

# A CONSIDERATION OF THE DIFFERENCE BETWEEN THE SPOKEN AND WRITTEN ENGLISH OF NATIVE SPEAKERS AND JAPANESE LEARNERS: A CORPUS-BASED STUDY

*Shin'ichiro Ishikawa*

## Abstract

It is often said that speeches and writings vary greatly with regard to vocabulary and grammar. However, how these differences can be seen in language use by English native speakers and non-native speakers has not been wholly elucidated. The current study, using the *International Corpus Network of Asian Learners of English* (ICNALE), quantitatively compares topic-controlled speeches and writings by native speakers and Japanese learners of English. Our learner-corpus-based analyses revealed that the difference is not as substantial as widely believed for native speakers in terms of highly frequent words, frequency of eleven textual indices, statistical positioning of individual samples, clustering structure of the indices, and the relationship between the production mode and the indices.

## Key words

speeches, writings, grammar, vocabulary, production modes, learner corpus

## 1 Introduction

Halliday (1985) suggests that speeches and writing are “not just alternative ways of doing the same things; rather, they are ways of doing different things” (ibid.: vii). It is widely believed that speeches and writings differ, but to which extent and in which ways have not necessarily been clarified. This is largely due to the fact that English studies have traditionally paid attention to the written mode. Halliday also insists that conventional grammar studies have been about “a grammar of written language” as something that exists. They do not appropriately outline a grammar of spoken language as something that happens or as a flow of conscious experience (Chafe 1994). Thus, as Carter and McCarthy (2006) state, most grammar books published to date “have had a bias towards the written language” and “the spoken language has been downgraded and has come to be regarded as relatively inferior to written manifestations” (ibid.: 9).

However, as mentioned by McEnery and Hardie (2012: 84-85), more and more linguists have come to be interested in the possibility that speeches have their own vocabulary and grammar system. Concerning vocabulary, Leech, Rayson and Wilson (2001) analyzed the *British National Corpus* (BNC) and

revealed that the frequencies of words vary greatly in speeches and writings. Lee (2001) also compared vocabularies in varied domains and genres in the *BNC Sampler Corpus* and concluded that spoken texts “use a fairly limited range of vocabulary, most of which come from a restricted selection from the common core of the language” (ibid.: 272). According to his analysis, vocabulary variability indices based on type/token ratios are 0.02 for speeches and 0.05 for writings, and the ratios of core lexical items are 90.52 per cent for speeches and 74.90 per cent for writings. Pointing out that speeches progress every moment and they are closely bound to situational contexts, Imura (2002) paid attention to the features of speech vocabulary. By analyzing lexical features seen in English film transcripts, which he regards as ‘pseudo-authentic’ spoken materials, he revealed that speech vocabulary is characterized by striking overuse of pronouns, contractions, interjections, and discourse markers, as well as by deviant use of several lexical phrases. Thus, he insists that in addition to written vocabulary, spoken vocabulary needs to be appropriately taught to EFL learners.

With regard to grammar, Brazil (1995) insisted that speeches are characterized most conspicuously by linearity in syntactic structure and situation dependency; thus, “a linear grammar” of speech can be regarded as an independent linguistic system.

“Speech is an activity that takes place in time: speakers necessarily say one word, follow it with another and then with another, and so on. There are obvious and well-recognized difficulties in reconciling this increment-by-increment presentation of speech with a hierarchical constituent-within-constituent account of how language is organized (ibid.: 4).”

In Brazil’s framework, speeches are essentially purposeful, interactive, and cooperative. In speeches, communicative value of lexis is determined by discourse (O’Grady 2010). Taking a similar standpoint, Carter and McCarthy (2006) suggested that learners should be familiar with English spoken grammar in addition to its written counterpart in order to understand how native speakers use English in speeches as well as in writings.

Meanwhile, some scholars pose a doubt on the concepts of speech vocabulary or speech grammar as an independent system. Admitting “striking differences of frequency” of several grammatical categories in spoken and written corpora, Leech (1998) nevertheless emphasized that speeches and writings form a kind of continuum and “one English grammar” can be applied both to speeches and to writings. He suggested that differences of frequency are mainly attributed to “the use of the grammar, rather than to the grammatical system itself.” In addition, recent grammar models such as pattern grammar (Hunston & Francis 2000) or

a linear unit model of grammar (Sinclair & Mauranen 2006), both of which are derived from the speech grammar model proposed by Brazil (1995), are seen as applicable to both speeches and writings.

With this in mind, do we really have speech vocabulary and speech grammar and do they need to be taught as independent systems in EFL contexts? In order to evaluate this, it is necessary to examine how speeches and writings are different for English native speakers and for learners of English. If the difference is clearly observed in language use by the former, and not in language use by the latter, this would give strong support to the teaching of speech vocabulary and grammar in classrooms.

When discussing differences between speeches and writings, objective criterion should be used for comparison. Therefore, in the analysis at hand, we pay attention to highly frequent lexical items as well as eleven kinds of textual indices, which concern lexical, syntactic, anaphoric, logical, and pragmatic dimensions of language use.

## **2 Research design**

### **2.1 Aims and research questions**

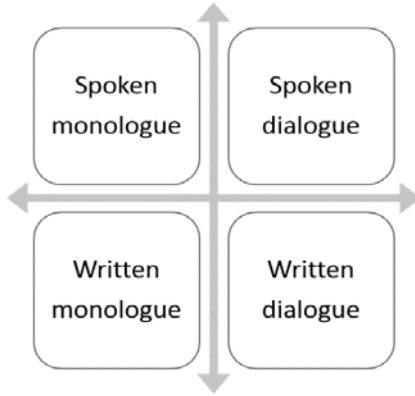
The principal aim of the current study is to examine the difference between speeches and writings by English native speakers (ENS) and Japanese learners of English (JLE) so as to reexamine the view that an independent lexical and syntactic system specific to speeches exists, and that it should be taught in EFL contexts. We compare speeches and writings based on the use of highly frequent lexical items and the eleven kinds of textual indices.

Five research questions are posed: Are speeches and writings by native speakers and Japanese learners of English different from each other in terms of highly frequent words? (RQ1); Are they different in terms of frequency of eleven textual indices? (RQ2); Are they different in terms of statistical positioning of individual samples based on frequency of the indices? (RQ3); Are they different in terms of clustering structure of the indices? (RQ4); and Are they different in terms of relationship between production mode and the indices? (RQ5)

### **2.2 Definitions of speeches and writings to be analyzed in the current study**

Before beginning the comparison, it is necessary to clarify what we mean by the words ‘speeches’ and ‘writings’. The current study only deals with spoken and written monologues among four types of possible linguistic productions: spoken monologues (e.g. reading aloud, speeches, and presentations), spoken dialogues (e.g. conversations, chats, discussions), written monologues (e.g. essays, diaries), and written dialogues (e.g. correspondence, online chats and discussion).

When discussing the difference between speeches and writings, many of the previous studies have compared spoken dialogues and written monologues. However, the analytical focus of the current study is not on the difference in communicative directionalities (i.e. monologues or dialogues) but on that in linguistic production modes (i.e. speeches or writings).



**Figure 1: Typology of speeches and writings**

It is clear that the current study deals with a limited type of spoken and written production data and therefore its findings may not be directly applicable to the traditional discussion about the difference between speeches and writings in general.

### **2.3 Data**

When comparing different sets of text samples, maximum attention should be paid to homogeneity in the data. If conditions are not appropriately controlled between speeches and writings and/or between native speakers and Japanese learners, it could be extremely difficult to interpret the results of comparisons. There is always a possibility that apparent differences in production modes are in fact due to the difference in topics and other parameters concerning linguistic production.

In order to control this, the *International Corpus Network of Asian Learners of English* (ICNALE) is used in the study. The ICNALE is a large collection of controlled speeches and essays by English native speakers and college students in ten Asian countries and areas. One striking feature of the ICNALE is that both speeches and writings are based on the same common topics: “It is important for

college students to have a part time job” and “Smoking should be completely banned at all the restaurants in the country.” Participants are required to speak or write about these two topics. The study at hand only examines the part-time job speeches in order to minimize the possibility of topic influence. Other conditions are also controlled in each of the spoken and written modules. For example, participants are required to speak as much as possible in the duration of sixty seconds, and to write 200 to 300 words in twenty to forty minutes. Too short or too long speeches and writings are excluded from the data. These careful controls in the data collection process make the ICNALE a highly reliable database for varied comparative studies (Ishikawa 2011, Ishikawa 2012, Ishikawa 2013, Ishikawa 2014).

The quantitative detail of the dataset used for the current analysis is shown in Table 1.

	ENS Speeches	ENS Writings	JLE Speeches	JLE Writings
Samples	50	100	80	80
Tokens	7,993	22,758	6,125	18,063
Types	957	2,094	554	1,323
Tokens/ Sample	159.86	227.58	76.56	225.79

**Table 1: Outline of the data used for the analysis**

Although there are no overlaps between speakers and writers, which is a measure taken in order to eliminate possible repetition effects, participants’ basic backgrounds are adjusted to be the same in both of the spoken and written modules. For example, native speakers’ mother tongues comprise approximately 60 per cent of American English and 40 per cent of British English. Likewise, L2 proficiency bands of Japanese learners, which are classified on the basis of the *Common European Framework of Reference* (CEFR), comprise 17.5 per cent of A2 (Waystage), 42.5 per cent of B1 lower (Threshold Low), 22.5 per cent of B1 upper (Threshold High), and 17.5 per cent of B2+ (Vantage or higher). Learners’ proficiency bands are determined from their scores in the TOEIC or TOEFL tests (Ishikawa 2013).

Although the use of the ICNALE enables us to conduct a highly reliable comparison of the linguistic production modes, we should also note that speeches and writings analyzed here might not be wholly natural and spontaneous.

## 2.4 Analytical procedure

To address the question posed in RQ1, the top twenty frequent words, which summarize varied features of individual texts, are analyzed.

To investigate RQ2, we examine eleven kinds of textual indices, which concern lexical, syntactic, anaphoric, logical, and pragmatic dimensions of texts. The outlines of the indices are shown in Table 2.

Indices	Definitions or Examples
<i>LEXICAL</i>	
Mean word length [WL]	Number of letters per word
Lexical variety [LV]	Herdan's C (log of types divided by log of tokens)
<i>SYNTACTIC</i>	
Mean sentence length [SL]	Number of words per sentence
<i>ANAPHORIC</i>	
Freq. of demonstrative pronouns [PN]	<i>this, that, these, those</i> (4 words)
Freq. of enumerative catch-all nouns [EN]	<i>approach, aspect, category, challenge, characteristic, circumstance, class, difficulty, event, experience, facet, fact, factor, etc.</i> (66 words)
<i>LOGICAL</i>	
Freq. of phrase-level conjunctions [CJ]	<i>and, but, or</i> (3 words)
Freq. of cohesion markers [CH]	<i>also, again, besides, furthermore, in addition, likewise, moreover, as well as, thus, therefore, as a result, consequently, for this reason, etc.</i> (95 words)
Freq. of complex prepositions [PR]	<i>according to, ahead of, along with, aside from, together with, because of, due to, contrary to, except for, instead of, prior to, subsequent to, etc.</i> (36 words)
<i>PRAGMATIC</i>	
Freq. of formal hedges [FH]	<i>about, according to, actually, apparent, apparently, approximate, approximately, broad, broadly, clear, clearly, comparative, etc.</i> (36 words)
Freq. of informal and conversational hedges [IH]	<i>almost, at all, at least, basically, dead, enough, few, hardly, just, little, only, pretty, quite, actually, in a way, kind of, maybe, like, etc.</i> (37 words)
Freq. of conversational exaggeratives [CE]	<i>absolute, absolutely, a lot, always, amazing, amazingly, awful, awfully, bad, badly, by all means, certain, certainly, clear, clearly, etc.</i> (68 words)

**Table 2: Textual indices analyzed in the current study**

The indices include three widely used quantitative measures (WL, LV, SL), which concern the overall use of the vocabulary, while the remaining indices concern the frequency of eight kinds of different lexical sets. Among these frequency-based indices, CH is based on Dorchies (2005), while the others are based on the work of Hinkel (2004).

The frequency-based analytical approach has several limitations. First, the five dimensions considered here cover many of the features concerning the difference between speeches and writings, but do not address all. Second, the words included in each lexical set, which are chosen from previous studies, are neither comprehensive nor mutually exclusive. Third, the occurrence of a word or a phrase does not necessarily mean that it is used with a particular intended function. However, measuring varied indices in a combination, without depending on a few indices, seems to assure a greater level of reliability.

## 2.5 Statistical measures

In the present study, we conduct three kinds of statistical analyses: correspondence analysis for observing the positioning of individual spoken and written samples (RQ3), cluster analysis to examine internal clustering structure of the eleven indices (RQ4), and binary logistic regression modelling to estimate the relationship between the production mode (namely speeches or writings) and the indices (RQ5).

Correspondence analysis is an exploratory method used to visualize the ‘correspondence’ between rows and columns of a two-way frequency cross-tabulation table. By reshuffling the rows and columns, the method identifies a few (often two) dimensions to maximize the correlation between rows and columns, and places all the variables and/or cases on the scatter plot, usually with the first dimension ( $Z_1$ ) as a horizontal axis and the second dimension ( $Z_2$ ) as a vertical axis. Correspondence analysis deals with the so-called item-category data: we regard the values of the eleven indices as item 1, and individual spoken and written samples as item 2.

Cluster analysis is another explanatory method, and its function is to visualize the relations among variables or cases. In the current study, we adopt cluster analysis to classify the eleven indices as variables based on their internal affinities. The initial distance is defined as the square root of  $(2-2r)$ , and the distance between clusters is calculated by the Ward method.

Binary logistic regression is a method to formularize the relationship between a set of independent variables and a binary dependent variable. In our analysis, the independent variables are eleven indices, and dependent variable is either 0 (spoken) or 1 (written). We adopt a stepwise method to choose significant independent variables and to obtain a best fitting model (Pin/Pout=.2).

### 3 Results and discussion

#### 3.1 Difference in speeches and writings in terms of highly frequent words (RQ1)

First, we examine the top twenty highest frequency words in speeches and writings by native speakers and Japanese learners. The list is shown in Table 3. As these twenty words cover 39.02 per cent and 39.17 per cent of all the words in speeches and writings by native speakers, and 45.08 per cent and 42.31 per cent of those in speeches and writings by Japanese learners, they are expected to reflect varied lexical and syntactic features of the whole texts.

	ENS Speeches	ENS Writings		JLE Speeches	JLE Writings
1	to	to		I	to
2	a	a		to	a
3	I	the		time	time
4	they	time		is	I
5	and	and		and	they
6	it	I		job	is
7	time	that		part	part
8	the	of		the	job
9	that	part		we	for
10	have	have		have	the
11	part	is		for	and
12	you*	job		they	have
13	in	in		it	students*
14	for	for		money	of*
15	students	students		so*	college*
16	job	college*		a	money
17	of	it		can*	that*
18	their	they		this*	it
19	is	work		important*	in*
20	work	their		think*	we

**Table 3: Top twenty words in speeches and writings by native speakers and Japanese learners**

NB: Words with asterisks are included in the top twenty words only in speeches or in writings. Several words are presented in the prompt: "It is important for college students to have a part time job."



An important finding that emerged from this data is that nineteen of the top twenty words are shared in speeches and writings by native speakers. This number of shared words decreases to fifteen in the case of Japanese learners. The result exemplifies that the difference between speeches and writings is not observed as clearly as generally expected for native speakers. In addition, it should be noted that the difference seems to be more salient for Japanese learners.

### 3.2 Difference in speeches and writings in terms of frequency of eleven textual indices (RQ2)

Next, we examine the frequency of the eleven indices. Results are shown in Table 4. The values, with the exception of WL, LV, and SL, represent the averages of the number of occurrences per one sample.

	Lexical		Syn.	Anaphoric		Logical			Pragmatic		
	WL	LV	SL	PN	EN	CJ	CH	PR	FH	IH	CE
ENS_Sp	4.28	0.88	27.93	4.06	1.44	5.92	4.26	0.28	1.08	2.90	3.62
ENS_Wr	4.26	0.88	27.80	6.61	2.31	7.66	4.71	0.26	1.25	3.84	5.65
JLE_Sp	4.08	0.87	19.84	1.80	0.93	2.94	1.38	0.11	0.26	1.05	2.66
JLE_Wr	4.23	0.86	14.30	4.66	3.34	7.13	5.04	0.28	1.09	2.55	6.26

**Table 4: Frequencies of the eleven indices in speeches and writings**

Here, we examine the average speeches/writings (S/W) ratios in five dimensions. Ratios close to 1.0 indicate that speeches and writings are highly similar, while ratios that deviate from 1.0 indicate that they are more different from each other. Five kinds of S/W ratios are shown in Figure 2.

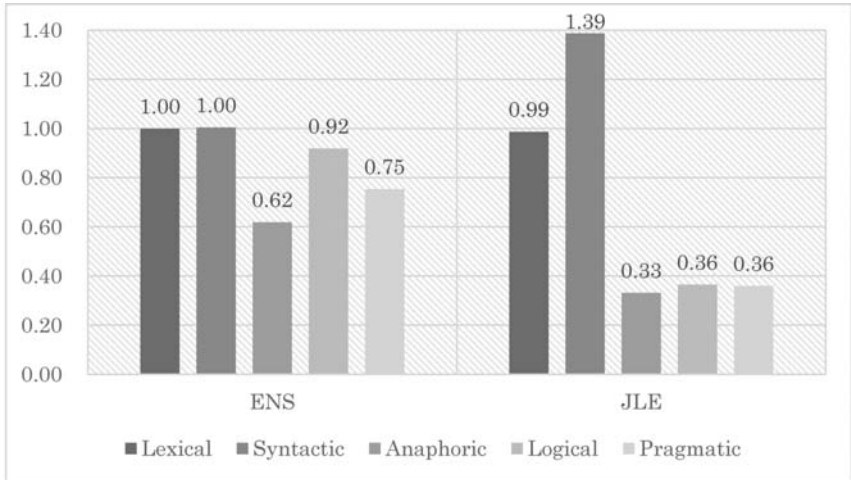


Figure 2: Average S/W ratios in five dimensions

As can be seen in Figure 2, in the case of native speakers, the S/W ratios are generally close to 1.0. Of particular note is that the difference is hardly seen in lexical, syntactic, and logical dimensions, and the overall average ratio reaches 0.84. Meanwhile, the ratios are much more deviant from 1.0 in the case of Japanese learners. In particular, the discrepancies are largest in anaphoric, logical, and pragmatic dimensions, and the average ratio is as low as 0.56.

Our analysis shows that the difference between speeches and writings is not so significant for native speakers in terms of the eleven indices concerning five major textual dimensions, as well as among the use of the highly frequent words examined in the previous section. However, the difference seems to be larger for Japanese learners.

### 3.3 Difference in speeches and writings in terms of statistical positioning of individual samples based on frequency of the indices (RQ3)

Our observations have consistently suggested that the difference between speeches and writings is rather ambiguous for native speakers. Following this, how are individual spoken and written text samples divided when based on frequency of the eleven indices? The results of correspondence analysis are shown in Figures 3 and 4. Two dimensions, Z1 and Z2, are both significant for native speakers as well as for Japanese learners ( $p < .001$ ).

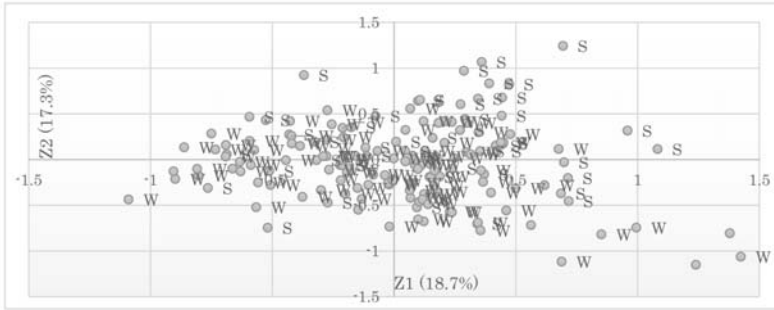


Figure 3: Scatter plot for speeches and writings by ENS

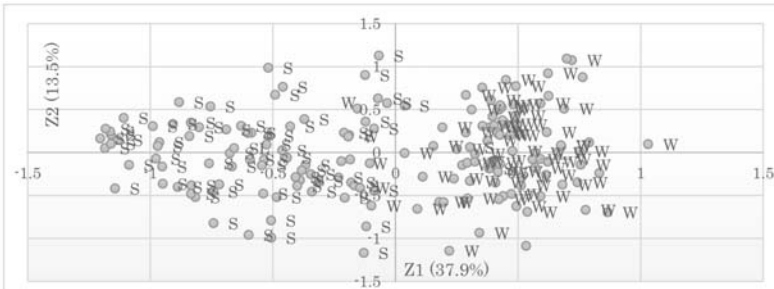


Figure 4: Scatter plot for speeches and writings by JLE

$Z_1$  and  $Z_2$  axes represent the first and the second staple factor in data classification. In the case of native speakers, speeches and writings are not clearly classified by the  $Z_1$  axis or by the  $Z_2$  axis. On the other hand, in the case of Japanese learners, speeches and writings are divided quite neatly into the left half and the right half of the scatter plot by  $Z_1$  axis, which explains 37.9 per cent of correspondence in the given dataset.

The two scatter plots corroborate the observation that speeches and writings are not comprised of two independent language systems but are formed from an inseparable continuum for native speakers.

### 3.4 Difference in speeches and writings in terms of clustering structure of the indices (RQ4)

How are the eleven indices mutually related and internally structured in speeches and writings by native speakers and Japanese learners? The results of cluster analyses are shown in Figures 5 to 8.

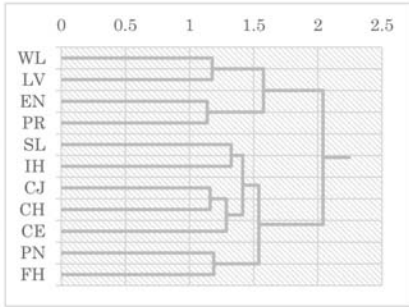


Figure 5: Dendrogram for speeches by NS

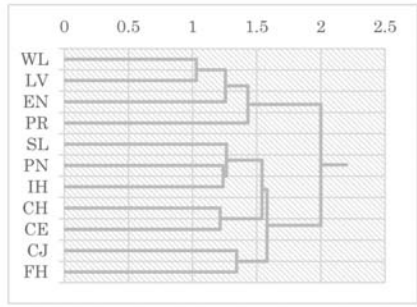


Figure 6: Dendrogram for writings by NS

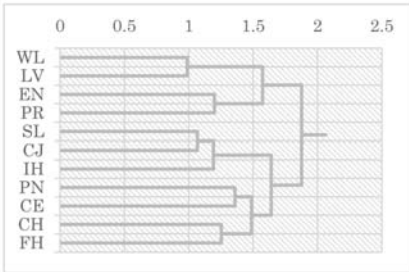


Figure 7: Dendrogram for speeches by JLE

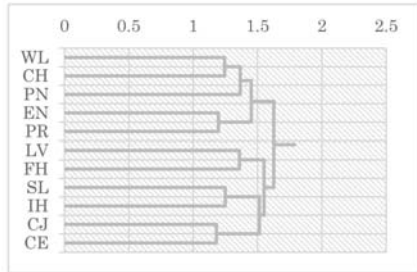


Figure 8: Dendrogram for writings by JLE

In interpreting the results of a cluster analysis, a cut-off point is usually determined at a longest branch in a dendrogram. Applying this rule to the data of native speakers, the eleven indices are classified neatly into two major clusters in both speeches and writings, and constituents of these two clusters are all identical: one includes WL, LV, EN, and PR, while the other includes the remaining seven indices. Such parallelism illustrates the similarity or oneness of speeches and writings by native speakers. Meanwhile, in the case of Japanese learners, such similar parallelism does not exist. The indices are divided into two clusters in speeches, while they are divided into much smaller clusters in writings.

### 3.5 Difference in speeches and writings in terms of the relationship between the production mode and the indices (RQ5)

Finally, focusing on logistic regression models, we investigate how the production mode can be related to the use of the indices. As the model did not converge well, we used ten indices, excluding CJ, as independent variables for

analysis of Japanese learner data. Six indices were chosen as significant in each of the two models, which are statistically significant ( $p < .001$ ).

$$(ENS) Y = 2.11WL + 0.40PN + 0.58EN + 0.40CJ + 0.17IH + 0.49CE - 17.09$$

$$(JLE) Y = -0.17SL + 1.22EN + 1.13CH - 1.75PR + 0.97IH + 0.68CE - 6.93$$

The accuracy rate of classification by the proposed models is 86.00 per cent for native speakers and 94.38 per cent for Japanese learners, which supports our finding that the difference between speeches and writings is less clear for native speakers than for Japanese learners.

Another interesting fact is that speeches and writings are characterized by different textual indices for native speakers and for Japanese learners. The overlap in chosen variables is seen only with EN, IH, and CE. Key variables determining the production mode also vary. The contribution of independent variables is usually measured by odds ratios (OR), which is defined as  $p / (1-p)$ . For native speakers, the key variable with the highest odds ratio is WL (OR=8.25), while it is EN (OR=3.38) and then CH (OR=3.09) for Japanese learners. This exemplifies the observation that writings are characterized by the use of longer words in the case of native speakers and by the use of more enumerative nouns and cohesion markers in the case of Japanese learners.

#### 4 Summary

In this study, we explored the difference between speeches and writings by native speakers and Japanese learners with a special focus on highly frequent words and eleven textual indices. Our findings are summarized below.

First, concerning RQ1 (highly frequent words), we found that the ratios of the top twenty words common to both speeches and writings are 95 per cent for native speakers, and 75 per cent for Japanese learners.

Second, concerning RQ2 (fluency of eleven textual indices), we found that the speeches per writings ratios are 0.84 for native speakers, and 0.56 for Japanese learners.

Concerning RQ3 (statistical positioning of individual samples), correspondence analyses demonstrated that speech samples and writing samples cannot be clearly separated for native speakers, but can be neatly divided in the case of Japanese learners.

With respect to RQ4 (clustering structure of the indices), cluster analyses showed that a high degree of parallelism is observed between speeches and writings in terms of internal affinity among the eleven indices for native speakers, but this is not so for Japanese learners.

Finally, concerning RQ5 (the relationship between the production mode and the indices), binary logistic regression modelling revealed that the difference between speeches and writings can be explained and estimated less accurately for native speakers. Additionally, the difference concerns the word length for native speakers, while degree of textual cohesion is a larger factor for Japanese learners.

A series of analyses have consistently proven that the difference between speeches and writings is not as significant as generally believed for native speakers, and it can be seen more clearly for Japanese learners. This observation might throw some doubt on the view that two kinds of English exist in a clearly separate way, namely spoken English and written English, and speech vocabulary and grammar should be taught independently even for novice and lower-intermediate learners in EFL conditions. Of course, we do not intend to deny the pedagogical importance of presenting varied types of English to learners, but it seems advisable for general L2 learners, especially those at a novice or an intermediate proficiency level, to pay their attention to continuity or wholeness of speeches and writings rather than to differentiate between the two.

What should be finally considered is why the difference between speeches and writings is more salient for Japanese learners in comparison to native speakers. There are several possible reasons for this. One is the limitation in overall L2 proficiency of Japanese learners who participated in the study. Pikulski and Templeton (2004) suggest that, for matured learners, a close relationship exists between vocabulary used in listening, speaking, reading, and writing. They also say that “people who have large speaking vocabularies generally tend to have large listening, reading, and writing vocabularies; likewise people who are limited in one of these aspects are likely limited in other aspects as well” (ibid.: 3). This conversely means that the pattern of vocabulary use and maybe grammar use as well tends to vary more easily according to the production modes for learners with relatively lower L2 proficiency.

Another is the limitation in the quantity of learners’ spoken productions. As shown in Section 2.3 above, native speakers produce an average of 159.86 words in speeches and 227.58 words in writings. The ratio of the amount of speeches to that of writings is 70.2 per cent. Meanwhile, Japanese learners produce an average of 76.56 words in speeches and 225.79 words in writings. The ratio is as low as 33.9 per cent, which may partly explain the relatively greater discrepancy between speeches and writings by Japanese learners.

Although the current study revealed many interesting facts about the difference between speeches and writings by native speakers and learners, there remains much to be improved in our analytical approach. Future research should

examine spoken dialogue in addition to spoken monologues in order to have a more complete picture of spoken English. Incorporating detailed semantic analysis into simple word count analysis would also be beneficial. These problems are addressed in future studies.

#### Note

This paper is based on the author's research projects supported by MEXT/JSPS KAKENHI Grant Numbers 25284104 and 24652120. This is a revised version of the manuscript read at the Sixth Brno Conference on Linguistics Studies in English 2014 (11-12, September, 2014) held at Faculty of Education, Masaryk University, Brno, Czech Republic. The author wishes to acknowledge two anonymous reviewers for their helpful comments to the manuscript.

#### References

- Brazil, D. (1995) *A Grammar of Speech*. Oxford: Oxford University Press.
- Carter, R. and McCarthy, M. (2006) *Cambridge Grammar of English: A Comprehensive Guide*. Cambridge: Cambridge University Press.
- Chafe, W. (1994) *Discourse, Consciousness, and Time. The Low and Displacement of Conscious Experience in Speaking and Writing*. Chicago: The University of Chicago Press.
- Dorchies, G. (2005) *Cohesion*. Retrieved from <http://www.clarkson.edu/writingcenter/docs/cohesion.pdf>.
- O'Grady, G. (2010) *A Grammar of Spoken English Discourse: The Intonation of Increments*. London and New York: Continuum.
- Halliday, M. A. K. (1985) *Spoken and Written Language*. Geelong, Victoria: Deakin University Press
- Hinkel, E. (2004) *Teaching Academic ESL Writing: Practical Techniques in Vocabulary and Grammar*. Mahwah, NJ and London: Lawrence Erlbaum Associates.
- Hunston, S. and Francis, G. (2000) *Pattern Grammar: A Corpus-Driven Approach to the Lexical Grammar of English*. Amsterdam: John Benjamins.
- Imura, M. (2002) 'Kopasu o katsuyoshita kogo eigo kyoiku kenkyu: Eiga serifu detabesu no kochiku to sono eigo kyoiku eno oyo.' Ph.D. Dissertation submitted to Osaka University. [A Study on Corpus-based Teaching of Colloquial English: Development of Application of Cinema Transcripts Database]
- Ishikawa, S. (2011) 'A new horizon in learner corpus studies: The aim of the ICNALE project.' In: Weir, G., Ishikawa, S. and Poonpon, K. (eds) *Corpora and Language Technologies in Teaching, Learning and Research*. Glasgow, UK: University of Strathclyde Publishing. 3-11.
- Ishikawa, S. (2012) *Beshikku Kopasu Gengogaku*. Tokyo: Hitsuji Shobo. [Basics of Corpus Linguistics].
- Ishikawa, S. (2013) 'The ICNALE and sophisticated contrastive interlanguage analysis of Asian learners of English.' In: Ishikawa, S. (ed.) *Learner Corpus Studies in Asia and the World, 1*. Kobe, Japan: Kobe University. 91-118.
- Ishikawa, S. (2014) 'Design of the ICNALE-Spoken: A new database for multi-modal contrastive interlanguage analysis.' In: Ishikawa, S. (ed.) *Learner Corpus Studies in Asia and the World, 2*. Kobe, Japan: Kobe University. 63-76.
- Lee, D. Y. (2001) 'Defining core vocabulary and tracking its distribution across spoken and written genres: Evidence of a gradience of variation from the British National Corpus.' *Journal of English Linguistics* 29/3, 250-278.

- Leech, G. (1998) 'English grammar in conversation.' In: Wilson, A. and Schmied, J. (eds) *Language Learning and Computers: Proceedings of the Chemnitz Symposium, 20-21 February 1998*. Retrieved from <https://www.tu-chemnitz.de/phil/english/sections/linguist/real/independent/llc/>
- Leech, G., Rayson, P. and Wilson, A. (2001) *Word Frequencies in Written and Spoken English: Based on the British National Corpus*. London: Routledge.
- McEnery, T. and Hardie, A. (2012) *Corpus Linguistics: Method, Theory, and Practice*. Cambridge: Cambridge University Press.
- Pikulski, J. J. and Templeton, S. (2004) *Teaching and Developing Vocabulary: Key to Long-Term Reading Success*. Boston, MA: Houghton Mifflin.
- Sinclair, J. McH. and Mauranen, A. (2006) *Linear Unit Grammar: Integrating Speech and Writing*. Amsterdam: John Benjamins.

**Shin'ichiro ISHIKAWA** is Professor of Applied Linguistics at Kobe University, Japan. He specialises in applied linguistics, corpus linguistics, and statistical analysis of language data. He is currently involved in the project to compile The International Corpus Network of Asian Learners of English (ICNALE), the aim of which is to collect controlled speeches and writings by varied learners of English in Asia. He has published many books and articles in the field of linguistics and language education.

**Address:** Shin'ichiro ISHIKAWA, Ph.D., School of Languages and Communication, Kobe University, 1-2-1, Tsurukabuto, Nada-ku, Kobe 657-8501, Japan. [e-mail: iskwhin@gmail.com]