# Paper vs. slides

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

The 10th International Scientific Conference
eLearning and software for Education
Bucharest, April 24-25, 2014
10.12753/2066-026X-14-000

# PAPER VS. SLIDES: DO THEY HAVE SIMILAR TEXTUAL TRAITS?

Lucia Larise Stavarache, Mihai Dascalu, Stefan Trausan-Matu
*Computer Science Department, University Politehnica of Bucharest, 313 Splaiul Independentei, Bucharest Romania*
*larise.stavarache@ro.ibm.com, mihai.dascalu@cs.pub.ro, stefan.trausan@cs.pub.ro*

Philippe Dessus
*Laboratoire des Sciences de l'Education, Université Grenoble Alpes, 38040 Grenoble CEDEX 9, France*
*philippe.dessus@upmf-grenoble.fr*

***Abstract**: As different learning methods and educational scenarios highly influence the corresponding outcomes, our aim is to highlight quantifiable discrepancies in terms of the complexity gap between presentations and hand-outs versus full documents (i.e. academic papers), expressed as concrete factors that directly influence the perceived difficulty. Although there are multiple dependant variables that affect the interpretation of a given topic (e.g., order of presented materials, difference in personal styles if materials originate from multiple authors), we limit the scope of our analysis to solely identifying textual traits that can be automatically extracted from conference papers and their corresponding slide presentations. Our approach represents the starting point for adapting MOOC (Massive Open Online Courses) materials to their target audience in terms of: textual complexity, learner comprehension and content reusability. Therefore, this study performs a detailed comparison using a wide variety of textual complexity metrics as background, ranging from surface, syntactic, morphological and semantic factors in order to grasp the specificities of each material. In other words, our goal consists of providing a set of required metrics for adapting learning materials in order to best suit the underlying educational activities. Preliminary results reflect a strong correlation between the two alternative presentation forms of the same material (papers and corresponding slides) and a similar degree of perceived textual complexity, emphasizing the strong and unitary writing characteristics of the author.*

**Keywords:** *textual complexity model; MOOC; comparative traits; discourse analysis.*

## I. INTRODUCTION

A central dimension in the development of Intelligent Tutoring Systems is focused on personalization. One way of doing it is improving the way materials are presented to learners in order to maximize comprehension and the learning outcomes. There are many alternate forms of presenting information to the audience, but a central issue for each author/presenter should consist of finding the best-suited method and form for presenting and adapting his or her materials to the public. Coming from an academic environment, we often pose the question: "What is the best way of presenting information to students and how can we predict likely gaps of understanding?" Therefore, we conducted a preliminary study to find in automatic manner possible objective answers for this question and key differences between the two most frequently used forms (textual materials from academic articles and their corresponding slides). In other words, we focused on the analysis of conference papers versus presentation slides, emphasising on the consistency of the textual complexity level among the analysed sets as we expect that both forms converge towards the same complexity level, without losing too much information. Moreover, the confirmation of the feasibility of our approach represents a fundament for our future studies regarding the textual complexity analysis of MOOCs, since they contain papers and slides as the most frequent learning materials.

As a general overview, our study targets to find boundaries of how different forms of presenting information can be linked, analysed, used and rated. The examples given above represent just the main threads of focus for this article; in terms of a detailed analysis, the goals are much more daring and wider, ranging up to an automated system able to measure the correlations between slides and their source materials from a qualitative point of view. From an academic perspective, the possibility to have such comparisons would add value to tutor-student communication and would facilitate the development of slides best aligned to the audience and designed to enhance comprehension. Moreover, this research path could improve the future development of quality MOOC platforms in terms of content personalization. In addition, this study considers that building a consistent, coherent and cohesive environment should be the actual target for developing viable eLearning platforms and, in order to achieve this, it is compulsory to define and assess qualitative boundaries while encouraging the use of automatic assessment tools based on natural language techniques.

As structure, the study performs a textual complexity comparison based on a multitude of surface, syntactic, morphological and semantic factors later described in detail. Of particular interest is that textual materials and the corresponding slides are strongly connected and all together address similar topics within the same domain. The following section presents the structure and the validation for the textual complexity model used in this study. The third section is an overview of the analysed corpus, describing the selection criteria for the materials and the results of the performed comparison. Moreover, it also details key factors that emerged from our experiments and their underlying interpretations. The final section is dedicated to conclusions and future work, aiming to highlight the key traits observed for the conducted paper vs. slides comparison and presents some key opportunities in terms of future research directions.

## II. THE TEXTUAL COMPLEXITY MODEL

### 2.1 Overview of the integrated model

Textual complexity analysis can be considered among the most subjective and difficult tasks for tutors, researchers or automated scoring tools. In order to address the usual problem of aligning materials and information to their target audience, we strive to identify relevant metrics and rules for assessing textual complexity. Let us take the example of a tutor – his role consists of finding the best alternative for presenting information to students in order to maximize the understanding of the materials. Moreover, multiple approaches concur in terms of the assessment of textual complexity, each employing different metrics and scoring methods [1]. Additionally there are many factors that influence the perception on the complexity degree like: user experience, user knowhow on the topic, the language used for the materials (native speakers *vs* non-native speakers), motivation or information structure [2].

Starting from these presumptions, the purpose of this study is to analyze from an objective perspective pairs of conference papers and their corresponding slides. The computational metrics used to measure textual complexity are distributed in the following categories: surface, morphology, syntax and aggregated. The aggregation is performed using Support Vector Machines (SVM) [3; 4]. Each category is further described in this paper, while precision results expressed in terms of exact and adjacent agreement [5; 6] between the manual classification and the automatic prediction for each dimension of computed factors are presented in Table 2.

Firstly, the *surface* category orbits around individual analysis elements (words, phrases, paragraphs) and employs simple statistics. The textual analysis factors from this category are based on Page's grading technique for automated scoring [7], simple readability formulas [8; 9], 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 [6].

Afterwards, the *syntactic and morphological* category, in opposition to surface analysis, changes the focus from a single component of analysis to the full parsing tree by finding the maximum depth and the size of the parsing structure [10]. Additionally, at this level we analyze the corpora from the perspective of each relevant part of speech: prepositions, adjectives, nous or verbs.

The most promising category is focused on *semantics,* in which lexical chains, entity-density features and co-reference inferences are used for identifying the referential relationships between terms, lexicon variety and cohesion [11; 12]. The evaluation of semantic similarity through cohesion plays an important role, mainly due to the fact that our study is centered on measuring the connectivity for each component of the set (paper *vs*. slides) and on the identification of key differences.

From a different perspective, w*ord complexity* analysis has been approached through the combination 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, the distance in the hypernym tree or the word polysemy count from WordNet [13].

### 2.2 Validation of our textual complexity model

In order to train our complexity model, we have opted to automatically extract English texts from Touchstone Applied Science Associates, Inc. (http://lsa.colorado.edu/spaces.html), using its Degree of Reading Power (DRP) score [14], considering six classes of complexity [15] of equal frequency (see Table 1).

Table 1. Ranges of the DRP scores used to define the six textual complexity classes [after 15].

| Complexity Class | Grade Range | DRP Minimum | DRP Maximum |
|---|---|---|---|
| 1 | K-1 | 35.38 | 45.99 |
| 2 | 2-3 | 46.02 | 51.00 |
| 3 | 4-5 | 51.00 | 56.00 |
| 4 | 6-8 | 56.00 | 61.00 |
| 5 | 9-10 | 61.00 | 64.00 |
| 6 | 11-CCR | 64.00 | 85.80 |

Based on the initial corpus, 3,000 documents (500 per complexity class) were extracted and later on processed by ReaderBench, our integrated analysis system [12; 16]. The large number of documents was needed in order to ensure a reliable and complete model and to reflect key points that could be useful for scoring purposes. Performance was measured through $k$-fold cross validation [17] and is reflected through exact agreement – *EA* and adjacent agreement – *AA* [5; 6] for presenting the correctness of the SVM's prediction. The measurements presented in Table 2 are taken as reference for this study and later on used within the comparison experiments. As it can be easily observed, most dimensions are good estimators of textual complexity, whereas the complete model has an extremely high precision score which arguments its relevance and adequacy.

Table 2. Textual complexity dimensions for $N$ =3.000 documents.

| Depth of metrics | Factors for evaluation | Avg. EA | Avg. AA |
|---|---|---|---|
| Surface Analysis | Readability Factors | .80 | .97 |
| | Fluency Factors | .36 | .55 |
| | Structure Complexity Factors | .81 | .97 |
| | Diction Factors | .86 | .94 |
| | Entropy Factors | .36 | .64 |
| Morphology & Syntax | Balanced CAF Factors | .86 | 1 |
| | Part of Speech Complexity Factors | .78 | .94 |
| | Parsing Tree Complexity Factors | .50 | .81 |
| Semantics | Named Entity Complexity Factors | .64 | .89 |
| | Co-reference Complexity Factors | .56 | .86 |
| | Lexical Chains Factors | .59 | .78 |
| | Discourse Factors | .69 | .92 |
| | Connectives Factors | .39 | .53 |
| | Word Information Factors | .47 | .72 |
| | Word Complexity Factors | .75 | .83 |
| *All Factors Combined* | | *.94* | *1* |

## III. COMPARISON BETWEEN PAPERS AND CORRESPONDING PRESENTATION SLIDES

### 3.1 Overview of the analysed corpus

The selected corpus for our study focused on respecting a validity model that would ensure the processing of a proper content with regards to the purpose of this study. The key requirements for selecting the corpus have been: consistency, integrity, cohesion and coherence. Finding full text coherent materials is not hard, but finding adequate corresponding slides proved to be a difficult task due to different presentation styles. In most cases, we found slides consisting only of pictures that are irrelevant for this analysis or slides with little information, elliptical formulations within bullet items and impossible to use in order to achieve an accurate comparison. The purpose of our paper is not to compare all kinds of slides with their underlying materials, but to define a particular pattern of slides that would help students retain key focus points and understand the purpose of the documents.

In addition, one of the main problems with slides is that sometimes they tend to present a multitude of key features in rather few words, without focusing on the auditorium background and capability of understanding. It happens often that students, who do not attend the courses and do not self-study the full materials, do not perform well at the exam due to their focus on learning solely key information from slides without having solid connections among the central concepts.

In this context, we opted to analyze a preliminary corpus containing four pairs of full papers and their corresponding slides. The sets have been selected from various technical areas, in order to offer variety and a wider analysis (computer science with emphasis on natural language processing and distributed computing, merged with educational sciences). Starting from the previous constraints for selecting the corpus, these sets have been validated against a set of pre-imposed rules in order to reduce unwanted noise, which consisted in: 1) human validations for each set focused on the tight correlation between the paper and corresponding slides in terms of consistency, content and style; 2) manual re-editing of the materials in order to eliminate non-descriptive sections or specific irrelevant elements for our textual analysis (e.g., images, quotes, special characters, formulas, tables and videos), and 3) slides had to cover at least 80% of the topics presented in the papers. All the materials have been afterwards pre-processed into ReaderBench's XML format.

In addition, in trying to have a consistent and coherent corpus, the difficulty of finding valid materials for the analysis significantly increased. Therefore, because the main content available in MOOCs failed in some extent to pass our restrictions, we decided not to include this kind of materials as an initial validation of our approach.

### 3.2 Comparison Results

Starting from the dimensions defined in section 2.1, different textual characteristics can be underlined using individual factors or through a combination of various metrics. For current experiments, the analysis was divided into two main layers. The focus of the first layer consisted in a macrolevel analysis in order to find common properties and trends for both slides and papers. In contrast, the second layer focused on a microanalysis of each individual set, mainly comparing paper versus corresponding slides to find traits and common characteristics.

The main traits noticed at a global level, after computing the textual complexity factors for the previously trained model, have been: 1) a consistent maximum complexity level (6) for all papers and slides, which reflects also the specificity of the domain, and 2) a high correlation factor ($r = .92$) between the average textual complexity measures per papers and per slides, but also among each pair ($r_{average} = .91$). In this context, we can clearly observe a consistency in terms of presentation between the two alternative forms, marking also that the correlation is constant throughout the sets and individual pair of paper-slides materials.

Table 3. Factors with similar results for slides and corresponding papers

| Depth of Metrics | Complexity factor | Average normalized difference |
|---|---|---|
| Surface Analysis | Readability FOG | 8% |
| | Readability Kincaid | 7% |
| | Average number of syllables per word | 3% |
| | Normalized number of sentences | 8% |
| | Normalized number of blocks | 2% |
| | Standard Deviation for words (letters) | 3% |
| | Standard Deviation for words (syllables) | 5% |
| | Word entropy | 3% |
| | Character entropy | 1% |
| Balanced CAF Factors | Lexical Sophistication | 5% |
| | Syntactic Diversity | 4% |
| Named Entity Complexity Factors | Average number of unique entities per sentences | 6% |
| | Percentage of nouns in total entities | 2% |
| Word Complexity Factors | Mean distance between words and corresponding stems | 6% |
| | Mean word polysemy count | 7% |
| | Mean word syllable count | 1% |

At a microscopic level, of particular interest was to highlight specific factors with a very low, insignificant average normalized difference (see Table 3) between the presentation forms, but also factors with a high discrepancy (see Table 4), marking in some extent the key similarities and differentiators between slides and their corresponding papers.

Table 4. Differentiators between slides and corresponding papers

| Depth of Metrics | Complexity factor | Average normalized difference |
|---|---|---|
| Surface Analysis | Average block size | 69% |
| Morphology | Average number of pronouns | 75% |
| | Average number of verbs | 67% |
| | Average number of adverbs | 87% |
| | Average number of prepositions | 69% |
| Named Entity Complexity Factors | Percentage of overlapping nouns per document | 67% |
| Co-reference Complexity Factors | Total number of co-reference chains per document | 80% |
| | Average co-reference chain span | 78% |
| | Number of co-reference chains with a big span | 83% |
| Lexical Chains Factors | Maximum span of lexical chains | 67% |
| | Number of lexical chains with more than 5 concepts | 77% |
| Discourse Factors | Average block score | 72% |
| | Overall document score | 67% |
| Connectives Factors | Causal relation | 95% |
| | Temporal relation | 97% |
| | Additive relation | 77% |
| | Logical relation | 74% |
| | Adversative and contrastive relations | 88% |
| | All connectives | 79% |
| Word Information Factors | First Person Plural Pronouns Count | 94% |
| | Second Person Pronouns Count | 100% |
| | Third Person Plural Pronouns Count | 78% |

Upon closer inspection, factors with similar results reflect the usage of a very similar vocabulary, whereas the differentiators highlight major changes in the underlying discourse structure and representation. Based on the two analysis dimensions, we can state that 1) the writing and presentation styles are consistent between slides and their corresponding papers and 2) there are specific factors that significantly reflect the major differences in structure between papers and slides.

## IV. CONCLUSIONS AND FUTURE WORK

Our goal was to observe how the same information can be transposed to different presentation forms and how these are interconnected. Moreover, we focused on identifying specific style markers and how the underlying factors impact the textual analysis. The obtained results confirm that the slides preserve a high connectivity with the sources from which they originated, conserving nevertheless the same complexity class and a high correlation with the textual material's complexity factors. Additionally, we observed that surface factors are consistent throughout papers and their corresponding slides, whereas the major differences are also completely justifiable by considering the differences between the length of a paper and its corresponding slides (a normal proportion would approximate that slides' content represents 10% of the corresponding paper).

Another conclusion that can be extracted shifts the perspective towards authors' writing styles, the structure of information and the evolution of ideas in providing tutors a comprehensive manner of measuring cohesion, coherence and the quality of their materials. From the tutor's perspective, the aim consists of presenting the information as clean, concise and coherent as possible.

As mentioned before, we aim to extend our study to a wider corpus extracted from MOOCs. Additionally, we want to explore beyond finding specific traits between papers and slides: the discovery of author's writing patterns and the identification of specific domain material characteristics from the presented information.

## Reference Text and Citations


[1] Nelson, J., Perfetti, C., Liben, D., & Liben, M. 2012. *Measures of text difficulty: Testing their predictive value for grade levels and student performance*, Editor Ed.^Eds., Council of Chief State School Officers. Washington, DC

[2] National Governors Association Center for Best Practices, Council of Chief State School Officers. 2010. *Common Core State Standards*, Editor Ed.^Eds., Council of Chief State School Officers. Washington D.C.

[3] Cortes, C., & Vapnik, V.N., 1995. Support-Vector Networks. *Machine Learning. 20*(3). Pag 273–297

[4] Hsu, C.-W., & Lin, C.-J., 2002. A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks. 13*(2). Pag 415–425

[5] François, T., & Miltsakaki, E., 2012. Do NLP and machine learning improve traditional readability formulas? In *First Workshop on Predicting and improving text readability for target reader populations*. ACL. Pag 49–57

[6] Dascalu, M., Trausan-Matu, S., & Dessus, P., 2012. Towards an integrated approach for evaluating textual complexity for learning purposes. In E. Popescu, R. Klamma et all. (Eds.), *ICWL 2012*. Springer. Pag 268–278

[7] Page, E., 1966. The imminence of grading essays by computer. *Phi Delta Kappan. 47*. Pag 238–243

[8] Kincaid, J.P., Fishburne, R.P., Rogers, R.L., & Chissom, B.S., 1975. *Derivation of New Readability Formulas for Navy Enlisted Personnel*, Chief of Naval Technical Training, Naval Air Station Memphis

[9] Flesch, R., 1948. A new readability yardstick. *Journal of Applied Psychology. 32*(3). Pag 221–233

[10] Gervasi, V., & Ambriola, V., 2002. Quantitative assessment of textual complexity In M. L. Barbaresi (Ed.), *Complexity in language and text*, Plus. Pisa, Italy. Pag 197–228

[11] Dascalu, M., Dessus, P., Trausan-Matu, S., Bianco, M., & Nardy, A., 2013. ReaderBench, an Environment for Analyzing Text Complexity and Reading Strategies. In H. C. Lane, K. Yacef, J. Mostow & P. Pavlik (Eds.), *16th Int. Conf. on Artificial Intelligence in Education (AIED 2013)*. Springer. Pag 379–388

[12] Dascalu, M., Dessus, P., Bianco, M., Trausan-Matu, S., & Nardy, A., 2014. Mining Texts, Learners Productions and Strategies with ReaderBench. In *Educational Data Mining: Applications and Trends*, Springer. Pag 335–377

[13] Miller, G.A., 1995. WordNet: A Lexical Database for English. *Communications of the ACM. 38*(11). Pag 39–41

[14] Carver, R.P., 1985. Measuring Readability using DRP Units. *Journal of Literacy Research. 17*(4). Pag 303–316

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

[16] Dascalu, M., 2014. *Analyzing Discourse and Text Complexity for Learning and Collaborating, Studies in Computational Intelligence*, Springer. Switzerland

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