A Corpus-Based Analysis of Korean EFL High School Students' Interlanguages: Focusing on Function Words

Eun-Yeon Lee

(Hanyang University)

Ⅰ Introduction

The goal of this study is to characterize the interlanguages of Korean high school students paying special attention to the function words they use. In order to investigate this issue, the present study has constructed three corpora of English writing samples:

ⅰ) samples by Korean English speakers at an intermediate level (KES1)

ⅱ) samples by Korean English speakers at an advanced level (KES2), and samples by native English speakers (NES). Each corpus consists of 30 essays on the same topic of ¡the reason and solution of poverty¡. Function words were extracted from the three corpora and were analyzed with Wordsmith, a computer text analysis tool.

(1) Research Questions

1. What are the differences in the distribution of function word classes across the three corpora: KES1 (Korean high school students at the intermediate level); KES2 (Korean high school students at the advanced level); and NES (Native English Speaking high school students)?

2. What are the differences in the use of individual function words across the three corpora?

3. How do the distribution of function word classes and the use of individual function words relate to the KES1 and KES2 learners' interlanguage development?

II Methods

Context	KES1, KES2: EFL learning context NES: Native English Speaker context
Medium	Written English
Genre	Argumentative short writing
Task Type	Free composition with a time limit and non-exam
Essay Topic	The reason and solution of poverty

Corpus Design

Subject Group	Number	School Year	Learner Background
KES 1 (Intermediate)	30	11th year	Level in the Nationwide Test of Sept.2005:60 -77%
KES 2 (Advanced)	30	11th year	Level in the Nationwide Test of Sept.2005:97-99.9%
NES (Native Writer)	30	12th year	British high school students

Characteristics of the Participants

III Results

1. Overall Comparison of the Three Corpora

	KES 1	KES 2	NES
Tokens	4,588	7,010	8,272
Types	726	1,062	1,638
Type/Token Raito	15.82	15.15	19.80
Ave. Word Length	4.40	4.41	4.53

Tokens, Types, and Average Word Length

2. Research Question One

What are the differences in the distribution of function word classes across the three corpora: KES1, KES2 and NES?

Personal pronouns are the highest in frequency and the next is primary verbs. This distribution is similar to that of the conversation register presented in LSWE[1], where the categories of function word classes are not identical.
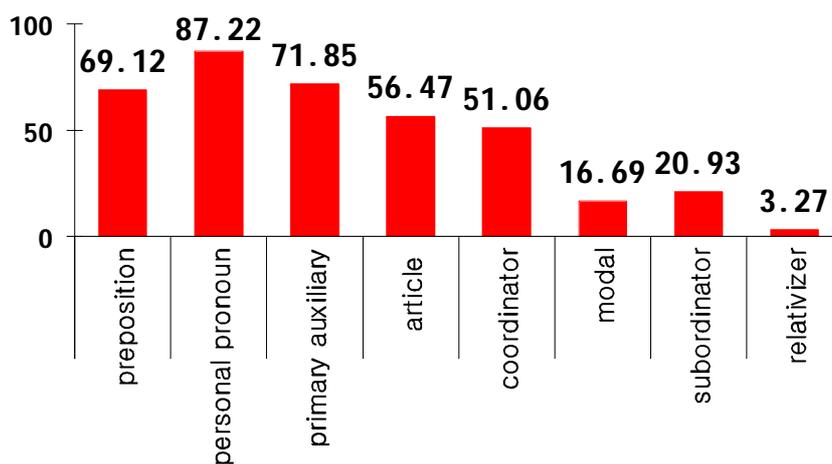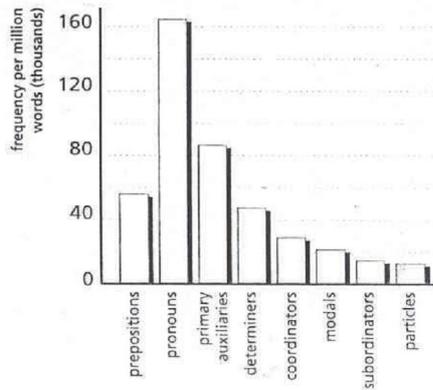

Figure 1 Distribution of Function Word Classes in KES1

This distribution of function words in the academic prose in LSWE shows similarity to their distribution in NES, as shown in Figure 3. It shows a high frequency of prepositions and articles. Figure 4 shows the distribution of function words in KES2. It may be concluded that KES2 has characteristics of both KES1 and NES: high frequencies of prepositions and articles (NES) and also relatively high frequencies of personal pronouns and primary verbs (KES1).

---

1) The Longman Spoken and Written English Corpus (the LSWE Corpus) contains over 40 million words of text. The LSWE corpus focuses particularly on the four registers of conversation, fiction, news, and academic prose. The LSWE includes both American English and British English (Biber et al., 1999).

Distribution of function word classes—conversation
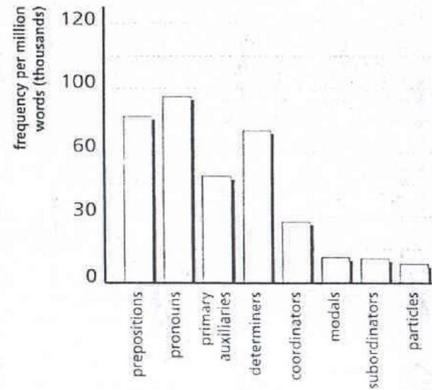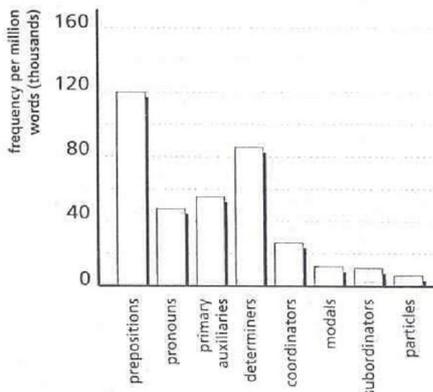
Distribution of function word classes—fiction

Figure 2 Distribution of Function Word Classes in LSWE - 1

Distribution of function word classes—news

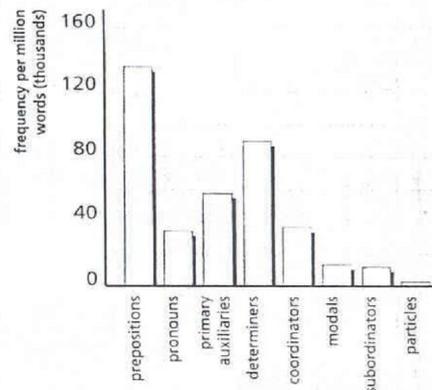## Distribution of function word classes—academic prose

Figure 2 Distribution of Function Word Classes in LSWE - 2

Figure 5 displays the distribution of function word classes in all the three corpora. The proportion of prepositions, articles, and relativizers increases consistently in the same order: KES1, KES2, and NES. The proportion of coordinators and personal pronouns decreases in the order of KES1, KES2, and NES. On the other hand, subordinators, primary verbs, and modal auxiliaries show a different tendency. That is, the proportion of primary verbs and modal auxiliaries is the highest in KES2, while the frequency of subordinators is the highest in NES.
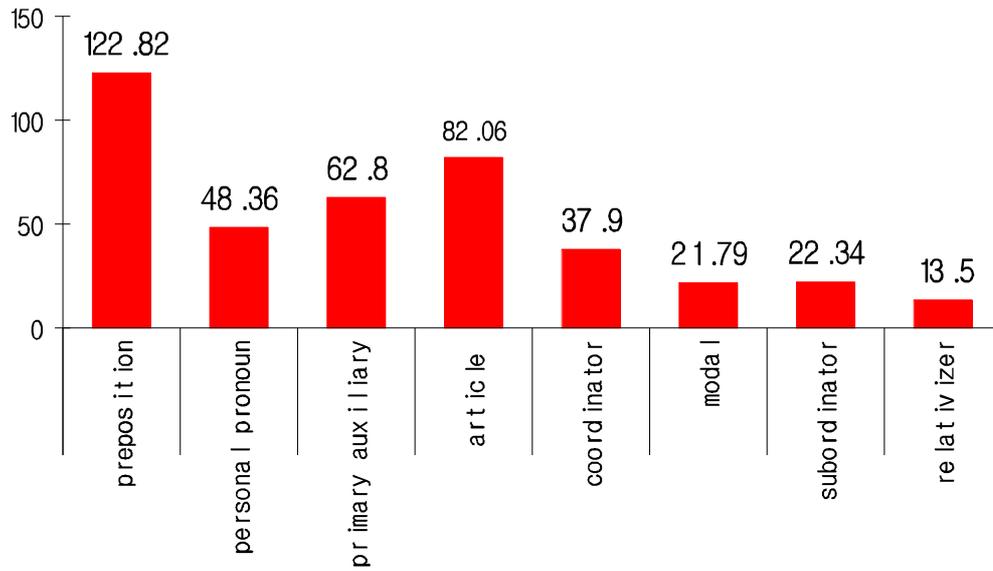
3. Research Question Two

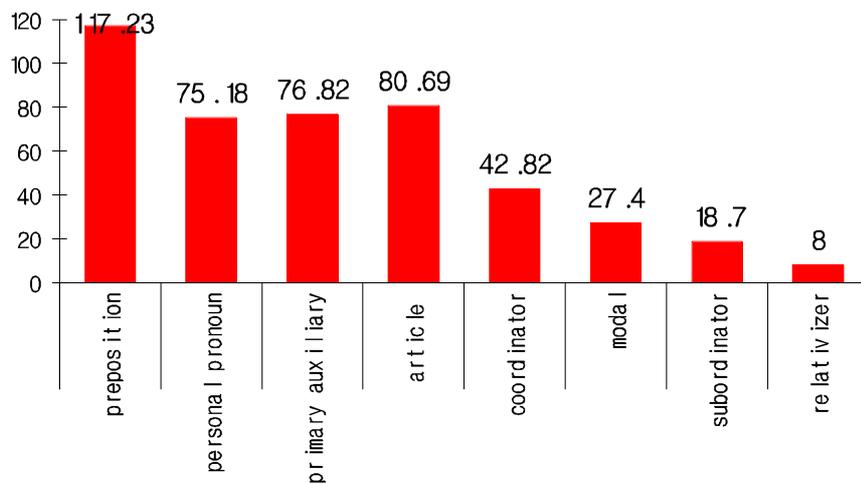**Figure 3. Distribution of Function word classes in   NES**

Figure 4 Distribution of Function Word Classes in   KES2

What are the differences in the use of individual function words across the three corpora?

	Word	KES1	KES2	Observation	Explanation
Under use	of	U		KES1: limited use of phrase	Lexis
	in	U	u	KES1 and KES2: clutch for one meaning -(ey)se(-)	Lexis
	are	U		KES1: limited use of "there are"/limited use of plural forms	Grammar
	been	U		KES1: limited use of the perfect tense	Grammar

	be	U		KES1: limited use of *to be* passive form	Grammar
	have had	U		KES1: limited use of perfect tense	Grammar
	as	U	u	KES1 and KES2: clutch for one meaning: (u)lose( ) and one function word class (preposition)	Lexis
	would	U	u	KES1 and KES2: limited use of external meaning	Lexis
	which/ where	U	u	KES1 and KES2: clutch for subject case	Grammar
	that		U	KES1: frequent use of *think that*	Grammar
Over use	but	O		KES1: clutch for one word (*but*) to express contrast	Style (spoken)
	so	O	O	KES1 and KES2: clutch for one usage (causative meaning)	Lexis
	because	O	o	KES1 and KES2: clutch for *because* to express causative meaning	Style (spoken)
	I, you (your)	O		KES1: learners's tendency to communicate with the readers	Style (spoken)
	they (their, them)	o	O	KES1 and KES2: clutch for human being property as a subject	Style
	can't	O	o	KES1 and KES2: overuse of all forms of contractions	Style (spoken)
	who		O	KES2: clutch for subject case in the relative clause	Grammar
	the	U	O	KES2: over-generalization	Lexis

Summary of Over-/Under-used Words
(U, u²⁾ = Underuse, O, o = Overuse)

In general, over/underused words in KES1 show appropriate use or the degree of over/underuse reduce in KES2, although there are some exceptions: that, so, who, the (the shaded ones). That is underused in KES2; so is overused in KES2 more than in KES1; *who* is overused in KES2; the is overused in KES2 but underused in KES1. However, more in-depth investigation showed that the underuse of *that* is not necessarily associated with the underdevelopment of a certain grammatical feature because numerous verbs are combined with *that* while one verb *think* is mainly combined with *that* in KES1. The overuse of *who* and *the* in KES2 also seems to show improvement compared to KES1 because learners go through an overgeneralization stage in

---

2) Capital U and small u indicate underuse: Capital O and small o, overuse. Small u / o means the degree of under/overuse has weakened compared to the other learner corpus.

the language development process. However, in the case of *so,* it seems to be necessary for the students to get explicit instruction in how to use other causative conjunctions.

3.1 The Use of as

	KES1	KES2	NES
as		0.43	3.27
		3.85	8.30

Frequencies of *as*

Both KES1 and KES2 learners underuse *as* in comparison with NES. *As* is mainly used as a conjunction, a preposition, or in phrases like *such as.*

An aspect to take notice of is that the percentage of uses of the subordinator *as* is the highest in the NES (39.13%) while it is null or meager
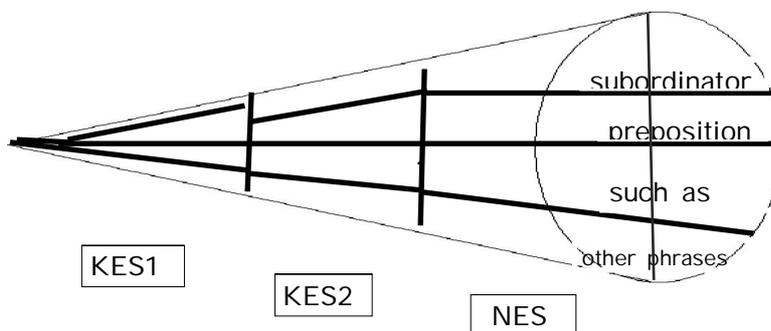

Figure 6   Development of *As*

in the KES corpora. On the other hand, the percentage of use of the preposition *as* is the highest in KES1 and decreases from KES1 to NES: KES1, (62.5%); KES2, (51.85%); NES, (18.84%). The frequency of such as is the highest in NES while the frequency of other phrases is the highest in KES2.

The use of as can be drawn as a cone (See Figure 6). That is, the frequencies lie on an increasing line and also the usages increase in diversity from KES1 to NES.

The preposition as, which is frequently used in KES1 and KES2, is mainly used in

one meaning, as shown in (1), indicating "what someone or something is or is thought to be, or what function they have" (Collins Cobuild), which could be translated into Korean as "- (u)lose".

Category	KES1	KES2	NES
Total Frequency	**1.74** (8)	**3.85** (27)	**8.30** (69)
Subordinator	N/ F	**0.43** (3) (11.11%)	**3.27** (27) (39.13%)
Preposition	**1.08** (5) (62.5%)	**2.00** (14) (51.85%)	**1.57** (13) (18.84%)
Phrase: *such as*	**0.44** (2) (25%)	**0.29** (2) (7.40%)	**2.54** (21) (30.43%)
Other Phrase	**0.22** (1) (12.5%)	**1.28** (8) (29.63%)	**0.97** (8) (11.59%)

Use of *As*

(1) ... they should restart their life *as* a worker...

The subordinator as, which is the most frequent in NES, is mainly used in two meanings, as shown in (2) and (3).

(2) with a meaning of "because"

They have no way of improving their situation *as* they are unable to receive ever basic education ...

(3) with a meaning of "in the same way"

Some people may be lazy, *as* is suggested in passage A but ... .

With regard to such as, which occurs the most frequently in NES, Altenberg & Tapper (1998) reported a similar result in a study comparing SWICLE (Swedish component of the ICLE corpus) and LOCNESS. They reported that *such as* is twice as common in native English speaker's essays as in Swedish learners' essays. An example is shown in (4).

(4) Large fund raising events *such as* comic Belief have an influence...

While NESs tend to use the phrase *such as,* KES2 learners tend to use other phrases: *same as* (2), *as long as* (2), *as if* (1), *as much as* (2), and *as well* (1)[3]

In summary, it seems to be L1 influence why KES learners stick to using the preposition as: i.e. they regard as as being equivalent to the Korean "(u)lose". On the other hand, NESs, often use the subordintaor as while Korean learners stick to because or so to represent causative meaning. In addition, NESs

---

3) The number in parenthesis indicates raw frequency.

prefer the phrase "such as", which is mainly used in written essays (Biber et al., 1999) whereas KES2 learners prefer to use "as" as a part of other different pre-fabricated expressions. Therefore, these differences in use may result in different writing styles; that is, Korean learners' writing bring an informal conversational impression.

4. Research Question Three

How do the distribution of function word classes and the use of individual function words relate to the KES1 and KES2 learners' interlanguage development?

		KES1 and KES2
Lexis		• grasp of only one particular meaning and structure of a word (*as, in, so, of*) • grasp of intrinsic meaning of *would* and *could* • limited use of idioms (*of, in, as*)
Grammar		• underuse of indefinite articles
Style	conversational register	• overuse of *coordinator so* and *subordinator because* • preference for coordinator *but* to connector *however* • frequent use of modal auxiliary *have to,* and contracted modal forms
	Korean-style	• sticking to human subject (overuse of *they*)

Similarities of KES1 and KES2

	KES1	KES2
Grammar	• errors in using *is* • limited use of *there are* • difficulty in distinguishing grammatical number→ overuse of *is* and underuse of *are* • difficulty in using "to + be	• N / F

	+p.p.(passive form)" → underuse of *be* • staying behind in the use of perfect tense → underuse of *have/been*	
	• only use of subject type relative clause	• use of subject type and object type of relative clause.
	• underuse of *the*	• over use of *the*
Style	• overuse of 1st and 2nd personal pronouns → informal style	• N/F
	• N / F	• lower frequency of *must* than *should* → :ormal style

Differences of KES1 and KES2

Learners' interlanguage can be characterized in the aspects of lexis, grammar, and style. KES1 and KES2 share common aspects in style. In other words, both of KES1 and KES2 take conversational register. The development of register seem to be inherently later than other features such as grammar, irrespective of one's L1 in second language acquisition. In case of lexis, both of KES1 and KES2 show similar tendency in that they stick to one usage or one word class, although KES2 presents a weaker tendency. The clutch for one meaning also influenced the sentence structure and writing style.

The crucial difference between KES1 and KES2 seems to be in grammar. It implies that KES1 high schoolers need to make an effort in studying grammar to improve their English competence.

Ⅳ Conclusion and Pedagogical Suggestions

The results carry some implications. First, the low frequencies of some functional word classes seem to be related to the difficulty of hierarchy. That is, the reason for the low frequencies of prepositions, articles, and relativizers is the complexities of semantic and syntactic features of these word classes. Second, the interlanguage characteristics according to different proficiency levels seem to be crucially differentiated by the degree of proficiency in grammatical features. Lastly, the limited use in functional lexis seem to

influence not only grammatical construction but also register, which is partially pointed out in previous studies.

   The present study makes the following pedagogical suggestions. Firstly, students' salient features in different interlanguages should be further identified. Secondly, proper education should be provided according to their different levels (clinical linguistics): 1) Intermediate level Korean high schoolers seem to stay behind in English structures compared to advanced counterparts. Thus, they need to focus more on English structures. 2) Both intermediate and advanced high schoolers need to improve their lexical repertoire, especially when one word or idiom can have diverse meanings. 3) Korean learners should be taught more about the differences of registers such as avoidance of contracted forms in writing formal essays.

Bibliography

Granger, S, ed. (1998). *Learner English on computer*. London: Longman
Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London: Longman.

Eun-Yeon Lee
e-mail : akpni@hanmail.net
Mobile : 019-460-9963