

## Design of a Learner Corpus for Listening and Speaking Performance

**Katsunori Kotani**

Kansai Gaidai University  
16-1 Nakamiyahigashino-cho, Hirakata,  
Osaka, Japan 573-1001  
kkotani@kansai-gaidai.ac.jp

**Takehiko Yoshimi**

Ryukoku University  
1-5 Yokotani, Seta, Otsu,  
Shiga, Japan 520-7729

### Abstract

A learner corpus is a useful resource for developing automatic assessment techniques for implementation in a computer-assisted language learning system. However, presently, learner corpora are only helpful in terms of evaluating the accuracy of learner output (speaking and writing). Therefore, the present study proposes a learner corpus annotated with evaluation results regarding the accuracy and fluency of performance in speaking (output) and listening (input).

### 1 Introduction

The linguistic properties of learners of English as a foreign language (EFL), which are different from those of native speakers, have been identified through the analysis of output compiled in learner corpora (Sugiura et al. 2007, Friginal et al. 2013, Barron and Black 2014). These properties have been used to statistically classify learners' output into a range of proficiency levels (Thewissen 2013). Thus, a learner corpus is a useful linguistic resource for developing assessment techniques that are implementable in a computer-assisted language learning (CALL) system.

Although the contribution of learner corpora is well acknowledged (Granger 2009), previous learner corpora are limited in that learners' outputs have only been examined in terms of linguistic accuracy. As noted by Housen et al. (2012), learners' performance should be analyzed in terms of both accuracy and fluency. On the one hand, it

is true that a proficient learner uses a target language accurately; however, on the other hand, a trade-off is often observed, as a learner speaks grammatically correct sentences (high accuracy), but does so at an unnaturally slow speech rate (low fluency) (Brand and Götz 2011, Chang 2012).

Another limitation is typically seen in the target skill. Most previous learner corpora cover output skills in spoken or written language. From the viewpoint of communicative competence in spoken language, learners need to be proficient not only in speaking (output), but also in listening (input). Although speaking proficiency is well correlated with listening proficiency, a gap between these proficiencies is also known to exist (Liao et al. 2010, Liu and Costanzo 2013), as a learner may comprehend some sentences containing lexical and syntactic items that are difficult for them to actually articulate.

Because previous learner corpora have been limited in terms of speaking, a spoken learner corpus that demonstrates accuracy and fluency in speaking and listening is needed. The present study proposes to build a spoken learner corpus by annotating relevant information on sentences that learners spoke and listened to, respectively.

Another limitation is seen in the scope of corpus data analysis. Although the learner corpus of Kotani et al. (2015) addressed listening, it was only capable of providing listening comprehension data for analysis at the text level, because that was the level at which comprehension was examined. However, identifying which linguistic properties affect listening comprehension through text-level analysis is difficult. To identify learners' linguistic

problem areas, language use in local domains such as sentences needs to be analyzed, similar to machine translation evaluation at the sentence level (Gamon et al. 2005, Stanojević and Sima'an 2014). Therefore, the present study proposes to annotate listening comprehension data for individual sentences, which is expected to offer a finer-grained analysis for the identification of learners' linguistic problem areas.

## 2 Related Learner Corpora

According to Izumi et al. (2004), most learner corpora have covered written but not spoken language; therefore, they proposed a speaking corpus for EFL learners. However, their corpus did not cover listening. As Prince (2014) suggested, the lack of a listening corpus for EFL learners might be due to the difficulty of compiling data that demonstrate how learners listen to sentences. However, Luo et al. (2010) and Kotani et al. (2015) identified several issues regarding listening corpora for EFL learners.

The objective of Izumi et al. (2004) was to construct a model of the developmental stages of speaking ability among EFL learners that could also be used to develop techniques for automatically identifying errors. Their learner corpus was compiled using interviews with 1,200 EFL learners classified into nine levels according to the Standard Speaking Test, which evaluates oral proficiency. In their study, learners performed an interview-response exercise in which learners started and finished with an informal discussion on general topics, such as the interviewees' job and hobbies; between these informal chats, they performed three task-based activities: picture description, role-playing, and storytelling. Their corpus was then annotated with errors in relation to sentence generation, but not listening comprehension.

The objective of Luo et al. (2010) was to develop an automatic assessment technique for phonetic recognition. Their learner corpus was composed of data from 32 EFL learners classified into three levels according to Test of English for International Communication (TOEIC) scores. In their study, learners performed an exercise in which they repeated 14 sentences articulated by a native-speaking English teacher. In addition to phonetic recognition, listening comprehension was

also examined at the text level. Comprehension questions were provided twice: once when learners finished listening to the material for the first time, and again after they listened to the material repeatedly until they felt they had reached full phonetic recognition.

The objective of Kotani et al. (2015) was to create a linguistic resource for analysis of pronunciation at the sentence level and listening comprehension at the text level. Their learner corpus was composed of data from 30 native English speakers and 90 EFL learners classified into three levels according to TOEIC scores. In their study, native speakers and learners performed reading aloud and listening comprehension exercises. In the former, native speakers and learners read 80 sentences from a set of four texts aloud. In the latter, they listened to 80 sentences from another set of four texts and answered five comprehension questions for each one.

## 3 Listening and Speaking Corpus

### 3.1 Objective

The objective of our learner corpus is to serve as linguistic resource for the development of techniques that can automatically assess performance, as well as material for listening and speaking exercises; this is described in greater detail in Section 4.

The target skills and exercises for compiling corpus data are summarized in Table 1. Our learner corpus demonstrates learners' performance in relation to listening and speaking skills. Listening is divided into phonetic recognition and comprehension, while speaking is divided into pronunciation and sentence generation. Listening data (phonetic recognition and comprehension) are compiled in a dictation exercise, while those of pronunciation and sentence generation are compiled in reading aloud and question-response exercises, respectively.

Compared with previous learner corpora (Izumi et al. 2004, Luo et al. 2010, Kotani et al. 2015), our corpus covers more skills, as shown in Table 2. The letter "X" indicates the presence of relevant data in a learner corpus.

Skill	Sub-skill	Exercise
Listening	Phonetic recognition	Dictation
	Comprehension	
Speaking	Pronunciation	Reading aloud
	Sentence generation	Question-response

Table 1: Target performance for EFL learners

	Izumi	Luo	Kotani	Ours
Phonetic recognition	---	X	---	X
Sentence comprehension	---	---	---	X
Text comprehension	---	X	X	X
Pronunciation	X	X	X	X
Sentence generation	X	---	---	X

Table 2: Comparison with existing corpora

Whereas the previous corpora were only capable of providing comprehension data for analysis at the text level, our corpus provides data for analysis at the sentence level, which offers a finer-grained analysis for the identification of learners’ linguistic problem areas.

### 3.2 Data to be Compiled

Our corpus consists of three-layer annotation data and phonetic data for speech sounds. The first layer consists of text data (txt) in the form of transcribed speech sounds, and visual representation data (prapic) such as spectrograms produced by the Praat phonetic analysis program (Boersma and Weenink 2013). The second layer consists of text analysis involving tagging and dependency relation by the Stanford parser (de Marneffe et al. 2006), as well as analysis of descriptive information such as word length, syntactic pattern density, word information and readability provided by a computer tool called Coh-Metrix (McNamara et al. 2014). This layer also consists of phonetic analysis regarding pitch, intensity, and formant contour, as well as visible pulses (Boersma and Weeink 2013). The third layer consists of evaluation results of learners’ performance.

Corpus data should cover accuracy and fluency in phonetic recognition, comprehension, pronunciation, and sentence generation, as shown in Table 3. Accuracy is represented in terms of a manual evaluation score, while fluency is

represented in terms of speech rate and ease of processing.

	Accuracy	Fluency
Phonetic recognition	Evaluator’s evaluation	Material speech rate & Ease of processing
Comprehension	Learner’s evaluation	Material speech rate & Ease of processing
Pronunciation	Evaluator’s evaluation	Learner speech rate & Ease of processing
Sentence generation	Evaluator’s evaluation	Learner speech rate & Ease of processing

Table 3: Corpus data regarding accuracy and fluency data

The accuracy of phonetic recognition is evaluated using phonetic recognition scores. These scores are calculated as the rate of correctly repeated words per total number of words in a sentence/chunk. The success/failure of phonetic recognition for each word is manually evaluated by native-speaking English teachers on a binary scale (correct or incorrect).

The accuracy of comprehension is self-evaluated by learners on a binary scale (comprehensible or incomprehensible). The validity of this method, which makes the evaluation of sentence-by-sentence comprehension possible, has been acknowledged (Ross 1998).

The accuracy of both pronunciation and sentence generation are evaluated in terms of linguistic properties by native-speaking English teachers. Accuracy regarding linguistic properties is evaluated based on a 5-point Likert scale (Poor, Fair, Average, Good, Excellent). Linguistic properties reported as common errors made by EFL learners (Bryant 1984) are summarized in Table 4.

The fluencies of phonetic recognition and comprehension are also subjectively evaluated on a 5-point scale for ease of processing. These fluencies are also evaluated in consideration of the speech rate that learners actually hear. The speech rate is calculated as the number of words articulated in a minute of speech time. As learners continue listening until they fully understand the material, listening fluency is also evaluated in consideration of the number of repetitions.

The fluencies of pronunciation and sentence generation are evaluated based on speech rate and

ease of processing among learners. The speech rate is calculated based on the speech time required for articulation. However, the speech time for sentence generation also includes the time during which questions are asked to learners, because learners start to consider their response at this time. Ease of processing among learners is subjectively evaluated on a 5-point scale.

Domain	Class	Instance
Pronunciation	Consonant	*/ð/ for /t/ in <i>Thames</i>
	Vowel	*/I/ for /ai/ in <i>bite</i>
	Silent consonant	*/saig/ for /sai/ in <i>sigh</i>
	Unstressed by schwa	*/me-mo-ry/ for /mem-(o)ry/
	Stress in word	*/tikét/ for /tíket/ in <i>ticket</i>
	Stress in sentence	*“ <b>the</b> project” for “ <b>the project</b> ”
	Vowel-elision	*/cho-co-late/ for /choc-late/
	Consonant-elision	*/un-known/ for /u-known/
Sentence generation	Word-form	* <i>chock</i> for <i>check</i>
	Inflection-form	* <i>runned</i> for <i>ran</i>
	Agreement in pronoun	* <i>he</i> for <i>she</i>
	Agreement with a modifier	* <i>each cars</i> for <i>each car</i>
	Agreement between subject and verb	* <i>he study</i> for <i>he studies</i>
	Inflectional agreement	* <i>has study</i> for <i>has studied</i>
	Case form	* <i>him</i> for <i>he</i>
	Determiner	* <i>boy</i> for <i>a boy</i>
	Preposition	* <i>look him</i> for <i>look at him</i>
	Verbal object	* <i>saw</i> for <i>saw it</i>
	Determiner-choice	* <i>a boy</i> for <i>the boy</i>
	Tense	* <i>is</i> for <i>was</i>
	Aspect	* <i>is having</i> for <i>has</i>
	Negation	* <i>think that ...not</i> for <i>don't think...</i>
	Word-choice	* <i>see</i> for <i>watch</i>

Table 4: Linguistic properties

### 3.3 Learners

We plan to compile our learner corpus using data from 120 university EFL learners classified into four levels according to TOEIC listening scores (range: 5-495) on the following scale (Liao 2010): beginner level (150-245); intermediate level (250-345); advanced level (350-425); and advanced-high level (430-495). EFL learners are divided equally among each level with respect to the number of learners (N = 30).

### 3.4 Tasks

In the experiment, EFL learners are first asked to perform a dictation exercise in which they listen to materials unit-by-unit and then write down what they hear. After completing each unit, learners subjectively evaluate the ease of phonetic recognition and comprehension on 5-point and binary scales, respectively. After learners listen to the material the first time, they listen again repeatedly until they are confident they have achieved full comprehension.

The second exercise is a reading aloud exercise in which they read the same texts from the listening materials aloud, sentence-by-sentence. After reading each sentence, they subjectively evaluate the ease of pronunciation on a 5-point scale.

The third exercise is a question-response exercise in which they answer five general questions regarding the listening material. After providing their answers, they evaluate the ease of sentence generation on a 5-point scale.

### 3.5 Materials

Since the target of this project includes beginner-level learners, the listening task should be fairly easy to complete. Therefore, listening material is obtained from the VOA (Voice of America) Learning English site (<http://learningenglish.voanews.com>).

This online resource was chosen due to the limited vocabulary (1,500 words), short sentences, and slower than natural speech rate in the material (VOA Special English 2009).

Four reports are chosen from the Level 1 (the easiest among the three levels) in order for learners at the beginner level to complete the tasks. The topic of the reports is education, which is a familiar topic for university students. This allows

differences in learners’ background knowledge on speaking and listening to be minimized, in contrast to specific news events. Both male and female voices are used for the reports. Two reports each are recorded in a male and female voice in order to minimize any influence from gender.

Each report is composed of less than 400 words. The linguistic properties of the material, including the length of each audio clip (sec), the number of sentences, the number of token (words), the number of word types, and the speech rate, are summarized in Table 5.

Report	A	B	C	D
Time	237	234	232	220
Sentence	25	25	15	15
Token	363	349	348	353
Type	196	182	195	187
Speech rate	91.9	89.5	90.0	96.3

Table 5: Linguistic properties of VOA texts

#### 4 Reliability of the task

Whether learners could complete the exercises, especially the dictation exercises, remained unclear; therefore, we confirmed the validity of the listening material before compiling corpus data. Even though learners do not have to dictate all sentences correctly, they do need to make an attempt. If the chosen material is too difficult, learners dictate nothing, thereby resulting in corpus data that fail to demonstrate how English sounds are recognized by learners.

Therefore, a preliminary experiment examining whether learners were able to write down some words in a sentence was conducted to confirm the appropriateness of VOA Learning English as a resource. In addition, in order to evaluate learners’ listening ability, the corpus data should include both recognizable and unrecognizable words; therefore, this experiment examined the accuracy of the dictation results.

Participants were 21 university EFL learners with beginner- to intermediate-level English proficiency (Test of English as a Foreign Language Institutional Testing Program scores: 383-463; average score = 431.3 (standard deviation = 25.0)). The dictation exercise carried out was that in Report A, as shown in Table 5. They listened to and transcribed the report sentence-by-sentence.

They were allowed to listen to the report three times. Although 21 learners participated, data from one learner were excluded because that learner did not complete the latter half of the exercise. Henceforth, the data analyzed in this paper were compiled from 20 learners.

The success/failure of the dictation was evaluated for each word, as shown in Table 6. The word ID shows both the sentence and the word number, and thus <s1.1> refers to the first word in the first sentence. The spoken words illustrate what learners listened to. Here, all the words, even proper nouns such as “VOA,” are lowercase, because capitalization was not taken into consideration during evaluation.

The response rate shows the proportion of learners who wrote down something for a spoken word. When a learner wrote something down for a spoken word, a response score of 1 was assigned. On the other hand, when a learner wrote nothing, a response score of 0 was assigned.

The correct rate shows the proportion of learners who correctly dictated a spoken word. When the dictation of a spoken word was correct, a correct score of 1 was assigned. On the other hand, when dictation of a spoken word was incorrect, a correct score of 0 was assigned.

Word ID	Spoken word	Response rate	Correct rate
s1.1	from	1.00	1.00
s1.2	voa	1.00	0.15
s1.3	learning	1.00	1.00
s1.4	english	0.95	0.95
s1.5	this	1.00	1.00
s1.6	is	1.00	1.00
s1.7	the	0.80	0.65
s1.8	education	1.00	1.00
s19.1	for	1.00	0.80
s19.2	voa	0.95	0.15
s19.3	learning	1.00	0.80
s19.4	english	0.95	0.95
s19.5	i'm	1.00	0.75
s19.6	mario	0.75	0.55
s19.7	ritter	0.70	0.00

Table 6: Response and correct rates for the dictation exercise

The descriptive statistics show that the response rate ranged from 0.05 to 1.00 (Table 7). That is, every word received a response by at least one learner. In addition, approximately 80% of the words ( $((68 + 35 + 24 + 25 + 21 + 11 + 17) / 267)$ ) achieved high response rates ( $<0.70$ ), and the greatest frequency ( $N = 68$ ) was found for a response rate of 1.00 (Figure 1). These results suggest that the use of listening material from VOA Learning English is appropriate and allows adequate collection of learners' dictation data.

	Response rate	Correct rate
Total number of words	267	267
Minimum	0.05	0.00
Maximum	1.00	1.00
Mean	0.78	0.54
Standard Deviation	0.24	0.31

Table 7: Descriptive statics for response and correct rates

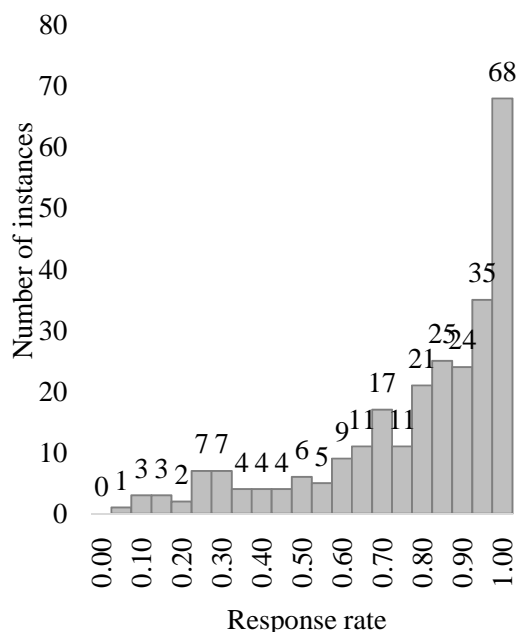


Figure 1: Frequency of the response rate

The descriptive statistics also show that the correct rate ranged from 0.00 to 1.00. In addition, as shown in Figure 2, the correct rate was evenly distributed. That is, no word was shown to be too

easy or too difficult for dictation. Hence, listening material from VOA Learning English allows the collection of both correct and incorrect dictation data, thereby suggesting the appropriateness of using VOA Learning English as listening material in compiling corpus data.

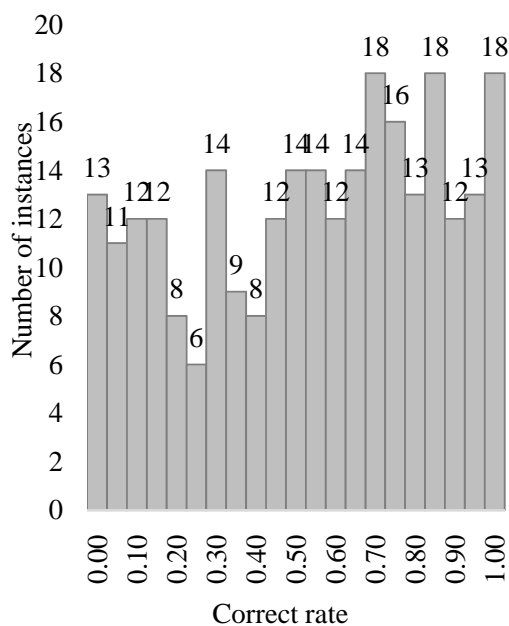


Figure 2: Frequency of the correct rate

Regarding minimum and maximum values in the response and correct rates, the minimum response rate of 0.05 was only found for the word <s15.6>, which is bolded here: “Georgetown University labor economist Anthony **Carnevale** says...” This word is a proper noun, and thus it seems unfamiliar to the learners. This unfamiliarity seems to decrease its associated response rate.

The minimum correct rate of 0.0 was found for 13 words. Among these incorrectly recognized words, an interesting example is the word <s7.1>, which is bolded here: “**Universities** say decreasing financial support...” This word should be frequently used by learners, and particularly familiar with the university learners in this experiment. This suggests that word familiarity is not related to a low correct rate. Upon further analysis, we found that most learners dictated the plural noun “universities” as a singular noun (“university”). The plural morpheme is pronounced unclearly, as illustrated by the dotted line in Figure

3. Hence, the presence of this morpheme should be made apparent due to the fact that the subsequent verb “say” would not follow the singular noun “university,” or that a singular noun needs a determiner such as “a” or “the.” Hence, learners fail to recognize this word due to a lack of syntactic knowledge or the failure to determine syntactic manipulation.

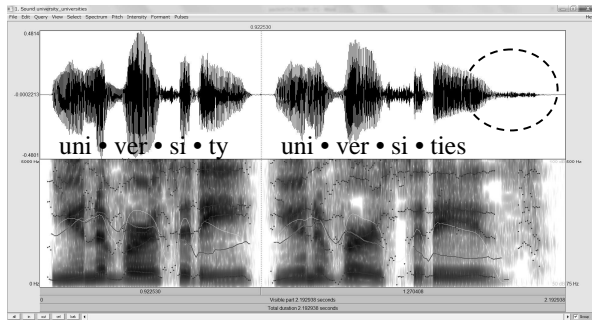


Figure 3: Spectrogram of “university” and “universities”

### 5 Application

An advantage of CALL-based learning is the use of “authentic” online materials such as news reports produced for native English speakers, but not designed for language learners. It is widely acknowledged that the use of authentic materials improves learners’ performance, particularly regarding practical communicative competence; however, the use of authentic materials can cause several problems.

One such problem concerns the assessment of performance among learners, because unlike textbooks, authentic materials are not designed to assess whether a learner’s language use is successful. Although CALL-based learning using authentic materials might be effective without it, assessment of performance certainly provides more effective learning because it enables the identification of linguistic problem areas among learners in daily communication.

Another problem concerns the difficulty of authentic materials, which, unlike textbooks, is uncontrolled, and thus increases the chance that a learner may lose motivation due to materials that are inappropriate or too difficult for their proficiency level. Hence, it is necessary to first assess the difficulty of authentic materials, and then to provide materials that are appropriate for

the individual learner’s proficiency, which is a rather burdensome task for language teachers. Therefore, automatic assessment of the materials’ difficulty supports both effective learning and effective teaching.

Such automatic assessment techniques result in a by-product that also improves CALL-based learning. In developing these assessment techniques, statistical models will be constructed for calculating proficiency-based optimal performance. If a CALL system demonstrated performance in listening and speaking exercises as well as what would be considered optimal performance, learners would then be able to assess their performance in terms of how it compares with optimal performance. This type of self-evaluation would allow learners to recognize gaps in their performance, and then to address these gaps by doing relevant practice exercises until their performance reaches or outperforms optimal performance; therefore, this type of system would promote autonomy among learners.

### 6 Conclusion

The present paper introduced the design of a new learner corpus for analyzing the accuracy and fluency of listening and speaking. This design differs from existing designs with respect to performance targets for learners. In addition, unlike the previous corpora, our learner corpus offers spoken language analysis at the sentence-level. This proposed learner corpus is expected to serve as a linguistic resource for the development of assessment techniques for both listening and speaking exercises in a CALL system.

### Acknowledgement

This work was supported by JSPS KAKENHI Grant Numbers, 22300299, 15H02940

### References

Anne Barron and Emily Black. 2014. Constructing small talk in learner-native speaker voice-based telecollaboration: A focus on topic management and backchanneling. *System*, 48: 112–128.

Paul Boersma and David Weenink. 2013. Praat: Doing Phonetics by Computer. Version 5.3.51, retrieved 2 June 2013 from <http://www.praat.org/>

- Christiane Brand and Sandra Götz. 2011. Fluency versus accuracy in advanced spoken learner language: A multi-method approach. *Errors and Disfluencies in Spoken Corpora*. Special Issue of *International Journal of Corpus Linguistics*, 16(2): 255–275.
- William H. Bryant. 1984. Typical errors in English made by Japanese ESL students. *JALT (Japan Association of Language Teachers) Journal*, 6(1): 1–18.
- Anna C.-S. Chang. 2012. Improving reading rate activities for EFL students: Timed reading and repeated oral reading. *Reading in a Foreign Language*, 24(1): 56–83.
- Eric Friginal, Man Li, and Sara C. Weigle. 2013. Revisiting multiple profiles of learner compositions: A comparison of highly rated NS and NNS essays. *Journal of Second Language Writing*, 23: 1–16.
- Michael Gamon, Anthony Aue, and Martine Smets. 2005. Sentence-level MT evaluation without reference translations: Beyond language modeling. In *European Association for Machine Translation (EAMT) 2005 Conference Proceedings*, pages 103–111.
- Sylviane Granger. 2009. The contribution of learner corpora to second language acquisition and foreign language teaching: A critical evaluation. In K. Aijmer. (ed.) *Corpora and Language Teaching*, pages 13–32, John Benjamins, Amsterdam and Philadelphia.
- Alex Housen, Folkert Kuiken, and Ineke Vedder. 2012. Complexity, accuracy and fluency: Definitions, measurement and research. In A. Housen, F. Kuiken, and I. Vedder. (eds.) *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA*, pages 1–20, John Benjamins, Amsterdam and Philadelphia.
- Emi Izumi, Kiyotaka Uchimoto, and Hitoshi Isahara. 2004. The overview of the SST speech corpus of Japanese learner English and evaluation through the experiment on automatic detection of learners' errors. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC-2004)*, pages 1435–1438.
- Katsunori Kotani and Takehiko Yoshimi. 2015 (to appear). Application of a corpus to identify gaps between English learners and native speakers. In *Proceedings of Eighth Workshop on Building and Using Comparable Corpora (BUCC)*.
- Chi-Wen Liao. 2010. TOEIC listening and reading test scale anchoring study. *TOEIC Compendium*, 5: 1–9.
- Chi-Wen Liao, Yanxuan Qu, and Rick Morgan. 2010. The relationships of test scores measured by the TOEIC listening and reading test and TOEIC speaking and writing tests. *TOEIC Compendium Study*, 10(13): 1–15.
- Jinghua Liu and Kate Costanzo. 2013. The relationship among TOEIC listening, reading, speaking, and writing skills. In D. E. Powers. (ed.) *The Research Foundation for the TOEIC Tests: A Compendium of Studies, Volume II*, pages 1–25, Educational Testing Service, Princeton, NJ.
- Dean Luo, Yutaka Yamauchi, and Nobuaki Minematsu. 2010. Speech analysis for automatic evaluation of shadowing. In *Proceedings of Speech and Language Technology in Education (SLaTE)*.
- Danielle S. McNamara, Arthur C. Graesser, Philip M. McCarthy, and Zhigang Cai. 2014. *Automated Evaluation of Text and Discourse with Coh-Metrix*. Cambridge University Press, Cambridge, MA.
- Marie-Catherine de Marneffe, Bill MacCartney and Christopher D. Manning. 2006. Generating typed dependency parses from phrase structure parses. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC-2006)*, pages 449–454.
- Peter Prince. 2014. Listening comprehension: Processing demands and assessment issues. In P. Leclercq, A. Edmonds, and H. Hilton. (eds.) *Measuring L2 Proficiency: Perspectives from SLA*, pages 93–108, Multilingual Matters, Clevedon.
- Steven Ross. 1998. Self-assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language Testing*, 15(1): 1–20.
- Miloš Stanojević and Khalil Sima'an. 2014. Fitting sentence level translation evaluation with many dense features. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 202–206.
- Masatoshi Sugiura, Masumi Narita, Tomomi Ishida, Tatsuya Sakaue, Remi Murao, and Kyoko Muraki. 2007. A discriminant analysis of non-native speakers and native speakers of English. In *Proceedings of the 2007 Corpus Linguistics Conference*.
- Jennifer Thewissen. 2013. Capturing L2 accuracy developmental patterns: Insights from an error-tagged EFL learner corpus. *The Modern Language Journal*, 97(1): 77–101.
- VOA Special English. 2009. *Word Book*. Voice of America, Washington, DC.