FACHLITERATUR
EDITION ROMIOSINI
ΕΠΙΣΤΗΜΗ

ICGL12

12th INTERNATIONAL CONFERENC
ON GREEK LINGUISTICS
16 – 19 SEPTEMBER 2015
FREIE UNIVERSITÄT BERLIN, CEMOG

Proceedings of the ICGL12

vol. 1

The International Conference on Greek Linguistics is a biennial meeting on the study and analysis of Greek (Ancient, Medieval and Modern), placing particular emphasis on the later stages of the language.

PROCEEDINGS OF THE ICGL12
ΠΡΑΚΤΙΚΑ ΤΟΥ ICGL12

Thanasis Georgakopoulos, Theodossia-Soula Pavlidou, Miltos Pechlivanos,
Artemis Alexiadou, Jannis Androutsopoulos, Alexis Kalokairinos,
Stavros Skopeteas, Katerina Stathi (Eds.)

PROCEEDINGS OF THE 12TH INTERNATIONAL CONFERENCE ON GREEK LINGUISTICS

ΠΡΑΚΤΙΚΑ ΤΟΥ 12ΟΥ ΣΥΝΕΔΡΙΟΥ ΕΛΛΗΝΙΚΗΣ ΓΛΩΣΣΟΛΟΓΙΑΣ

VOL. 1

EDITION
ROMIOSINI

Vertrieb und Gesamtherstellung: Epubli (www.epubli.de)
Satz und Layout: Rea Papamichail / Center für Digitale Systeme, Freie Universität Berlin
Gesetzt aus Minion Pro
Umschlaggestaltung: Thanasis Georgiou, Yorgos Konstantinou
Umschlagillustration: Yorgos Konstantinou

ISBN 978-3-946142-34-8
Printed in Germany

Online-Bibliothek der Edition Romiosini:
www.edition-romiosini.de

Στη μνήμη του Gaberell Drachman (†10.9.2014)
και της Αγγελικής Μαλικούτη-Drachman (†4.5.2015)
για την τεράστια προσφορά τους στην ελληνική γλωσσολογία
και την αγάπη τους για την ελληνική γλώσσα

ΣΗΜΕΙΩΜΑ ΕΚΔΟΤΩΝ

Το 12ο Διεθνές Συνέδριο Ελληνικής Γλωσσολογίας (International Conference on Greek Linguistics/ICGL12) πραγματοποιήθηκε στο Κέντρο Νέου Ελληνισμού του Ελεύθερου Πανεπιστημίου του Βερολίνου (Centrum Modernes Griechenland, Freie Universität Berlin) στις 16-19 Σεπτεμβρίου 2015 με τη συμμετοχή περίπου τετρακοσίων συνέδρων απ᾽ όλον τον κόσμο.

Την Επιστημονική Επιτροπή του ICGL12 στελέχωσαν οι Θανάσης Γεωργακόπουλος, Θεοδοσία-Σούλα Παυλίδου, Μίλτος Πεχλιβάνος, Άρτεμις Αλεξιάδου, Δώρα Αλεξοπούλου, Γιάννης Ανδρουτσόπουλος, Αμαλία Αρβανίτη, Σταύρος Ασημακόπουλος, Αλεξάνδρα Γεωργακοπούλου, Κλεάνθης Γκρώμαν, Σαβίνα Ιατρίδου, Mark Janse, Brian Joseph, Αλέξης Καλοκαιρινός, Ναπολέων Κάτσος, Ευαγγελία Κορδώνη, Αμαλία Μόζερ, Ελένη Μπουτουλούση, Κική Νικηφορίδου, Αγγελική Ράλλη, Άννα Ρούσσου, Αθηνά Σιούπη, Σταύρος Σκοπετέας, Κατερίνα Στάθη, Μελίτα Σταύρου, Αρχόντω Τερζή, Νίνα Τοπιντζή, Ιάνθη Τσιμπλή και Σταυρούλα Τσιπλάκου.

Την Οργανωτική Επιτροπή του ICGL12 στελέχωσαν οι Θανάσης Γεωργακόπουλος, Αλέξης Καλοκαιρινός, Κώστας Κοσμάς, Θεοδοσία-Σούλα Παυλίδου και Μίλτος Πεχλιβάνος.

Οι δύο τόμοι των πρακτικών του συνεδρίου είναι προϊόν της εργασίας της Εκδοτικής Επιτροπής στην οποία συμμετείχαν οι Θανάσης Γεωργακόπουλος, Θεοδοσία-Σούλα Παυλίδου, Μίλτος Πεχλιβάνος, Άρτεμις Αλεξιάδου, Γιάννης Ανδρουτσόπουλος, Αλέξης Καλοκαιρινός, Σταύρος Σκοπετέας και Κατερίνα Στάθη.

Παρότι στο συνέδριο οι ανακοινώσεις είχαν ταξινομηθεί σύμφωνα με θεματικούς άξονες, τα κείμενα των ανακοινώσεων παρατίθενται σε αλφαβητική σειρά, σύμφωνα με το λατινικό αλφάβητο· εξαίρεση αποτελούν οι εναρκτήριες ομιλίες, οι οποίες βρίσκονται στην αρχή του πρώτου τόμου.

Η Οργανωτική Επιτροπή του ICGL12

ΠΕΡΙΕΧΟΜΕΝΑ

2ος Τόμος

FEATURE EXTRACTION AND ANALYSIS IN GREEK L2 TEXTS IN VIEW OF AUTOMATIC LABELING FOR PROFICIENCY LEVELS

Maria Giagkou[1], Giorgos Fragkakis, Dimitris Pappas[1] & Harris Papageorgiou[1]
[1]Institute for Language and Speech Processing / R.C. ATHENA
mgiagkou@ilsp.gr; fragakis@sch.gr; dpappas@ilsp.gr; xaris@ilsp.gr

Περίληψη

Στο άρθρο διερευνάται ένα σύνολο γλωσσικών χαρακτηριστικών κειμένων που απευθύνονται σε μαθητές της Ελληνικής ως Γ2 και εξετάζεται η σχέση των εν λόγω χαρακτηριστικών με το επίπεδο γλωσσομάθειας για το οποίο θεωρούνται κατάλληλα τα κείμενα αυτά. Στόχος είναι να διερευνηθεί ποια χαρακτηριστικά παρουσιάζουν επαρκή διακριτική ικανότητα μεταξύ των επιπέδων, ώστε να αξιοποιηθούν σε μια προσέγγιση αυτόματης κατηγοριοποίησης σε επίπεδα γλωσσομάθειας. Προς αυτό το σκοπό αξιοποιείται ένα σώμα κειμένων που συγκροτήθηκε από εγχειρίδια της Ελληνικής ως Γ2. Τα αποτελέσματα αναδεικνύουν τη σημαντική επίδραση, μεταξύ άλλων, χαρακτηριστικών που ποσοτικοποιούν την περιπλοκότητα των συντακτικών δέντρων εξαρτήσεων, της γενικής πτώσης και των επιθετικών προσδιορισμών.

Keywords: L2 reading, text complexity, linguistic features, proficiency levels, automatic labelling

1. Introduction

The last two decades have seen increasing interest in modelling text difficulty, i.e. readability. Automatic readability estimation systems are intended to assess whether a text retrieved from a large collection, such as a repository or the web, is appropriate for a given group of readers, according to their abilities in L1 or by taking into account the readers' special needs (e.g. learning difficulties). Readability estimation is particularly relevant for second language (L2) learners as well. From the L2 perspective, the aim is to automatically identify or retrieve a text given the proficiency level of the learner or group of learners.

To this end, recent studies attempt to grade L2 texts according to proficiency levels in order to facilitate reading in L2 or as an aid to the selection of assessment material (e.g. Centre for the Greek Language 2013, Tzimokas and Tantos 2014, François and Fairon 2012, Ott and Meurers 2010, Pilán et al. 2014, Vajjala and Meurers 2012). In a similar approach, the development of productive skills in L2 (mainly writing) is investigated in view of an automated evaluation of L2 writing (e.g. Lu 2010, 2011, Vyatkina 2012, Giagkou et al. 2015).

The long tradition of L1 readability assessment, dating back to the early 20th century (see DuBay 2006), has bequeathed readability formulas (e.g., *Flesch Reading Ease Score*, *Flesch-Kincaid Grade Level*, *Fog index, SMOG,* etc.) that assign a difficulty grade or level to a text by relying on surface linguistic features such as sentence and word length, as simple proxies for syntactic complexity and vocabulary burden, respectively. More recently, advances in NLP have boosted readability research. That is, new resources (electronically available texts) and new tools (taggers, parsers, semantic treebanks, etc.) have made it feasible to apply machine learning techniques in large training corpora and to quantify more thorough and linguistically sound text features. Semantic and discourse features are investigated e.g., named entities (Barzilay & Lapata 2008) and lexical cohesion (Pitler & Nenkova 2008). Shallow syntactic complexity indicators, such as average sentence length, are combined with the height of syntactic trees (see also Heilman et al. 2008). Instead of simple proxies of vocabulary burden, N-gram Language Models (LM) are used for predicting the grade level of texts (Callan and Eskenazi 2007, Petersen & Ostendorf 2009, Schwarm and Ostendorf 2005).

In this paper, we present an investigation of linguistic features of texts addressed to learners of Greek as a second language (L2). The goal of this study is to identify the textual properties that indicate the development of reading skills in Greek L2 with the aim of employing these properties as parameters for automatic proficiency level labelling. The set of features investigated in the current study draws on the traditional readability research combined with NLP-enabled features and machine learning techniques for text classification, as this merging was found to result in performance gain (François & Miltsakaki 2012).

The paper is organized as follows: Section 2 provides information on the corpus used and the features identified, selected and computed, in order to form the dataset for the analysis. In Section 3, the analysis applied on the features is presented and the results are analyzed. We conclude with a summary of the main findings and their implications to the directions of future work in view of automatic proficiency level classification for Greek L2.

2. Datasets

2.1. Corpus

For the purposes of this investigation, a Greek L2 text set that is labelled for proficiency levels in an objective and qualified way, and can thus be considered as gold-standard, deemed necessary. Such dataset was retrieved from the Greek L2 textbooks published by the Centre of Intercultural and Migration Studies (E.DIA.M.ME.) and freely available online. These textbooks are addressed to Greek migrants living abroad, from preschoolers (aged 6) to 18 year-olds, learning Greek as a second or foreign language. E.DIA.M.ME. employs five proficiency levels aligned to the Greek educational system grades and to CEFR levels (Council of Europe 2001) as presented in Table 1.

Age	**School grade**	**E.DIA.M.ME. level**	**Language content**	**CEFR level alignment**
6	Preschool	1	Pre-reading, reading	A1
7	1			
8	2			
9	3	2	Speaking and writing consolidation	A2
10	4			
11	5	3	Further practice in speaking and writing	B1
12	6			
13	7	4	Independent writing	B2 & C1
14	8			
15	9			

16	10	5	Greek language and literature	C2
17	11			
18	12			

Table 1 | E.DIA.M.ME. proficiency levels (Damanakis 2004: 76) and their alignment to CEFR levels (E.DIA.M.ME. 2014)

Only prose texts were extracted from the textbooks, while poems, lyrics, exercises, and guidelines to the exercises were excluded. The selected texts belong to different genres (mainly narrative, descriptive, expository, and procedural) and types (letters, announcements, instructions, diary entry, etc.). Dialogues were also included, as they are very frequently used as educational material in L2 textbooks, though the role/name of the speaker was removed.

The final corpus employed in this investigation comprises 753 texts and a total of 112.169 tokens (Table 2). Each individual text inherited the proficiency level assigned to the textbook it was retrieved from, e.g. a text drawn from a textbook labeled as level 5, was considered as addressed to level 5 learners.[1]

Grouped levels	**EDIAMME levels**	**Texts**	**Sentences**	**Tokens**
1 (CEFR A1-A2)	1	24	136	720
	2	295	4.552	33.636
2 (CEFR B1-C1)	3	108	1.263	8.780
	4	147	2.305	19.272
3 (CEFR C2)	5	179	3.356	49.761
Totals:		753	11.612	112.169

Table 2 | Corpus description

[1] It should be noted that this decision imposes a degree of "noise" to the data, as, although a low level textbook is not expected to include a text addressed to higher levels, the reverse is not equally unlikely. E.g. certain texts retrieved from a level 5 textbook can actually address lower level learners.

The texts were automatically annotated for morphological types, syntactic dependencies and phrase structure using the Institute for Language and Speech Processing NLP tools pipeline (Prokopidis et al. 2011, Prokopidis and Papageorgiou 2014).

2.2 Feature selection and computation

The set of features investigated as indices of the proficiency level was selected on the basis of previous research on L1 and L2 readability assessment, as well as on second language acquisition and development. These features capture morphological, syntactic, lexical/semantic, and other attributes of the text that are salient to the target proficiency level discrimination and prediction task.

In total, 303 text features were identified and computed. These fall grossly into the following categories:

a) **Surface features**: word and sentence length (e.g. average word length), number of characters, punctuation marks, numbers, etc.
b) **Lexical/semantic**: lexical density (i.e. content to functional words), lexical variation (e.g. type/token ratio, hapax/dis-legomena), including noun and verb variation measures, text entropy, lexical richness, etc.
c) **Morphological**: frequencies and ratios of the different parts of speech, including their forms, e.g. ratio of passive verbs to verbs, ratio of nouns in the genitive case to nouns, ratio of 1st person personal pronouns to pronouns, etc.
d) **Syntactic**: frequencies and ratios of the different syntactic roles (e.g. subjects to verbs ratio), measures of the dependency trees (e.g. depth and height of syntactic trees), phrase structure (e.g. length of noun, verb and adjectival phrases), subordination and apposition (e.g. average number of coordinating and subordinating conjunctions per sentence), etc.
e) **Discourse-based features**, e.g. use of relative pronouns as an index of the degree of anaphora density, frequency of present and past tenses as indices of temporality and narrativity, etc.

The defined features were computed with a specialized software, the ILSP FeatExt tool, developed in Python. The input of FeatExt is any corpus of Greek texts, automatically annotated for Part of Speech, syntactic dependencies and phrase structure. It calculates the values of raw surface features (frequencies of words, sentences, nouns, verbs, etc.) and computes their standardized values (i.e. meaningful ratios). In order to cater for zero values, MinMaxScaler transformation is applied to all raw features. The output is a table of extracted feature values, preferably in CSV format. Settings can be modified through an optional configuration file to define, among others, the set of features to be computed, the corpus location, or additional feature-relevant data such as a list of words to be counted (e.g. functional words, basic vocabulary for a specific proficiency level or topic, etc.).

3. Analysis and results

In order to investigate the underlying associations of text features with the proficiency level, correlation analysis was applied between all the extracted features and the grouped proficiency levels. Table 3 reports the twenty features that exhibited the highest absolute values of Spearman's rho correlation coefficient in descending order (p<0,05).

Among the best performing features, the average number of noun phrases in the genitive case per sentence was found to exhibit the highest correlation coefficient (rho=0,542). The association of the genitive case with the text's level is also evidenced by the performance of two more features, i.e. the average number of adjectival phrases in the genitive case per sentence (rho=0,473) and the average length of adjectival phrases in the gen. case (rho=0,448). Complementing and looking at these results from a different angle, the influence of phrase structure, especially of the length and relative frequency of nominal phrases is apparent. Out of the 20 best performing features, six are indices of phrase structure (features in ranks 1, 6, 8, 12, 15 and 16 in Table 3). The frequency of use of modifiers, namely of adjectives, also seems to be highly correlated to the proficiency level: the more adjectives used in a text the more likely it is that the text is addressed to higher level learners. This is evidenced by the average number of adjectival phrases and of adjectives per sentence.

Another important finding is highlighted by the performance of features that attempt to quantify syntactic dependencies. These include the width and height of dependency trees (rho=0,495 and 0,486, respectively), as well as the number of leafs and governor nodes (rho=0,490 and 0,485, respectively). Their emergence in the top ranks of Table 3, qualifies them as key predictors of the proficiency level.

	Feature	**Spearman's rho**	**EDIAMME grouped level-pairs**		
			1vs2	**2vs3**	**1vs3**
1	Av. # of Noun Phrases in gen. case per sentence	0,542	■	■	■
2	Av. Width of dependency trees	0,495	■	■	■
3	Av. # of Leafs in dependency trees	0,490		■	■
4	Av. Height of dependency trees	0,486	■	■	■
5	Av. Sentence Length	0,485		■	■
6	Av. # of Adjectival Phrases per sentence	0,485		■	■
7	Av. # of governor nodes in dependency trees	0,485		■	■
8	Av. # of Noun Phrases per sentence	0,480		■	■
9	% of sentences with length>20 words	0,477	■	■	■
10	Av. # of Adjectives per sentence	0,474		■	■
11	Av. Word Length	0,474	■	■	■
12	Av. # of Adjectival Phrases in gen. case per sentence	0,473		■	■
13	% of sentences with length>10 words	0,470		■	■
14	Terminal punctuation to total characters ratio	-0,461		■	■
15	Av. Length of adjectival phrases in gen. case	0,448	■	■	■
16	Av. # of Adjectival Phrases in acc. case per sentence	0,446		■	■
17	% of sentences with length>30 words	0,443		■	■
18	Av. # of Passive Verbs per sentence	0,442	■	■	■
19	Relative pronouns to Pronouns ratio	0,439		■	■
20	Av. # of prepositions per sentence	0,438		■	■

Table 3 | Top-20 features highly correlated with EDIAMME grouped levels and post hoc multiple comparisons between level-pairs

Different aspects of syntactic complexity are also highlighted by the average number of passive verbs and prepositions per sentence. As expected, passive constructions are rarely used in lower levels, while learners encounter them more and more frequently in textbooks as their reading skills develop. The same is true for prepositions, a feature that indicates that higher proficiency level texts employ more complex-compound sentences.

The statistically significant correlation performed by the ratio of relative pronouns to pronouns (rho=0,439), signifies the role of anaphora. As anaphora resolution is considered a linguistically and cognitively demanding task during reading, anaphoric structures are rare in lower levels, but significantly more frequent in upper levels. As a result, the use of relative pronouns can be considered as a successful discriminator of proficiency levels.

The list of the best performing features also includes some more "traditional" indices of text complexity, such as word and sentence length. The average sentence length appears in rank 5 in Table 3 (rho=0,485), while relevant features that quantify sentence length from a different perspective are also present (the percentage of sentences with more than 10, 20 and 30 words). Additionally, the presence of the ratio of terminal punctuation to total characters, should be also interpreted as an inverse to sentence length. Regarding lexical features, it is noticeable that among the various features investigated (lexical diversity, density, etc.), only the average word length is present in the top performers (rho=0,474).

A more thorough investigation of the above features employed one-way ANOVA for means comparison across levels, which resulted in statistically significant main effects for all of the 20 features. Since, however, this type of analysis cannot determine whether the mean values of a feature are statistically different between all possible level pairs, post-hoc multiple comparisons (Bonferroni tests) were also applied. The results are presented in Table 3: statistically different means for each feature are indicated for each level combination separately. These comparisons indicate that all features can successfully discriminate group 3 (i.e. EDIAMME level 5, CEFR C2) from lower levels (both from group 2 and group 1). However, some of the features were not as successful in discriminating group 1 (i.e. EDIAMME levels 1 and 2, CEFR A1, A2) from group 2 (i.e. EDIAMME levels 3, 4, CEFR B1-C1). Poor performers in discriminating levels group 1 from group 2, were all the features relevant to sentence length, with the exception of the proportion of sentences with more than 20 words. This implies that a group 1 text is unlikely to include lengthier sentences, thus imposing a possible threshold for the transition from CEFR A2 to B1 level.

4. Conclusions and discussion

The current investigation highlighted a number of textual features, automatically extracted from a morphologically and syntactically annotated Greek L2 corpus. With the aim of identifying indices of text difficulty that are directly associated with the proficiency level, we employed statistical analysis and put forward the best performing features. These can be regarded as potential predictors of the proficiency level of a previously unseen text in an automatic labelling/classification approach.

The results highlight the influence of syntactic features on the characterization of proficiency level: with the exception of average word length, the rest of the best performing features are directly or indirectly related to syntactic complexity. This finding is in line with previous research where syntax-related features consistently appear in the best-performing prediction models (e.g. Pitler and Nenkova 2008, Schwarm and Ostendorf 2005, Callan and Eskenazi 2007, Kate et al. 2010, Kotani et al. 2008). The frequencies of the genitive case, of adjectives and prepositions were additionally identified as successful discriminators. Surface features used in traditional readability formulas, such as sentence and word length, were found to be significantly correlated to proficiency levels. Similar recent research in Greek has also highlighted the influence of such surface features on proficiency level classification (Tzimokas and Tantos 2014). It is interesting to notice that some of the features put forward by Georgatou (2016) as the most informative, i.e. sentence length, passive verbs and adjectives, are confirmed by the current study as well, thus qualifying them as reliable of indices of Greek texts difficulty level.

When the best performing features were tested for their discriminatory power between all possible level pairs, they proved to be highly discriminative of the upper proficiency level. This finding implies a significant shift in L2 reading skills during the transition from C1 to C2 level, and this shift can successfully be measured by the features investigated herein. On the contrary, the transition from A2 to B1 seems to go in hand with the acquisition of language skills not depicted in the features that emerged from the current analysis.

It is true that the current investigation is subject to limitations imposed by the corpus at hand, which comprised texts drawn from textbooks of a single publisher. As such, the findings may be influenced by the publisher's choices regarding the types and topics of texts, and the linguistic descriptors of proficiency levels the editor has adopted. To cater for this limitation, the work described herein is continued and expanded in order to exploit a larger corpus of Greek L2 texts from different publishers. Proficiency level labelling for this expanded corpus does not rely exclusively on the publisher's labelling. Rather, three independent experts in Greek L2 teaching have judged each text to determine the CEFR proficiency level. The expert's judgements is treated as the dependent variable in a machine learning approach for the automatic labelling of previously unseen texts which has already yielded significant results.

Reading comprehension is a key skill in L2 development, and reading is an integral part of L2 instruction and assessment. In this view, an automated approach to matching L2 learners to texts suitable for their proficiency level is expected to facilitate selection of reading material both for learners and teachers. It is at the same time an anticipated aid in assessment procedures, by providing an objective measurement for the estimation of level-appropriateness of items included in diagnostic, placement or achievement language tests.

References

Barzilay, Regina, and Mirella Lapata. 2008. "Modeling Local Coherence: An Entity-based Approach." *Computational Linguistics* 34(1):1–34.

Centre for the Greek Language. 2013. "Logismiko Anagnosimotitas." Accessed March 1, 2017. http://www.greek-language.gr/certification/readability

Council of Europe 2001. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR)*. www.coe.int/lang-CEFR

Damanakis, Michalis, ed. 2004. *Theoritiko Plaisio kai Programmata Spoudon gia tin Ellinoglossi Ekpaideusi sti Diaspora*. Rethymno: E.DIA.M.ME. http://www.ediamme.edc.uoc.gr/diaspora2/indexphp?id=23,65,0,0,1,0

DuBay, William H. 2006. *The Classic Readability Studies*. Impact Information, Costa Mesa, California.

E.DIA.M.ME. 2014. *Epipeda Glossomatheias kai Ekpaideutiko Yliko.* http://www.ediamme.edc.uoc.gr/ellinoglossi/index.php/el/ekp-yliko-kepa

François, Thomas, and Cédrick Fairon. 2012. "An "AI readability" Formula for French as a Foreign Language." In *Proceedings of the 2012 Joint Con-*

ference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, 466–477. Association for Computational Linguistics.

François, Thomas, and Eleni Miltsakaki. 2012. "Do NLP and Machine Learning Improve Traditional Readability Formulas?" In *Proceedings of the First Workshop on Predicting and improving text readability for target reader populations*, 49–57. Montréal.

Georgatou, Spyridoula. 2016. "Approaching Readability Features in Greek School Books." Master thesis. University of Tübingen.

Giagkou, Maria, Kantzou, Vicky, Stamouli, Spyridoula, and Maria Tzevelekou 2015. "Discriminating CEFR Levels in Greek L2: A Corpus-based Study of Young Learners' Written Narratives." *Bergen Language and Linguistics Studies* 6:153–169.

Heilman, Michael, Kevyn Collins-Thompson, and Maxine Eskenazi. 2008. "An Analysis of Statistical Models and Features for Reading Difficulty Prediction." In *The Third Workshop on Innovative Use of NLP for building Educational Applications. Proceedings of the Workshop*, 71–79. ACL.

Callan, Jamie, and Maxine Eskenazi. 2007. "Combining Lexical and Grammatical Features to Improve Readability Measures for First and Second Language Texts." In *Proceedings of HLT-NAACL'07*, 460–467. Association for Computational Linguistics.

Lu, Xiaofei. 2010. "Automatic Analysis of Syntactic Complexity in Second Language Writing." *International Journal of Corpus Linguistics* 15(4):474–496.

Lu, Xiaofei. 2011. "A Corpus-Based Evaluation of Syntactic Complexity Measures as Indices of College-Level ESL Writers' Language Development." *TESOL Quarterly* 45(1):36–62.

Ott, Niels, and Detmar Meurers. 2010. "Information Retrieval for Education: Making Search Engines Language Aware." *Themes in Science and Technology Education* 3(1–2):9–30.

Petersen, Sarah E., and Mari Ostendorf. 2009. "A Machine Learning Approach to Reading Level Assessment." *Computer Speech and Language* 23(1):89–106.

Pilán, Ildikó, Volodina, Elena, and Richard Johansson. 2014. "Rule-Based and Machine Learning Approaches for Second Language Sentence-Level Readability." In *Proceedings of the Ninth Workshop on Innovative Use of NLP for Building Educational Applications,* 174–184. Association for Computational Linguistics.

Pitler, Emily, and Ani Nenkova. 2008. "Revisiting Readability: A Unified Framework for Predicting Text Quality." In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing,* 186–195. Honolulu: ACL.

Prokopidis, Prokopis, and Harris Papageorgiou. 2014. "Experiments for Dependency Parsing of Greek". In *Proceedings of the First Joint Workshop on Statistical Parsing of Morphologically Rich Languages and Syntactic Analysis of Non-Canonical Languages*, 90–96. Dublin, Ireland.

Prokopidis, Prokopis, Georgantopoulos, Byron, and Harris Papageorgiou. 2011. "A suite of NLP tools for Greek". In *The 10th International Conference of Greek Linguistics,* 373–383. Komotini, Greece.

Schwarm, Sarah E., and Mari Ostendorf. 2005. "Reading Level Assessment Using Support Vector Machines and Statistical Language Models." In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05),* 523–530. Ann Arbor, Michigan.

Tzimokas, Dimitrios, and Sotirios Tantos. 2014. "Logismiko Anagnosimotitas Ellinikon Keimenon: Ena Neo Ergaleio gia ti Didaskalia tis Ellinikis os Ksenis/Deuteris Glossas kai tin Pistopoiisi Ellinomatheias." Paper presented at *Diethnes Synedrio gia ti Didaskalia kai tin Pistopoiisi tis Ellinikis os Ksenis/Deuteris Glossas*. Thessaloniki, October 25. http://speakgreek.gr/el/images/pdf/tzimwkas.pdf.

Vajjala, Sowmya, and Detmar Meurers. 2012. "On Improving the Accuracy of Readability Classification Using Insights from Second Language Acquisition." In *Proceedings of the Seventh Workshop on Building Educational Applications Using NLP,* 163–173. Association for Computational Linguistics.

Vyatkina, Nina. 2012. "The Development of Second Language Writing Complexity in Groups and Individuals: A Longitudinal Learner Corpus Study." *The Modern Language Journal* 96(4):576–598.