The Inner Circle vs. the Outer Circle or

British English vs. American English

Yong-hun Lee
Chungnam Nation University
99 Daehak-ro, Gung-dong, Yuseong-gu
Daejeon 34134, S. Korea

yleeuiuc@hanmail.net

Ki-suk Jun
Hannam University
70 Hannam-ro, Ojoeng-dong, Daedeok-gu
Daejeon 34430, S. Korea

mango0322@naver.com

Abstract

In this paper, the use of two modals (*can* and *may*) in four varieties of English (British, India, Philippines, and USA) was compared and the characteristics of each variety were statistically analyzed. After all the sample sentences were extracted from each component of the ICE corpus, a total of twenty linguistic factors were encoded. Then, the collected data were statistically analyzed with R. Through the analysis, the following facts were observed: (i) India and Philippine speakers used *can* more frequently than natives, (ii) Three linguistic factors interacted with CORPUS, and (iii) The distinctions between American and British were more influential than those of the Inner Circle vs. the Outer Circle.

1   Introduction

As English has spread worldwide, new varieties of English have emerged and they got independent status accordingly. In order to systematically classify them, Kachru (1992) introduced the three concentric circles as way of conceptualizing this pluri-centricity. There should be a distinction between American English (AmE) and British English (BrE) as well.

Out of the varieties of English, we chose four different ones and statistically analyzed their properties. To this end, we picked out four components of the International Corpus of English (ICE; Greenbaum, 1996), which are the varieties of British, India, Philippines, and USA. Then, all the

sentences with two modal auxiliaries *can* and *may* were extracted. Then, a total of twenty linguistic factors were encoded to the extracted ones, and the encoded data were statistically analyzed with R, with the theoretical basis of Competition Model (Bates and MacWhinney, 1982, 1989). In addition, two statistical analysis methods were adopted. One was a logistic regression with which the properties of each component were closely investigated. The other was a Behavior Profile (BP) analysis where the four components were clustered by their similarity.

In short, we selected two modal auxiliaries *can* and *may* for comparison for the following reasons. As several of the previous studies (Leech, 1969, Coates, 1983; Collins, 2009) pointed out, these two modal verbs have similar meanings, and the native speakers interchange them in similar contexts. However, the distributions of these two are systematic, even in native speakers' writings. Then, what happens in non-native speakers' counterparts and how can the phenomena be explained? We are to present one possible type of answer to these questions.

2   Previous Studies

2.1   World Englishes

The term 'World Englishes', not 'World English', refers to emerging localized/indigenized varieties of English, especially the varieties which have developed in territories influenced by the United Kingdom (Great Britain) or the United States. The primary goals of World Englishes are (i) to identify

30th Pacific Asia Conference on Language, Information and Computation (PACLIC 30)
Seoul, Republic of Korea, October 28-30, 2016

339

the varieties of English in diverse sociolinguistic contexts and (ii) to analyze how the sociolinguistic factors (histories, multi-cultural backgrounds and contexts of function) influence the use of English in different regions of the world.

There are several theoretical models to explain the spread of English, but the three concentric circles model by Kachru is probably the most influential one. In this model, the spread of English is classified and grouped into three different categories of regional varieties of English. These three categories are called the Inner Circle, the Outer Circle, and the Expanding Circle (Kachru, 1992:356). Figure 1 illustrates the three concentric circles.
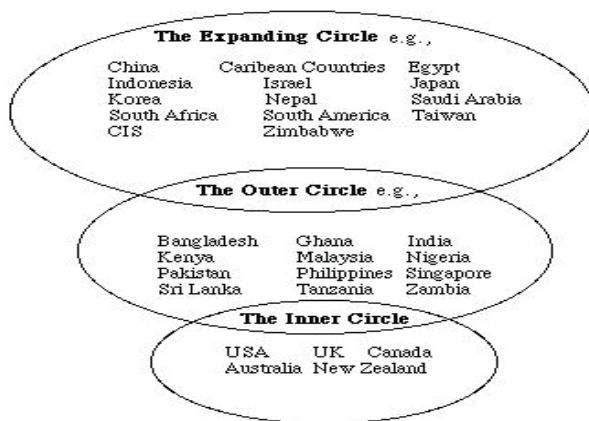

Figure 1: The Three Concentric Circles

The English varieties in each circle have their own characteristics.

The Inner Circle of English took shape first and spread across the world in the first diaspora. In this early spread of English, speakers from England carried the language to the colonies, such as Australia, New Zealand, North America, and so on. The English language in this circle represents the traditional historical and sociolinguistic bases in the regions where it is now used as English as the Native Language (ENL): the United Kingdom, the United States, Australia, New Zealand, Ireland, Canada, South Africa, and some of the Caribbean territories. In these countries, English is the native language or mother tongue for most people. The total number of English speakers in this circle is estimated to be as many as around 380 million.

The Outer Circle of English was made during the second diaspora of English, which diffused the language through the expansion of Great Britain.

In the areas such as Asia and Africa, English is not the native language, but it serves as a useful lingua franca between various ethnic and language groups. Some people with higher education, the legislature and judiciary, national commerce, and others may speak English for practical purposes. The countries in this circle include India, Nigeria, Bangladesh, Pakistan, Malaysia, Tanzania, Kenya, non-Anglophone South Africa, the Philippines and others. The total number of English speakers is estimated to range from 150 million to 300 million.

The Expanding Circle includes the countries in which English plays no historical or governmental role but is widely used as a medium of international communication. This includes much of the rest of the world's population not categorized as either of the other two circles: China, Russia, Japan, most of Europe, Korea, Egypt, Indonesia, etc. It is difficult to estimate the total number of people in the Expanding Circle, but the estimates range from 100 million to one billion.

2.2 British English and American English

In addition to the three concentric circles in Kachru (1992), one of the most influential classifications of English is that of British English and American English.

British English (BrE) refers to the form of English primarily used in the Great Britain, but it includes all the dialects used in other areas which were the former colonies of Great Britain. Likewise, American English (AmE) is the form of English mainly used in the United States, but it includes all the dialects used in other areas like the former colonies of the United States.

As the Great Britain expanded its territories by colonization, the United States of America (USA) also established a few colonies in Asian countries. Accordingly, English in these countries was influenced by its superpower. Nowadays, as the influences of the USA increased in many other countries, the importance of AmE increased as well.

English in Australia, Canada, Ireland and New Zealand belongs to BrE. In addition, most of Africa (including Egypt and South Africa), South Asia (Pakistan, India, and Bangladesh), Malta, some countries in Southeast Asia (Myanmar, Singapore, Malaysia, and Thailand), and Hong Kong still use BrE. On the other hand, most of Eastern Europe (including Russia), most East

Asian countries excluding Hong Kong (China, Japan, and Korea), Philippines, most American countries (except Canada, Jamaica and the Bahamas), and some African countries (Liberia and Namibia) still use AmE.

There have been quite a few studies on the differences between BrE and AmE (McArthur, 2002; Tottie, 2002; Crystal, 2003; Hargraves, 2003; Peters, 2004; Algeo, 2006; Trudgill et al. 2013). The differences between these two types of English cover various areas including phonetics, phonology, morphology, syntax, semantics, and so on. However, most of the previous studies were focused on lexical differences and did not adopt any statistical methods in their analyses.

2.3 Competition Model

The Competition Model (CM), on which this paper is theoretically based, is a psycholinguistic theory of language acquisition and sentence processing. This model was developed by Elizabeth Bates and Brian MacWhinney. The most important idea of the CM is that the meaning of a language must be and can be interpreted by comparing a number of linguistic factors within a sentence. In addition, a language is acquired and/or learned through the competition of basic cognitive mechanisms with a rich linguistic environment.

The CM claims that human beings understand the meaning of a sentence by taking into account various factors, such as word order, morphology, and semantic characteristics (e.g. animacy), and so on. Thus, when people articulate a sentence, they unconsciously calculate the probabilities of each meaning and choose the one with the highest value.

We adopted this model as a theoretical basis because two modal auxiliaries *can* and *may* occur in similar linguistic environments and that they compete with each other. As a result of the competition, one of them is chosen as a winner in the given linguistic environments. The winner has more probability than the other in the given environments. Then, the question is which factor would decide the winner. We investigated the decision mechanisms with a statistical analysis.

3 Research Method

3.1 Research Procedure

Our research proceeded as follows. First, four corpora were selected from the ICE: British, India, Philippines, and USA. Each corpus included about 1 million of word tokens, and the composition of each corpus was nearly identical. They are listed as in Table 1. Next, all the sentences with the two modal auxiliaries were extracted from the four corpora, using NLPTools (Lee, 2007).

	The Inner Circle	The OuterCircle
BrE	Britain	India
AmE	USA	Philippines

Table 1: Classification of Four Corpora

Since there were so many sentences in each variety, we extracted 1,000 sentences per each corpus with random sampling. Then, twenty different linguistic factors were manually encoded into them, following Deshors (2010) and Deshors and Gries (2014). Lastly, a statistical analysis of the corpus data was done with the help of R (R Core Team, 2016).

3.2 Encoding Variables

Table 2 illustrates the encoded factors, used in this paper. Following Atkins (1987), each linguistic factor and its level are called ID tag and ID tag levels.

ID Tag Type	ID Tag	ID Tag Levels
Data	CORPUS	Britain, India, Philippines, USA
Morphological	FORM	*can*, *may*
	ELLIPTIC	yes, no
	VOICE	active, passive
	ASPECT	simple, progressive, perfect
	MOOD	indicative, subjunctive
	SUBJMORPH	adj., adv., common noun, proper noun, relative pronoun, noun phrase, etc.
	SUBJPERSON	1, 2, 3
	SUBJNUMBER	singular, plural
	SUBJREFNUMBER	singular, plural
Syntactic	NEG	affirmative, negated
	SENTTYPE	declarative, interrogative
	CLTYPE	main, coordinate, subordinate
Semantic	SENSE	epistemic, deontic, dynamic
	SPEAKERPRESENCE	weak, medium, strong
	VENDLER	accomplishment, achievement, process, state

341

	VerbSemantics	abstract, general action, action incurring transformation, action incurring movement, perception, etc.
	RefAnim	animate, inanimate
	AnimType	animate, floral, object, place/time, mental/emotional, etc.
	Use	idiomatic, literal, metaphorical

Table 2: Encoded Factors and Predictors

The variables were used in the statistical analysis.[1]

3.3 Statistical Analysis

We also carried out a multi-factorial analysis, in which not only the effects of each factor but also the interactions among the factors are statistically analyzed. The multi-factorial analyses of linguistic data are supported by many studies in cognitive linguistics. Langacker (2000:3) mentioned that "to conceive of [linguistic] entities in connection with one another (e.g., for the sake of comparison, or to assess their relative position), not just as separate, isolated experiences. This is linguistically important because relationships figure in the meaning of almost all expressions, many of which (e.g., verb, adjectives, prepositions) designate relationships." Gries (2003) also conducted the multi-factorial analysis to analyze the distributions of particle placement in native speakers' English. Deshors (2014:11) also mentioned that "The multi-factorial approach also helps the authors make a connection between degrees of grammatical complexity of speakers' utterances and learners' lexical choices during second language production. For instance, they observe that *can* rather than *may* is more frequently used by French English learners (compared to native speakers) in more complex grammatical environments such as negated or subordinated linguistic contexts."

As a multi-factorial approach, we used a Generalized Linear Model (GLM) with logistic regression, since it is one of the simplest and most widely-adopted analyses. For regression analysis, Deshors (2014:11) mentioned that "Binary logistic regression is a confirmatory statistical technique that allows the analyst to identify possible correlations between the dependent and the independent factors/variables. Ultimately, this

statistical approach allows us to see what factors influence learners' choices of *may* and *can*."

During the analysis process, a stepwise model selection procedure was adopted as follows. First, an initial model was constructed with all of the factors and their interactions. Second, a new model was constructed in which only one factor or one interaction was deleted from the previous model. Third, the newly constructed model was compared with the previous one with an ANalysis Of VARiance (ANOVA). Fourth, an optimal model was chosen according to some criteria such as significance testing or information ones: If a model m_1 contained a factor f or an interaction i but a model m_2 did not contain f or i, and (i) when the p-value of the ANOVA test was significant ($p<.05$), it implied that the factor f or an interaction i must NOT be deleted from the model and the model m_1 was selected consequently, and (ii) when the p-value of ANOVA was NOT significant ($.05<p$), it implied that the factor f or an interaction i can safely deleted from the model and the model m_2 was selected accordingly. The processes continued until all the factors and their interactions were scrutinized.

We also adopted another multi-factorial analysis, a Behavioral Profile (BP) analysis. It was developed by Gries and Otami (2010) and Gries (2010a), and it is a statistical method to examine the behavioral properties of each linguistic factor. The analysis represents the similarity or dissimilarity of the components with a dendrogram (the hierarchical agglomerative cluster analysis). It was originally used to analyze the synonymy and/or the antonymy in lexical semantics. However, the same method can also be used here, since the use of the modal constructions in the EFL learners' writings can be classified on a basis of the behavioral properties of linguistic factors.

4 Logistic Regression

4.1 The Analysis

The first step for the (binary) logistic regression was to set up an initial model. Table 3 shows the initial model of our study.

FORM~CORPUS+NEG+SENTTYPE+CLTYPE+SUBJ MORPH+SUBJPERSON+SUBJNUMBER+VOICE+ASP ECT+MOOD+SUBJREFNUMBER+SENSE+SPEAKER

---

[1] This process is called *operationalization*.

PRESENCE+USE+VERBSEMENATICS+REFANIM+ANIMTYPE+CORPUS:NEG+CORPUS:SENTTYPE+CORPUS:CLTYPE+CORPUS:SUBJMORPH+CORPUS:SUBJPERSON+CORPUS:SUBJNUMBER+CORPUS:VOICE+CORPUS:ASPECT+CORPUS:MOOD+CORPUS:SUBJREFNUMBER+CORPUS:SENSE+CORPUS:SPEAKERPRESENCE+CORPUS:USE+CORPUS:VERBSEMANTICS+CORPUS:REFANIM+CORPUS:ANIMTYPE

Table 3: Initial Model

Then, model selection procedures were applied (cf. Section 3.3) and the final (optimal) model was selected. Table 4 shows the final model.

FORM~CORPUS+SUBJMORPH+MOOD+SENTTYPE+CLTYPE+VENDLER+CORPUS:SUBJMORPH+CORPUS:SENTTYPE+CORPUS:VENDLER

Table 4: Final Model

As seen in Table 3 and Table 4, the six main factors and three interactions with CORPUS survived in the final model.

4.2 Analysis Results

With the final model obtained, all the main factors and their interactions with CORPUS were statistically analyzed as in Table 5. Here, '×' (not significant) is used when $0.1<p$; '.' (marginally significant) when $p<0.1$; '*' (significant) when $p<0.05$; '**' (very significant) when $p<0.01$; and '***' (highly significant) when $p<0.001$.

	df	deviance	AIC	LRT	p	
<none>			29195	29575		
CORPUS	3	1352.7	1470.7	40.15	9.926e-09	***
ELLIPTIC	1	1313.7	1435.7	1.16	0.2816880	
VOICE	1	1312.5	1434.5	0.00	0.9696053	
ASPECT	3	1316.6	1434.6	4.06	0.2549911	
MOOD	1	1323.9	1445.9	11.36	0.0007513	***
SUBJMORPH	8	1315.7	1423.7	3.21	0.9202972	
SUBJPERS	2	1313.9	1433.9	1.37	0.5034411	
SUBJNUM	1	1313.3	1435.3	0.83	0.3623101	
SUBJREFNUM	1	1312.8	1434.8	0.25	0.6186114	
NEG	1	1315.7	1437.7	3.14	0.0765925	.
SENTTYPE	2	1324.7	1444.7	12.22	0.0022183	**
CLTYPE	2	1320.0	1440.0	7.53	0.0231573	*
SENSE	2	1972.1	2092.1	659.55	<2.2e-16	
VENDLER	3	1324.8	1442.8	12.25	0.0065658	**
VERBSEM	8	1323.0	1431.0	10.50	0.2318564	
REFANIM	1	1313.1	1435.1	0.55	0.4579886	
ANIMTYPE	20	1332.1	1416.1	19.56	0.4854248	
USE	1	1312.5	1434.5	0.02	0.8965201	
CORPUS:ELLIPTIC	3	23068	23442	0.0	1	
CORPUS:VOICE	3	22852	23226	0.0	1	
CORPUS:ASPECT	6	24293	24661	0.0	1	
CORPUS:MOOD	2	23573	23949	0.0	1	
CORPUS:SUBJMORPH	13	40801	41155	11606.1	<2e-16	***
CORPUS:SUBJPERS	6	24438	24806	0.0	1	
CORPUS:SUBJNUM	3	26744	27118	0.0	1	
CORPUS:SUBJREFNUM	3	24726	25100	0.0	1	
CORPUS:NEG	3	23140	23514	0.0	1	
CORPUS:SENTTYPE	3	41594	41968	12399.0	<2e-16	***
CORPUS:CLTYPE	6	27321	27689	0.0	1	
CORPUS:SENSE	6	1156	1524	0.0	1	
CORPUS:VENDLER	8	37557	37921	8362.1	<2e-16	***
CORPUS:VERBSEM	19	25375	25717	0.0	1	
CORPUS:REFANIM	3	21554	21928	0.0	1	
CORPUS:ANIMTYPE	36	1169	1477	0.0	1	
CORPUS:USE	0	29195	29575	0.0		

Table 5: Analysis Results

The table demonstrates that five main factors and three interactions with CORPUS were statistically significant in the model. It also shows that one factor (SUBJMORPH) survives in the final model because of its interactions with the factor CORPUS.

Since we obtained the final model, it was possible to investigate how the speakers' use of *can* and *may* was different in the four components of the ICE corpus, with graphic representations.

Among the main factors, only one factor (i.e., CORPUS) was examined with a graphic tool. Figure 2 illustrates the association plot for CORPUS. As shown in the figure, the effects of a factor are represented by the baseline (the dotted line) and rectangles above and below it. Here, the baseline refers to the expected frequency of each value for a given factor. The width of the rectangle is proportional to the square root of the expected frequency, and the width of the rectangle to the standardized residual.
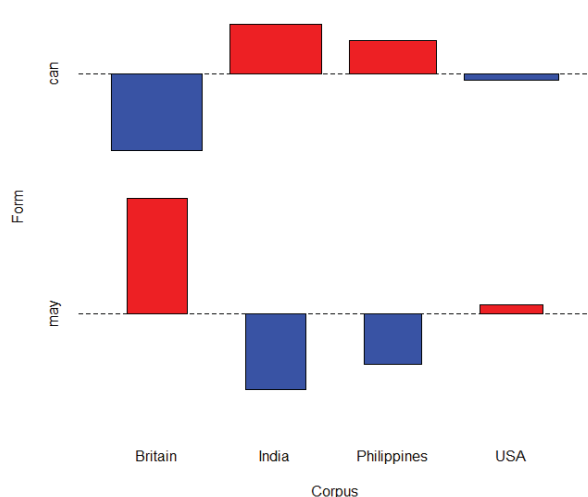

Figure 2: Association Plot for CORPUS

As this plot indicates, the ENL speakers (Britain and USA) use *may* more often and *can* less often than the ESL speakers (India and Philippines). In other words, the ESL speakers use *may* less frequently and *can* more frequently than the ENL speakers.

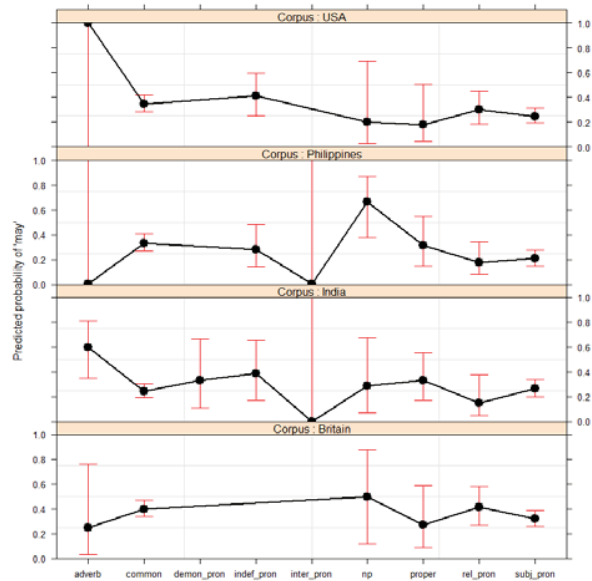Figure 3 illustrates the effect plot for CORPUS: SUBJMORPH.

Figure 3: Effect Plot for CORPUS:SUBJMORPH

This plot demonstrates several facts about the use of *can* and *may* by different groups of speakers. When the subject is an 'adverb' (i.e., *here* or *there* [existential constructions]), USA and India use *may* more frequently than *can*, while Britain and Philippines demonstrate the opposite tendency. When the subject contains a 'common noun', all the groups of speakers prefer to use *can*. When the subject includes an 'NP', the Philippines learners prefer to use *may*, while the other three groups of speakers prefer to use *can*. For the three types of pronouns ('demon_pron (demonstrative pronoun)', 'indef_pron (indefinite pronoun)', and 'inter_pron (interrogative pronoun)'), only the Indian ESL speakers used all of them, whereas all the other speakers employed only some of them. When the subject contains a 'proper noun', a 'relative pronoun', or a 'subject (personal) pronoun', all the groups of speakers prefer to use *can*.

Figure 4 demonstrates the effect plot for CORPUS:SENTTYPE. As you can observe, in both types of sentences, the ENL speakers and the ESL speakers prefer to use *can* rather than *may*, but the probabilities of *may* increase when SENTTYPE is 'declarative', in both groups of speakers.
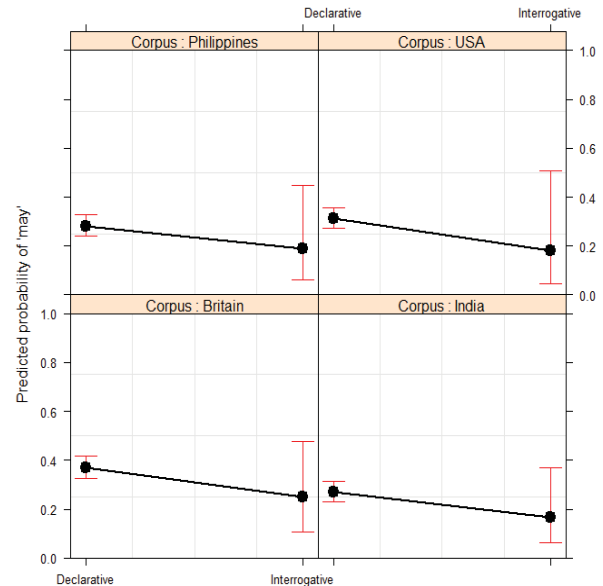

Figure 4: Effect Plot for CORPUS:SENTTYPE

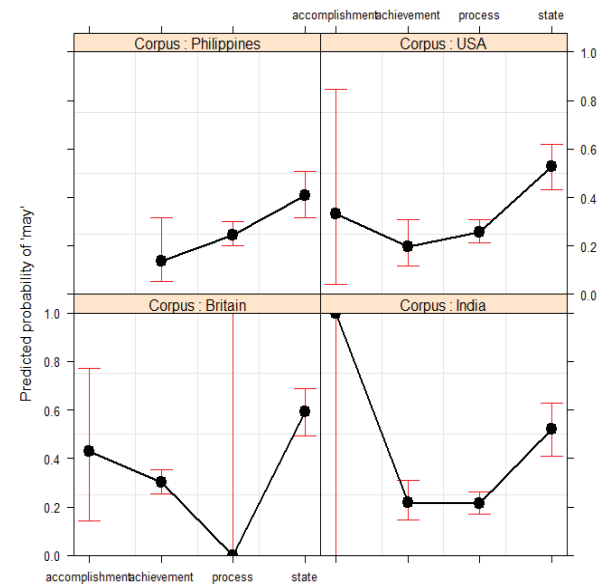Figure 5 shows the effect plot for CORPUS: VENDLER.

Figure 5: Effect Plot for CORPUS: VENDLER

This plot illustrates that all the groups of speakers prefer to use *may* more when the verbs represent 'accomplishment' or 'state' but that they prefer to

use *can* when the verbs represent 'achievement' or 'process'.

5   The BP Analysis

As the analysis results in Section 4 show, four groups of speakers demonstrated different characteristics in using two modal auxiliaries *can* and *may*. Then, the question was whether the Inner/Outer distinctions influenced more or the AmE/BrE distinctions influenced more. To get the answer, a BP analysis was performed.

Among the factors in Table 2, the combination of CORPUS and FORM were chosen as a dependent variable and the other factors as independent ones. Figure 6 illustrates the dendrogram resulting from the analysis (multiscale bootstrap resampling clustering).

Here, the horizontal lines represent which component(s) is/are grouped with which component(s), and the vertical lines indicate the distance between these two groups. Two numeric values in the dendrogram refer to AU (approximately unbiased) *p*-value and BP (bootstrap probability) value for each cluster, respectively.
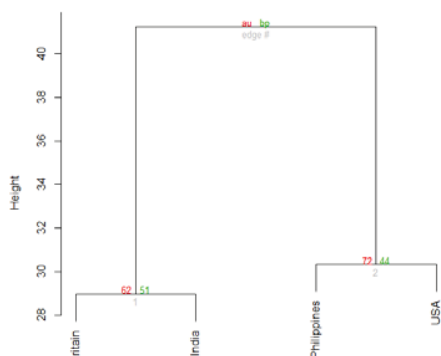

Figure 6: BP Analysis Result

This dendrogram represents which one is closer to which one.

As you can see, Britain and India were grouped together first. Likewise, Philippines and USA were grouped together first. Then, the two groups were combined together, to be represented as {{Britain, India}, {Philippines, USA}}. Though more complicated statistical analysis is necessary, the analysis result shows us the fact that the AmE/ BrE distinctions were more powerful than those of the Inner/ Outer Circle.

6   Discussion

In this paper, the use of two modal auxiliaries *can* and *may* was compared on a basis of the data extracted from the four components of the ICE corpus. Twenty linguistic factors were encoded to the sentences, and they were analyzed with a logistic regression and a BP analysis.

The analysis results in Section 4 and Section 5 reveal several facts about the use of two modal auxiliaries *can* and *may* in the four components.

The association plot in Figure 2 demonstrates the fact that the ENL speakers (British and USA) use *may* more often and *can* less often than the ESL speakers (India and Philippines). Namely, the ESL speakers use *may* less frequently and *can* more frequently than the ENL speakers. It also illustrates the possibility that the Inner/Outer Circle distinctions might be sharper than those of the BrE/AmE.

The analysis results in Figure 5 and the effect plots in Figure 3, Figure 4, and Figure 5 indicate that each component of the ICE corpus had its own characteristics, and three interactions with CORPUS (i.e., CORPUS:SUBJMORPH, CORPUS:SENTTYPE, and CORPUS:VENDLER) made each component unique in the use of the two modal auxiliaries.

The BP analysis in Figure 6 demonstrates that the AmE/BrE distinctions were more clear-cut than those of the Inner/Outer Circle. Note that the grouping of the components was made as {{Britain, India}, {Philippines, USA}}. If the Inner/Outer Circle distinctions were stronger than those of AmE/BrE, the grouping of the components would be made as {{Britain, USA}, {India, Philippines}}. The grouping of Figure 6 clearly shows that the AmE/BrE distinctions were more important than those of the Inner/Outer Circle in the four components of the ICE corpus.

7   Conclusion

In this paper, the sentences with two modal auxiliaries (*can* and *may*) were extracted from the four components of the ICE corpus (British, India, Philippines, and USA), and their uses were examined. After twenty linguistic factors were encoded to the sentences, the collected data were statistically analyzed with R.

Two statistical methods were adopted in the analysis. One was a logistic regression by which the properties of each ICE component were closely

investigated. The other was a BP analysis where the four components were clustered with the similarity.

Through the analysis, the following three facts were observed: (i) India and Philippine speakers used *can* more frequently than natives, (ii) Three linguistic factors interacted with CORPUS, and (iii) The AmE vs. BrE differences were more influential than those of the Inner vs. Outer Circle.

References

Beryl Atkins. 1987. Semantic ID Tags: Corpus Evidence for Dictionary Senses. In Proceedings of the Third Annual Conference of the UW Centre for the New Oxford English Dictionary, 17-36.

Braj Kachru. 1992. The Other Tongue: English across Cultures. University of Illinois Press, Urbana, IL.

David Crystal. 2003. The Cambridge Encyclopedia of the English Language. Cambridge University Press, Cambridge.

Deshors, Sandra. 2014 Constructing Meaning in L2 Discourse: The Case of Modal Verbs and Sequential Dependencies. In Glynn, Dylan and M. Sjo..lin (eds.) Subjectivity and Epistenicity: Stance Strategies in Discourse and Narration, 329-348. Lund University Press, Lund.

Elizabeth Bates and Brian MacWhinney. 1982. Functionalist Approaches to Grammar. In Eric Wanner and Lila Gleitman (eds.) Language Acquisition: The State of the Art, 73–218. Cambridge University Press, Cambridge.

Elizabeth Bates and Brian MacWhinney. Functionalism and the Competition Model. In Brian MacWhinney and Elizabeth Bates (eds.) The Cross-linguistic Study of Sentence Processing, 3–73. Cambridge University Press, Cambridge.

Geoffrey Leech. 1969. Towards a Semantic Description of English. Indiana University Press, Bloomington.

Gunnel Tottie. 2002. An Introduction to American English. Wiley- Blackwell, Oxford.

Jennifer Coates. 1983. The Semantics of the Modal Auxiliaries. Croom Helm, London.

John Algeo. 2006. British or American English? Cambridge University Press, Cambridge.

Orin Hargraves. 2003. Mighty Fine Words and Smashing Expressions. Oxford University Press, Oxford.

Pam Peters. 2004. The Cambridge Guide to English Usage. Cambridge University Press, Cambridge.

Peter Collins. 2009. Modals and Quasi-modals in English. Rodopi, Amsterdam.

Peter Trudgill and Jean Hannah. 2013. International English: A Guide to Varieties of Standard English. 5th Edition. Routledge, London.

R Core Team. 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.

Ronald Langacker. 2000. Grammar and Conceputalization. Mouton, Berlin.

Sandra Deshors and Stefan Greis. 2014. A Case for the Multifactorial Assessment of Learner Language: The Uses of *May* and *Can* in French-English Interlanguage. In Dylan Glynn and Justyna Robinson (eds.), Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy, 179-204. John Bejamins, Amsterdam.

Sandra Deshors. 2010. Multifactorial Study of the Uses of *May* and *Can* in French-English Interlanguage. Ph.D. dissertation, University of Sussex.

Sidney Greenbaum. 1996. Comparing English Worldwide: The International Corpus of English. Claredon, Oxford.

Stefan Gries and Naoki Otani. 2010. Behavioral Profiles: A Corpus-based Perspective on Synonymy and Antonymy. ICAME Journal 34:121-150.

Stefan Gries. 2003. Multifactorial Analysis in Corpus Linguistics: A Study of Particle Placement. Continuum Press, London.

Stefan Gries. 2010a. Behavioral Profiles: A Fine-grained and Quantitative Approach in Corpus-based Lexical Semantics. The Mental Lexicon 5(3):323-346.

Stefan Gries. 2010b. Behavioral Profiles 1.01: A Program for R 2.7.1 and Higher.

Tom McArthur. 2002. The Oxford Guide to World English. Oxford University Press, Oxford.

Yong-hun Lee. 2007. Corpus Analysis and Their Applications Using NLPTools: Applications to Study of Language, English Education, and Development of English Textbooks. Cambridge University Press, Seoul.