the complexity of the read text. We transpose the evaluation process of textual complexity into measurable factors, identify linearly independent variables and combine multiple perspectives to obtain a holistic approach, addressing lexical, syntactic and semantic levels of textual analysis. Also, the proposed evaluation model combines statistical factors and traditional readability metrics with information theory, specific information retrieval techniques, probabilistic parsers, Latent Semantic Analysis and Support Vector Machines for best-matching all components of the analysis. First results show a promising overall precision (>50%) and near precision (>85%).

Keywords: textual complexity, Latent Semantic Analysis, readability, Support Vector Machines

1 Introduction

Measuring textual complexity is in general a difficult task because the measure itself is relative to the reader and high differences in the perception for a given lecture can arise due to prior knowledge in the specific domain, familiarity with the language or to personal motivation and interest. Readability ease and comprehension are related to the readers' education, cognitive capabilities and background experiences. Therefore a cognitive model of the reader must be taken into consideration and the measured complexity should be adapted to this model. Additionally, software implementing such functionalities should be 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. Fortunately, the target texts processed in this paper were accessible from a syntactical and vocabulary viewpoint by primary school pupils, and the required level of knowledge to grasp the read story was also on the same range.

Assessing the textual complexity of the material given to pupils is a common task that teachers encounter very often. However, this assessment cannot be performed without taking into account the actual pupils' reading proficiency and this point makes it time-consuming. Moreover, the impact of textual complexity on instruction

E. Popescu, R. Klamma, H. Leung, & M. Specht (Eds.), Advances in Web-based Learning (ICWL 2012) (pp. 268–278). New York: Springer, LNCS 7558.

and learning is important: pupils read faster and learn better if textual materials are not too complex nor too easy. A web-based system can help teachers select and calibrate the appropriate texts presented to students and also help the latter attain their own learning objectives in selecting not too simple nor too difficult reading materials.

Our aim is to design and implement a system that automatically gives a measure of the complexity of texts read by children by studying the relations between human vs. computer measures of this textual complexity, similar to some extent to [9]. Although there are numerous applications that give estimates regarding textual complexity, they often do not rely on cognitive models of human reading or use only simple lexical or syntactic factors. An example of a complex system covering multiple perspectives of discourse analysis is *Coh-Metrix 2* [1] that automatically calculates the coherence of texts and determines how text elements and constituents are connected for specific types of cohesion. Besides lexical and syntactic factors, POS tagging and Latent Semantic Analysis (LSA), used also within *Coh-Metrix 2*, we provide an integrated approach through Support Vector Machines (SVMs) that covers in a novel manner all previous dimensions, tightly connected to the implemented discourse model. Therefore, our model is capable of automatically adapting its categories based on the training corpus, enabling us to personalize even further the classification process and the assigned weights to each evaluation factor.

Due to the fact that textual complexity cannot be determined by enforcing a single factor of evaluation, we propose a multitude of factors, categorized in a multilayered pyramid [3], from the simplest to the more complex ones, that combined provide relevant information to the tutor regarding the actual "hardness" of a text. The first and simplest factors are at a surface (*word*) level and include readability metrics, utterance entropy at stem level and proxies extracted or derived from Page's [4] automatic essay grading technique. At the *syntactic* level, structural complexity is estimated from the parsing tree in terms of max depth and size of the parsing structure [6]. Moreover, normalized values of occurrences of specific parts of speech (mostly prepositions) provide additional information at this level. *Semantics* is addressed through topics that are determined by combining *Tf-Idf* with cosine similarity between the utterance vector and that of the entire documents. The textual complexity at this level is expressed as a weighted mean of the difficulty of each topic, estimated in computations as the number of syllables of each word. Moreover, textual complexity is evaluated in terms of semantic cohesion determined upon social networks analysis metrics applied at macroscopic level upon the utterance graph [5]. Discourse markers, co-references, rhetorical schemas and argumentation structures are also considered, but are not included in current experiments.

By considering the disparate facets of textual complexity and by proposing possible automatic methods of evaluation, the resulted measurement vectors provide tutors valuable information regarding the hardness of presented texts. The remainder of this paper details the various metrics. All the measures used to evaluate textual complexity are then unified into a single result by using SVMs in the effort to best align automatic results to the classes manually assigned by teachers. The paper ends with conclusions and future improvements

2 Surface Analysis

Surface analysis addresses lexical and syntactic levels and consists of measures computed to determine factors like fluency, complexity, readability taking into account lexical and syntactic elements (e.g., words, commas, phrase length, periods).

2.1 Readability

Traditional readability formulas [7] are simple methods for evaluating a text's reading ease based on simple statistical factors as sentence length or word length. Although criticized by discourse analysts [8] as being weak indicators of comprehensibility and for not closely aligning with the cognitive processes involved in text comprehension, their simple mechanical evaluation makes them appealing for integration in our model. Moreover, by considering the fact that reading speed, retention and reading persistence are greatly influenced by the complexity of terms and overall reading volume, readability formulas can provide a viable approximation of the complexity of a given text, considering that prior knowledge, personal skills and traits (e.g. intelligence), interest and motivation are at an adequate level or of a similar level for all individuals of the target audience. In addition, the domain of texts, itself, must be similar because subjectivity increases dramatically when addressing cross-domain evaluation of textual complexity. Starting from simple lexical indicators, numerous mathematical formulas were developed to tackle the issue of readability. The following three measures can be considered the most famous:

1. The **Flesch Reading Ease** Readability Formula scores texts on a 100 point scale, providing a simple approach to assess the grade-level of pupils: the higher the score, the easier the text is to read (not necessarily to be understood).
2. The **Gunning's Fog Index** (or **FOG**) Readability Formula estimates the number of years of formal education a reader of average intelligence needs to understand the given text on the first reading. Although considering that words with more than two syllables are complex can be seen as a drawback, we chose this estimation due to its high precision and simplicity.
3. The **Flesch Grade Level** Readability Formula uses the same factors as the first readability metric and rates texts based on U.S. grade school levels.

2.2 Trins and proxes

Page's initial study was centered on the idea that computers can be used to automatically evaluate and grade student essays using only statistically and easily detectable attributes, as effective as human teachers [4, 5]. In order to perform a statistical analysis, Page correlated two concepts: *proxes* (computer approximations of interest) with human *trins* (intrinsic variables - human measures used for evaluation) for better quantifying an essay's complexity. A correlation of 0.71 proved that computer programs could predict grades quite reliably, similar to the inter-human correlation. Starting for Page's metrics of automatically grading essays and taking

into consideration Slotnick's method of grouping proxes based on their intrinsic values, the following categories were used for estimating complexity (see Table 1).

Table 1. Surface analysis proxes

Quality	Proxes
Fluency	Normalized number of commas
	Normalized number of words
	Average number of words per sentence
Diction	Average word length
	Average number of syllables per word
	Percent of hard words (extracted from FOG Formula)
Structure	Normalized number of blocks (paragraphs)
	Average block (paragraph) size
	Normalized number of utterances (sentences)
	Average utterance (sentence) length

Normalization is inspired from Data-Mining and our results improved by applying the logarithmic function on some of the previous factors in order to smooth results, while comparing documents of different size. All the above proxes determine the average consistency of sentences and model adequately their complexity at surface level in terms of the analyzed lexical items.

2.3 Complexity, Accuracy and Fluency

Complexity, accuracy, and fluency (CAF) measures of texts have been used in linguistic development and in second language acquisition (SLA) research. *Complexity* captures the characteristic of a learner's language, reflected in a wider range of vocabulary and grammatical constructions, as well as communicative functions and genres [2]. *Accuracy* highlights a text's conformation to our experience with other texts, while *fluency*, in oral communication, captures the actual volume of text produced in a certain amount of time. Similar to the previous factors, these measures play an important role in automated essay scoring and textual complexity analysis. Schulze [2] considered that selected complexity measures should be divided into two main facets of textual complexity: sophistication (richness) and diversity (variability of forms). The defined measures depend on six units of analysis: letter (l), word form (w), bigram (b – groups of two words) and period unit (p), word form types (t) and unique bigrams (u). Additionally, textual complexity is devised into *lexical* and *syntactic* complexity:

Lexical Complexity:

- *Diversity* is measured using Carroll's Adjusted Token Type Ratio (Eq. 1) [2]:

$$v_1 = \frac{t}{\sqrt{2w}}, with\ \frac{1}{\sqrt{2w}} \leq v_1 \leq \sqrt{\frac{w}{2}} \quad (1)$$

- *Sophistication* estimates the complexity of a word's form in terms of average number of characters (Eq. 2) [2]:

$$v_2 = \frac{l}{w}, with\ 1 \leq v_2 \leq l \quad (2)$$

Syntactic Complexity:

- *Diversity* captures syntactic variety at the smallest possible unit of two consecutive word forms. Therefore Token Type Ratio is also used, but at a bigram level (Eq. 3) [2]:

$$v_3 = \frac{u}{\sqrt{2b}}, with\ \frac{1}{\sqrt{2b}} \leq v_3 \leq \sqrt{\frac{b}{2}} \quad (3)$$

- *Sophistication* is expressed in terms of mean number of words per period unit length and it's intuitive justification is that longer clauses are, in general, more complex than short ones (Eq. 4) [2]:

$$v_4 = \frac{w}{p}, with\ 1 \leq v_4 \leq p \quad (4)$$

All the previous measures can be integrated into a unique measure of textual complexity at lexical and syntactic levels. Following this idea, these factors were balanced by computing a rectilinear distance (Raw Complexity - RC) as if the learner had to cover the distance along each of these dimensions [2]. Therefore, in order to reach a higher level of textual complexity, the learner needs to improve on all four dimensions (Eq. 5) [2]:

$$RC = \left|v_1 - \frac{1}{\sqrt{2w}}\right| + |v_2 - 1| + \left|v_3 - \frac{1}{\sqrt{2b}}\right| + |v_4 - 1| \quad (5)$$

Afterwards, CAF is computed as a balanced complexity by subtracting the range of the four complexity measures (max - min) from the raw complexity measure (Eq. 6):

$$CAF = RC - (\max(v_1, v_2, v_3, v_4) - \min(v_1, v_2, v_3, v_4)) \quad (6)$$

The ground argument for this adjustment is that if one measure increases too much, it will always be to the detriment of another. Therefore, the measure of raw complexity is decreased by a large amount if the four vector measures vary widely and by a small amount if they are very similar. Moreover, the defined measure captures lexical and syntactic complexity evenly, provides two measures for sophistication and two measures for diversity and, in the end, compensates for large variations of the four vector measures.

2.4 Entropy

Entropy, derived from Information Theory, models the text in an ergodic manner and provides relevant insight regarding textual complexity at character and word level by ensuring diversity among the elements of the analysis. The presumption of induced complexity pursues the following hypothesis: a more complex text contains more information and requires more memory and more time for the reader to process. Therefore, disorder modeled through entropy is reflected in the diversity of characters and of word stems used, within our implemented model, as analysis elements. The use

of stems instead of the actual concepts is argued by their better expression of the root form of related concepts, more relevant when addressing diversity at syntactic level.

3 Part of Speech Tagging and Parsing Tree Structure

Starting from different linguistic categories of lexical items, our aim is to convert morphological information regarding the words and the sentence structure into relevant metrics to be assessed in order to better comprehend textual complexity. In this context, parsing and part of speech (POS) tagging play an important role in the morphological analysis of texts, in terms of textual complexity, by providing two possible vectors of evaluation: the normalized frequency of each part of speech and the structural factors derived from the parsing tree. Although the most common parts of speech used in discourse analysis are nouns and verbs, our focus was aimed at prepositions, adjectives and adverbs that dictate a more elaborate and complex structure of the text. Moreover, pronouns, that through their use indicate the presence of co-references, also indicate a more inter-twined and complex structure of the discourse. On the other hand, multiple factors can be derived from analyzing the structure of the parsing tree: an increased number of leafs, a greater overall size of the tree and a higher maximum depth indicate a more complex structure, therefore an increased textual complexity. Our implemented system uses the log-linear Part of Speech Tagger publicly available from Stanford University [11] with the ”bidirectional-wsj-0-18.tagger” package as English tagger, therefore ensuring an accuracy of 97.18% on WSJ 19-21 concepts and of 89.30% on unknown words.

4 Semantic Analysis – Coherence through LSA

Coherence is a central issue to text comprehension and comprehension is strongly related to textual complexity. In order to understand a text, the reader must first create a well-connected representation of the information withheld. This connected representation is based on linking related pieces of textual information that occur throughout the text. Coherence is determined and maintained through the links identified within the utterance graph [5]. Our implemented discourse model characterizes the degree of semantic relatedness between different segments through means of Information Retrieval (Term Frequency – Inverse Document Frequency) reflected in word repetitions [3] and of LSA, capable of measuring similarity between discourse segments through concepts of the vector space [1].

The power of computing semantic similarity with LSA comes from analyzing a large corpus from which LSA builds relationships between a set of documents and terms contained within. The main assumption is that semantically related words will co-appear throughout documents of the corpus and will be indirectly linked by concepts after the Single Value Decomposition specific to LSA.

Coherence is determined using the utterance graph modeled as a Directed Acyclic Graph (DAG) of sentences, ordered sequentially through the ongoing discourse. The first thing that needs to be addressed is semantic cohesion between two sentences

which is seen as the degree of inter-connection among them [3]. This similarity is computed by combining *repetitions* of stems and *Jaccard similarity* as measures of lexical cohesion, with semantic similarity computed by means of LSA (Eq. 7 and 8):

$$coh(u,v) = |repetions| \times \frac{|stems\ in\ common\ u,v|}{|stems\ in\ u\ or\ v|} \times cos(vector(u), vector(v)) \quad (7)$$

$$vector_k(u) = \sum_i (1 + |word_i \in u|) \times \left(\frac{|D|}{|word_i \in D|}\right) \times U_k[word_i] \quad (8)$$

where $U_k[word_i]$ is the vector of $word_i$ in the U_k matrix from LSA.

After determining all possible connections between the sentences of a text through the previous equations, the utterance graph is built by selecting the links that have their corresponding values above a threshold (in our experiments, the best empirical value was the mean value of all viable cohesion values determined for any possible link within our initial text). Overall, coherence is evaluated at a macroscopic level as the average value for all links in the constructed utterance graph. Co-references and other specific discourse analysis methods (e.g., argumentation acts) will be used to further refine the previous DAG.

5 Support Vector Machines

All the measures previously defined capture in some degree different properties of the analyzed text (readability, fluency, accuracy, language diversity, coherence, etc.) and therefore can be viewed as attributes that describe the text. In order to use these attributes to estimate the complexity of the text, we have used a classifier that accepts as inputs text attributes and outputs the minimum grade level required by a reader to comprehend the specified text. Therefore multiple Support Vector Machine (SVM) classifiers are used to achieve the desired result. A SVM is typically a binary linear classifier that maps the input texts seen as *d*-dimensional vectors to a higher dimensional space (hyperspace) in which, hopefully, these vectors are linearly separable by a hyperplane.

Due to the fact that binary classifiers can map objects only into two classes, our multiclass problem can be solved using multiple SVM, each classifying a category of texts with different predefined classes of complexity. A one-versus-all approach implementing the winner-takes-all strategy is used to deal with the problem of multiple SVM kernel returning 1 for a specific text (the classifier with the highest output function assigns the class).

LIBSVM [10] was used to ease the implementation of the classifier. An RBF with degree 3 was selected and a *Grid Search method* was enforced to increase the effectiveness of the SVM through the parameter selection process. Exponentially growing sequences for C and γ were used ($C \in \{2^{-5}, 2^{-3}, \dots, 2^{13}, 2^{15}\}$, $\gamma \in \{2^{-15}, 2^{-13}, \dots, 2^{1}, 2^{3}\}$) and each combination of parameter choices was checked using a predefined testing corpus. In the end, the parameters with the best obtained precision were picked.

6 Preliminary validation results

The preliminary validation experiments were run on 249 reading assignments given by teachers to pupils ranging from 1st grade to the 5th grade (29 – C1, 41 – C1, 66 – C3, 71 – C4, 42 – C5). All the previous measurements have been automatically extracted for each these texts and were later used as inputs for the SVM. The split between the training corpus and the testing one was manually performed by assigning 166 texts as training set and the remaining for evaluation. Additional to data normalization that was previously performed, all factors were linearly scaled to the [-1; 1] range. Two types of measures were used to evaluate the performance of our model: *Precision* (P), as the percent to which the SVM predicted the correct classification for the test input, and *Near Precision* (NP), as the percent to which the SVM was close in predicting the correct classification (i.e. answered n^{th} grade instead of $(n+1)^{th}$ grade). NP was introduced due to the subjectivity of the evaluation of the corresponding grade level and also due to the fact that the complexity of the finishing n^{th} year text may be very close to one from the beginning of the $(n+1)^{th}$ year.

Table 2 presents in detail the optimum C and γ parameters determined for the SVMs via Grid Search method, precision and near precision for all classes, average and weighted average of both precision and near precision.

Table 2. Precision (P%) and Near Precision (NP%) for all evaluation factors

Factor	C	γ	C1 P / NP	C2 P / NP	C3 P / NP	C4 P / NP	C5 P / NP	Avg. P/NP	Weig. Avg. P/NP
Readability Flesch	128	0.125	90 / 90	14 / 64	27 / 91	78 / 100	14 / 86	*44.6 / 86.2*	*44.5 / 88.2*
Readability FOG	32	0.5	80 / 90	29 / 64	32 / 100	83 / 91	7 / 79	*46.2 / 84.8*	*47.4 / 86.8*
Readability Kincaid	32768	2	90 / 90	7 / 57	45 / 95	57 / 100	14 / 57	*42.6 / 79.8*	*42.2 / 83.2*
Normalized no. of commas	0.5	2	0 / 80	29 / 57	14 / 100	83 / 91	0 / 71	*25.2 / 79.8*	*32.2 / 83.1*
Avg. sentence length	32	2	80 / 80	7 / 93	55 / 82	61 / 100	36 / 50	*47.8 / 81*	*48.5 / 83.3*
Normalized no. of words	2048	0.008	50 / 100	79 / 86	0 / 91	83 / 83	0 / 71	*42.4 / 86.2*	*42.5 / 85.6*
Avg. word length	2048	0.008	80 / 80	43 / 93	50 / 86	61 / 96	29 / 50	*52.6 / 81*	*51.9 / 83.2*
Avg. no. of words/ sentence	128	0.031	100/ 100	0 / 21	5 / 95	100/ 100	0 / 93	*41 / 81.8*	*41.5 / 84.5*
Avg. no. of syllables/word	2	0.125	60 / 60	0 / 71	41 / 100	87 / 100	0 / 79	*37.6 / 82*	*42.7 / 87*
Percent of complex words	0.5	0.5	70 / 70	0 / 86	41 / 91	83 / 100	0 / 79	*38.8 / 85.2*	*42.7 / 88.3*
Word Entropy	0.125	8	80 / 100	57 / 100	23 / 95	65 / 83	0 / 64	*45 / 88.4*	*43.3 / 87.8*
Character Entropy	8	0.5	70 / 70	7 / 14	18 / 100	87 / 96	0 / 86	*36.4 / 73.2*	*38.9 / 78.8*

Lexical diversity	2	0.125	50 / 100	79 / 100	27 / 100	74 / 87	0 / 79	*46 / 93.2*	*47.1 / 92.8*
Lexical sophistication	8	8	40 / 60	14 / 64	45 / 86	65 / 100	14 / 50	*35.6 / 72*	*39.8 / 77.3*
Syntactic diversity	0.5	2	40 / 40	0 / 64	45 / 100	91 / 100	0 / 93	*35.2 / 79.4*	*42.5 / 85.9*
Syntactic sophistication	32768	8	80 / 90	0 / 93	50 / 82	61 / 96	21 / 50	*42.4 / 82.2*	*43.5 / 83.3*
Balanced CAF	512	0.5	70 / 100	79 / 100	50 / 100	57 / 91	0 / 71	*51.2 / 92.4*	*50.7 / 92.5*
Average number of nouns	2	0.125	0 / 0	0 / 79	68 / 100	74 / 100	0 / 71	*28.4 / 70*	*39.1 / 80*
Average no. of pronouns	128	2	40 / 60	14 / 71	50 / 91	39 / 96	7 / 64	*30 / 76.4*	*32.5 / 81*
Average no. of verbs	128	0.5	80 / 80	0 / 36	5 / 77	96 / 96	0 / 93	*36.2 / 76.4*	*38 / 78.7*
Average no. of adverbs	2	8	40 / 60	14 / 64	27 / 95	61 / 91	29 / 79	*34.2 / 77.8*	*36.4 / 82*
Average no. of adjectives	32	0.125	30 / 30	0 / 93	55 / 100	70 / 100	0 / 71	*31 / 78.8*	*38 / 85.8*
Average no. of prepositions	128	8	30 / 70	50 / 79	36 / 91	65 / 96	14 / 64	*39 / 80*	*42.2 / 83.4*
Average POS tree depth	2048	8	80 / 90	14 / 93	55 / 82	65 / 96	14 / 50	*45.6 / 82.2*	*47.1 / 83.3*
Average POS tree size	512	2	80 / 80	0 / 93	45 / 77	57 / 100	36 / 64	*43.6 / 82.8*	*43.6 / 84.3*
Avg. doc. cohesion	2	2	40 / 50	0 / 64	32 / 95	74 / 100	0 / 71	*29.2 / 76*	*34.2 / 82*
Comb. lexical-syntactic	**512**	**0.002**	**100/ 100**	**43 / 93**	**41 / 82**	**48 / 100**	**36 / 86**	***53.6 / 92.2***	***49.4 / 91.7***
Combined POS	**2**	**0.125**	**80 / 80**	**14 / 93**	**55 / 82**	**52 / 100**	**29 / 71**	***46 / 85.2***	***45.9 / 86.9***
Combined semantic	**2048**	**0.008**	**70 / 70**	**7 / 86**	**45 / 91**	**70 / 96**	**0 / 79**	***38.4 / 84.4***	***41.2 / 87.1***
Combined all	**32768**	**0.008**	**100/ 100**	**57 / 100**	**55 / 86**	**35 / 91**	**50 / 64**	***59.4 / 88.2***	***54 / 87.7***

Taking into consideration the previous experiment, we have obtained a promising overall precision (>50%) and an excellent near precision (>85%), taking into consideration the difficulty and the subjectivity of the task at hand. Moreover, as expected, the effectives of our method increased by combining multiple factors and, although simple in nature, readability formulas, average sentence length, average word length and balanced CAF provided the best alternatives at lexical and syntactic level. Also, character entropy proved to be a lesser relevant factor than word entropy that reflects vocabulary diversity. In term of parts of speech tagging, prepositions had the highest correlation of all types of parts of speech, whereas depth and size of the parsing tree provided a good insight of textual complexity. In contrast, semantic factors had lower scores because the evaluation process at this level is based on the links between sentences; but texts used in educational environments are characterized by a low variance between different classes in terms of the computed semantic cohesion function (a text belonging to C3 does not necessarily have to be more

cohesive than a text from C1). Also, the most difficult class to identify was the last one because, in general, there are relatively small changes in comparison to the previous class and it's difficult to highlight the differences.

7 Conclusions and Future Development

By combining different factors as readability, lexical and syntactic complexity, accuracy and fluency metrics, part of speech evaluation and characteristics of the parsing tree with LSA embedded within the discourse model, we obtained an elaborate and multi-perspective model capable of providing an overall balanced measure for textual complexity. In order to fine-tune even further the results, additional investigations and experiments are to be conducted to find the best parameters for the SVM, making predictions more reliable, whereas additional coherence measurement techniques will be included for enriching the semantic perspective of our analysis. Our research can be easily extended to any online materials and can be considered a cornerstone in developing an adaptive system for proposing personalized reading materials. This adaptive system would also assess the relation between textual complexity and pupils' understanding, as measured by online questionnaires.

References

1. McNamara, D.S., Louwerse, M.M., McCarthy, P.M., Graesser, A.C.: Coh-Metrix: Capturing linguistic features of cohesion. Discourse Processes, 47(4):292–330 (2010)
2. Schulze, M.: Measuring textual complexity in student writing. Proc. AAAL 2010, Atlanta (2010)
3. Trausan-Matu, S., Dascalu, M., Dessus, P.: Considering textual complexity and comprehension in Computer-Supported Collaborative Learning. In: Cerri, S. A., Clancey, W. J, Papadourakis, G., Panourgia, K. (eds.) ITS 2012, pp. 352–357. Springer, Berlin, (2012)
4. Page, E.: The imminence of grading essays by computer. Phi Delta Kappan, 47:238–243 (1966)
5. Trausan-Matu, S., Rebedea, T.: A polyphonic model and system for inter-animation analysis in chat conversations with multiple participants. In: Gelbukh, A. (ed.) CICLing 2010, pp. 354–363. Springer, New York (2010)
6. Gervasi, V., Ambriola, V.: Quantitative assessment of textual complexity. In: Merlini Barbaresi, L. (ed.): Complexity in language and text, pp. 197–228. Plus, Pisa (2002)
7. Brown, J. D.: An EFL readability index. JALT Journal, 20, 7–36 (1998)
8. Davison, A., Kantor, R.: On the failure of readability formulas to define readable texts: A case study from adaptations. Reading Research Quarterly, 17, 187–209 (1982)
9. Petersen, S.E., Ostendorf, M.: A machine learning approach to reading level assessment. Computer Speech and Language, 23:89–106, (2009)
10. Chang, C.-C., Lin, C.-J.: LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2(3):1–27 (2011)
11. Klein, D., Manning, C.D.: Accurate Unlexicalized Parsing. Proceedings of the 41st Meeting of the Association for Computational Linguistics, pp. 423–430 (2003)