Modern English Education, Vol. 10, No. 1, Spring 2009

Corpus Evidence of the Features of Children's English Literature in an EFL Context

Kyu-Hwa Kim
Cheongju National University of Education

Kim, Kyu-Hwa. (2009). Corpus evidence of the features of children's English literature in an EFL context. *Modern English Education, 10*(1), 66-92.

This study aims at investigating the language used in children's English literature. Two small corpora were created referring to the BNC corpus. In order to collect and analyze the data, WordSmith 3.0, kfNGram and CLAWS tagset programs were used. The results reveal that: (1) there is a higher degree of nominalization in a corpus of imaginative fiction written for Korean children (hereafter, "IFKC"), (2) the IFKC offered much more reference to people and things, (3) the IFKC texts contains fantasy, while the imaginative fiction written for an adult audience (hereafter, "IFAA") texts are more realistic and sophistical, (4) the two fiction corpora portray the lexical verbs used in the key processes of a narrative, (5) The adjectives in the two fiction corpora reveal the typical characteristics of narrative, (6) the two fiction corpora to contain N-grams denotes times and places where narrative events unfold, and locations and flows of events, (7) the IFKC represents the '*I*'-centered world view (8). Most examples in the IFKC, *'in time'* indicates a deadline or the completion of an action. In sum, the dialogues in the elementary school English textbooks in Korea must be reconstructed by actively and positively adopting the strengths of children's literature, which reflects the feelings and emotions of children.

[corpus/comparison of linguistic features/코퍼스/언어적 특징 비교]

Ⅰ. INTRODUCTION

1. Rationale of the Study

Since the mid-1990s, there has been a growing interest in learning English in Korea. The great desire of many Koreans to learn the English language resulted in the preponderance of English books in Korea; the English education market in

the private sector grew rapidly, and the younger English learners were able to read English books more frequently and easily. Considering the foregoing, it can be said that it is now time to conduct more corpus-based studies on English literature in the EFL context, and to make the results of such studies available in bookstores and Websites; recently, several number of corpus-based studies already have been conducted in Korea (Dong-Ju Lee, 2008; Jae-hwang Shim, 2008).

There were two main strands of the corpus-based research; one was to discover how language learners respond directly to corpus-based approaches in learning about language and the other was to explore features of the corpus itself. The latter would be a starting point for the former and there is a need to collect the examples of English language drawn from English texts that Korean children in Korea might be familiar with or would likely read. Thus, a specialized corpus, prepared for Korean children in Korea, should be created for pedagogic application, but this raised some questions about the corpus itself as Thompson and Sealey (2007) have mentioned as below:

> To what extent does such a corpus represent a 'scale-down' version of 'language in general', simplified to be made accessible to these younger readers, and to what extent is it distinctive in its linguistic properties? If corpus-based teaching were to be extended to this population on a larger scale, would a corpus such as this one be adequate for pedagogical purposes, or would the range of investigations that teachers and learners could conduct with it be restricted because of the properties of the language used in the texts it contains? (p. 2)

Based on the questions above, linguistic analysis of the corpus needs to be conducted. Thus establishing the distinctive features of texts from English literary books aimed at a child audience in an EFL context, compared to that of a general audience (adults readers in an L1 context).

2. Research Questions

This study aims to unveil the distinctive features of the children's English literature corpus. Specifically, the relative frequencies within the three corpora are addressed, in the frequencies of words as well as in the sequences of words.

Research question 1: What similarities and differences are there in the overall frequencies of words, part of speech, and word and POS sequences in the three

corpora?

First and foremost, whether the language that is used to portray the world and human experience in writing to a child reader is different from that used to portray the same to an adult reader has to be looked into. Specifically, an investigation has to be made whether certain lexical items are used in a particular way when representing the world to child readers. Furthermore, the portrayal of the relationship between the protagonist(s) and the world to child readers must be compared with the portrayal of the same relationship to adult readers.

Thus the following questions are offered:

Research question 2: Are there similarities and differences between the children's English literature and Adults' English literature corpora in the use of particular lexical items?

Research question 3: Is the discourse of the children's English literature corpus distinctive in its representation of the self in the world and of the world by the self?

The characteristics of English learning materials should be different; they should depend on the target of English teaching. As for elementary English education, the linguistic characteristics of children's literature in English, which reflects the feelings and emotions of children, should first be understood, and such characteristics should be reflected in the dialogues and reading materials to be included in the English textbooks. The results of this study are expected to have important implications on the development of English learning materials for children-centered English education.

Ⅱ. LITERATURE REVIEW

1. Corpus Authenticity

Authenticity may be a factor that is brought into the classroom, but as Widowson (2000) has pointed out, an argument against corpus authenticity still remains. A pedagogic context authenticity is not an objective characteristic of texts, but one which the reader must willingly bring to the text, by one's own capability. Widowson's objection against corpus data, which is the lack of authenticity for students, rests on the fact that it is impossible to replicate the

original contextual conditions of the language in any place other than where the original context was uttered. In addition to supporting this statement, McCarthy (2001) also criticized Widowson for seeing only in the perspective of an intellectual linguist, and not in the eyes of the learner: "What we really need to test authenticity and relevance is learners' own reports." (2001, p. 138). Regardless of how precise and interpretable an observer's reports may be, it is far more important and realizable to speculate on the learner's response. The views taken by McCarthy and Widowson has a surprising 'equivalence' in translational theory, not defined as a linguistic similarity, but as 'an equivalent response in the target audience.' Unfortunately little comparative research based on such response remains.

2. Usefulness of Corpora: Lexico-grammatical Profile

A corpus strategy, when looking at concordance lines, is to create a word's 'lexico-grammatical profile' and its contexts of use (O'Keeffe, McCarthy, & Carter, 2007). A lexico-grammatical profile illuminates typical contexts in terms of:

1) Collocates: most frequently used word(s) with statistical significance (i.e. not just by random occurrence) in the word's environment.

2) Chunk/idioms: word forming a part of a recurrent chunk, word that is likely an idiom, and a certain type (i.e. binomial or trinomial such as *rough and ready,* or *ready, willing and able*).

3) Syntactic restrictions: any systemic patterns which restrict the word (i.e. any prepositions that go with the word), its typical clause-positions (initial/medial /final), and any tense/aspect restrictions.

4) Semantic restrictions: any semantic restrictions that appear (i.e. word/phrase is applied to humans only, or is never used with an intensifier).

5) Prosody: Louw (1993) presented the term 'Semantic prosody', which means that words, as well as having typical collocates, usually appear in specific environments, in a way that its meaning, especially their connotative and attitudinal meanings, stretch over several words. For example, innumerable words go with positive or negative situations. According to Stubbs (1995), more than 90% of the collocates of the word *cause* are negative: *accident, cancer, commotion, crisis, delay.* As a positive word, the example of the word *provide* is given: *care, food, help, jobs, relief, support.* Before the use of computerized language analysis, this remarkable insight has never been properly codified in terms of actual usage.

6) Other relevant or recurring features.

3. Style and Corpora

The place of corpora in regards to the literary studies is ambiguous. First off, the discipline, stylometrics, which is devoted to the statistical study of literary texts (Burrows, 1992; Holmes, 1998), exists. This measures the style and for centuries have been used to establish authorship of equivocal texts as well as trace developments in literary style. The world of literary criticism, however, adamantly frowns upon the use of corpora as an aid to the interpretation of individual literary texts (Louw, 1997, p. 241).

When examining the different features of literary styles, Barnbrook (1996) and Louw (1997) both recommend using a large, general corpus to establish a median for comparison. Louw remarks, "there is general agreement that the difference between the norm and features of the text in responsible for many of the 'devices' which give the reader so much pleasure," (1997, p. 244). He further states that by comparison with a large corpus, a 'hidden meaning' in a literary work might be revealed.

According to Burgess (2000), Lawson (2000) and Barnbrook (1996), certain corpus investigation techniques can be used to bring forth a writer's style. They use corpora consisting of a single work as well as show how the corpus investigation techniques, such as word frequency lists, concordance lines, and collocations can be used to portray certain aspects of a writer's style. Moreover, the writers illustrate a project in which a corpus of narrative texts, such as fiction, newspaper reports, and biography, is defined in terms of the representation of speech and thought (Leech & Short, 1981). This annotation is important to the study of narrative and literature, because how a writer expresses his/her thoughts depends on his/her perspective and focus (Simpson, 1993).

4. Children's Literature and Corpora

Researchers of children's fiction, observing its role in socialization, have concluded that the texts were filled with inevitable ideology (Hunt, 1992; Lesnik-Oberstein, 1994; Sealey, 2000; Stephens, 1992; Wall, 1991). Only a few have taken the corpus linguistic approach to analysis, Stubbs (1996) incorporated the corpus technique to analyze the differences in gender based texts addressed to boys verses girls. He cites Baker and Freebody's (1989) analysis of the different

usages of lemmas '*girl*,' '*boy*' and '*child*,' as well as their collocates in beginner level reading books. Furthermore, using the 'critical linguistics' tradition, Knowles and Malmkjaer's study (1996) states how "an awareness of patterns of textual structure and of language choices may provide information about how the author wants his/her readers to view society" (p. 263). They applied the use of concordance in some parts of their study to analyse collocational patterns. An emphasis was placed in revealing the ideological functions of linguistic expression:

> ... culturally formative, and of massive importance educationally, intellectually, and socially. Perhaps more than any other texts, they [novels for children] reflect society as it wishes to be, as it wishes to be seen, and as it unconsciously reveals itself to be. (Hunt, 1990, p. 2)

This approach, which engrossed many researchers, is focused on how the world is represented to the child reader. Wall (1991), for example, proposes that "... the narrator-narratee relationship ... is the distinctive marker of a children's book" (p. 9). She provides an extensive study of how adults writing for children and the children's concerns have been alternating for the last two centuries. The adult writer, unlike the child reader, stays irrelative throughout the writing to the world and the subject matter at hand. Additionally, the writers of children's fiction have a wide range to choose their authorial stance. As Wall and others have indicated, the different authorial stances include: first-person narrators who are ostensibly children themselves; the narrators who present the story from their own childhood reminiscence; anthropo-morphic animal narrators with whom child readers can see from the imaginative lens of an adult, as well as the animals themselves; narrators who, though detached themselves, "allow the narratee to stand in [a child character]'s shoes and to move as she moves" (Wall, 1991, p. 197).

Ⅲ. METHODS AND METHODOLOGICAL ISSUES

1. Corpora in This Study

This study aims at investigating the language used in children's English literature. The analysis discussed here are part of an exploratory study of the

potential for using corpus evidence with Korean children in their learning about English in an EFL context. For the purpose of the study, a small corpus was created of texts written for a child audience who learn English in Korea. Texts were taken from English literary books that parents and children can purchase at the bookstores or on the Internet in Korea. The corpus was given the acronym 'IFKC', which stands for 'imaginative fiction written for Korean children'. It contains 250 texts from 250 English literary books, in which most of them are imaginative fiction.

The IFKC corpus and a comparison corpus of imaginative fiction written for an adult audience (abbreviated to IFAA), composed of 317 texts from the BNC, were created; the IFAA corpus data in this study was quoted from Thompson and Sealey's study (2007). Comparisons between the two corpora would answer the question: 'Does writing for a child audience, expecially a Korean child in an EFL context, demonstrate different linguistic properties from writing for adults in an L1 context?' However, although any differences found should be largely attributable to the child/adult or EFL/L1 variable, the restriction to imaginative fiction excludes other kinds of differences. Thus, as an additional comparison, a third corpus, of Elementary school English textbooks in Korea (abbreviated to EETK), was created in order to contrast the features of imaginative fiction writing in general with those of English textbooks that Elementary school students apply in the English classes in Korea. It helps to understand the differences between the English language patterns young Korean readers encounter in their private English reading as well as the different types of English language patterns they acquire in their English classes at schools.

Table 1 shows the three corpora used in this study: IFKC, IFAA, and EETK.

Table 1

The Three Corpora Used in This Study

IFKC	imaginative fiction written for a child audience, available at bookstores and on the Internet in Korea	250 texts	138,819 tokens
IFAA	imaginative fiction written for an adult audience, from BNC	317 texts	12,869,883 tokens
EETK	Elementary school English textbooks in Korea	4 books	2,886 tokens

Different types of corpora is created for its specific purposes. A corpus generally falls into one of these eight categories: specialized, general, comparable, parallel, learner, pedagogic, historical or diachronic, and monitor (Hunston, 2002).

Among these, the IFKC corpus in this study falls into the category of a specialized corpus, considering it is of a particular type of texts (children's English literature), and it aims to be representative of the given type of text.

2. Data Collection and Analysis Methods

The basic word lists according to its type were assembled, in regards to its relative frequency of different morphological forms (for example, *work, working, worked*). Frequency information in parts of speech was also gathered to show its overall recurrent usage as well as the most repeated tokens. Frequency lists for 4-grams were also compiled thus revealing word sequences in comparison to the frequency list.

The manner in which the POS information was extracted was so that the separate files for each corpus for each part of speech can be amassed. All occurrences of a particular POS tag and the string of letters following it were extracted (using the kfNGram). Then the tags were stripped and word lists were constructed. The word lists and other statistical information were developed using Oxford Wordsmith Tools, Version 3. Identification of n-grams and of POS-grams was executed using two applications, called kfNgram and CLAWS. tagset.

The sizes of the three corpora in number of tokens were unequal. Therefore, all of the frequency counts were made consistent to show the frequencies as percentages, in order to compare the figures.

Ⅳ. RESULTS AND DISCUSSION

1. Frequency of Types

Table 2 shows the ten most frequent types in the three corpora (see Appendix A). It can be seen as a relatively similar profile between the IFKC and IFAA corpora, with mostly the same types appearing in the top ten. The most frequently occurring type is '*the*' between the two imaginative corpora (IFKC=6.23%; IFAA=5.06). The EETK, on the other hand, does not have '*the*' in the list, which suggests there is little degree of nominalization.

The interesting fact is that the profiles show the similar appearance in the conjecture '*and*' (IFKC=3.55% (2nd), IFAA=2.60% (3rd), and EETK=0%).

Meaning, the two imaginative texts have much more complex sentences than the EETK texts and the EETK texts mostly contain simple sentences. This is because EETK texts were written for the beginners in English. Furthermore, a striking point is in the frequency of '*I*'. The EETK texts have a higher use of 'I' (4.78%) than the two imaginative texts (IFKC=1.61%; IFAA=1.72%), which may relate to the selection of the topic of the English textbooks for early English learners: from personal or individual life to social life.

Table 2

The Ten Most Frequently Occurring Types in Each of the Three Corpora

	IFKC			IFAA			EETK		
N	Word	Freq	%	Word	Freq	%	Word	Freq	%
1	the	8,647	6.23	the	651,685	5.06	is	202	7.00
2	and	4,925	3.55	to	335,442	2.61	you	151	5.23
3	a	3,435	2.47	and	335,074	2.60	I	138	4.78
4	to	3,024	2.18	a	284,278	2.21	it	92	3.19
5	I	2,240	1.61	of	272,255	2.11	oh	52	1.80
6	of	1,928	1.39	I	221,448	1.72	this	46	1.59
7	in	1,914	1.38	he	208,592	1.62	what	46	1.59
8	he	1,859	1.34	was	202,970	1.58	do	45	1.56
9	you	1,720	1.24	she	190,574	1.48	am	44	1.52
10	is	1,605	1.16	in	180,394	1.40	a	43	1.49

2. Frequency of Parts of Speech

Table 3 shows the frequent occurrences of different parts of speech in the three corpora, stated in percentages (see Appendix B & C). The most notable fact in the figures is that the profiles between the first two columns for each POS appear to be very similar. This represents the percentage figures for the two imaginative corpora, IFKC and IFAA. However, the IFKC corpus contains slightly higher proportions of nouns and prepositions, while the IFAA corpus carries more adverbs. This suggests a higher degree of reference to people and things in the IFKC, but overall the two profiles are very much alike especially in comparison to the EETK corpus. The EETK corpus appears with high proportions of nouns and verb be, while there are much lower uses of conjecture, preposition, and lexical verbs. Accordingly, the sentences of the EETK texts are

relatively short and simple as well as the nouns being mostly characters' names (according to Table 4). The EETK corpus is also distinguished in a way that the high usage of the proportions of Lexical verbs, conjectures and prepositions indicate a more dominant role in human intercession and action in the fiction corpora.

Table 3

Relative Frequency of Different Parts of Speech in Each of the Three Corpora

N	POS	IFKC (%)	IFAA (%)	EETK (%)
1	Article	10.53	7.82	3.63
2	Adjective	6.90	5.91	8.71
3	Adverb	1.67	7.68	3.68
4	Conjecture	6.67	5.72	0.82
5	Possessive	2.53	2.69	2.72
6	Determiner	1.15	2.74	6.40
7	Noun	20.76	16.60	35.57
8	Proper Noun	3.59	3.89	5.72
9	Preposition	12.04	7.51	4.50
10	Pronoun	6.22	3.89	2.22
11	Infinitive to	4.03	1.68	3.13
12	Verb be	4.28	4.23	11.16
13	Verb do	0.50	0.88	1.45
14	Verb have	1.02	1.87	0.54
15	Moral Verb	1.81	1.65	2.59
16	Lexical Verb	16.31	13.74	7.17

3. Frequency Data: Nouns

Table 4 shows the top ten nouns in the three corpora. Although the three corpora don't appear to be overlapped considerably to the eye, one can find great similarities between the IFKC and IFAA corpora. A very different list exists as well in EETK in terms of the properties of a core vocabulary for narrative texts. The most frequent items in the IFKC and IFAA corpora denote aspects of the experience of fictional protagonists: embodied people or humanized animals (*bear, mouse, frog, mother, man, eyes, face, head, hand, people*); settings where events take place (*house, home, door, room*); and items relating to chronology (*day, time*). It is important to note however, that frequency data should be used with

caution. It is also inadvisable to draw any definite conclusion of the meanings of these items without first examining the context in which they come about.

The ten most frequent nouns of the IFKC corpus contain animals such as bear, mouse, and frog. It reflects that the IFKC texts were written intended for children and these types of nouns give children fantasy; the bear, the mouse, and the frog can speak and act like human beings. On the other hand, the IFAA texts were intended for adults and they are more realistic and sophistical than the IFKC texts. The EETK texts were developed for the pedagogical purpose and the most occurring words are the names of the characters in the dialogues in English textbooks as already mentioned above.

Table 4

The Ten Most Frequent Nouns in the Three Corpora

	IFKC			IFAA			EETK		
N	Word	Freq	%	Word	Freq	%	Word	Freq	%
1	day	284	0.20	time	23,446	1.12	time	15	0.52
2	water	259	0.19	man	18,014	0.86	day	12	0.42
3	time	243	0.18	way	17,005	0.81	Julie	12	0.42
4	bear	239	0.17	eyes	16,670	0.79	Minsu	12	0.42
5	house	235	0.17	face	13,058	0.62	Jinho	11	0.38
6	way	204	0.15	head	12,349	0.59	Zeeto	11	0.38
7	mother	202	0.15	door	11,297	0.54	Joon	10	0.35
8	mouse	192	0.14	hand	11,135	0.53	Mom	10	0.35
9	frog	186	0.13	room	10,444	0.50	Nami	10	0.35
10	home	181	0.13	people	10,144	0.48	Smith	10	0.35

4. Frequency Data: Lexical Verbs

Table 5 shows the top ten lexical verb forms in the three corpora. The pattern indicates that the IFKC and IFAA corpora continue to show a very similar profile, with the English textbook corpus providing a useful contrast. The highest frequency of '*said*' is explicable by its language-specific property; the verb 'said' has (an) agent(s) that is(are) human being(s) or (an) animated animal(s) or (a) thing(s), it introduces the protagonist(s) of stories to the readers, and it is a typical marker of indirect speech. In addition, we can say that the past tense form '*went*' belongs to '*go*', '*came*' to '*come, 'got'* to *'get',* and *'thought'* to *'think'*.

In the case of the lexical verbs, then, ten words in the IFKC list also occur in the IFAA list, with the only differences being '*like*', '*made*', *'think (thought)*', and '*looked*'. It is controversial to state that the verbs in the list represent the core processes with which narrative is primarily concerned. Thus the marked similarity between the children's and adult's fiction is only natural. The protagonists in stories are with whom the readers connect. These protagonists come alive in their fictional world and readers are invited to partake in these characters' adventures and perceptions.

All the words (types) in the EETK list, on the other hand, are the present tense forms and it suggest that English language materials for beginners in English are presented in the present tense or 'now' situations.

Table 5

The Ten Most Frequent Lexical Verbs in the Three Corpora

	IFKC			IFAA			EETK		
N	Word	Freq	%	Word	Freq	%	Word	Freq	%
1	said	1,405	1.01	said	64,513	3.71	thank	26	0.90
2	like	389	0.28	know	26,015	1.50	let	23	0.80
3	went	348	0.25	see	20,345	1.17	like	20	0.69
4	see	296	0.21	think	18,048	1.04	look	18	0.62
5	go	285	0.21	get	17,530	1.01	help	17	0.59
6	came	248	0.18	go	17,455	1.01	want	17	0.59
7	get	231	0.17	looked	16,611	0.96	meet	13	0.45
8	come	206	0.15	thought	15,696	0.90	see	13	0.45
9	made	201	0.14	come	15,126	0.87	watch	10	0.35
10	know	200	0.14	got	14,599	0.84	play	8	0.28

5. Frequency Data: Adjectives

Table 6 shows the ten most frequent adjectives. The two fiction lists are much more similar to each other than to the English textbook corpus. The adjectives in the two fiction corpora are used partly for describing protagonists' appearance or things or time (*little*, *big*, *old*, *good*, *long*, *small*, *young*); such descriptions are one of the typical characteristics of narrative or stories. Another similarity between the two fictions is the use of opposite words such as little/big/small, old/young/new, and some/other; which allows the fiction to have binary structures,

being the main characteristics of stories. The EETK list, on the other hand, is from a relatively small size of the corpus and the EETK texts were created with being focused on communicative functions and their exponents. Thus, the EETK list reflects this feature. For example, '*good*' and '*nice*' are used as a part of greeting expressions such as '*Good morning*' and '*Nice to meet you*'.

Table 6

The Ten Most Frequent Adjectives in the Three Corpora

	IFKC			IFAA			EETK		
N	Word	Freq	%	Word	Freq	%	Word	Freq	%
1	little	588	0.42	good	12,071	1.62	good	31	1.07
2	big	302	0.22	other	10,702	1.43	nice	18	0.62
3	some	267	0.19	old	10,246	1.37	sorry	10	0.35
4	old	205	0.15	little	8,494	1.14	right	9	0.31
5	good	201	0.14	small	6,783	0.91	great	9	0.31
6	right	162	0.12	sure	6,773	0.91	happy	7	0.24
7	long	159	0.11	long	6,502	0.87	much	7	0.24
8	more	150	0.11	young	5,965	0.80	new	7	0.24
9	white	147	0.11	new	5,693	0.76	beautiful	6	0.21
10	great	144	0.11	right	5,515	0.74	now	6	0.21

6. N-gram Data

A remarkable pattern to be noted is the similar profile in the two imaginative corpora, especially when compared to the English textbook corpus. To further test this similarity, the most common N-grams known as lexical bundles or clusters in the three corpora were observed. The 4-grams were identified as the most revealing, on the basis that there are too many 3-grams, to make them tractable to careful analysis, and greatly minimal 5-grams. Following this preference, the most frequent 4-grams were calculated, using the kfNGram program, and the top 20 4-grams for each corpus are shown in Table 7.

The third column in Table 7 shows, firstly, the position of the 4-gram given in the first column in the IFKC corpus, and secondly, the ranking of the same 4-gram in the IFAA corpus. The appearance of letter X indicates that the given 4-gram does not appear in the first 20 of the 4-grams in the IFAA corpus.

Apparently, there are 12 X's, which means that 8 of the 4-grams appear in the

top 20 of both lists, implicating a relatively high degree of similarity. None of the common N-grams, by contrast, appear in the English textbook corpus; most 4-grams in the English textbook corpus are independent sentences such as '*Nice to meet you.*', '*What time is it?*', '*This is for you.*', '*Will you help me?*', and '*Can you help me?*', and parts of the exponents of communicative functions such as '*What do you want* - ' and '- *do you want to* - '. Going into details, the two fiction corpora to contain N-grams denote times and places where narrative events unfold, and this indeed seems to be the case: '*at the same time*', '*the top of the*'. Also prevalent in this set of data are expressions denoting locations and flows of events: '*the side of the*', '*in front of*', and '*at the end of*'.

Table 7

The Top 20 4-grams in Each of the Three Corpora

IFKC		IFKC-IFAA	IFAA		EETK	
4-gram	Freq		4-gram	Freq	4-gram	Freq
said the little girl	33	1-X	the end of the	1041	nice to meet you	12
the top of the	15	2-11	for the first time	963	what time is it	7
for a long time	14	3-X	the rest of the	905	this is for you	5
I've run from a	14	4-X	at the end of	836	to meet you to	5
the rest of the	13	5-3	in the middle of	733	will you help me	5
and he ran and	12	6-X	at the same time	673	can you help me	4
down the little chute	12	7-X	I don't want to	669	it's time for lunch	4
for the first time	12	8-2	the back of the	654	I'm sorry I can't	4
the end of the	12	9-X	the edge of the	645	what do you want	4
the side of the	12	10-13	the top of the	556	do you want to	3
in the middle of	11	11-5	in front of the	515	go straight and turn	3
little old man and	10	12-X	the middle of the	511	It's my pencil case	3
out of the ground	10	13-X	the side of the	501	right at the bank	3
the little old man	10	14-X	he was going to	490	thank you very much	3
the little old woman	10	15-X	the other side of	488	this is my teacher	3
and the little old	9	16-X	what do you mean	481	turn right at the	3
at the end of	9	17-4	I don't know what	475	what did you do	3
all over the world	8	18-X	he shook his hand	453	what will you do	3
at the same time	8	19-6	she shook her head	431	you help me please	3
in front of the	8	20-11	on the other side	427	you want to do	3

7. Words Denoting Parts of the Body

In the fiction corpora, lexical elements denoting the parts of the body are common. Based on many observations, words found in children fiction have more of a literal meaning rather than a figurative or a metaphorical meaning. Consequently, the relative use of figurative expressions involving names of the parts of body is necessary, as well as the different experience of the world through the body represented in each fiction corpus.

The first word chosen for such an investigation was '*neck*.' This was chosen in preference to the other frequent words such as *'head'* or *'eye,'* because those would have appeared in too many instances to be able to examine in detail. In the IFKC corpus the word '*neck*' appears 76 times and in the COMP corpus, 1897 times. The frequent left collocates are displayed in Table 8.

The left collocates of '*neck*,' excluding '*my,*' occur in a similar proportion in both the IFKC corpus and the IFAA corpus. The word '*my*,' being the left collocate of *'neck'* is also a part of the IFKC corpus, but has a relatively higher proportion compared to others. This is because in most children literature, the world tends to be focused on '*I*.' This world revolving around '*I*' is expressed as '*my*.' Thus, in the sentence, "So he would run the clippers on the back of my neck," the focus of the story is '*I*' or the person telling the story is based on child reader '*I*.' A child reader, who has read children literature told in first person, can relate to his or herself to the '*I*' in the book as well as the '*I*' in the real world. The reader is able to have an interaction to the story, maintaining his or her personal outlook of the world.

Table 8

The Most Frequent Left Collocates of NECK in the IFKC and IFAA Corpora

	IFKC	IFAA
His	20 (26.3%)	533 (28.1%)
Her	22 (28.9%)	516 (27.2%)
The	10 (13.2%)	285 (15.0%)
My	12 (15.8%)	151 (8.0%)
(a)round *neck	12 (15.8%)	409 (21.6%)

As shown in Table 9, the second word chosen in the body part was '*finger*.' According to the context in which '*finger*' appeared, an examination was made to

categorize as figurative or literal. In the figurative sense, this word was not found to be prominent in the IFAA corpus. Nonetheless, in comparison to the IFKC corpus, it showed to be occurring more. The IFKC corpus showed 12.9%, whereas the IFAA corpus showed 18.8%. In order to make a much more precise investigation, however, a larger children's corpus is needed, being that the difference in proportions may be due to the small number of texts in the existing IFKC corpus.

Table 9

The Most Frequent Left Collocates of FINGER in the IFKC and IFAA Corpora

	IFKC	IFAA
literal	27 (87.1%)	101 (81.2%)
figurative	4 (12.9%)	19 (18.8%)

8. Time

As noted previously, protagonists may be represented slightly different in children's stories compared to an adult fiction. As size is relative, time also may be taken in differently from a child viewer. Different perspectives change how size or time appears to be. Thus different experiences may result in the physical, literal and/or socio-cultural constituents, as can be seen in the following example. Creatures or objects that are no threat at all to adults may seem like a giant from a little child's perspective. Likewise, time may seem to pass slowly to children when it may not be so to others. The following example of '*in time*' generates the following set of concordance lines from the IFKC corpus:

Evaporation continues, but condensation slows down. **In time,** the uncovered drops will disappear. How
one knows for sure what happened. But **in time** new animals took the place of the
"However am I going to finish it **in time**?" "We have had a good idea." said
colors of joy. Than he married Nicoletta. **In time** he became famous, and mice from all
spreads through-out the room as vapor. **In time**, the drops disappear. Inside the glass,
was worried. She had to get back **in time** for an after-school karate class. The rest
fence under the beehives and got back **in time** for dinner. Sea story primrose and
Ananse," said the leopard, "You are just **in time** to be my lunch." Ananse replied. "As
Turkey Luckey. "You will never get there **in time**." said Foxy Loxy. "Come with me and

Below is an example taken from the IFAA corpus in comparison:

no doubt it will all come out **in time**. Is there anybody who can vouch for
Well, not at this moment **in time**, perhaps. But you never know
was imputing to me, but saw **in time** that to do so would be to rise
be pleaded, though if caught **in time**, the house could have been converted
would occur in the soul and, **in time**, the fripperies of science would
Everything would change **in time**. Things changed faster as time
his head from side to side, **in time** to the beat. He lay face down
were already moving off **in time** to take their appropriate places
Tummel Catchment Area. Just **in time** to preserve the high quality of
But if all goes well, **in time** we should be able to sell our will
They will see **in time**, when she is over this, she will
that particles move backwards **in time**, when she is over this, she will

Similarities as well as differences can be detected from the examples above. To start, in most of the IFKC examples, '*in time*' refers to the meeting of a deadline of the accomplishment of an action that must be completed before a penalty or unwanted outcome. In the IFAA corpus, a similar use can be seen, as '*in time*' also refers to a sense of local time (as in line 3, "saw in time"). However, in other lines, '*in time*' is expressed in a larger scale, as in centuries for example (line 6, "Everything would change in time"), or in terms of a gradual passing of time (line 10, "in time we should be able to...").

Ⅴ. CONCLUSION AND PEDAGOGICAL IMPLICATIONS

1. Conclusion

Since the basis of this analysis was based on such limited texts, especially in the case of the IFKC corpus, one must draw the following conclusion with much care and consideration. If there had been a resourceful heterogeneous corpus for children literature on a much larger scale, it would have been possible to conclude diverse inquiries. Furthermore, a more thorough analysis of the subjects discussed would have been made if a greater extensive corpus for children‘s fiction had been available. A much needed area for corpus development exists, as well as the potential for a wide range of applications, including those in the pedagogical content.

Nevertheless, this analysis has shed light on a number of interesting discoveries. Without deploying the POS-gram approach that makes a novel perception of frequency of complex prepositional phrase possible, there remains

room for further investigation. For example, such phrases play certain roles within different forms of discourse, in their expression of vast relations (temporal, causal, spatial, etc.) as well as their role in grammar.

On the basis of investigation about the distinctive linguistic properties in children fiction, the IFKC corpus displayed a close similarity with the two other corpora. This similarity was in terms of the overall frequency lists and different parts of speech. Therefore, one can suggest that narrative fiction, regardless of the audience's age level, may be characterized by linguistic characteristics. A detailed analysis of the concordance output for particular words and phrases, however, reveals how the world and the humans' relationship with it are exhibited in the literature written for children.

2. Pedagogical Implications

As shown in the research, the present English textbooks in Elementary schools are based on communicative functions and exponents. Thus, a vocabulary focused teaching method based on authentic materials is not utilized. Furthermore, reading and writing in English textbooks for Elementary schools are focused on communication using conversational listening skills, which in effect, reflect the wide use of phonetics in the classroom. It is now time to find and take measure in improving the problems, while maintaining the conversational focused method.

According to the Elementary and Middle School curriculum proposal (Ministry of Education, Science and Technology, 2008), third and fourth graders undergo two hours of English class, whereas fifth and sixth graders take three hours of English class. Based on the lengthened English class time, the range of vocabulary and the level of language increased significantly in comparison to the present curriculum. Despite the quantitative expansion of such linguistic materials, there is a need to reconsider the insistence on the inclusion of dialogues and reading materials that focus on everyday-life-related communication skills in the current English textbooks in Korea. It seems that the English dialogues and reading materials in the current English textbooks in the country do not match the language use trends and style of elementary school children, who operate based on pleasure and desire. Therefore, the dialogues and reading materials to be developed based on the revised curriculum should properly reflect the characteristics of children's literature, which reflects the feelings and emotions of children. Through such efforts, it is believed that the new English textbooks will be able to make the children interested in and excited to learn English.

REFERENCES

Baker, C., & Freebody, P. (1989). *Children's first school books.* Oxford: Blackwell.

Barnbrook, G. (1996). *Language and computers: A practical introduction to the computer analysis of language.* Edinburgh: Edinburgh University Press.

Burgess, A. (2000). Corpus analysis in the service of literary criticism: Goethe's 'Die Wahlverwandtsschafen' as a model case. In B. Dodd (Ed.), *Working with German corpora* (pp. 40-68). Birmingham: University of Birmingham Press.

Burrows, F. (1992). Computers and the study of literature. In C. Butler (Ed.), *Computers and written texts* (pp. 167-204). Oxford: Blackwell.

Holmes, I. (1998). The evolution of stylometry in humanities scholarship. *Literary and Linguistic Computing, 13*, 111-117.

Hunston, S. (2002). *Corpora in applied linguistics.* Cambridge: Cambridge University Press.

Hunt, P. (Ed.). (1990). *Literature for children: Contemporary criticism.* London: Routledge.

Hunt, P. (Ed.). (1992). *Children's literature: The development of criticism.* London: Routledge.

Knowles, M., & Malmkjaer, K. (1996). *Language and control in children's literature.* London: Routledge.

Lawson, A. (2000). 'Die schőne Geschichte': A corpus-based analysis of Thomas Mann's Joseph und seine Brűder. In B. Dodd (Ed.), *Working with German corpora* (pp. 161-180). Birmingham: University of Birmingham Press.

Leech, G., & Short, M. (1981). *Style in fiction*. London: Longman.

Lee, Dong-Ju. (2008). Computer learner corpora and their pedagogical applications. *Modern English Education, 9*(3), 83-101.

Lesnik-Oberstein, K. (1994). *Children's literature: Criticism and the fictional child.* London: Clarendon Press.

Louw, B. (1993). Irony in the text or insincerity in the writer? the diagnostic potential of semantic prosodies. In M. Bake & E. Tognini-Bonell (Eds.), *Text and technology in honour of John sinclair* (pp. 157-176). Amsterdam: John Benjamins.

Louw, B. (1997). The role of copora in critical literary appreciation. In A.

Wichmann., S. Fligelstone., T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 140-251). London: Longman.

McCarthy, M. (2001). *Issues in applied linguistics*. Cambridge: Cambridge University Press.

Ministry of Education, Science and Technology. (2008). Elementary and Middle School curriculum proposal. Seoul: Author.

O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching.* Cambridge: Cambridge University Press.

Sealey, A. (2000). *Childly language: Children, language, and the social world.* Harlow: Longman.

Shim, Jae-hwang. (2006). An analysis of corpus in EFL middle school English class. *Modern English Education, 7*(1), 53-72.

Simpson, P. (1993). *Language, ideology and point of view*. London: Routledge.

Stephens, J. (1992). *Language and ideology in children's fiction.* Harlow: Addison Wesley Longman.

Stubbs, M. (1995). Collocations and semantic profiles: On the cause of the trouble with quantitative studies. *Functions of Language, 2*, 23-55.

Stubbs, M. (1996). *Text and corpus analysis.* Oxford: Blackwell.

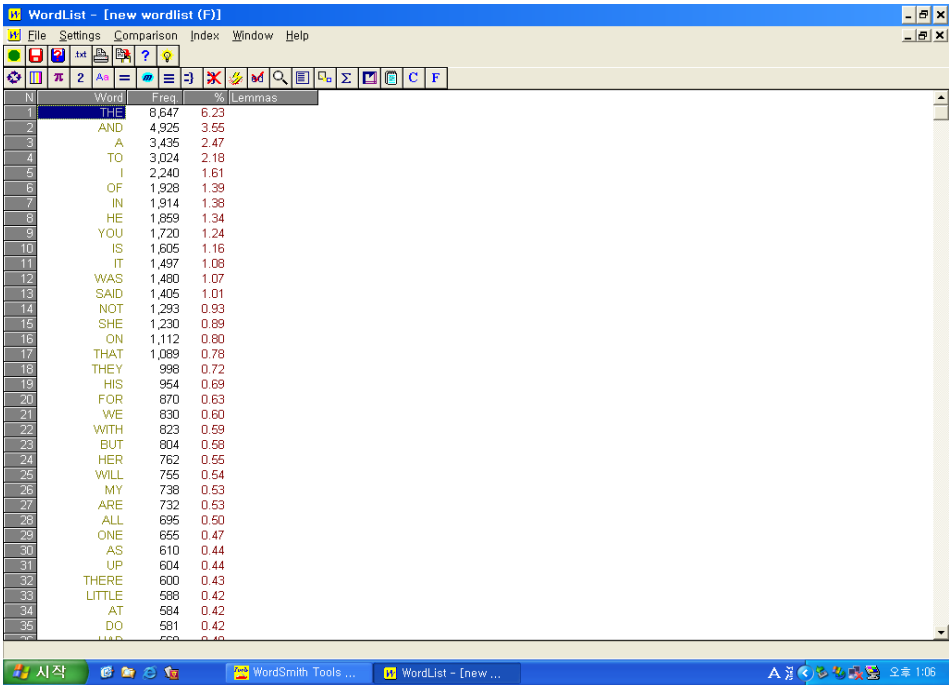
Thompson, P., & Sealey, A. (2007). Through children's eye? Corpus evidence of the features of children's literature. *International Journal of Corpus Linguistics, 12*(1), 1-23.

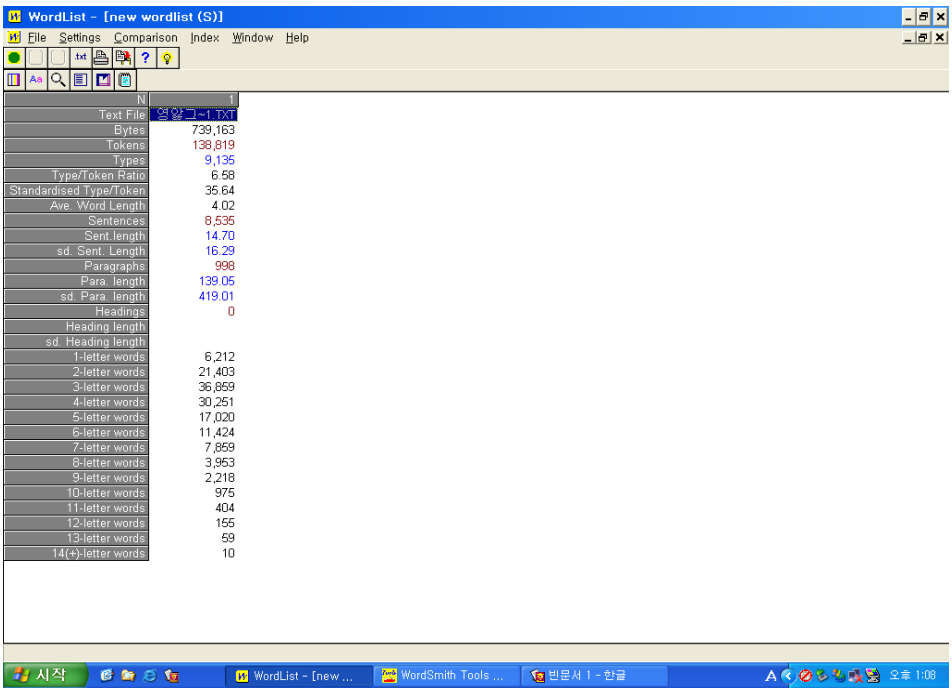
Wall, B. (1991). *The narrator's voice: The dilemma of children's fiction.* London: Macmillan.

Widdowson, H. (2000). On the limitations of linguistics applied. *Applied Linguistics, 21*(1), 3-25.

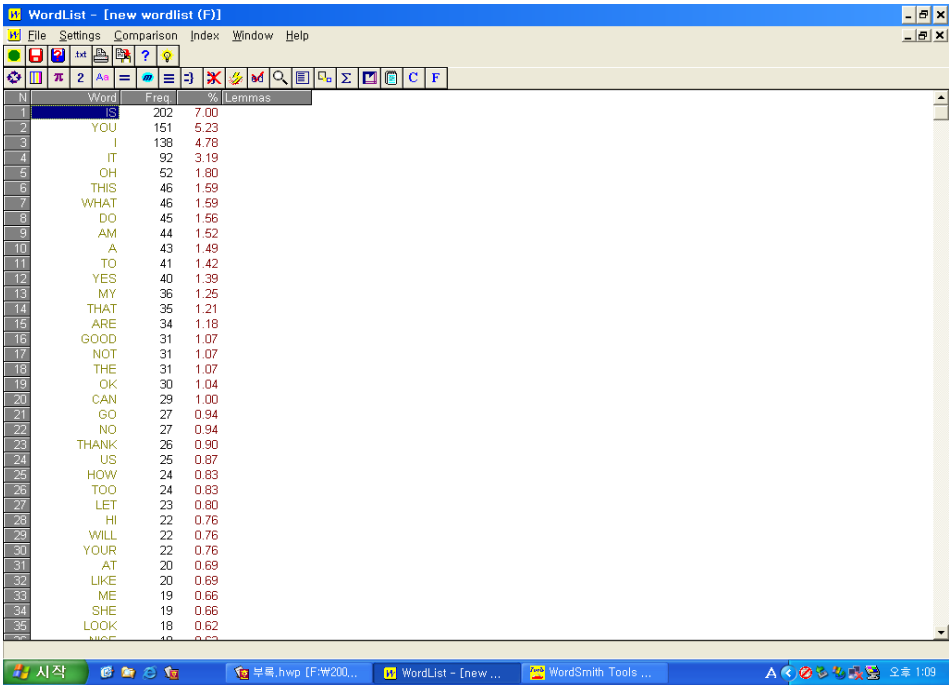
APPENDIX A

Part of the Wordlists of the IFKC corpus

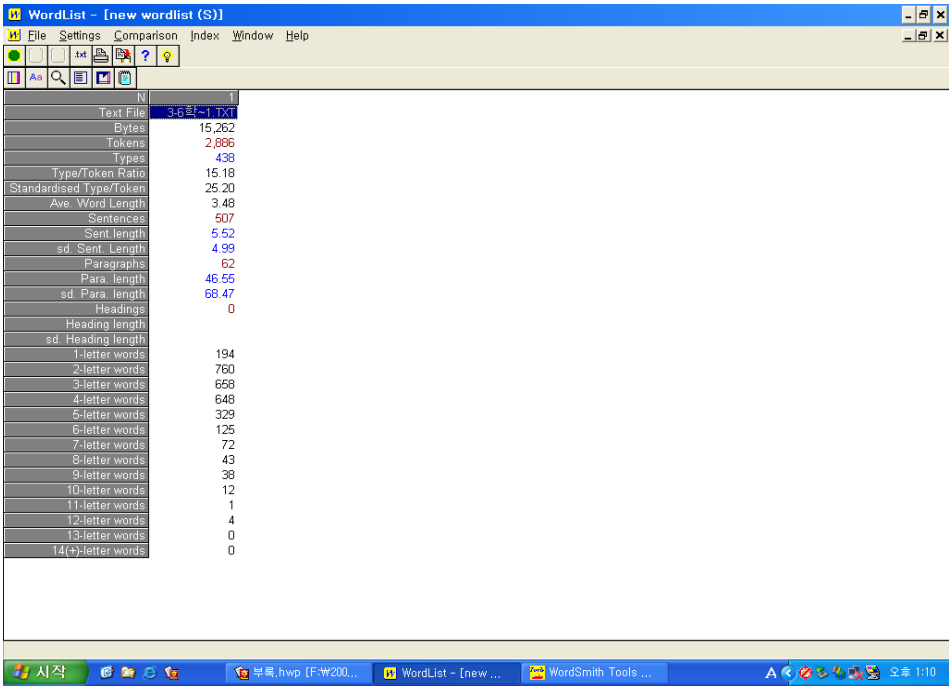




Part of the Wordlists of the EETK corpus



N	Word	Freq	%	Lemmas
1	IS	202	7.00	
2	YOU	151	5.23	
3	I	138	4.78	
4	IT	92	3.19	
5	OH	52	1.80	
6	THIS	46	1.59	
7	WHAT	46	1.59	
8	DO	45	1.56	
9	AM	44	1.52	
10	A	43	1.49	
11	TO	41	1.42	
12	YES	40	1.39	
13	MY	36	1.25	
14	THAT	35	1.21	
15	ARE	34	1.18	
16	GOOD	31	1.07	
17	NOT	31	1.07	
18	THE	31	1.07	
19	OK	30	1.04	
20	CAN	29	1.00	
21	GO	27	0.94	
22	NO	27	0.94	
23	THANK	26	0.90	
24	US	25	0.87	
25	HOW	24	0.83	
26	TOO	24	0.83	
27	LET	23	0.80	
28	HI	22	0.76	
29	WILL	22	0.76	
30	YOUR	22	0.76	
31	AT	20	0.69	
32	LIKE	20	0.69	
33	ME	19	0.66	
34	SHE	19	0.66	
35	LOOK	18	0.62	



N	1
Text File	3-6학~1.TXT
Bytes	15,262
Tokens	2,886
Types	438
Type/Token Ratio	15.18
Standardised Type/Token	25.20
Ave. Word Length	3.48
Sentences	507
Sent length	5.52
sd. Sent. Length	4.99
Paragraphs	62
Para. length	46.55
sd. Para. length	68.47
Headings	0
Heading length	
sd. Heading length	
1-letter words	194
2-letter words	760
3-letter words	658
4-letter words	648
5-letter words	329
6-letter words	125
7-letter words	72
8-letter words	43
9-letter words	38
10-letter words	12
11-letter words	1
12-letter words	4
13-letter words	0
14(+)-letter words	0

APPENDIX B

UCREL CLAWS7 tagset

APPGE	possessive pronoun, pre-nominal (e.g. my, your, our)
AT	article (e.g. the, no)
AT1	singular article (e.g. a, an, every)
BCL	before-clause marker (e.g. in order (that),in order (to))
CC	coordinating conjunction (e.g. and, or)
CCB	adversative coordinating conjunction (but)
CS	subordinating conjunction (e.g. if, because, unless, so, for)
CSA	as (as conjunction)
CSN	than (as conjunction)
CST	that (as conjunction)
CSW	whether (as conjunction)
DA	after-determiner or post-determiner capable of pronominal function (e.g. such, former, same)
DA1	singular after-determiner (e.g. little, much)
DA2	plural after-determiner (e.g. few, several, many)
DAR	comparative after-determiner (e.g. more, less, fewer)
DAT	superlative after-determiner (e.g. most, least, fewest)
DB	before determiner or pre-determiner capable of pronominal function (all, half)
DB2	plural before-determiner (both)
DD	determiner (capable of pronominal function) (e.g any, some)
DD1	singular determiner (e.g. this, that, another)
DD2	plural determiner (these,those)
DDQ	wh-determiner (which, what)
DDQGE	wh-determiner, genitive (whose)
DDQV	wh-ever determiner, (whichever, whatever)
EX	existential there
FO	formula
FU	unclassified word
FW	foreign word
GE	germanic genitive marker - (' or's)
IF	for (as preposition)
II	general preposition
IO	of (as preposition)
IW	with, without (as prepositions)
JJ	general adjective
JJR	general comparative adjective (e.g. older, better, stronger)
JJT	general superlative adjective (e.g. oldest, best, strongest)
JK	catenative adjective (able in be able to, willing in be willing to)
MC	cardinal number,neutral for number (two, three..)
MC1	singular cardinal number (one)
MC2	plural cardinal number (e.g. sixes, sevens)
MCGE	genitive cardinal number, neutral for number (two's, 100's)
MCMC	hyphenated number (40-50, 1770-1827)
MD	ordinal number (e.g. first, second, next, last)

MF	fraction,neutral for number (e.g. quarters, two-thirds)
ND1	singular noun of direction (e.g. north, southeast)
NN	common noun, neutral for number (e.g. sheep, cod, headquarters)
NN1	singular common noun (e.g. book, girl)
NN2	plural common noun (e.g. books, girls)
NNA	following noun of title (e.g. M.A.)
NNB	preceding noun of title (e.g. Mr., Prof.)
NNL1	singular locative noun (e.g. Island, Street)
NNL2	plural locative noun (e.g. Islands, Streets)
NNO	numeral noun, neutral for number (e.g. dozen, hundred)
NNO2	numeral noun, plural (e.g. hundreds, thousands)
NNT1	temporal noun, singular (e.g. day, week, year)
NNT2	temporal noun, plural (e.g. days, weeks, years)
NNU	unit of measurement, neutral for number (e.g. in, cc)
NNU1	singular unit of measurement (e.g. inch, centimetre)
NNU2	plural unit of measurement (e.g. ins., feet)
NP	proper noun, neutral for number (e.g. IBM, Andes)
NP1	singular proper noun (e.g. London, Jane, Frederick)
NP2	plural proper noun (e.g. Browns, Reagans, Koreas)
NPD1	singular weekday noun (e.g. Sunday)
NPD2	plural weekday noun (e.g. Sundays)
NPM1	singular month noun (e.g. October)
NPM2	plural month noun (e.g. Octobers)
PN	indefinite pronoun, neutral for number (none)
PN1	indefinite pronoun, singular (e.g. anyone, everything, nobody, one)
PNQO	objective wh-pronoun (whom)
PNQS	subjective wh-pronoun (who)
PNQV	wh-ever pronoun (whoever)
PNX1	reflexive indefinite pronoun (oneself)
PPGE	nominal possessive personal pronoun (e.g. mine, yours)
PPH1	3rd person sing. neuter personal pronoun (it)
PPHO1	3rd person sing. objective personal pronoun (him, her)
PPHO2	3rd person plural objective personal pronoun (them)
PPHS1	3rd person sing. subjective personal pronoun (he, she)
PPHS2	3rd person plural subjective personal pronoun (they)
PPIO1	1st person sing. objective personal pronoun (me)
PPIO2	1st person plural objective personal pronoun (us)
PPIS1	1st person sing. subjective personal pronoun (I)
PPIS2	1st person plural subjective personal pronoun (we)
PPX1	singular reflexive personal pronoun (e.g. yourself, itself)
PPX2	plural reflexive personal pronoun (e.g. yourselves, themselves)
PPY	2nd person personal pronoun (you)
RA	adverb, after nominal head (e.g. else, galore)
REX	adverb introducing appositional constructions (namely, e.g.)
RG	degree adverb (very, so, too)
RGQ	wh- degree adverb (how)
RGQV	wh-ever degree adverb (however)
RGR	comparative degree adverb (more, less)

RGT	superlative degree adverb (most, least)
RL	locative adverb (e.g. alongside, forward)
RP	prep. adverb, particle (e.g about, in)
RPK	prep. adv., catenative (about in be about to)
RR	general adverb
RRQ	wh- general adverb (where, when, why, how)
RRQV	wh-ever general adverb (wherever, whenever)
RRR	comparative general adverb (e.g. better, longer)
RRT	superlative general adverb (e.g. best, longest)
RT	quasi-nominal adverb of time (e.g. now, tomorrow)
TO	infinitive marker (to)
UH	interjection (e.g. oh, yes, um)
VB0	be, base form (finite i.e. imperative, subjunctive)
VBDR	were
VBDZ	was
VBG	being
VBI	be, infinitive (To be or not... It will be ..)
VBM	am
VBN	been
VBR	are
VBZ	is
VD0	do, base form (finite)
VDD	did
VDG	doing
VDI	do, infinitive (I may do... To do...)
VDN	done
VDZ	does
VH0	have, base form (finite)
VHD	had (past tense)
VHG	having
VHI	have, infinitive
VHN	had (past participle)
VHZ	has
VM	modal auxiliary (can, will, would, etc.)
VMK	modal catenative (ought, used)
VV0	base form of lexical verb (e.g. give, work)
VVD	past tense of lexical verb (e.g. gave, worked)
VVG	-ing participle of lexical verb (e.g. giving, working)
VVGK	-ing participle catenative (going in be going to)
VVI	infinitive (e.g. to give... It will work...)
VVN	past participle of lexical verb (e.g. given, worked)
VVNK	past participle catenative (e.g. bound in be bound to)
VVZ	-s form of lexical verb (e.g. gives, works)
XX	not, n't
ZZ1	singular letter of the alphabet (e.g. A,b)
ZZ2	plural letter of the alphabet (e.g. A's, b's)

APPENDIX C

Part of the CLAWS7 tagset of the IFKC corpus

Rog_NP1 ran_VVD up_II the_AT path_NN1 to_II Toad_NP1 's_GE house_NN1 ._.
He_PPHS1 knocked_VVD on_II the_AT front_JJ door_NN1 ._.
There_EX was_VBDZ no_AT answer_NN1 ._.
"_" Toad_VV0 ,_, Toad_NP1 ,_, "_" shouted_VVD Frog_NN1 ,_, "_" wake_VV0 up_RP ._.
It_PPH1 is_VBZ spring_NN1 !_! "_"
"_" Blah_UH ,_, "_" said_VVD a_AT1 voice_NN1 from_II inside_RL the_AT house_NN1 ._.
"_" Toad_VV0 !_!
Toad_VV0 !_! "_" cried_VVD Frog_NN1 ._.
"_" The_AT sun_NN1 is_VBZ shining_VVG ._.
The_AT snow_NN1 is_VBZ melting_VVG ._.
Wake_VV0 up_RP ._. "_"
"_" I_PPIS1 am_VBM not_XX here_RL ,_, "_" said_VVD the_AT voice_NN1 ._.
Frog_NN1 walked_VVD into_II the_AT house_NN1 ._.
it_PPH1 was_VBDZ dark_JJ ._.
all_DB the_AT shutters_NN2 were_VBDR closed_VVN ._.
"_" Toad_VV0 ,_, where_RRQ are_VBR you_PPY ?_? "_" called_VVD Frog_NN1 ._.
"_" Go_VV0 away_RL ,_, "_" said_VVD the_AT voice_NN1 from_II a_AT1 corner_NN1 of_IO the_AT room_NN1 ._.
Toad_NN1 was_VBDZ lying_VVG in_II bed_NN1 ._.
He_PPHS1 had_VHD pulled_VVN all_DB the_AT covers_NN2 over_II his_APPGE head_NN1 ._.
Frog_NN1 pushed_VVD Toad_NP1 out_II21 of_II22 bed_NN1 ._.
he_PPHS1 pushed_VVD him_PPHO1 out_II21 of_II22 the_AT house_NN1 and_CC onto_II the_AT front_JJ porch_NN1 ._.
Toad_NP1 blinked_VVD in_II the_AT bright_JJ sun_NN1 ._.
"_" Help_VV0 !_! "_" said_VVD Toad_NP1 ._.
"_" I_PPIS1 can_VM not_XX see_VVI anything_PN1 ._. "_"
"_" Do_VD0 not_XX be_VBI silly_JJ ,_, "_" said_VVD Frog_NN1 ._.
"_" What_DDQ you_PPY see_VV0 is_VBZ the_AT clear_JJ warm_JJ light_NN1 of_IO April_NPM1 ._.
And_CC it_PPH1 means_VVZ that_CST we_PPIS2 can_VM begin_VVI a_AT1 whole_JJ new_JJ year_NNT1 together_RL ,_, Toad_NP1 ._.
Think_VV0 of_IO it_PPH1 "_" said_VVD Frog_NN1 ._.
"_" We_PPIS2 will_VM skip_VVI through_II the_AT meadows_NN2 and_CC run_VVN through_II the_AT woods_NN2 and_CC swim_VV0 in_II the_AT river_NN1 ._.
In_II the_AT evenings_NNT2 we_PPIS2 will_VM sit_VVI right_RR here_RL on_II this_DD1 front_JJ porch_NN1 and_CC count_VV0 the_AT stars_NN2 ._. "_"
"_" You_PPY can_VM count_VVI them_PPHO2 ,_,
Excuse_VV0 me_PPIO1 ._.
Where_RRQ is_VBZ Hagkuk_NP1 Elementary_JJ School_NN1 ?_?

```
Over_II there_RL ._.
I_PPIS1 am_VBM going_VVG there_RL too_RR ._.
Let_VV0 us_PPIO2 go_VVI together_RL ._.
Thanks_NN2 ._.
I_PPIS1 am_VBM Tan_JJ ._.
What_DDQ is_VBZ your_APPGE name_NN1 ?_?
My_APPGE name_NN1 is_VBZ Jinho_NP1 ._.
Hello_UH ,_, Mrs._NNB Smith_NP1 ._.
Hi_UH ,_, Jinho_NP1 ._.
How_RRQ are_VBR you_PPY ?_?
I_MC1 am._RA fine_JJ ,_, thank_VV0 you_PPY ._.
How_RRQ about_II you_PPY ?_?
Very_RG well_RR ,_, thanks_NN2 ._.
Jinho_NN1 ,_, this_DD1 is_VBZ Ann_NP1 ._.
She_PPHS1 is_VBZ from_II New_NP1 York_NP1 ._.
Hi_UH ,_, Ann_NP1 ._.
I_PPIS1 am_VBM Jinho_NP1 ._.
Nice_JJ to_TO meet_VVI you_PPY ,_, Jinho_NP1 ._.
Nice_JJ to_TO meet_VVI you_PPY ,_, too_RR ._.
Oh_UH ,_, What_DDQ is_VBZ your_APPGE name_NN1 ?_?
My_APPGE name_NN1 is_VBZ Joon_NP1 ._.
Nice_JJ to_TO meet_VVI you_PPY ,_, Joon_NP1 ._.
Nice_JJ to_TO meet_VVI you_PPY ,_, too_RR ._.
Oh_UH ,_, do_VD0 you_PPY like_VVI Basket_NN1 ball_NN1 ?_?
Yes_UH ,_, I_PPIS1 do_VD0 ._.
How_RRQ about_II you_PPY ?_?
I_PPIS1 love_VV0 basketball_NN1 ._.
Really_RR ?_?
Umm_UH ._.
Wow_UH !_!
What_DDQ day_NNT1 is_VBZ it_PPH1 today_RT ?_?
```

Kyu-Hwa Kim
Department of English Education
Cheongju National University of Education
Wooju Dongbaek Apartment 201-106
Janggwan-ri Jincheon-eup Jinchen-gun Chungcheongbuk-do 365-802, Korea
Tel: (043) 533-8026/ C.P.: 011-9836-8026
Email: kimkh5004@hanmail.net

Received 31 January 2009
Reviewed 14 February 2009
Revised version received 1 April 2009