A Listenability Measuring Method for an Adaptive Computer-assisted Language Learning and Teaching System

Katsunori Kotani
Kansai Gaidai University / Osaka, Japan
kkotani@kansaigaidai.ac.jp

Shota Ueda
Ryukoku University / Shiga, Japan

Takehiko Yoshimi
Ryukoku University / Shiga, Japan

Hiroaki Nanjo
Ryukoku University / Shiga, Japan

Abstract

In teaching and learning of English as a foreign language, the Internet serves as a source of authentic listening material, enabling learners to practice English in real contexts. An adaptive computer-assisted language learning and teaching system can pick up news clips as authentic materials from the Internet according to learner listening proficiency if it is equipped with a listenability measuring method that takes into both linguistic features of a news clip and the listening proficiency. Therefore, we developed a method for measuring listening proficiency-based listenability. With our method, listenability is measured through multiple regression analysis using both learner and linguistic features as independent variables. Learner features account for learner listening proficiency, and linguistic features explain lexical, syntactic, and phonological complexities of sentences. A cross validation test showed that listenability measured with our method exhibited higher correlation ($r = 0.57$) than listenability measured with other methods using either learner features ($r = 0.43$) or other linguistic features ($r = 0.32$, $r = 0.36$). A comparison of our method with other methods showed a statistically significant difference ($p < 0.003$ after Bonferroni correction). These results suggest the effectiveness of learner and linguistic features for measuring listening proficiency-based listenability.

1 Introduction

Listening practice using authentic materials is necessary for learners of English as a foreign language (EFL) who have little or no chance to use English in their daily life because these materials let them immerse themselves in real-life settings. Since authentic materials are not usually constrained by the ease of listening comprehension or listenability (Chall & Dial 1948), teachers have to select materials according to learner listening proficiency; otherwise, too difficult or easy materials reduce the learning effect or spoil learner motivation (Hubbard 2004, Petrides 2006). An adaptive computer-assisted language learning and teaching system can pick up news clips as authentic materials from the Internet according to the listening proficiency if it is equipped with a listenability measuring method that takes into both linguistic features of a news clip and the listening proficiency. Therefore, we propose an automatic method that statistically measures listenability for EFL learners. This method is useful for learning and teaching English by showing listenability levels of authentic listening materials such as news clips. It also helps to create a computer-based self-learning environment because EFL learners can select appropriate materials with this method.

28th Pacific Asia Conference on Language, Information and Computation pages 387–394

Although listenability of authentic listening materials can be measured with readability measuring methods using lexical, syntactic, and discourse features (Flesch 1950, Graesser et al. (2004), and Shen et al 2013), our listenability measuring method uses phonological features as well as lexical and syntactic features. Phonological feature accounts for listenability in terms of speech rate and phonological modification. The natural speech rate for native speakers reduces listenability for learners because learner processing speed is slow due to the lack of automation of mental language processing. Phonological modification refers to sound change such as the elision observed in the second vowel sound of "chocolate" (Roach 2001). Phonological modification has been reported to increase listenability for native speakers, but reduce it for learners (Henricksen 1984).

In addition to linguistic features such as lexical, syntactic, and phonological, our method also uses learner features, which account for the listening proficiency. Unlike native speakers, the listening proficiency greatly differs among individuals (Saville-Troike 2006). That is, listening material can be appropriate for a learner but not for another. Therefore, it is necessary to measure listenability based not only on linguistic features but also learner features.

2 Relevant Study

Fang (1966) developed a listenability measuring method for native speakers based on a linguistic feature showing the presence of multiple-syllable words. With this method, a sentence including more multiple-syllable words is judged as more difficult. The effect from single-syllable words is suppressed because such words are assumed to be ineffective for listenability.

Unlike Fang (1966), Messerklinger (2006) took into account individual differences of background knowledge in measuring listenability for native speakers. According to Messerklinger (2006), the following features should also be taken into account in measuring listenability: speech rate, length of pause, sentence length, repairing, accent, and intensity.

Similarly to Messerklinger (2006), Kiyokawa (1990) also took into account properties of a listener. What this study focused on was not background knowledge, but the overall proficiency of EFL learners. This method measured listenability for learners at the intermediate level based on Kiyokawa's vocabulary list, which defines words that intermediate-level learners should have learnt. Words not listed in this list were regarded as difficult for intermediate-level learners. In addition, Kiyokawa (1990) used sentence length as another linguistic feature.

Although Kiyokawa's listenability method was developed for intermediate-level learners, what has not been thoroughly examined in the previous studies is the listenability for learners at different proficiency levels. Fang (1966) and Messerklinger (2006) discussed listenability for native speakers of English, assuming that listening proficiency does not differ much among native speakers. However, learners have different proficiencies; thus, individual differences of listening proficiency should be considered. Therefore, we address this remaining problem.

3 Features for Measuring Listenability

3.1 Learner Feature

Learner features must show the listening proficiency. This study uses scores of English language tests for determining the listening proficiency. The English language test used in this study was the Test of English for International Communication (TOEIC) because this test is a major English language test for university learners in the country where the experiment takes place.

Because TOEIC consists of a listening section and reading section, a learner acquires two scores. Our method uses TOEIC listening scores (the range of scores: 5-495) as a learner feature.

3.2 Linguistic Feature

Linguistic features must show the lexical, syntactic, and phonological complexity of a sentence. We used linguistic features, i.e., mean length of words, sentence length, presence of multiple-syllable words, speech rate, difficulty of words, and presence of phonological modification. The linguistic features used with our method, except phonological, which explains the presence of phonological modification, were originally used in the previous studies (Fang 1966, Kiyokawa 1990, Messerklinger 2006).

Type (Description)	Condition for phonological modification
elision (elimination of phonemes)	(i) vowel sound immediately following stressed syllable such as second "o" sound in "chocolate" (ii) consonant followed by similarly articulated sound such as (a) continuous same sound as in "unknown," (b) continuous plosive sound as in "c" sound and "t" sound of "doctor," and (c) plosive sound followed by nasal sound as in "suddenly"
reduction (weakening sound by changing vowel to schwa)	vowel sound in functional words such as personal pronouns, interrogative pronouns, auxiliaries, modals, prepositions, articles, and conjunctions
contraction (combining pair of words)	(i) pair of subject noun with (a) be-auxiliary, (b) have-auxiliary, or (c) modal (ii) pair of interrogative pronoun with (a) be-auxiliary, (b) have-auxiliary, or (c) modal (iii) pair of negative adverb "not" with (a) be-auxiliary, (b) have-auxiliary, or (c) modal
linking (connecting final sound of word with initial sound of following word)	(i) words between word starting with vowel and (a) word ending with "n" sound as in "in an hour" or (b) word ending with "r" sound as in "after all" (ii) word followed by (a) indefinite article, (b) preposition, or (c) conjunction
deduction (elimination of sounds between words)	(i) words sharing same sound between final sound of word and initial sound of following word as in "good day" (ii) words between word ending with plosive sound and word starting with plosive, affricative, fricative, nasal, or lateral sound as in "next chance"

Table 1: Condition for phonological features

The presence of phonological modification is automatically measured as follows. Because phonological modification is supposed to occur under a certain condition, it is measured as the ratio of conditions for phonological modification to the total number of words in a sentence. Table 1 summarizes the type of phonological modification, its description, and condition for phonological modification. These phonological features are extracted with the procedures shown in Table 2.

Type	Feature extraction procedure
elision	a. convert to phonetic symbol b. search conditions (i) and (ii) c. count number of words in sentence d. calculate number of identified conditions per number of words in sentence
reduction	a. parse part of speech (Schmid 1994) b. search condition c. count number of words in sentence d. calculate number of identified conditions per number of words in sentence
contraction	a. count number of apostrophes* b. calculate number of apostrophes per number of words in sentence *Contraction has written form using apostrophe such as "I've."
linking	a. convert to phonetic symbol b. search conditions (i) and (ii) c. count number of words in sentence d. calculate number of identified conditions per number of words in sentence
deduction	a. convert to phonetic symbol b. search conditions (i) and (ii) c. count number of words in sentence d. calculate number of identified conditions per number of words in sentence

Table 2: Extraction procedure for phonological features

4 Training/test Data Collection

4.1 Data Outline

To develop a listenability measuring method with multiple regression analysis, it is necessary to collect training/test data consisting of dependent and independent variables.

Dependent variables are scores for listenability of a sentence. Listenability is scored based on a five-point Likert scale of ease of listening comprehension judged by learners as 1: easy, 2: somewhat easy, 3: average, 4: somewhat difficult, or 5: difficult.

Independent variables consist of learner and linguistic features. As described in Section 3, learner features show the listening proficiency, and linguistic features show the lexical, syntactic, and phonological complexities of a sentence.

4.2 Learners

Ninety university EFL learners (males: 48 and females: 42) took part in the data collection task. They were paid for their participation. The mean age was 21.5 years (standard deviation (S.D.) 2.6). The learners were asked to submit valid TOEIC scores, taken that year or the year before. Learners were equally divided into three groups of TOEIC scores: low score group (below 475), middle score group (from 480 to 725), and high score group (above 730). That is, 30 learners were chosen for each group. TOEIC scores were used as the proficiency benchmark, because the EFL learners were recruited not only for this study but also for another study on measurement of readability (Kotani et al. 2012, 2013). The EFL learners were also confirmed for basic computer literacy such as typing with a keyboard and controlling a mouse because they needed to use a computer in the data collection task.

The mean TOEIC listening score for the 90 learners was 334.8 (S.D. 97.6). Figure 1 shows the distribution of the number of learners for TOEIC listening scores, which follows a double-peaked distribution at scores between 200 and 249 (n = 17) and scores above 450 (n = 16). The distribution was skewed due to the small number of learners below a score of 200. Kolmogorov-Smirnov test showed that the distribution did not follow the normal distribution (K=1.24, p=0.04). An investigation into the effect on measurement error due to the skewed distribution is for future study.

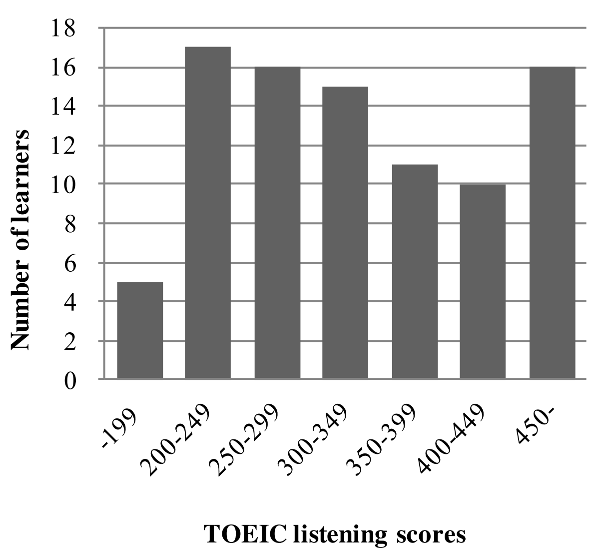

Figure 2: Distribution of TOEIC listening scores

4.3 Materials

The materials used in this study were news clips because they are often used as listening practice materials for university EFL learners. Each news clip included five multiple-choice comprehension questions to let learners work on the listening task as they would in an actual English language test. These questions were made in the format of Nation & Malcher (2007): two true questions to choose a correct description about the article; two false questions to choose an incorrect description about the article; and one content question to choose a correct brief description of the article.

The news clips were chosen from the two types of sections in the Voice of America (VOA) site (http://www.voanews.com): the special section for English learners and the editorial section. News clips in the special section were made for learners, while news clips in the editorial section were made for native speakers of English. The former news clips consisted of short, simple sentences using the 1,500 basic vocabulary of VOA, and avoiding idiomatic expressions. By contrast, the editorial section's news clips were made without any restriction on vocabulary and sentence construction as long as they were appropriate as news clips for native speakers of English. The speech rate of special section's news clips was two-thirds slower than the editorial section's news clips, which were read aloud at a natural speech rate, approximately 250 syllables per minute, according to Robb & Gillon (2007).

	Elision	Reduction	Contraction	Linking	Deduction
Mean	0.13	0.38	0.00	0.04	0.19
S.D.	0.14	0.10	0.02	0.07	0.17
Minimal	0.00	0.13	0.00	0.00	0.00
Maximal	0.63	0.75	0.17	0.40	0.75
Occurrence (n)	63	80	2	32	69

Table 3: Descriptive statistics of linguistic features for phonological modification

The linguistic features for phonological modification in the materials are summarized in Table 3. The features for phonological modification are the ratio of conditions for phonological modification as described in Section 3.2. The mean value is calculated by summing the ratio of conditions for phonological modification, and dividing the sum by the number of sentences (n = 80). Among the 80 sentences, phonological modification was observed in 63 sentences for elision, 80 sentences for reduction, 2 sentences for contraction, 32 sentences for linking, and 69 sentences for deduction.

4.4 Task

Each learner was asked to listen to the four news clips sentence-by-sentence only once, using a headphone. After listening to each sentence, the learner assigned a listenability score for the sentence from the five-point Likert scale. After listening to a news clip, the learner answered five multiple-choice comprehension questions.

Each learner used a data collecting tool, which displayed on a computer screen several icons to move on to the next sentence, and to select a choice from multiple choice items for listenability score and comprehension questions. The data collecting tool also recorded the learner's choices.

The learners were asked to complete a listening task as fast as possible during the allotted time (8 minutes for each news clip), and to stop working either when the task was completed or the experimenter and the data collecting tool alerted them of the end of the allotted time. They were prohibited to use dictionaries or any other reference books. The data collecting tool did not allow learners to return to a sentence for listening again after moving on to another sentence.

4.5 Listenability Score

Although the training/test data should consist of 7,200 instances (90 learners × 80 sentences) for a valid listenability score, 6,804 instances were used in developing our listenability measuring method. 396 instances were regarded as invalid, because no listenability score was recorded. Each instance consisted of a listenability score, a learner feature in terms of a TOEIC listening score, and linguistic features. The mean listenability score was 2.83 (S.D. 1.32).

Figure 2 shows how listenability scores distribute according to the listening proficiency level. Learners were classified into three proficiency levels based on TOEIC listening scores: 34 advanced (score range: 365-495), 40 intermediate (score range: 240-360), and 16 beginner (score range: 130-235). As expected, the distribution of listenability scores followed the proficiency levels. Advanced learners tended to judge listening as easy, intermediate learners tended to judge listening as moderate, beginner learners tended to judge listening as difficult.

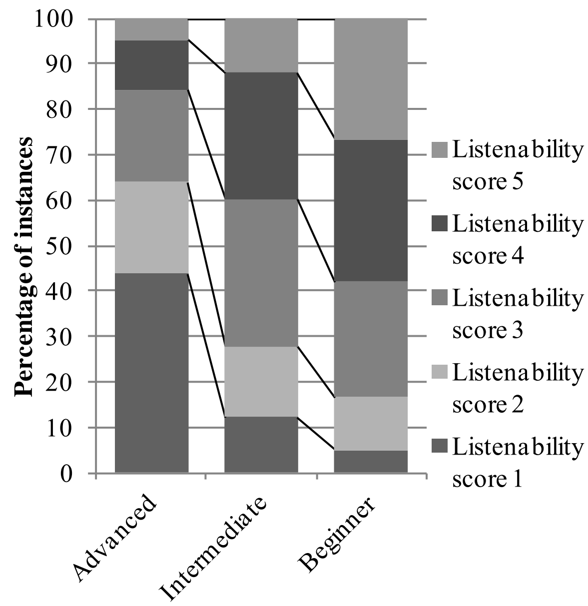

Figure 2: Distribution of listenability scores

5 Experiment

5.1 Development of Our Method

We conducted a multiple regression analysis for developing our method. The independent variables were the learner and linguistic features described in Section 3, which show the listening proficiency and lexical, syntactic, and phonological complexities of a sentence. The dependent variable was listenability scores, as described in Section 4.5.

Before carrying out the multiple regression analysis, the learner and linguistic features were examined with respect to the presence of multiple-collinearity by calculating the variance inflation factor (VIF) (Neter et al. 1996), and a multiple-collinearity of more than 10 was not found ($1.14 < VIF < 8.17$).

The linear combination of learner and linguistic features was significantly related to the listenability scores, $F(11, 6{,}792) = 292.83$, $p < 0.01$. The sample multiple correlation coefficient adjusted for the degrees of freedom was 0.57, indicating that approximately 32% of the variance of the listenability scores in the sample could be accounted for by the linear combination of learner and linguistic features. The standardized partial regression coefficients are summarized in Table 4.

Type	Feature	Standardized partial regression coefficient
learner feature	TOEIC listening score	–0.43**
linguistic feature	mean length of words	–0.08**
	sentence length	0.05
	difficulty of words	0.07*
	presence of multiple-syllable words	0.09**
	speech rate	0.25**
	elision	0.02
	reduction	0.01
	contraction	–0.10**
	linking	0.03**
	deduction	–0.06**

Table 4: Standardized partial regression coefficients
(one asterisk: $p < 0.05$, two asterisks: $p < 0.01$)

5.2 Evaluation of Our Method

Our listenability measuring method was examined in a leave-one-out cross validation test by comparing with sample methods (Method I-III) that were developed by using some of the features in our method. The features used in each method are marked in Table 5. In the cross validation test, the methods were examined n times (n = 6,804) by taking one instance as test data and n -1 instances as training data.

Each method was examined by comparing listenability scores assigned by learners and listenability scores measured with one of the methods. Spearman's correlation coefficients are also summarized in Table 5. The correlation coefficients in Table 5 were statistically significantly different from zero ($p < 0.01$). The difference in correlation coefficients between our method and the other methods was examined using the Meng-Rosenthal-Rubin method (Meng et al. 1992). The results showed that there was a statistically significant difference between our method and the other methods I-III ($p < 0.003$ after Bonferroni correction for three comparisons). Thus, our method was marked with the highest correlation. This result suggests that proficiency-based listenability is affected by both learner and linguistic features.

	Our method	Method I	Method II	Method III
TOEIC listening score	●			●
Mean length of words	●	●		
Sentence length	●	●		
Difficulty of words	●	●		
Presence of multiple syllable words	●	●		
Speech rate	●		●	
Elision	●		●	
Reduction	●		●	
Contraction	●		●	
Linking	●		●	
Deduction	●		●	
Correlation coefficient	0.57**	0.32**	0.36**	0.43**

Table 5: Feature and correlation coefficients
(two asterisks: $p < 0.01$)

Measurement errors from the cross validation test results are plotted in Figure 3. Measurement error was calculated as an absolute value of the difference between a listenability score measured with a method and a listenability score assigned by a learner. Our method had more instances in the ranges of small measurement error (0.0 and 0.1-1.0) than the other methods, as seen in Figure 3.

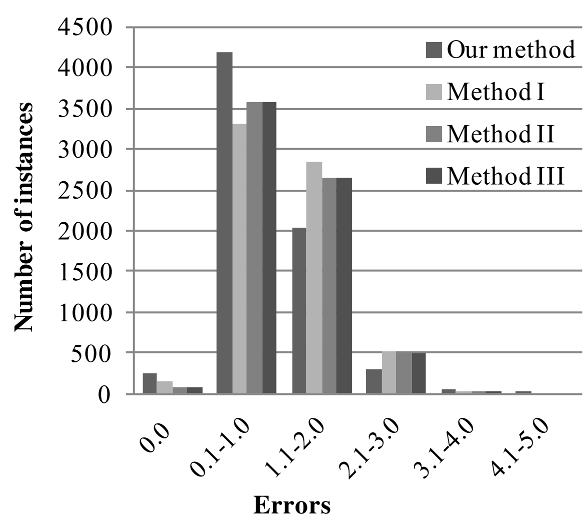

Figure 3: Error distribution

Table 6 summarizes the ratio of the number of instances with error values 0.0-1.0 to the number of instances (6,804). As expected from the results in Table 5 and Figure 3, our method had more instances with error up to 1.0 than other methods. These results also suggest the effectiveness of our method.

	Our method	Method I	Method II	Method III
Ratio	0.65	0.51	0.54	0.54

Table 6: Ratio of error up to 1.0

5.3 Discussion on Our Method

The standardized partial regression coefficients in Table 4 show that listenability mostly depends on the TOEIC listening score. This result suggests that the TOEIC listening score is useful in measuring listenability for learners.

Among the five phonological features for phonological modification, elision and reduction had no statistically significant effect, contrary to our expectation. As Henricksen (1984) suggested, it was expected that phonological modification reduces listenability for learners as seen in the positive effect from linking. However, the presence of a negative effect from contraction and deduction suggests that phonological modification can increase listenability for learners as well as native speakers.

The standardized partial regression coefficients showed the unexpected effect of sentence length. Sentence length is a well known linguistic feature for explaining syntactic complexity of a sentence and has been used for measuring listenability as well as readability. However, sentence length had no statistically significant effect on listenability. An unexpected effect was also observed in the mean length of words. Assuming that longer words convey complex meanings, word length is a primary linguistic feature for measuring readability (Flesch 1950). However, as the negative value of the standardized partial regression coefficient shows, longer words increase listenability. We believe that this divergence between readability and listenability arises from the different recognition styles. Reading requires letter recognition, while listening requires sound recognition (Rayner & Reichle 2010, Vandergrift 2011). Hence, learners may not fail in letter recognition, but fail in sound recognition. However, the learners did not fail in sound recognition probably due to longer words. These results suggest that listenability is not parallel with readability.

6 Conclusion

We proposed a method for automatically measuring listenability for EFL learners. Unlike the previous studies on listenability, our method directly takes into account the listening proficiency as well as linguistic features, which consist of mean length of words, sentence length, presence of multiple-syllable words, speech rate, difficulty of words, and presence of phonological modification (elision, reduction, contraction, linking, and deduction)

In an experiment, our method showed higher correlation between listenability scores assigned by learners and scores measured using other methods, which partially used learner and linguistic features.

With our method, linguistic features for phonological modification were extracted from transcriptions of news clips. When transcription is unavailable, our method must use automatic speech recognition. Thus, we need to examine the validity of our method when using speech recognition for future work.

Acknowledgments

This work was supported in part by Grant-in-Aid for Scientific Research (B) (22300299).

References

Chall, J. S. and Dial, H. E. 1948. Listener Understanding and Interest in Newscasts. Educational Research Bulletin. 27(6): 141–153+168.

Fang, I.E. 1966. The Easy Listening Formula. Journal of Broadcasting & Electronic Media, 11(1): 63–68.

Flesch, R. 1950. Measuring the Level of Abstraction. Journal of Applied Psychology, 34: 384–390.

Graesser, A. C., Danielle S. M., Max M. L., and Zhiqiang C. 2004.. Coh-Metrix: Analysis of Text on Cohesion and Language. Behavior Research Methods, Instruments, and Computers, 36: 193-202.

Henricksen, L. 1984. Sandhi Variation: a Filter of Input for Learners of ESL. Language Learning, 34: 103–126.

Hubbard, P. 2004. Learner Training for Effective Use of CALL. In S. Fotos and C. Browne (Eds), New Perspectives in CALL for Second Language Classrooms: 45–67, Lawrence Eribaum, Mahwah, NJ.

Kiyokawa, H. 1990. A Formula for Predicting Listenability: the Listenability of English Language Materials 2. Wayo Women's University Language and Literature, 24: 57–74.

Kotani, K., Yoshimi, T., Nanjo, H. and Isahara, H. 2012. Applicability of Readability Formulae to the Measurement of Sentence-level Readability, Proceedings of International Conference of Education, Proceedings of 5th International Conference of Education, Research and Innovation (ICERI): 6023–6031.

Kotani, K., Yoshimi, T. and Isahara, H 2013. Application of Reading Data in an Integrated Learner Corpus, Procedia–Social and Behavioral Sciences, 95: 513–521.

Meng, X.L., Rosenthal, R. and Rubin, D.B. 1992. Comparing Correlated Correlation Coefficients. Psychological Bulletin, 111(1): 172–175.

Messerklinger, J. 2006. Listenability. Center for English Language Education Journal, 14: 56–70.

Nation, P. and Malarcher, C. 2007. Reading for Speed and Fluency. Compass Publishing, Seoul, Korea.

Petrides, J. R. 2006. Attitudes and Motivation and their Impact on the Performance of Young English as a Foreign Language Learners. Journal of Language and Learning, 5(1): 1–20.

Rayner, K. and Reichle, E. D. 2010. Models of the Reading Process. Wiley Interdisciplinary Reviews: Cognitive Science, 1(6): 787–799.

Roach, P. 2001. English Phonetics and Phonology. Cambridge University Press, Cambridge.

Robb, M. P. and Gillon, G. T. 2007. Speech Rates of New Zealand English- and American English-speaking children. Advances in Speech-Language Pathology, 9(2): 1–8.

Saville-Troike, M. 2006. Introducing Second Language Acquisition. Cambridge University Press, Cambridge.

Schmid, H. 1994. Probabilistic Part-of-Speech Tagging Using Decision Trees. In Proceedings of International Conference on New Methods in Language Processing: 44–49. Manchester, UK.

Shen, W., Williams, J., Marius, T., and Salesky, E. 2013. A Language-Independent Approach to Automatic Text Difficulty Assessment for Second-Language Learners. In Proceedings of the 2nd Workshop on Predicting and Improving Text Readability for Target Reader Populations: 30–38. Sofia, Bulgaria.

Vandergrift, L. 2011. Second Language Listening: Presage, Process, Product and Pedagogy. In E. Hinkel (Ed) Handbook of Research in Second Language Teaching and Learning: 455–471. Routledge, New York.