Journal of Modern Languages
Vol. 29, (2019)
https://doi.org/10.22452/jml.vol29no1.2

Constructing a longitudinal learner corpus to track L2 spoken English

Abe Mariko
abe.127@g.chuo-u.ac.jp
Chuo University, Japan

Yusuke Kondo
yusukekondo@waseda.jp
Waseda University Japan

Abstract

The main purposes of this article are to provide an overview of a research project on a longitudinal learner spoken corpus and to share procedures related to the transcription of learners' utterances from audio files using automated speech recognition (ASR) technology (IBM Watson Speech-to-text). The data of the corpus were collected twice or thrice a year for three consecutive years from 2016, creating eight data collection points altogether. They were gathered from 120 secondary school students who had been learning English in an English as a Foreign Language context for three years. The students were asked to take a monologue speaking test, the Telephone Standard Speaking Test, consisting of various tasks. The overall discussion of the article focuses on the details of this project and highlights how a methodological approach of combining electronic learner language data and ASR technology is useful in constructing learner spoken corpora.

Keywords: longitudinal data, learner corpus, L2 spoken English, corpus construction, automated speech recognition technology

1. Introduction

As Meunier (2015) stated, there has been a dramatic increase in learner corpora over the last two decades, yet the majority have been cross-sectional or pseudo-longitudinal in design. Thus, they fail to shed light on complex, and often unpredictable, developmental patterns in language learning and acquisition. Abe (2007, 2014), for example, investigated the largest spoken learner corpus in Japan, the National Institute of Information and Communications Technology Japanese Learner English (NICT JLE) Corpus, on which the dialogue test the Standard Speaking Test is based. Abe's study revealed that various types of linguistic features can be used to distinguish differences in oral proficiency levels. However, the NICT JLE Corpus consists of cross-sectional data, and with such data it is impossible to observe learning trajectories. In other words, it is difficult to examine how each individual learner progresses or regresses in their L2 learning and acquisition over time. Accordingly, it is crucial to conduct additional studies using adequate amounts of longitudinal learner data (Larsen-Freeman & Cameron, 2008). Learner corpora have the potential to increase our understanding of interlanguage, but few attempts have been made to provide systematic data about spoken language use that can be applied to language teaching and assessment materials (Pendar & Chapelle, 2008).

Learner corpora are relatively recent, with the construction of such corpora increasing in the 1990's (Granger, 1998), including the well-known International Corpus of Learner English (ICLE) (Granger *et al*., 2002; Granger *et al*., 2009), which was created as a part of the International Corpus of English (ICE) made up of national and regional varieties of English that were used for comparative studies of English. Several corpora based on the language production of Asian learners of English have been also developed. For example, the Japanese EFL Learner Corpus (JEFLL) (Tono, 2007) consists of essays written by junior and senior high school students, the Nagoya Interlanguage Corpus of English (NICE) (Sugiura, 2008) consists of essays written by university students. Furthermore, the International Corpus Network of Asian Learners of English (ICNALE) Written (Ishikawa 2013) and ICNALE Spoken (Ishikawa 2014) are based on the same corpus construction principle, so that researchers can compare the performance of different production modes and proficiency groups of Asian learners of English.

However, these learner corpora cannot be used to track how each individual learner's performance changes from one data collection point to another. Accordingly, in order to fill this knowledge gap in learner corpus research, the newly-developed Longitudinal Corpus of L2 Spoken

English (LOCSE) was designed to directly grasp L2 developmental patterns in a literal sense, not only on a whole group level, but also on an individual basis. The study has collected the same English learners' task performances an average of three times per year for the three consecutive years beginning in 2016, creating a total of eight data collection points (See Table 1). This is, to our knowledge, the largest longitudinal spoken corpus of lower-level speech samples produced by learners of English in the world. As a result, this corpus contains the potential for new insights regarding Learner Corpus Research (LCR) and Second Language Acquisition (SLA) research.

Table 1: Data collection points and numbers of participants

1	2016. 7/16 – 7/26	122 students
2	2016. 12/3 – 12/11	120 students
3	2017. 3/1 – 3/14	122 students
4	2017. 7/1 – 7/10	119 students
5	2017. 10/26 – 11/5	113 students
6	2018. 1/18 – 1/28	114 students
7	2018. 3/10 – 3/20	109 students
8	2018. 6/22 – 6/27	108 Students

The main purposes of this paper are to provide an overview of this innovative research project with a specific focus on the compilation of a longitudinal learner spoken corpus and to share procedures related to the transcription of learners' utterances from audio files using automated speech recognition (ASR) technology (IBM Watson Speech-to-text). In the following section, a description of the quality and quantity of the LOCSE data is provided. First, general information about the non-English-speaking learners is introduced. Second, the speaking test, which the LOCSE corpus is based on, is described in detail.

2. Method

2.1. Participants

Samples were collected from a group of 122 upper-secondary school students (52 males and 70 females), who agreed to participate in this research project. They were public senior-high school students, aged 15 when at the start of data collection. They spoke Japanese as their first language,

and they had no long-term experience in English speaking countries. They had been learning English in an English-as-a-Foreign Language (EFL) context for three years at the time of the first data collection. In the typical EFL context of Japan, secondary school students have limited opportunities to speak the target language inside and outside of the classroom. According to our questionnaire introduced at the start of data collection, most students had hardly any opportunity to receive informal foreign language education outside of the classroom. Thus, their out-of-class English use was limited. In other words, there was an inadequate amount of exposure to the target language. In addition, there was no necessity or opportunity to use English in their daily lives, and English remained an academic subject that was unlikely to become a tool for communication for many of the participants. English was primarily used by students as an instrument for gaining academic success by way of passing entrance examinations, and the learning of English vocabulary and grammar was typically strongly focused on this limited goal. To our surprise, however, the participants of our research project were given sufficient speaking tasks to apply newly learned grammatical forms to real communication inside of their English classes. The same three teachers taught all students using the same syllabi and same EFL textbook, which focused on four English skills until the time of their graduation from secondary school in three years. Two different types of English classes (described below) included adequate follow-up speaking tasks to apply newly learned grammar points to real communication. To sum up, they were studying the target language under similar learning setting.

Two different types of English classes (i.e., Content and Language Integrated Learning, Oral Communication) included adequate follow-up speaking tasks to apply newly learned grammar points and useful expressions to real communication. The main objectives of the "Content and Language Integrated Learning" were three-fold. First, students were encouraged to improve their critical thinking skills in second language (L2) English based on a range of familiar topics such as geography, psychology and sociology — a concept comparable to Content and Language Integrated Learning (CLIL). Second, students were encouraged to take the initiative in interacting with their peers while discussing the topics in L2 English. Third, students were encouraged to be aware of the CLIL-based assessments. In this approach, their final grade was decided based not only on their L2 English proficiency, but also on their understanding and performance related to the topics covered in class. To these ends, the students had a number of opportunities to engage in writing, speaking, reading and listening throughout the term.

The "Oral Communication" portion of the programme mainly focused on developing speaking and writing skills. First, students were encouraged to work on improving their recognition and use of the target language while they participated in conversation and discussion practice. They practised oral communication in pairs and groups through a variety of topics and communication tasks, such as agreeing, apologizing, and explaining. Teachers aimed to develop students' communication skill in terms of accuracy and fluency through these activities. Second, the course focused on encouraging students to understand and use a variety of English expressions in different contexts. Students learned how to respond appropriately in verbal communication using the target language. They were also encouraged to build a sense of confidence and achievement in using their English language skills. While teachers continued to focus on correct pronunciation, stress, and intonation, students were asked to record their speech using a digital voice recorder to check their performance. Third, reinforcement and enhancement of writing skills was encouraged by having students closely examine various components of the writing aspects, such as sentence structure, grammar, word choice, linking words, and organization. Students were also required to learn how to write English essays and summarize English texts.

The content and foci of each class were mainly based on the textbook materials in each unit (i.e., dialogue, reading / listening passages, grammar / vocabulary exercises). However, some exercises were reduced or eliminated to put more focus on writing and oral communication practice. Viewing teachers' annual lesson plans, it was clear that they focused on the lessons which were most likely to fit students' interest, so that students could maintain their motivation to learn the target language. Most classes that we observed were not merely composed of explaining grammar points and translating reading passages from English to Japanese. Instead, they were spent on a variety of activities, such as small talk (e.g., "What is your favourite drink?"), dictation, discussion, question and answers, crossword puzzles, pronunciation practice, and reading aloud practice. Teachers were eager to support students during the listening and reading comprehension activities by giving adequate feedback for speaking activities. Judging from our classroom observations, teachers paid sufficient attention to both transmitting knowledge and applying newly learned linguistic items. Figure 1 shows a video clip from a classroom observation which was held on February 8th, 2017 that supports this position.

Figure 1: A video clip from a classroom observation (February 8th, 2017)

Information on scores from other English examinations are valuable in understanding the general English proficiency level of the non-English-speaking learners and for checking the effectiveness of the speaking test, which this research project is based on. It can be said that the oral proficiency levels of learners are at the early beginner level. Table 2 shows the results of EIKEN Test, which was held in March, 2017.

Table 2. Scores on the EIKEN Examination by the Test-Takers

EIKEN	Grade 1	Grade Pre-1	Grade 2	Grade Pre-2	Total
Year 1	0	0	29	76	124
Year 2	1	2	78	30	132
Year 3	0	5	49	44	110

Note.

1. *EIKEN Grade 1 can understand and use the English necessary to participate effectively in a wide range of social, professional, and educational situations.*
2. *EIKEN Grade Pre-1 can understand and use the English necessary to participate effectively in social, professional, and educational situations.*
3. *EIKEN Grade 2 can understand and use English at a level of taking part in social, professional, and educational situations. It is aimed at Japanese senior high school graduates.*
4. *EIKEN Grade Pre-2 can understand and use English at a level of taking part in general aspects of daily life. This level is aimed at second-year Japanese senior high school students.*

2.2. Speaking Test

The students were asked to take a monologic speaking test, the Telephone Standard Speaking Test (ALC Press, 2016), which consists of various tasks (e.g., description, comparison, and reasoning), and their utterances were compiled to create the LOCSE data. This speaking test is an automated telephone-based English-speaking test. It consists of 10 recorded questions, and test-takers are required to respond to each question in 45 seconds without any planning time or the use of reference material. Three certified raters gave a holistic score to each speech sample based on various criteria, such as function-based ability, sentence structure, accuracy, and content. Test-takers scores are divided into 9 levels.

Figure 2 shows the result of the Telephone Standard Speaking Test, which was held in July 2016. The number of learners at each oral proficiency level (Level 1: 0, Level 2: 8, Level 3: 62, Level 4: 47, Level 5: 5, Level 6: 0, Level 7: 0, Level 8: 0, Level 9: 0) is distinct. The number of learners in level 3 (intermediate low) was considerably higher than the other levels, while the number of learners in levels 1, 2, 8, and 9 were relatively low. In the three batches of data collected over one year, the holistic scores spanned across five oral proficiency levels out of an eight-point scale and the learners' overall score tended to rise across the year.

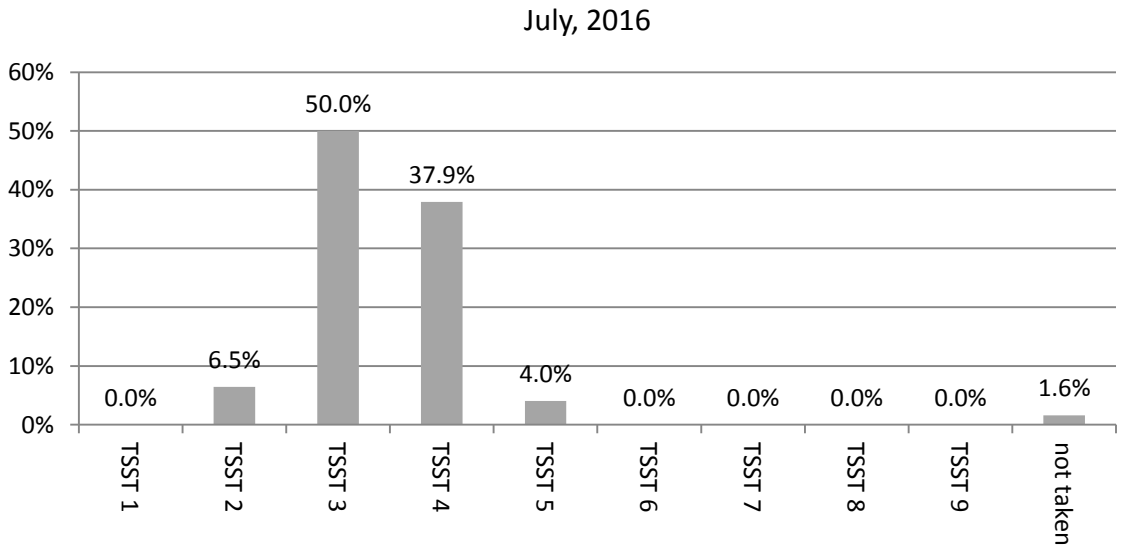

Figure 2: Test-takers at each oral proficiency level (N = 122).

2.3. Data Collection

We checked whether students have not erased recorded speaking performances until they can produce a version that they are pleased with. Speaking practice activities were included in learners' weekend homework, and speaking skills were included in their end-of-term test. Each learner owned a digital voice recorder to practise speaking English, and they were asked to take the Telephone Standard Speaking Test as a homework assignment. Test-takers were given a score report (Figure 3) whenever they took the test.

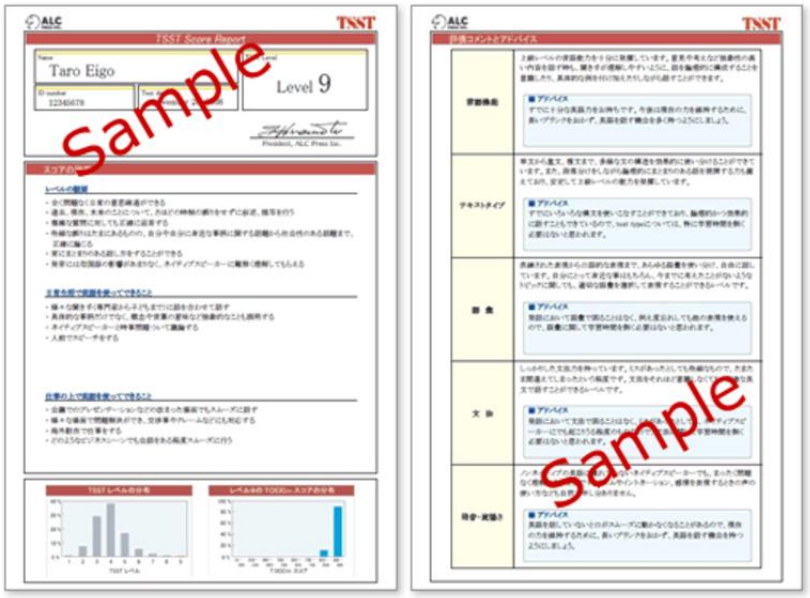

Figure 3: A sample of score report given to each test-taker

Students were also given more detailed feedback (e.g., total word counts in each data collection point) from the researchers with special comments. We counted the number of words, sentences, types, tokens, and silent pauses and filled pauses and calculated the average number of words in a sentence in individual utterances based on the transcriptions. We offered these measures to each learner who provided us with a speech sample. We assume that learners can be motivated when they capture the development in their speech through these detailed measures. Table 3 provides an example of feedback data that we offered to a particular learner.

Table 3: An example of feedback data

	July 2016	December 2016	March 2017
Score	5	5	5
No. of tokens	682	764	855
No. of types	196	197	242
No. of sentence	37	29	25
Average number of words in a sentence	18.43	26.34	34.20
No. of fillers	24	93	49

Along with the corpus development process, an abundance of relevant metadata was collected and added to the texts to make full use of this new longitudinal spoken learner corpus. With this design, we can gain new insights into learner language development. For example, what impact individual differences (e.g., motivation, personality, and learning style), English use, task type, and oral proficiency may have on the speech produced by learners of English. Figure 4 shows how we have organized this research project.

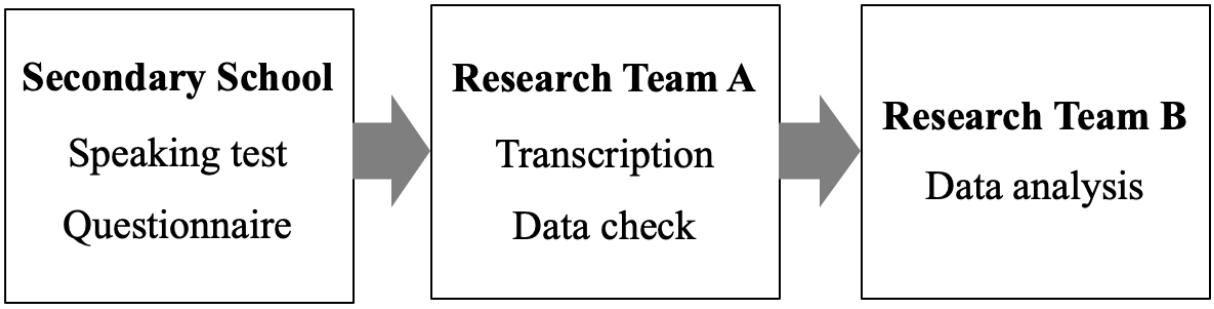

Figure 4: Project organization and workflow for constructing LOCSE

2.3. Corpus Construction

This corpus construction project aimed to investigate the language use of individual learners through large-scale data collection. Therefore, the study focuses specifically on the quantity and speed of transcription and tag-annotations. The development of a spoken corpus requires the collection of all relevant information, and the maintenance of high levels of consistency (Thompson, 2005). Consequently, it is necessary to establish clear guidelines for transcription and a thorough procedure for checking each transcription to reduce the risk of inconsistency.

In order to reduce potential inconsistency clear rules for the transcription and checking process were cautiously prepared. To ensure that the correct procedure was followed, each transcriber checked the work of other transcribers, and the researchers monitored the work of three transcribers. As the resources are to be made publicly available to other researchers in the future, procedures of the data collection and coding schemes have been documented. We have kept records of problematic cases, and additional information regarding discussions of deviations from the guidelines that were needed to solve problematic cases. After adopting these processes, the guideline samples were modified. This process was repeated until the guideline was completed.

The corpus construction methodology of Izumi, Uchimoto, and Isahara (2004) was referred to when constructing the LOCSE corpus. Considering the interchangeability of the resource, XML format was chosen for mark-up of the transcribed texts. The guidelines cover the following components: utterances, pauses, Japanese filler, English filler, overlaps, repetitions (e.g., he he he), self-corrections (e.g., He don't doesn't), and non-verbal phenomena (e.g., laughter, sigh, and cough). In order to increase the speed and accuracy of corpus construction, the transcripts were automatically annotated, and then manually annotated. The following lists the guidelines for transcribing learners' spoken performance.

1. Use American style English to spell out the speech.
2. If Japanese is used in test-takers' talk, spell it out as a Japanese word.
   ex) dou, sou
3. Do not add any period nor space for an abbreviation.
   ex) OK / CD / TV / BBC / TOEFL / UCLA / US / USA
4. Do not use any numbers nor codes (e.g., \, $, &, @), but spell it out.
   ex) October tenth, tenth of October
   ex) one thousand yen
   ex) thirty five hundred
5. Use a period, comma or question mark at the end of phrases and sentences. Even if a

sentence is not completely finished, add a period at the end of speech. Leave one white space after a period and comma. (Transcribers decided sentence boundaries on the basis of speech characteristics such as silence and filled pauses and syntactic information.)

6. Do not use the followings: exclamation mark, single quotation, double quotation.
7. Spell out the filler as it is pronounced.
   ex) Japanese fillers (e.g., aa, ee, and etto)
   ex) English fillers (e.g., ah, er, and um)
8. Use (..) for a 2 to 3 seconds' pause, and use (…) for a pause longer than 3 seconds. When transcribers use (..) or (…) between the word, add one white space between (..) or (…). When the speech end with a 2 to 3 seconds' pause (..) or a pause longer than 3 seconds (…), add one space between them to end the speech. If the speech ends or beginning with a pause, use (..) or (…) to begin or end the sentence.
   ex) Um, Christmas season. … Um, ah big Christmas trees and next ee snow … um … so … Christmas party and presents. … So .. I like snow because Christmas season.
9. Even if the learners are trying to correct their utterances, spell it out as they have said.
   ex) It would be a kind of trash so it's a it's waste it's a kind of waste.
   ex) It's difficult to keep the plants vivid vividly.
   ex) I close closed the door when I left school.
   ex) I'm I am a high school student.
   ex) It's pla planning it's planned by my teacher.
10. Even if the learners are repeating the same words or phrases, spell it out as they have said.
   ex) I think he is a very ka kind person.
   ex) My pe parents don't allow me to live in Tokyo.
   ex) Um I I like skei skating I like to skating,
   ex) My mother doing doing tea Japanese teacher. Ah she she .. she speak she speak .. su English well. And she li she like .. she like English English. Then I wan I like um her.
11. Even if the grammar is irregular, spell it out as it is spoken. Even if the word does not exist, spell it out as it is pronounced.
   ex) So other country movies is so very dynamics.
   ex) She speaked a lot to her childs.
   ex) My cafeteria name's is <?>Kan'etsu</?>. .. <?>That</?> .. there place is um .. ah I I ... I often use .. I often go to ... there. ... Um there's food is delicious.
   ex) My cafeteria name's is <?>Kan'etsu</?>. .. <?>That</?> ..
12. If transcribers were able to understand mispronounced words (e.g., right and light) from the context, they should be corrected. When transcribers encountered a problem identifying a word, single-question-mark tags (<?>…</?>) were inserted; when utterances were not clear enough to understand, double-question-mark tags (<??></??>) were inserted; when utterances were impossible to understand.
   ex) Um the the day the day before yesterday, I went to uh small party, then then I drank ah drank ah <?>spirituous</?>.
   ex) Ee so .. I .. <?>parsi</?> .. good life um.
   ex) ... Um .. I ... um ... um I have never .. sco <?>scowat</?> from my my parents.

ex) I'm <?>waying</?> a school uniform because um .. because .. I went to .. school ...
ex) <??></??> should be very beautiful.

13. Even if the word is incorrectly pronounced, spell it out as the correct spelling; if transcribers can understand the word from the context.
    ex) McDonald's
14. If the word is incorrectly pronounced and cannot be understood without special attention of the transcriber, use the <pro></pro> tag to indicate it.
    ex) There are many <pro>birds</pro> on the tree.
    ex) .. Um I I respect my father .. because my father i is <pro>working</pro> for for <pro>us</pro>.
    ex) I have two things. First, I think ah ... big city is useful .. when .. we moved <pro>better</pro> place.
    ex) My father didn't <pro>allow</pro> me to buy a motorcycle.
15. If the transcriber has confidence in their understanding of the utterances of the learner, but it is difficult to judge which word should be used, they should make their best guess, and use the <un></un> tag to indicate it.
    ex) Olympic is world um .. ah a lot of sports. ... World player ... attack ... <un>four</un> years. (for or four)
    ex) Um because um .. I I like ... <un>there</un> .. <un>there</un> master. .. She is woman. Um she is very .. fun and .. kind. So .. when I then I go to .. there. (there, their, they're)
    ex) No, I usually don't sleep because ee .. I must .. I must study .. every day <un>to</un> four hours. So .. a and I ah I must a lot of homework every. (to, two)
16. If the learner is laughing, use the <laughter></laughter> tag to indicate it.
17. Use the following tags to show the non-verbal phenomena.
    ex) <nvs>laughter</nvs>
    ex) <nvs>sigh</nvs>
    ex) <nvs>cough</nvs>
    ex) <nvs>sniff</nvs>
    ex) <nvs>yawn</nvs>
    ex) <nvs>burp</nvs>
    ex) <nvs>sneeze</nvs>
    ex) <nvs>click</nvs>
18. Use <slip></slip> tag to show the slip of tongue, which does not occur from the lack of vocabulary knowledge.
    ex) Take brand goods is very bad. Ah take <slip>grand</slip> goods ... makes me angry.

2.5. Information about the Data

The learners responded to 10 items in a single data collection point. The learners' utterances were transcribed and stored in text files by item. Therefore, 10 text files were created for each learner in a single data collection point. Items are different in terms of the grammar items used, task type,

item difficulty, and topic, and the learners' utterances were scored by human raters. Along with the transcribed speech, these pieces of information were stored as file names as shown below:

201806_001_01_4_PJ_26.txt

The first six digits denote the year and the month when the utterance was collected; the second three digits indicate the learner's ID; the third two digits denote the item number; the forth digit is the score given to the utterance; the next two alphabetic characters specify the grammar items targeted in this item and task type (e.g., description); the last two digits refer to the topic (e.g., friends). The file name above shows that the learner with ID 001 responded to item 01 whose topic is days of the week in June, 2018, the task type is 04, the learner is supposed to use a particular grammar point (i.e., PJ), and his/her utterance should be given under topic 26. The following texts are transcription samples of the LOCSE.

> *I had adjust myself into new environment when I was a first first student ah when I was a first hi second grade student in my high school because I <pro>lived</pro> in dormitory school dormitory it is was first experience for me. So I had to adjust myself into the new environment. Ah in my in school dormitory I have I had to share my share my life and share equipment a lot of equipment. So it was very hard to me and and as a result I couldn't adjust myself aa adjust myself in. (201806_010_07_6_ED_43.txt)*

> *It is different of .. like of food. .. My parents like cheese cake, but I don't like cheese cake. .. I .. don't like cheese food because I think che chesse's smell is aw awful. But .. my. (201806_039_09_3_CO_19.txt)*

2.6. Transcription by IBM Watson Speech-to-Text Technology

In the initial stage of the transcription in this corpus project, one of the human transcribers transcribed speech, then the transcriptions were checked by one of the other two human transcribers, and lastly the transcriptions were examined by another human transcriber. However, in order to reduce the burden to the human transcribers, we introduced IBM Watson to our transcription procedure.

We used IBM Watson Speech-to-text technology to transcribe learners' utterances through a Python client package, watson-developer-cloud on Python 3.4. By using this technology, learners' utterances were automatically transcribed. The following text is an example of transcribed speech sample based on this automated method.

of course yes and one day in but in English scale when %HESITATION that's yeah I went to scuttle well it's good draped I stay at my host how meets house I have to speak English did talk waste my post from me I don't want to I can't speak English way I do want to speak English

In the transcriptions by IBM Watson, no punctuation marks are found and filled pauses are coded as "%HESITATION". The transcriptions by IBM Watson were manually corrected by a human transcriber, and the corrected versions were checked by another transcriber, and the mutually-checked versions of the transcriptions were carefully examined by another human transcriber.

To investigate the impact the introduction of IBM Watson had on the transcription process and how its potential to reduce the burden placed on the human transcribers, we randomly chose 50 speech samples and transcribed them using two different transcription procedures. In Procedure A, a human transcriber transcribed each speech sample, then the transcriptions were checked and rechecked by other human transcribers. In Procedure B, on the other hand, IBM Watson transcribed speech samples and then the transcriptions were checked and rechecked by human transcribers. Through these two procedures, we obtained six different transcriptions, which are depicted in Figure 5 below.

H: Human transcription
HH: Checked human transcription
HHH: Rechecked human transcription
W: Machine transcription
WH: Checked machine transcription
WHH: Rechecked machine transcription

Figure 5: Six types of transcriptions obtained in this study

To examine the differences between these six transcriptions, we adopted an analysis that implemented word error rate (WER). WER is calculated using the formula listed below. We used the Python code in Thoma (2013) as the basis for these calculations:

S: The number of substitutions

D: The number of deletion

I: The number of insertion

N: The number of words in the reference

Below are the examples of substitution, deletion, and insertion.

Substitution
Reference: He is a good teacher
Target: He is a good pitcher

teacher is substituted with pitcher.

Deletion
Reference: He is a good teacher
Target: He is a teacher

good is deleted.

Insertion
Reference: He is a good teacher
Target: He is a very good teacher

very is inserted.

Before calculating WER, the tags in the human transcriptions were removed, and non-lexical fillers such as “err” and “um” were replaced with %HESITATION because these tags appear only in human transcriptions and non-lexical fillers transcribed as %HESITATION in the machine transcriptions. Thus, this modification helped ensure consistency between transcription methods.

Firstly, we compared the WER between the machine transcriptions (W) and the checked machine ones (WH) with that between the human transcriptions (H) and the checked human transcriptions (HH) to examine how effective the machine transcription could be in reducing the burden on our human transcribers. If we obtained no difference in the WERs between the W and WH and the H and HH, that would indicate that the machine transcription served as a sufficient transcriber. Secondly, we compared the rechecked machine transcriptions (WHH) with the rechecked human transcriptions (HHH) using WER. If this WER was large, that would mean that the machine had its influence on our human transcriptions. Table 4 summarizes the results of WER.

Table 4. The basic statistics of word error rates

Statistics	**H and HH**	**W and WH**	**WHH and HHH**
Count	50	50	50
Mean	0.04	0.56	0.13
Std	0.04	0.22	0.11
Min	0.00	0.13	0.00
25%	0.01	0.37	0.06
50%	0.03	0.58	0.10
75%	0.05	0.72	0.19
Max	0.14	0.96	0.48

When speech was transcribed by the machine, over 50 percent of the transcription was corrected by human transcribers during the checking process. On the other hand, only 4 percent of the transcription by the human transcribers was corrected in the checking process. Although our human transcribers felt that the introduction of the machine transcription reduced their workload as a whole, when we adopted the machine transcription, our human transcribers were still needed to correct the machine transcriptions since half of these were found to be incorrect. However, our human transcribers thought that transcribing speech was more time-consuming than correcting erroneous transcriptions. Another important point we considered was the difference between the rechecked human transcriptions and the rechecked machine transcriptions. This difference indicates how much influence the machine transcription of speech had on the checking and the rechecking procedures by our human transcribers. The average difference is 13 percent, which represents a relatively minor influence. However, when we look at the transcriptions with larger differences between the rechecked human transcription and the rechecked machine transcription, many tags that indicate that the transcriber could not identify words were found. The following texts are examples with the larger differences.

WHH
Erm I I think I think ee peo mocha is mm ah I think that <??></??> but ee <?>official</?> <?>life</?> is ee eight eight <?>smoke</?> ... uh there are mm there is eight <?>smoke</?> <?>smoke</?> eight <?>smoke</?> people ee ... ee ... .

HHH
Aa I I think .. I think ee <?>good</?> smoker is mm the I think <??></??> but ee official guide is aa <??></??> <??></??> smoke. .. So there are m there is <??></??> smokers smokes <?>bad</?> smokes people the ... .

As the examples show, the differences in these two transcriptions are apparently caused by the differences between the confidence of transcribers (The tag containing "?" and "??" means that transcribers could not identify the words.). If the transcriptions with the larger differences are excluded, the average difference rates would decrease to a large degree. Therefore, although the number of samples is quite small, the influence of the introduction to the machine translation on our transcription procedure may be considerably small.

3. Initial Findings from the First Five Data Collection Points

In this corpus, disfluency features of learners' utterances are marked by various types of tags, such as silent pauses, non-lexical fillers, and non-verbal sounds. In addition to these tags, the utterances include a number of repetitions. To find out how many meaningful English words learners utter in each task, we categorized the tags for disfluency and deleted certain elements. We deleted the tags if the words between the tags were actually uttered, such as <pro> </pro> (pronunciation error), <laughter> </laughter> (including non-verbal sound), and <JP> </JP> (Japanese words), and deleted Japanese words. Furthermore, we replaced silent pauses and filled pauses with one single symbol _p_, and deleted repeated words. The first text listed below is an example that illustrates the results of this process.

Before processing

.. Ee well ... I think that ... I <pro>should</pro> ee <?>cling</?> it, .. or ee ee ee because ... if .. I can ... ee ee.

After processing

*well _p_ i think that _p_ i should _p_ cling it _p_ or _p_ because _p_ if _p_ i can _p_*

Notes.
.. (two dots) : long pause (2 to 3 seconds' pause)
... (three dots): very long pause (a pause longer than 3 seconds)
<?></?>: the transcriber's confidence is low because utterances were not clear enough to understand

After the processing, we calculated the average number of utterances and words in each task and the average number of words per utterance in the initial five consecutive data collection points form July, 2016 to October, 2017. Figures Y1, Y2, and Y3 show the development of learners' fluency over this time period. In this analysis, utterance is defined as word sequence between pause or filled pause.

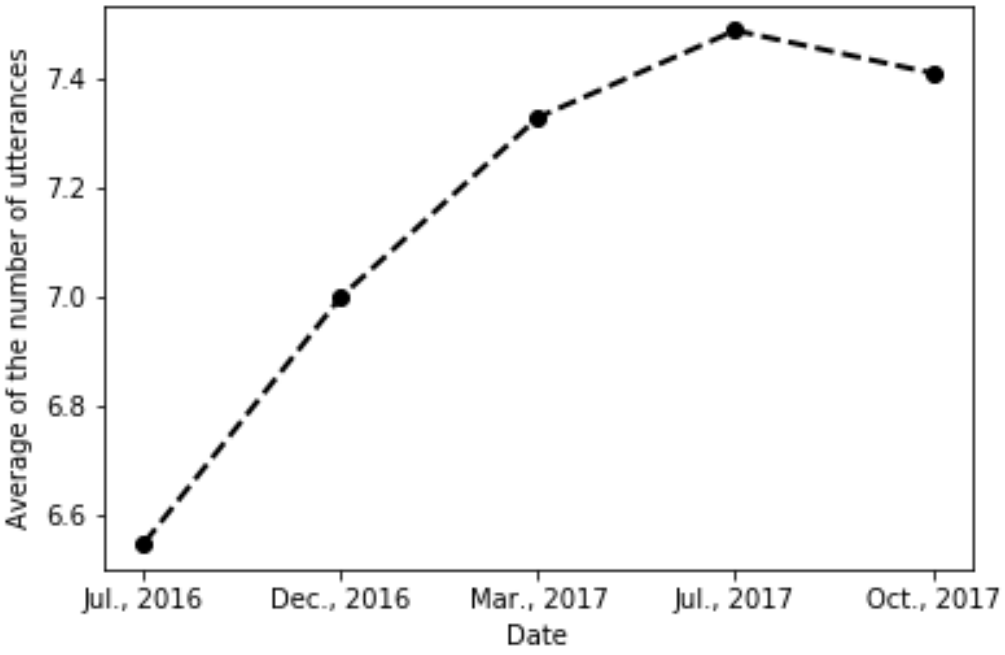

Figure Y1: Average Number of Utterances in a Task

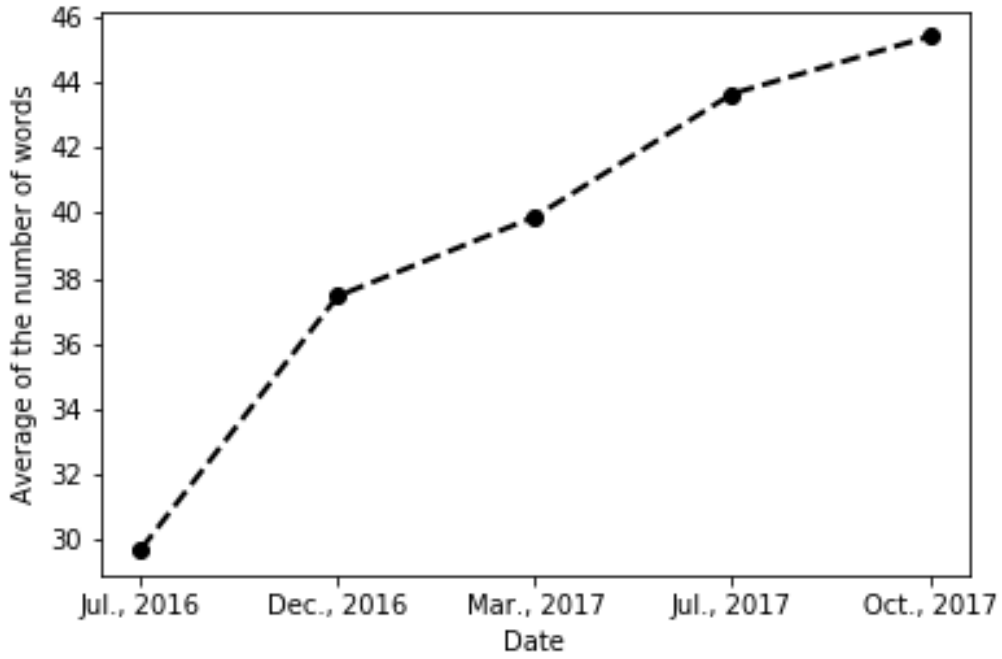

Figure Y2: Average Number of Words in a Task

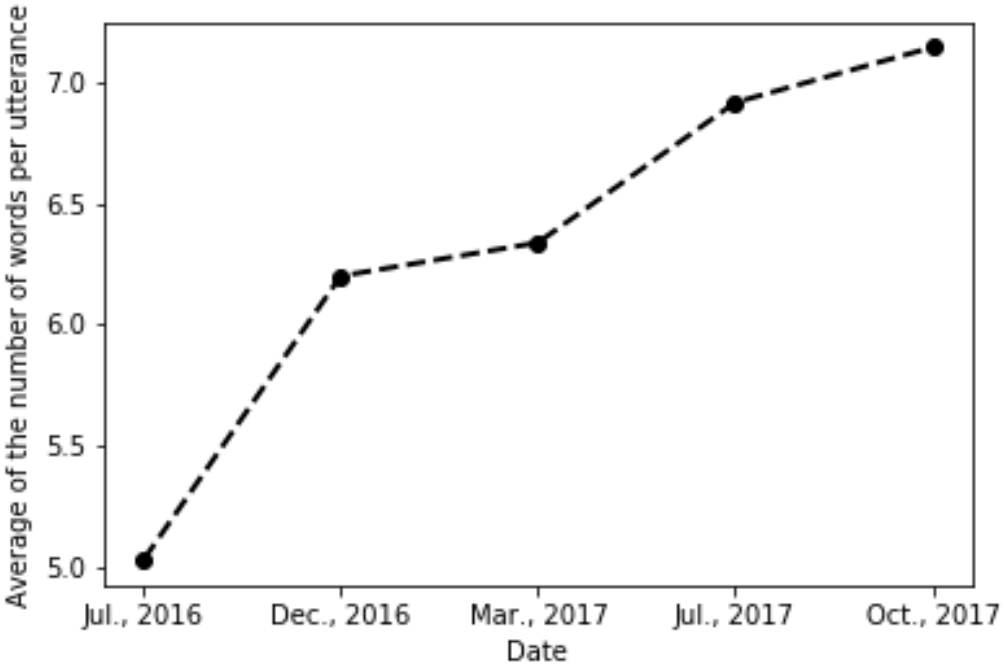

Figure Y3: Average Number of words in an utterance

Through the analysis presented above, we were able to capture the development of learners' fluency from the first data collection point to the fifth. The average number of utterances in each task did not increase drastically (Figure Y1). The number of utterances in each task increased only by 0.8 from July, 2016 to Oct. 2017. The number of words, on the other hand, increased by about 15 on average (Figure Y2). As Figure Y3 shows, therefore, the average number of words in an utterance increased. That indicates that the number of words that the learners could utter without pauses increased over time.

Learners utterances are described with various kinds of tags in this corpus. In this analysis, we reduced the type of disfluency markers, such as non-lexical pauses and repetitions, and found that the learners developed their fluency through the first data collection point to the fifth.

4. Conclusion

Understanding how learner language differs in eight data collection points is significant because this understanding can contribute to more effective language learning and language teaching. During the process of language learning, it is useful for learners to understand the norms of the target language and the characteristic differences between the interlanguage and the target language. It is also necessary for teachers and researchers to understand these differences to develop appropriate teaching and language assessment materials. It was also shown that a methodological approach of combining electronic learner language data and automated speech recognition (ASR) technology (IBM Watson Speech-to-text) is useful in constructing learner spoken corpora.

Acknowledgement

This work was supported by JSPS KAKENHI Grant Number JP16H03455.

References

Abe, M. (2007). A corpus-based investigation of errors across proficiency levels in L2 spoken production. JACET Journal, 44, 1-14.

Abe, M. (2014). Frequency change patterns across proficiency levels in Japanese EFL learner speech. Journal of Applied Language Studies, Special issue on "Learner language and learner corpora", 8(3), 85-96.

ALC Press (2016). Telephone Standard Speaking Test (TSST). Retrieved from https://tsst.alc.co.jp/biz/en/

Granger, S. (Ed.). (1998). Learner English on computer. London: Addison Wesley Longman.

Granger, S., Dagneaux, E., & Meunier, F. (2002). The international corpus of learner English: Handbook and CD-ROM. Louvain-la-Neuve: Presses Universitaires de Louvain.

Granger, S., Dagneaux, E., Meunier, F., & Paquot, M. (2009). International corpus of learner English. (2nd version). Louvain-la-Neuve: Presses Universitaires de Louvain.

Ishikawa, S. (2011). A new horizon in learner corpus studies: The aim of the ICNALE project. In G. Weir, S. Ishikawa, & K. Poonpon (Eds.), Corpora and language technologies in teaching, learning and research (pp. 3-11). Glasgow: University of Strathclyde Press.

Ishikawa, S. (2013). The ICNALE and sophisticated contrastive interlanguage analysis of Asian learners of English. In *Learner Corpus Studies in Asia and the World 1*, S. Ishikawa (ed.), 91-118. Kobe, Japan: Kobe University.

Ishikawa, S. (2014). Design of the ICNALE-Spoken: A new database for multi-modal contrastive interlanguage analysis. In *Learner Corpus Studies in Asia and the World 2*, S. Ishikawa (ed.), 63-75. Kobe, Japan: Kobe University.

Izumi, E., Uchimoto, K., & Isahara, H. (2004). 日本人 1200 人の英語スピーキングコーパス [A speaking corpus of 1200 Japanese learners of English]. Tokyo, Japan: ALC Press.

Larsen-Freeman, D., & Cameron, L. (2008). Complex systems and applied linguistics. Oxford: Oxford University Press.

Meunier, F. (2015). Developmental patterns in learner corpora. In S. Granger, G. Gilquin & F. Meunier (Eds.). The Cambridge handbook of learner corpus research. Cambridge: C UP, 378-400.

Pendar, N., & Chapelle, C. (2008). Investigating the promise of learner corpora: Methodological

issues. CALICO Journal, 25(2), 189-206. doi:10.1558/cj.v25i2.189-206

Sugiura, M. (Ed.). (2008). 英語学習者のコロケーション知識に関する基礎的研究 [Basic research on collocation knowledge of L2 English learners] Nagoya: Nagoya University.

Thoma, M. (2013, November, 15). Word error rate calculation. [Blog post] Retrieved from https://martin-thoma.com/word-error-rate-calculation/

Thompson, P. (2005). Spoken language corpora. In M. Wynne (Ed.). Developing linguistic corpora: A guide to good practice. Oxford: Oxbow Books, 59-70.

Tono, Y. (2007). (Ed.). 日本人中高生一万人の英語コーパス ”JEFLL Corpus” [JEFLL Corpus, a corpus of 10,000 Japanese EFL learners: Analysing written composition of Japanese junior and senior high school students]. Tokyo, Japan: Shogakukan.

About the Authors

Mariko Abe is a Professor of the Faculty of Science and Engineering at Chuo University, Japan. She received her Doctoral degree in Education from Temple University. She has published works on multivariate analyses of L2 spoken and written development. Her current research interests include identifying key linguistic characteristics that distinguish learners of different proficiency levels and applying the findings of corpus analysis to learner material development. She is also interested in computer-aided error analysis and automated scoring.

Yusuke Kondo works as Associate Professor at Waseda University. He received the degrees of BA, MEd, and PhD from Waseda University. His research interests are in language testing and machine learning. Now he is developing an automated scoring system for L2 learners' speech and writing.