



# Data Collection and Annotation of Relatively Spontaneous and Relatively Extended Elicited Utterances by English Learners in Undergraduate Japanese Courses

Kawamura, Mayumi

Yokomori, Daisuke

Suzuki, Masanori

Harada, Yasunari

---

**(Citation)**

Learner Corpus Studies in Asia and the World, 1:179-197

**(Issue Date)**

2013-03-23

**(Resource Type)**

departmental bulletin paper

**(Version)**

Version of Record

**(JaLCD0I)**

<https://doi.org/10.24546/81006682>

**(URL)**

<https://hdl.handle.net/20.500.14094/81006682>



# Data Collection and Annotation of Relatively Spontaneous and Relatively Extended Elicited Utterances by English Learners in Undergraduate Japanese Courses

Mayumi KAWAMURA<sup>1</sup>, Daisuke YOKOMORI<sup>2</sup>,

Masanori SUZUKI<sup>3</sup> & Yasunari HARADA<sup>4</sup>

<sup>1</sup>*Language Annotator*, <sup>2</sup>*JSPS & Nagoya University*, <sup>3</sup>*Pearson Knowledge Technologies*,

<sup>4</sup>*Waseda University*

## Abstract

In this paper we report on an eight-year project for collecting spoken utterances by English learners enrolled in Japanese university English language classes. Most of our efforts so far have focused on data collection, with some rudimentary manual transcriptions of a very limited part of the collected utterances. The raw audio and video recordings of students interacting in English and Japanese come to a total of 200 hours or so per semester, as seven to eleven groups of students engage in the activity for 25 to 30 minutes with a 90-minute class that convene 15 times in a semester. The core data corresponds roughly to one third or one fifth of these raw materials. The portion manually transcribed so far represents very roughly about 1/20 of the core data to be compiled and the transcription came to 150,000 words including 10% of filler-like expressions. This suggests that when we transcribe the entire core data among the collected spoken utterances, it will come to something like 3,000,000 words. In this presentation, we describe our data collection and audio transcription interface and then touch upon potentials for new research that can be conducted on the basis of the collected data. We propose a new way to look at some of the natures of spontaneously spoken English by Japanese learners in interactive settings.

## Keywords

Relatively spontaneous utterances, Elicited utterances, Data collection, Transcription annotation, Distributed annotation, Browser interface, Learners of English, Phrase-final vowel lengthening,

## I Introduction

In this paper we report on an eight-year project for collecting spoken utterances by English learners enrolled in Japanese university English language classes. Most of our efforts so far have focused on data collection, with some rudimentary manual transcriptions of a very limited part of the collected utterances. The raw audio and video recordings of students interacting in English and Japanese come to a total of 200 hours or so per semester, as seven to eleven groups of students engage in the activity for 25 to 30 minutes with a 90-minute class that convene 15 times in a semester. The core data corresponds roughly to one third or one fifth of these raw materials. The portion manually transcribed so far represents very roughly about 1/20 of the core data to be compiled and the transcription came to 150,000 words including 10% of filler-like expressions. This suggests that when we transcribe the entire core data among the collected spoken utterances, it will come to something like 3,000,000 words. In this presentation, we describe our data collection and audio transcription interface and then touch upon potentials for new research that can be conducted on the basis of the collected data. We propose a new way to look at some of the natures of spontaneously spoken English by Japanese learners in interactive settings.

Learner corpora, or collection of data produced by learners of a given language, are drawing greater attention in recent years. For the past twenty years or so, increasingly greater emphasis is being placed on cultivating 'communicative competences' on the part of learners in English-language education within Japanese school systems, but it is often difficult to fathom, for English-language teaching faculty at universities, what their students know about and can do with English. This is partly because learning experiences and mastery of English at the time when they enter college differ greatly in depth and coverage, as the Japanese Ministry of Education, or MEXT, which is short for Ministry of Education, Culture, Sports, Science and Technology, is trying to diversify educational institutions, systems and curricula in Japan.

In the research project outlined below, our current focus is on obtaining relatively spontaneous and relatively extended elicited utterances that are produced by learners of English studying English at the university level. Since Japanese learners of English do not utter English sentences completely spontaneously and completely on the fly, we need to provide them an environment in which those learners would be invited or prompted to express themselves in English. One such device is providing them with questions in English to be read by their peer learners and giving a restricted time-frame in which to respond. Although they are constrained in the topics, or the questions they are expected to answer to, and as to the length of the responses, there is still much room for freedom and flexibility in what can actually be said.

## II Data Collection Design, Equipment and Tools

### 2.1 Objectives and Characteristics of our Data Collection

One noteworthy aspect of this project is that this data collection procedure is embedded in general English classes for the first-year students and focuses on relatively spontaneous utterances in face-to-face oral communication. Another important feature in this project is that the utterance data can be linked to proficiency levels of the students who uttered them. Moreover, as the data collection in this project is embedded in students' in-class activities aimed at improving performance at specific tasks, the data accumulated would eventually witness potential longitudinal changes or improvements of the students' performance and proficiency, while the data collection activities in turn are expected to help the students integrate the four basic skills of listening, speaking, reading and writing into integrated interactional activities and thus turn out a pedagogically effective means for helping those students learn to communicate in spoken and written English. Thus, the data collection is not limited to the audio recordings of oral response practices. Written works from students, including essays (in Word files), presentation slides (in PowerPoint files), and records of their extensive reading activities (in Excel files) are collected and accumulated.

### 2.2 Classroom Procedures for Data Collection

For each class, ten questions pertaining to one particular topic are prepared in advance and printed on business-card size pieces of paper. The questioner picks up one of those ten question cards and reads the question aloud to the respondent twice. The respondent has ten seconds to think and formulate the answer and 45 seconds to speak whatever comes to her/his mind. The time-keeper prompts the respondent by saying "Start!" ten seconds after the question is read the second time, and says "Stop!" 45 seconds later. After the response is given, the questioner and the time-keeper give a score to the response based on a rubric given to the students and write the score onto a peer-review sheet for the respondent. Then, the three students change their respective roles and go on to the next question. Usually, 20 to 25 minutes are devoted to this activity in a session of 90 minutes and students go on to start writing a 500-word essay on the topic in odd-numbered sessions or peer-review the essays written a week earlier in even-numbered sessions.

The English classes in which those data collection activities are conducted convene in computer cluster rooms once a week for 15 weeks in a semester, so that many of the students' activities are digitally recorded at various stages. For instance, students are expected to read a picture book, a chapter book, a graded reader, or a paperback novel of their choice, from among the several hundred volumes the principal researcher brings to the classroom for the session every week and report on the title, number of pages read

and time spent for reading the material in an Excel file. Students either write an essay or revise the one written one week earlier and submit an initial version written in class, a revised version completed as homework and a final version revised after peer-review and peer-evaluation in class, all in Word files. In the latter half of the year, students also engage in small-group presentations on what they discussed during the oral interaction practices in odd-numbered sessions or on what they saw on the web news sites in even-numbered sessions and the PowerPoint files are collected. These electronic files are easier to collect, and perhaps also to analyze, while the students' face-to-face oral interactions are ephemeral and much harder to store and analyze, which this project aims to perform.

### 2.3 Equipment for Data Collection

The numbers of students in the first-year English classes where the data collections take place are currently roughly around 24 to 30 and definitively at most 36, partly because of the curriculum design and partly because of facility constraints. As the students are organized into groups of three (or less) during the oral interaction practices, a maximum of 12 tracks had to be recorded simultaneously. No substantial overlap is expected to occur between the end of the question and the beginning of the response so one microphone per group seemed to suffice in order to record the question read twice and the response spontaneously given. Some other practical issues were also taken into consideration when making decisions regarding the audio-recording device, such as:

- recording quality: linear PCM with highest sampling rate / highest bit rate possible
- storage and post-processing of sound files to be handled on Windows machines
- equipment to be carried, deployed and used in any classroom

Eventually, we decided on the following basic configuration of our digital audio recorder:

- Alesis ADAT HD24 XR: 24-Track Hard Disk Recorder
- Alesis MultiMix 12R: 8ch microphone fader (amplifier/mixer)
- Sony ECM-360: electret-condenser microphone
- microphone cable
- portable container on wheels for the equipment

Along with this audio recording device, the current data collection utilizes 8 to 12 sets of one Sony DCR-SR100, a video camera with 30GB internal hard-drive plus one Sony HCM-HW1, a wireless Bluetooth microphone for the video camera, to be used by each group of students. The internal hard drives of those video cameras are recognized as external drives when connected to Windows machines via USB 2.0 cables and the time stamps of those files keep track of when the segment was recorded. As long as all the

video cameras' internal clocks are synchronized, it is relatively easy to tell when a given file was shot from the time stamp of the file. In addition, the video cameras come with 5.1 channel surround sound recording systems, with the center channel assigned to the Bluetooth wireless microphone when attached. With a reasonably decent stereo playback system, you can tell which direction a given voice is coming from, which may be of help in identifying the speaker of a given piece of utterance.

#### 2.4 Annotation Tools and Working Environments

The utterances collected in this project are either mainly in English or mainly in Japanese depending on when the utterance is taking place within a group session, although we also find portions where the languages switch as students encounter difficulties. For the part of audio data where the language is supposed to be English, a given segment may represent either a question prepared and printed by the principal researcher and read aloud by a student during oral interaction practices, followed by a relatively spontaneous and unprepared response by another student to the questions just read. On the other hand, the recorded audio and video data also covers students' intra- and inter-group interactions in their native Japanese language. The intra-group interactions often reflect students' coordination or confirmation of roles during the oral interaction practices, support and help when one of the student encounters some difficulty in reading or answering a question, or their chatting that could be related or unrelated to the topics of questions in practice. Inter-group interactions may reflect the teacher and/or TA giving instructions, students asking questions to the teacher and/or TA and interactions among students in groups mostly in their native language.

#### 2.5 Browser-interface for query and annotation.

Currently, we are using a web-browser interface for searching for a particular audio segment, or responses, with specific information as to the question being answered, the date of the utterance, the identification tag and/or the proficiency levels of the student who is producing the response, etc. The same web-browser interface is used to assign transcription annotation work to student annotators, as shown in figure 1 on the next page. The annotators can just click on icons for the sound file and use audio-playback software to listen to the segment and type in the transcription annotation into the web pages.

While annotators can only see listings of the audio segments assigned and cannot see actual student names or listen to the segments where the speaker is identifying her/himself, administrators such as master annotator or principal researcher can see a listing of all audio segments, as shown in figure 2 on the next page. For administrators, the same utterance segment may be transcribed by multiple annotators, so that one record represents one transcription rather than one audio segment.

書き起こしツール ASPD (書き起こし用)

Date of 2013/07/02 22:44:02 GMT+9  
 URL: http://www.pfu.ac.jp/~language/17/annotation.asp, administrator: OHPQuestion, ADV: Annotation-Disable Version

sound	no	track_id	track	SPP	SPP	P_L	UT	AN	GU	ADV	
328	4780	08								on	
329	4780	08								on	
330	4780	08								on	
331	4780	08								on	
332	4780	08	21	E	s			Q-ma01	080501	study at home using computer em because I don't have much money, so I can't study	
333	4780	08	6	E	r			Q-ma01	080501	study at home using computer em because I don't have much money, so I can't study	
334	4780	08								on	
335	4780	08	10	E	s			Q-ma02	080502	Draw your hands when you live in a country.	
336	4780	08								on	
337	4780	08								on	
338	4780	08								on	
339	4780	08								on	
340	4780	08	21	E	r			Q-ma02	080501	[laughter] To tell the truth, I don't know computer software, but I think here it has 1	
341	4780	08								on	
342	4780	08	6	E	s			Q-ma03	080503	Needless to say, success comes from progress.	

Figure 1 data listing for student annotators

書き起こしツール ASPD (書き起こし用)

Date of 2013/07/02 22:53:51 GMT+9  
 URL: http://www.pfu.ac.jp/~language/17/annotation.asp, administrator: OHPQuestion, ADV: Annotation-Disable Version

1頁/100件表示

sound	no	track_id	track	SPP	SPP	P_L	UT	AN	GU	ADV	
328	4780	08								on	
329	4780	08								on	
330	4780	08								on	
331	4780	08								on	
332	4780	08	21	E	s			Q		My name is [redacted] My number is 21.	
333	4780	08	21	E	s			Q-ma01	080501	study at home using computer em because I don't have much money, so I ca	
334	4780	08	6	E	r			Q		[redacted] number 6.	
335	4780	08	6	E	r			Q-ma01	080501	study at home using computer em because I don't have much money, so I ca	
336	4780	08								on	
337	4780	08	10	E	s			Q		My name is [redacted] My number is 10.	
338	4780	08	10	E	s			Q-ma02	080502	Draw your hands when you live in a country.	

Figure 2 data listing for administrators

Annotators can click on the icons for the sound file they are going to work on and type in their transcriptions in the text field in Figure 3 and submit the data. They can also search from among the audio segments assigned.

アノテーションデータ記入・修正 ASPD

これは no 305 のデータです。 may

track\_id 4780  
 start 105  
 end 135  
 file\_id [redacted]  
 track 08  
 Speaker1 5  
 Speaker2 [redacted]  
 Speaker3 [redacted]  
 student\_id st\_id,tmp  
 PrimaryLanguage E  
 UtteranceType r  
 Annotator [redacted]  
 Question Q-ma01  
 ADV 080501  
 Text study at home using computer em because I don't have much money, so I can't study  
 note None

[記入・修正する]

Figure 3.3 annotation data input





### III Spoken English Ability upon Entrance into University

#### 3.1 Expected Improvement in Spoken English Ability

Since MEXT proposed a strategic plan known as “Japanese with English Abilities” in 2002, developing communicative English skills in workplaces has been recognized as an imminent educational problem from elementary schools through universities. In fact, since 1990’s, the courses of study designed by MEXT have asked foreign language classes (*i.e.*, English, practically speaking) to emphasize communicative activities. A further strategic move was the introduction of a listening section in the English subject within the University Entrance Center examinations, which more than 500,000 prospective entrants to universities across Japan take in mid January each year. These marked shifts in the focus in the curricula and in the university entrance exams should supposedly help improve Japanese learners’ spoken English skills but even with the revised curricula and efforts, degrees of improvements are at best questionable.

#### 3.2 Significance of Accumulation of Objective Data

Due to the prevailing use of email, reading and writing skills have acquired renewed significance in workplace everyday communication today. However, listening and speaking skills still play a critical role in foreign language learning. Although the university entrance center examination has added the listening component in its English examination, there exist few objective data available as to how well university first-year students can listen and speak English because a speaking test has not been introduced yet. It is a necessary first step to understand students’ spoken English skills in order to effectively train university students to the level at which they can effectively use English for work purposes. Furthermore, it is also critical to have a clear picture of students’ English levels as early as possible after they enter university such that their limited university time can be maximized to improve their spoken English skills.

#### 3.3 Automated Spoken English Test: Versant English Test

Versant English Test (VET) is an automated listening and speaking test developed by Pearson Knowledge Technologies. The test can be taken on the phone or via a computer and is scored automatically by the computerized scoring system including speech processing technologies. One of the salient differences between VET and conventional spoken English tests is that VET does not require an interviewer or a rater. In the current project, approximately 80 students are asked to take this spoken English test four times every year.

VET is designed to measure *facility in spoken English*, which is considered to be at the core of listening and speaking activities and measures how well the test-taker can understand spoken English and respond in real time. Items used in VET conform to the most frequent 8,000 words in a spoken English corpus called Switchboard Corpus. All VET

items are written in relatively easy language; however, these items are recorded by native English speakers at a conversational pace, requiring the test-taker to process each spoken item in real-time.

Each test-taker needs a test paper and a land-line telephone to take a test. When the test-taker is ready, she/he dials the phone number printed on the test paper using telephone keypads. When prompted by the recorded instructions, the test-taker starts to hear the instructions and then the test starts. The test completes in approximately 15 minutes and the scoring system automatically scores the test. Usually, scores become available in a few minutes and the score reports can be retrievable. In the score report, the Overall score and four sub-scores of sentence mastery, vocabulary, fluency and pronunciation are provided on a scale of 20-80. A web-based score management tool allows the score users (teachers or HR people) to view the scores of their test-takers in one page and to download all the scores in to an Excel spreadsheet. In addition, the score users can listen to part of the test-takers' responses through the score management tool. VET consists of six tasks as shown in Table 1.

Table 1 Test Structure of Versant English Test

<b>Part</b>	<b>Task</b>
Part A	Reading
Part B	Sentence Repeat
Part C	Short Answer Questions
Part D	Sentence Builds
Part E	Story Retellings
Part F	Open Questions

The sentences that are asked by the spoken instructions to read out loud in *Part A: Reading* are printed on the test paper. From Part B to Part F, all test items are presented only aurally. For these sections, the test paper presents only the test instructions and sample items. The responses from Parts A to E are scored automatically and Part F is not scored at this moment.

The test reliability for the Overall scores was estimated through two different methods: the split-half method and test-retest method. Both methods reported a coefficient of 0.97, indicating that VET scores are highly reliable and consistent. (See Pearson 2003 for details.) Several concurrent validity studies demonstrated high correlations with well-established English exams such as the speaking section of the TOEFL-iBT ( $r=0.75$ ,  $n=130$ ), IELTS Speaking test ( $r=0.76$ ,  $n=130$ ), TOEIC listening scores ( $r=0.76$ ,  $n=171$ ). (See Pearson 2007 for details.)

Furthermore, Pearson's internal linking study with Common European Framework of Reference (CEFR) results in the concordance table, as summarized in Table 2.

Table 2 Versant English Test and CEFR

Versant	CEFR
20-25	> A1
26-35	A1
36-46	A2
47-57	B1
58-68	B2
69-78	C1
79-80	C2

### 3.4 Data Analysis

Descriptive statistics from the 2006 to 2011 administrations are summarized in Table 3 below.

Table 3 Descriptive statistics by year

Year	N	Mean	SD	Variance	Max	Min
2006	90	37.27	6.95	48.31	62	21
2007	74	37.26	8.27	68.39	66	23
2008	74	36.01	6.43	41.41	51	23
2009	71	37.11	5.90	34.82	49	25
2010	86	37.30	7.47	55.86	80	23
2011	97	37.43	6.00	36.02	53	20

The mean scores from all test administrations show remarkable consistency (around 36 and 37). One-way ANOVA confirmed that there is no statistical significance among the mean scores [ $F(5,486) = 0.45, n.s.$ ]. Figure 6 below page shows cumulative density functions (CDF) for the six different administration years. The figure also depicts remarkable similarities in the score distributions year after year.

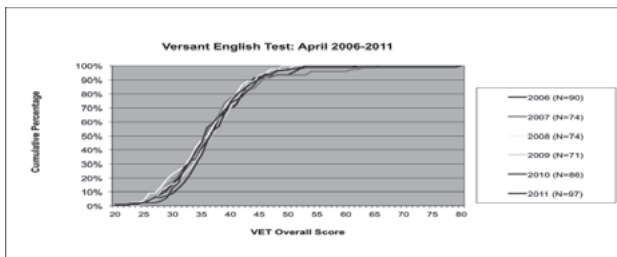


Figure 6 CDFs for VET by year

### 3.4.1 Consideration from CEFR perspective

When we compare the mean scores of 36 or 27 with the CEFR levels, the scores correspond to the A2 level. According to Pearson's previous study with CEFR, the VET Overall scores up to 46 fall into the A2 level. Then, if the score of 46 is examined in the CDFs in Figure 1, it can be seen that approximately 90% of the students are A2 or below on CEFR.

CEFR describes language abilities in practical language use contexts by including functional elements and situational elements. These descriptions are presented as Can-Do statements. We now know that the vast majority of the students in this study belong to either A1 or A2. According to CEFR, the A level is broadly defined as "Basic User". The specific Can-Do statements for A1 and A2 are summarized in Table 4.

Table 4 Can-Do Statements for  
CEFR A1 and A2 levels

A2	Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g., very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.
A1	Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.

Some expressions used to describe these levels are "familiar", "immediate relevance/need", "simple routine", "basic", etc. In other words, language users at the A level can deal with immediate personal topics, using frequently-used basic expressions. However, they cannot understand or speak about complex and/or abstract ideas. Moreover, approximately 40% of the students each year received a score of 35 or below on VET, which corresponds to A1 or below on CEFR. At A1, language users cannot process language spoken at a conversational speed and need cooperation from their interlocutor such as slowed speech or enunciated pronunciation. It is self-evident that students' spoken English skills are far from the level of MEXT's target in "the Japanese with English ability".

### 3.4.2 Consideration from "Situated" communication

Successful communications in situated contexts can be considered contingent upon a number of independent factors and their interactions. For example, interlocutors,

familiarity with topics, contexts, seniority, information amount and complexity, are among such possible factors. However, as theories such as Speech Act Theory (Austin 1962) and Cooperative Principles (Grice 1975) maintain, linguistic communications require the speaker to instantly assess these “external” factors and to relate such assessment of the communicative context to the appropriate language use. When this situated communication is considered in the context of the first language, the speaker’s basic language processing ability is usually not taken into account because it is the speaker’s native language and it is assumed that the speaker already possesses the basic psycholinguistic language processing skills and communication problems are usually ascribed to the speaker’s deficiency in or lack of awareness about sociolinguistic or pragmatic aspects of language use.

However, in the context of language use as a second language, the speaker’s psycholinguistic competence becomes one of the factors that contribute in situated communications. For instance, in communicative contexts of English as a second or foreign language, in addition to the external factors discussed earlier, the speaker’s “internal linguistic factor”, that is, “real-time language processing ability” can be presumed to play a crucial foundational role in effectively assessing and incorporating external factors to achieve successful communications.

As described in 2.1, the test construct assessed in VET is the ability to understand and use basic lexical items and sentence structures. VET, therefore, is designed to measure how well the test-taker can process relatively simple spoken language in real time. Given this, VET appears to assess the speaker’s internal linguistic ability (factor). The mean VET scores in Table 3 suggest that it may be reasonable to think that university first-year students allocate most of their attention to language processing and decoding such that they are unable to execute smooth communication while paying attention to external factors at the same time.

### 3.4.3 Longitudinal Assessment

VET was administered to the same group of students three times during one academic year to track the students’ progress in spoken English: in April 2006 (at the time of entrance), June 2006 (end of the first semester), and January 2007 (end of the academic year). In the April administration, a total of 90 students participated; however, the number of students who took VET all three times was 78 students. In Table 5, the descriptive statistics were summarized for these 78 students.

Table 5 Summary of descriptive statistics of VET scores of students in academic year 2006 over three administrations

Administration	Mean	SD	Variance	Max	Min
April 2006	37.43	7.07	50.22	62	21
June 2006	39.68	5.71	32.64	53	29
January 2007	40.90	6.17	38.02	58	25

It can be seen that the mean scores gradually improved every administration time. A score improvement of 3.47 was observed from April 2006 to January 2007 and the paired t-test showed that the difference was statistically significant ( $t(81)=-6.89$ ,  $p<0.01$ ). However, the mean score of 40.90 is still at the A2 level on CEFR and the students' English proficiency level on average is not at the level that is sufficient to have real-time conversations efficiently. How to help students reach the next level (B1/B2, Independent User) seems to be the urgent agenda in university English classes.

## IV Phrase-final Vowel Lengthening in spontaneous utterances by Japanese learners of English

### 4.1 Phrase-final Vowel Lengthening

Based on the data that we collected in "oral response practices" conducted in class, we observe a phenomenon we call Phrase-final Vowel Lengthening, or PfVL, at the end of words that occur in the course of relatively spontaneous speech by Japanese college learners of English<sup>1</sup>. For example, in sentences (1) and (2) below, stressed/inserted vowels are transcribed with curly brackets and vowel lengthening is marked by colons (:). Underscores mark words with PfVL.

- (1) There is{u}: mountain and{o}: sea.
- (2) ... but I{t}: think{u}: it is more important to  
have{u}: fun with friends.

PfVL may impair the intelligibility of the learners' speech. In a word level, it may cause a deviation from proper syllable structure and stress pattern of the word. In a phrase/clause/sentence level, it may result in a deviation from proper rhythm and/or intonation structure of the phrase. Thus, PfVL should be regarded as a target of remedy in English language education.

There are some apparently similar and related phenomena in Japanese EFL learners' speech: (A) epenthesis, (B) "L1-English-like" vowel lengthening, and (C) vowel lengthening in the middle of a word. First, epenthesis is a phenomenon in which the speaker adds a vowel after a closed syllable. For example, a word "bed" is pronounced as /beddo/ instead of /bed/ and "match" is pronounced as /matchi/ not as /match/. This phenomenon is caused by the Japanese phonological structure where there is basically no closed syllable. Interestingly, even learners who have already learned how to pronounce closed syllables may speak with PfVL in their spontaneous speech. Secondly, as Fox Tree & Clark (1997) have shown, even native speakers of English do lengthen

vowels at the end of some words, especially when they do not speak fluently for some reason.

- (3) “and when you come when you come to look at thiv, thuh literature, - I mean you know thuh actual statements” (Fox Tree & Clark 1997: 152)

If the vowel lengthening sounds like “L1 English,” we do not regard it as an instance of PfVL. Thirdly, Japanese ESL learners lengthen vowels in a non-final position of the word.

- (4) I feel more confident tha{a}::n whē{e}::n I did it

This phenomenon may also impair proper pronunciation and it is clearly related to PfVL, but we excluded them from the present study.

#### 4.2 Distribution Pattern of PfVL

Based on audio recordings and transcriptions of face-to-face interactions, we investigate the conditions under which such prolongations of word-final vowels may occur in actual learners’ speech, more specifically, (1) whether it tends to occur more frequently in reading aloud written texts or in spontaneous utterances and (2) what kinds of word tends be involved in this process and explore possible explanations for these. For the present pilot study, we use data recorded in a class during the first week of the 2007 spring semester.

##### 4.2.1 Reading-out vs. Spontaneous speech

The result of the investigation shows that learners who can pronounce English words properly when they read out printed sentences may speak with PfVL in their spontaneous speech. On the other hand, learners who occasionally speak with PfVL in their spontaneous speech rarely do so when they read out printed sentences. Thus, we conclude that (1) PfVL is a characteristic phenomenon in spontaneous speech; (2) PfVL is not just a matter of phonological knowledge and/or skills of pronunciation.

Utterance type	PfVL
Reading-out (Question)	0
Spontaneous (Answer)	46

#### 4.2.2 Syntactic Status of Words with PfVL

It is shown that words often pronounced with PfVL are Subject Nouns, Transitive Verbs, BE-Verbs, Auxiliaries, and Conjunctions, while words pronounced without PfVL are Object Nouns, Intransitive Verbs, Adverbs, Adjectives, Prepositions, and Articles. This distributional tendency suggests that PfVLs tend to occur at a word before a syntactically obligatory element.

syntactic status	frequency	items
subject pronoun	19 (41.3%)	<i>I</i> (19)
transitive verbs	10 (21.7%)	<i>have</i> (2), <i>has</i> , <i>study</i> , <i>keep</i> , <i>like</i> , <i>try</i> , <i>think</i> , <i>work</i> , <i>read</i>
conjunctions	6 (13.0%)	<i>and</i> (5), <i>because</i>
<i>be</i> -verbs	5 (10.9%)	<i>is</i> (4), <i>am</i>
auxiliaries	4 (8.7%)	<i>will</i> (2), <i>has</i> , <i>not</i>
others	2 (4.3%)	<i>of</i> (of in "a lot of"), <i>as</i> (as in "as for")

#### 4.2.3 Japanese prosodic phenomenon in spontaneous speech

There is a Japanese prosodic phenomenon in spontaneous speech as a possible source. In spontaneous Japanese speech, vowels at the end of some phrases (smaller chunks than a sentence) are often pronounced with prosodic prominence and/or prosodic lengthening (Sadanobu, 2006; 2007).

(5) M : Shigoto shi-ta atoni{i}:

work do-PST after  
*After (I) finished (my) work,*

W : Un

*yeah*

M : go-ji kara-sa{a}:

5 o'clock from-PRT  
*from 5 o'clock,*

W : Un

*yeah*

M : Juu-ji made{e}:

10 o'clock to  
*to 10 o'clock,*  
[M continues to tell a story]

(Sadanobu, 2007)



This phenomenon occurs when a speaker has to produce chunks of utterances bit by bit, especially in order to perform a “large-sized” speech act, such as storytelling or explaining, which needs multiple clauses. In spoken Japanese, lengthening and strengthening phrase-final vowels is not necessarily perceived as a disfluent speech behavior. Rather, it can be an effective strategy for marking certain discourse context.

#### 4.3 PfVL as a characteristic phenomenon in spontaneous speech

As reported above, PfVL should be understood as a characteristic phenomenon in spontaneous speech. In spontaneous speech, a speaker has to formulate what to say and how to say in real-time (Levelt, 1989). Moreover, utterance formulation/production should be done under a time-pressure (e.g., Silence should be avoided), so that speakers often start talking before they formulate enough (Clark, 1996). The distributional tendency of PfVLs could be explained in this light.

##### 4.3.1 Interpretation of the distributional pattern

We have seen that PfVLs tend to occur at a word right before a syntactically obligatory element. From the viewpoint of discourse structure, such a position in a sentence may be also a boundary between a discourse-topic/old-information/presupposition and a focus/new-information/assertion (cf. Lambrecht, 1994). Namely, phrases pronounced with PfVL convey information which is easy to access or formulate, while phrases after PfVL convey information hard to access or formulate. The fact that large portion of the PfVLs appear at a subject pronoun ‘I’ may be a consequence of the contextual bias for the students’ responding speech: most questions in the interaction practice are about respondents’ college life, personal experience, or opinion, so that a pronoun ‘I’ may be a word which conveys information easy to access or formulate in the discourse context. When a student face some difficulty with formulating utterance, (s)he put something easy to formulate first with PfVL and the syntactic feature of the word with PfVL signals “more to come.” In conclusion, we suggest that when Japanese learners of English face a difficulty with on-line formulation of the utterance, they adopt Japanese prosodic strategy, resulting in PfVLs.

#### 4.4 Summary of findings

Japanese learners of English occasionally speak with PfVLs, which cause problems in intelligibility of their speech. Although PfVLs are phenomena concerning issues of pronunciation, teaching proper phonological knowledge and/or pronunciation skills may not suffice for an effective solution to the problem. Some facts from a sample data study show that (1) Japanese prosodic phenomenon in spontaneous speech as a possible source, (2) PfVL is a characteristic phenomenon in spontaneous speech, and (3) PfVL

tend to occur at particular kinds of syntactic/discursive slot. Thus, for Japanese learners of English, PfVL can be seen as a reasonable strategy for constructing talk on-line, which is pervasive in their first language. Namely, when the speakers face a difficulty with on-line formulation of the utterance, they adopt Japanese prosodic strategy, resulting in PfVLs.

## V Conclusion

In this paper, we report on our project for collecting spoken utterances by English learners enrolled in Japanese university language courses. We described our rationale for such data collection efforts and described our equipments for speech recordings and transcription annotation tools. We then discussed potential year-to-year differences and within-year changes among students proficiency of spoken English. Then we presented a case study of some particular aspects of relatively spontaneous utterances by the Japanese learners, which we believe reflect more strongly discourse strategies among native speakers of Japanese than pronunciation characteristics of Japanese learners of English, although the two are in ways intermingled..

## Acknowledgments

The data-collection efforts reported here were made possible in part by financial supports of JSPS/MEXT Grant-in-Aid Basic Research (B) Project Number 21320109 entitled Additional Data Collection and Exploratory Analysis of Elicited Utterances by Japanese Learners of English (April, 2009 through March 2014) and Grant-in-Aid Basic Research (B) Project Number 18320093 entitled Construction and Analysis of Spoken Corpus of Profiled Japanese Learners of English (April, 2006 through March 2009). Assemblage of the digital audio recording system was made possible by Waseda University Grant for Special Research Projects 2004A-033 entitled Study on Learner Profiling with Standardized Tests for Improving Undergraduate English Education (July 2004 through March 2005). Experience in use of this equipment was obtained through a pilot project partially funded by Waseda University Grant for Special Research Projects 2005B-022 entitled Study on Learner Profiling and Multimodal Corpus for Improving Undergraduate English Education (July 2005 through March 2006). The principal researcher gained much insight into the nature of spoken language tests and importance of face-to-face oral interaction practice through a joint research project between Media Network Center of Waseda University and KDDI group mainly conducted in 1999 through 2002.

## Notes

<sup>1)</sup> Other educators and researchers discuss similar or related issues found in spontaneous speeches by Japanese learners of English. See Rose (2008) and Hosoda (2007), among others.

## References

- Austin J. L. (1962). *How to Do Things with Words*. London: Oxford University Press.
- Clark, H. (1996). *Using Language*, Cambridge: Cambridge University Press.
- Harada, Y., Maebo, K., & Kawamura. (2007). VALIS 2.0: Transcription of What was (not) Uttered. *IPSJ SIG Technical Reports 2007-CE-90 (1)*, 2007 (69), 1-8.
- Harada, Y., Maebo, K., Kawamura, M., Maeno, J., Kusumoto, N., Suzuki, Y. & Suzuki M. (2007). VALIS: {Vast | Versatile} {Accumulative | Autonomous | Academic} {Learner | Learning | Language} {Information | Interaction} {Storage | System}. *IPSJ SIG Technical Reports 2007-CE-88*, 2007 (12), 169-176.
- Harada, Y., Maebo, K., Kawamura, M., Suzuki M., Suzuki, Y., Kusumoto, N., & Maeno, J. (2008). Toward Construction of a Corpus of English Learners' Utterances Annotated with Speaker Proficiency Profiles: Data Collection and Sample Annotation. In T. Tokunaga and A. Ortega (Eds.), *LKP 2008, Lecture Notes in Artificial Intelligence (LNAI) 4938* (pp. 171-178). Berlin Heidelberg: Springer-Verlag.
- Hosoda, Y. (2007). Response Opportunity Space in Japanese Conversation and Its Transfer to Second Language Conversation (in Japanese). *Jimbun Kenkyu*, 163, 201-226.
- Huang, C. R., Cheung, W., Harada, Y., Hong, H., Skoufaki, S. & Chen, H. K. Y. (2010). English Learner Corpus: Global Perspectives with an Asian Focus. In T. Kao & Y. Lin (Eds.), *A New Look at Language Teaching and Testing: English as Subject and Vehicle: Selected Papers from the 2009 LTTTC International Conference on English Language Teaching and Testing March 6-7* (pp. 85-117). Taipei: The Language Training & Testing Center.
- Grice P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and Semantics, 3: Speech Acts* (pp. 22-40). New York: Academic Press.
- Kawamura, M., Harada, Y., Maebo, K., Kusumoto, N. & Maeno, J. (2008). VALIS: Distributed Network Support for Annotation and Constrained Sharing of Utterance Data Obtained from Japanese College Learners of English (in Japanese). *IPSJ SIG Technical Reports 2008-CE-93*, 2008 (13), 155-162.
- Lambrecht, K. (1994). *Information Structure and Sentence Form: Topic Focus and the Mental Representation of Discourse Referents*. Cambridge: Cambridge University Press.

- Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. Cambridge: The MIT Press.
- Pearson (2003). *Versant English Test Can-Do Guide*, Palo Alto, CA: Pearson
- Pearson (2007). *Versant English Test - Test Description and Validation Summary*. Palo Alto, CA: Pearson.
- Pearson (2007). *Consistency of Versant English Test over Multiple Administrations*. Unpublished manuscript, Palo Alto, CA: Pearson.
- Rose, R. (2008). Filled Pauses in Language Teaching: Why and How. *Bulletin of Gunma Prefectural Women's University*, 29, 47-64.
- Sadanobu, T. (2006). Bunsetsu to Bun no Aida: Matsubiage o Megutte. In Onsei Bunpo Kenkyukai (Ed.), *Bumpo to Onsei 5*, (pp. 107-133). Tokyo: Kuroshio.
- Sadanobu, T. (2007). Shizen Hatsuwa Deeta ni motozuku Nihongo Modoshitsuki Matsubiage no Kansatsu. Beijing China-Japan Forum on Education and Culture, September 22nd.
- Tono, Y., Izumi, E. and Kaneko, T. (2004). *The NICT JLE Corpus: the final report*. JALT 2004 Conference Proceedings. JALT.
- Tono, Y., Kaneko, T., Isahara, H., Saiga, T. and Izumi, E. (2001). The Standard Speaking Test (SST) Corpus: A 1 million-word spoken corpus of Japanese learners of English and its implications for L2 lexicography. In S. Lee (Ed.), *ASIALEX 2001 Proceedings: Asian Bilingualism and the Dictionary: The Second Asialex International Congress, August 8-10, 2001, Yonsei University, Korea*, (pp. 257-262).
- Tree, F., and Clark, J. E.; H. (1997) Pronouncing "the" as "thee" to signal problems in speaking. *Cognition*, 62, 151-167.