

A Corpus-based Analysis of English Newspaper Headlines

Moon Jee-young

Department of English Education

Graduate School of Education, Chonnam National University

I. Introduction

Recently using English newspapers as a useful teaching tool for English has emerged among educators due to its easy accessibility and wide range of content. In this stream, in particular, English headlines seem to be catchy in getting learners interested and motivated. In general, English headlines are said to have a unique style of language. There are some generalizations made on this language, among which the following three are most frequently mentioned: Headlines often omit articles and copular or auxiliary *be*; tenses are simplified to the present tense; conjunctions are also often excluded.

This study examines how much truth there is in the three generalizations on the headlines by examining an actual sample corpus drawn from *The Korea Herald* Headlines Corpus (KHC). In addition to this, other corpora, the British National Corpus (BNC) Sampler Written, the British National Corpus (BNC) and a sample corpus drawn from *The New York Times* Headlines Corpus (NHC) by Chung (2006) were referred to for the purpose of comparisons.

II. Methodology

1. Data

The data in this paper is drawn from a corpus (50 files, 180 headlines, 1,095 words) compiled from *The Korea Herald* headlines that appeared between January and June 2007. The headlines were drawn arbitrarily but they were mainly from recent politics, economics, culture, science, and entertainment. Admittedly, the corpus, KHC (*The Korea Herald* Headlines Corpus), is of a small size, but the size does not make much significant difference since the patterns in headlines are not so varied.

2. Tagging Tools and Procedure

I analyzed the KHC raw data file using Wmatrix, a software tool for corpus analysis and comparison. Wmatrix first processed part-of-speech tagging using CLAWS POS tagger. The raw data file of the KHC was tagged word by word.

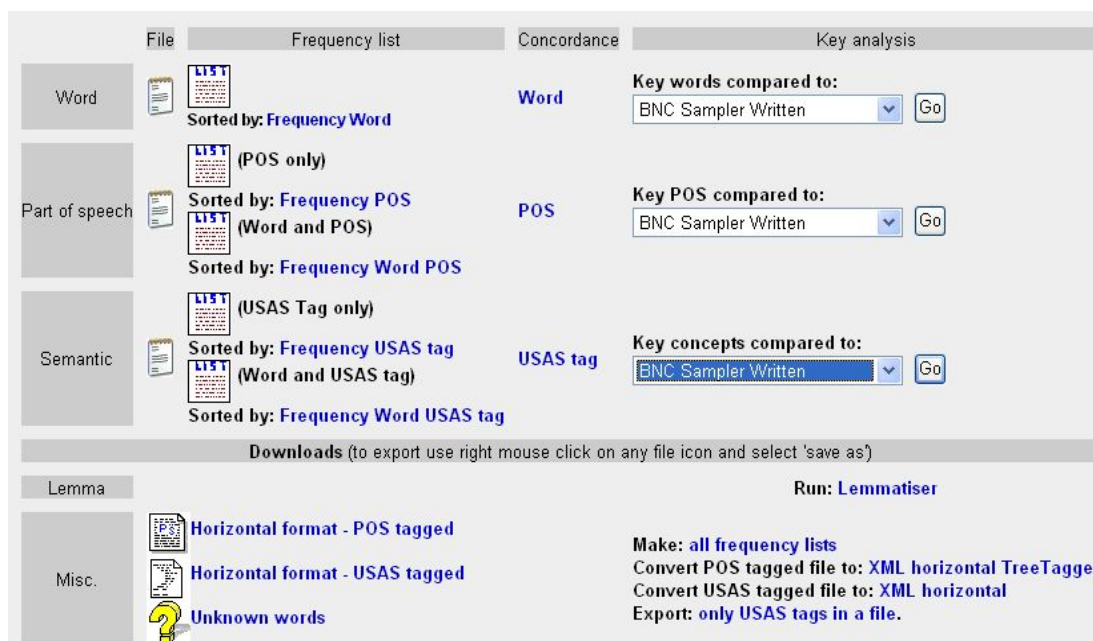
Then, I conducted semantic tagging using UCREL Semantic Analysis System (USAS). USAS undertakes the automatic semantic analysis of text. It has 21 major

discourse fields and contains 13,153 single words, 4,444 proper names, and 713 MWE (Multi-Word-Expression) templates. However, this semantic analysis system could not identify some proper names in the KHC and classified them into "Z99" which stands for unmatched words.

In addition, Wmatrix makes it possible to compare the frequency lists against other larger normative corpus such as the BNC Sampler. In this study, the BNC Sampler Written was compared against the KHC. The key comparison results are sorted on the LL (log-likelihood) field which shows how significant the differences are. The items with a '+' code shows overuse in the text as compared to the standard English corpora. Here, I only referred to the items with a LL value over 7 since it shows the statistical significance.

Figure 1 indicates the view of the KHC and shows the result of analysis by Wmatrix. It mainly consists of three parts, Word, Part of speech, and Semantic field. Also, it has a Frequency list, Concordance, and Key word analysis.

Figure 1. View of the KHC analysis



III. Results and Discussion

1. Results

1) 20 Most Frequently Used Words in the KHC

Figure 2 shows the top 20 high frequent words in the KHC. The most frequent words are *to*, *in*, *for*, *of*, *on*. From the top to 4th are all function words and pronouns such as *Korea*, *U.S.*, *Seoul*, *Korean* are also high rank.

Figure 2. 20 Most Frequently Used Words in the KHC

Word	Frequency	Relative Frequency	
to	42	3.84	Concordance
in	31	2.83	Concordance
for	21	1.92	Concordance
of	18	1.64	Concordance
on	17	1.55	Concordance
Korea	10	0.91	Concordance
U.S.	10	0.91	Concordance
new	8	0.73	Concordance
's	7	0.64	Concordance
at	7	0.64	Concordance
Seoul	7	0.64	Concordance
talks	6	0.55	Concordance
korean	6	0.55	Concordance
this	5	0.46	Concordance
two	5	0.46	Concordance
rules	5	0.46	Concordance
a	5	0.46	Concordance
may	5	0.46	Concordance
rise	4	0.37	Concordance
Roh	4	0.37	Concordance
N.K.	4	0.37	Concordance
FTA	4	0.37	Concordance
the	4	0.37	Concordance
faces	4	0.37	Concordance
more	4	0.37	Concordance
banks	4	0.37	Concordance
medical	4	0.37	Concordance
China	4	0.37	Concordance
with	4	0.37	Concordance
firms	4	0.37	Concordance
'	4	0.37	Concordance

2) POS Frequency in the KHC

Figure 3 shows the POS frequency in the KHC. According to it, the most frequent 10 POSs in the KHC are NN1 (singular common noun), NN2 (plural common noun), JJ (general adjective), NP1 (singular proper noun), II (general preposition), VVZ (-s form of lexical verb), VVI (infinitive), VV0 (base form of lexical verb), TO (infinitive marker), IF (preposition *for*) in order.

Figure 3. POS Frequency in the KHC

POS	Frequency	Relative Frequency	
NN1	279	25.48	Concordance List
NN2	130	11.87	Concordance List
JJ	121	11.05	Concordance List
NP1	97	8.86	Concordance List
II	79	7.21	Concordance List
VVZ	62	5.66	Concordance List
VVI	39	3.56	Concordance List
VV0	37	3.38	Concordance List
TO	32	2.92	Concordance List
IF	21	1.92	Concordance List
MC	19	1.74	Concordance List
IO	18	1.64	Concordance List
VVG	16	1.46	Concordance List
VVN	14	1.28	Concordance List
VVD	12	1.10	Concordance List
GE	11	1.00	Concordance List
NNT1	9	0.82	Concordance List
MD	8	0.73	Concordance List

News deals with mostly current issues or events and describes the here/now happening. There is a comparison of the verbs which show the rate of present tense versus past tense in the KHC. The base form of lexical verbs or -ing participle of lexical verbs do not have any tense so that only VVZ which is -s form of lexical verbs and VVD which is past tense of lexical verbs are compared here to show the rate of the tense.

Figures 4 and 5 show the distribution of present and past tense forms through concordance in the KHC respectively. When comparing, the rate of frequency between VVZ and VVD, the difference is quite overt. The rate of VVZ is overwhelmingly larger than that of VVD. Strictly speaking, VVD is observed about five times more than VVZ.

Figure 4. Concordance of VVZ

Wrote 62 occurrences.			
id to North Korea Assembly panel	South Korea resumes	passes	rice aid to North Korea Assembly
ses pension reform bill S. Korea	develops	denies	pension reform bill S. Korea deve
ix-party negotiations Spy agency	declines	Panel rules	cutting-edge infantry fighting ve
year Foreign holdings in market	declines	Panel rules	reports of Kim Jong-il 's health
oldings in market declines Panel	rules	Panel rules	Roh violated election
			Roh violated election law again H

Figure 5. Concordance of VVD

Wrote 12 occurrences.			
market declines Panel rules Roh	violated		election law again Hankuk Univer
products start July 1 S. Korea	ranked		11th in 2006 military spending H
decline Two more medical groups	investigated		Three Koreans abducted in Nigeri
roups investigated Three Koreans	abducted		in Nigeria GNP leadership patche
se to bar beatings Thai workers	found		dead in house fire Mortgage rate
ree trade talks 55 by-elections	slated		for April 25 Anti-FTA protester

However, the frequency of the article is significantly lower in the KHC. Article (AT) only occurred 5 times in the KHC and also singular article (AT1) occurred 5 times as well.

Conjunction is also said to be often excluded in the headline. CLAWS7 Tagset has 7 kinds of conjunction; CC (coordination conjunction), CCB (adversative coordinating conjunction), CS (subordinating conjunction), CSA (*as* as conjunction), CSN (*than* as conjunction), CST (*that* as conjunction), CSW (*whether* as conjunction) and among these only CC and CSA appeared in the KHC. However, conjunction occurred only twice in the KHC and in detail they are *nor* and *as*.

3) Word and POS Frequency

As figure 6 shows, "TO" which is the to-infinitive marker is distinguished from "II" which is the general preposition, though their form is exactly same. In detail, the infinitive marker *to* appeared 32 times and the preposition *to* was used 10 times. According to the POS frequency, NN1 is the top rank in the KHC. However, in terms of Word-Pos frequency, the 1st rank is not NN1 but the to-infinitive marker. Other high rank words and their POS are mostly prepositions such as *in*, *for*, *of*, *on*.

Figure 6. Word and POS Frequency List

Word	POS	Frequency	Relative Frequency	
to	TO	32	2.92	Concordance
in	II	31	2.83	Concordance
for	IF	21	1.92	Concordance
of	IO	18	1.64	Concordance
on	II	17	1.55	Concordance
U.S.	NP1	10	0.91	Concordance
Korea	NP1	10	0.91	Concordance
to	II	10	0.91	Concordance
new	JJ	8	0.73	Concordance
's	GE	7	0.64	Concordance
at	II	7	0.64	Concordance
Seoul	NP1	7	0.64	Concordance
korean	JJ	6	0.55	Concordance
two	MC	5	0.46	Concordance
this	DD1	5	0.46	Concordance
talks	NN2	5	0.46	Concordance

4) Word and USAS Tag Frequency

Here is the Word and USAS Tag Frequency List and it is possible to see concretely each word and its semantic tag. According to UCREL Semantic Tagset, Z5 stands for "Grammatical bin", Z8 is "Pronouns", Z2 is "Geographical names", T3- is "Time: New and young", I1.1 is "Money and pay", Q2.1 is "Speech: Communicative", N1 is "Numbers", and M6 means "Location and direction" respectively.

Figure 7. Word and USAS Tag Frequency List

Word	Semtag	Frequency	Relative Frequency	
to	Z5	42	3.84	Concordance
in	Z5	31	2.83	Concordance
for	Z5	21	1.92	Concordance
of	Z5	18	1.64	Concordance
on	Z5	17	1.55	Concordance
Korea	Z2	10	0.91	Concordance
U.S.	Z8	10	0.91	Concordance
new	T3-	8	0.73	Concordance
's	Z5	7	0.64	Concordance
Seoul	I1.1	7	0.64	Concordance
at	Z5	7	0.64	Concordance
korean	Z2	6	0.55	Concordance
talks	Q2.1	6	0.55	Concordance
a	Z5	5	0.46	Concordance
two	N1	5	0.46	Concordance
this	M6	5	0.46	Concordance

2. Discussion

1) Comparison between the KHC and NHC

There is some significant discrepancy in the frequency rank between the KHC and NHC. At first, the definite article *the* ranks 19th in the KHC but 5th in the NHC. Also, the preposition and infinitive marker *to* is the 1st in the KHC but only ranks 13th in the NHC. In addition to this, the indefinite article *a* takes 3rd place in the NHC but 14th in the KHC.

Table 1. 20 Most Frequently Used Words in the KHC and NHC

KHC			NHC		
word	token	frequency (%)	word	token	frequency (%)
to	42	3.84	in	75	3.58
in	31	2.83	of	47	2.24
for	21	1.92	a	42	2.01
of	18	1.64	the	38	1.86
on	17	1.55	to	37	1.76
Korea	10	0.91	for	36	1.72
U.S.	10	0.91	is	33	1.58
new	8	0.73	on	23	1.1
's	7	0.64	and	21	1
at	7	0.64	at	20	0.96
Seoul	7	0.64	new	18	0.86
talks	6	0.55	U.S	17	0.81
Korean	6	0.55	to	16	0.76
this	5	0.46	Bush	15	0.72
two	5	0.46	says	15	0.72
rules	5	0.46	be	11	0.53
a	5	0.46	from	11	0.53
may	5	0.46	Iraq	11	0.53
rise	4	0.37	with	11	0.53

It is surprising that articles take the highest rank in the NHC since this runs against the generalization that articles tend to undergo deletion in the headline. In the KHC, however, articles do not take such a high place. Therefore it is necessary to note that there is some discrepancy between KHC and NHC even though they are both headline corpora.

The frequency of the article is significantly lower in the KHC than in the NHC. As mentioned already in the generalizations of the headlines, articles are said to be generally deleted and are actually shown rarely in the KHC. It is an expected result; on the other hand, the NHC shows the high frequency of articles, especially the indefinite ones.

Thus, it is necessary to compare other headline corpora to find more about the frequency and deletion of articles in the headlines. Although the results of the KHC do not correspond with the results of the NHC, it is still meaningful in that two headline corpora were first compared.

Another interesting fact observed above is that in the KHC, as well as the NHC, the 5 most frequently used words are all function words and, in particular, they are prepositions, articles or *to*-infinitive markers. The most frequent word in the KHC is *to* which is either a preposition or a *to*-infinitive marker.

When comparing two corpora, there is some similarity in that nouns, verbs, adjectives, and prepositions all rank high. Especially, the preposition *in* ranks 2nd in the KHC and 1st in the NHC. However, despite this similarity, there is some significant discrepancy in the POS frequency rank between the KHC and the NHC.

The article ranks 17th in the KHC, but 5th in the NHC. This is interesting since the NHC apparently runs against the generalization that articles tend to undergo deletion in the headline. In contrast to the NHC, the KHC proves one of the generalizations on headlines that articles usually do not appear in the headline.

Also, it is needed to analyze the contrast between NN1, singular common noun, and NN, common noun, since the Wmatrix couldn't read NN correctly in some cases. For example, *rice* was included in the NN1 though it is neutral for number. Also *Song* which is a family name in Korea was regarded as NN1 as if it were lyrics even though its POS is NP1, singular proper noun. By the way, the NHC classified POS into 13 categories and never mentioned the germanic genitive marker so that it is impossible to compare it to the KHC.

What needs to be explained is the different context of each corpus; that is, English usage of each country could be varied. The NHC is based on the English used in the U.S.A while the KHC is based on the English used in Korea and this could have affected the POS frequency. Thus, it is necessary to study more on the different usage of English in each country in the further study.

2) USAS Tag Frequency

An interesting point of the USAS tag frequency list is that the three most frequent semantic tags in the KHC are all "Z" which means 'names and grammar'.

Figure 8. USAS Tag Frequency List

Semtag	Frequency	Relative Frequency	
Z5	177	16.16	Concordance List
Z99	60	5.48	Concordance List
Z2	47	4.29	Concordance List
M6	24	2.19	Concordance List
Q2.2	23	2.10	Concordance List
N1	21	1.92	Concordance List
T1.3	19	1.74	Concordance List
M1	19	1.74	Concordance List
I1.1	18	1.64	Concordance List
G2.1	18	1.64	Concordance List
Z1	17	1.55	Concordance List
S5+	17	1.55	Concordance List
P1	15	1.37	Concordance List
S7.1+	15	1.37	Concordance List

In detail, the top rank Z5 represents 'Grammatical bin'. When looking over the concordance of Z5, it is possible to find what it means. That is, it mainly contains prepositions, genitive markers, or articles. These words are all function words and do not have semantic meaning so that they were put in 'Grammatical bin'.

Another interesting and significant fact shown in the tag frequency result is about the second rank Z99 which means 'Unmatched'. Z99 includes most Korean family names, city names, and abbreviations like *Jong-il*, *KOSPI*, *ROH*, *Hankuk*, *Choi*, *PGA*, *Daegu*, *FTA*. That is, the semantic tagset could not read and identify a

semantic field which is related to Korea and Korean in the KHC. Thus, it is necessary to develop a semantic tagset which could distinguish and classify unfamiliar or uncommon words such as proper nouns correctly for reliability.

3) Comparison between the KHC and BNC Sampler Written

In Figure 9, O1 is observed frequency in the KHC and O2 is observed frequency in the BNC Sampler Written. '+' indicates overuse in O1 relative to O2 and '-' indicates underuse in O1 relative to O2.

Figure 9. Key Words compared to BNC Sampler Written

	Item	O1	%1	O2	%2	LL
Concordance	Korea	10	0.91	3	0.00 +	121.68
Concordance	U.S.	10	0.91	9	0.00 +	109.45
Concordance	Seoul	7	0.64	3	0.00 +	82.79
Concordance	korean	6	0.55	4	0.00 +	67.98
Concordance	FTA	4	0.37	0	0.00 +	54.29
Concordance	N.K.	4	0.37	0	0.00 +	54.29
Concordance	Roh	4	0.37	0	0.00 +	54.29
Concordance	China	4	0.37	2	0.00 +	46.65
Concordance	S._Korea	3	0.27	0	0.00 +	40.72
Concordance	Samsung	3	0.27	0	0.00 +	40.72
Concordance	ceo	3	0.27	0	0.00 +	40.72
Concordance	talks	6	0.55	104	0.01 +	35.09
Concordance	rules	5	0.46	84	0.01 +	29.54
Concordance	Hyundai	2	0.18	0	0.00 +	27.14
Concordance	Koreas	2	0.18	0	0.00 +	27.14
Concordance	N._Korea	2	0.18	0	0.00 +	27.14

Figure 9 illustrates the overuse of proper noun such as *U.S.*, *Korea*, *Seoul* in the KHC which means the close relationship between U.S. and Korea. Also, it indicates the fact that the KHC deals with international news that focuses on Korea. This could be the unique characteristics of *The Korea Herald* since the NHC does not have a proper noun in its high frequency words list. While in the BNC Sampler Written, *talks* and *rules* are more common than proper nouns such as *Korea*, *U.S.* That is, the frequency of *talks* is 104 while that of *Korea* is only 3.

4) POS Comparison between the KHC and BNC Sampler Written

Figure 10 illustrates the comparison between the KHC and the BNC Sampler Written. According to LL value, we can notice that "VVZ", -s form of lexical verbs, is overused in the KHC compared to the BNC Sampler Written. This is because the headline uses the present tense overwhelmingly rather than using the past tense. This means that headlines tend to use a present tense much more than a past tense and it also shows the fact that most tenses are simplified to the present tense in the headline.

However, the past tense does exist although it is less frequent than the present tense. Also, the "TO", infinitive marker, shows that headlines tend to use the infinitive marker especially when expressing future events.

Figure 10. Key POS compared to BNC Sampler Written

	Item	01	%1	02	%2	LL
List Concordance	VVZ	62	5.66	7602	0.79 +	137.81
List Concordance	NN1	279	25.48	147395	15.22 +	62.74
List Concordance	NN2	130	11.87	55665	5.75 +	54.37
List Concordance	NP1	97	8.86	39079	4.04 +	46.83
List Concordance	VV0	37	3.38	11012	1.14 +	31.43
List Concordance	JJ	121	11.05	74927	7.74 +	13.66
List Concordance	TO	32	2.92	14041	1.45 +	12.59
List Concordance	ND1	5	0.46	700	0.07 +	9.99
List Concordance	IF	21	1.92	8765	0.91 +	9.34
List Concordance	NPM1	6	0.55	1231	0.13 +	8.30

Therefore this comparison table is useful to show the characteristics of headlines comparing to general written corpus. Except for "VVZ, TO" other overused POSs which have significant differences according to LL value are NN1 (singular common noun), NN2 (plural common noun), NP1 (singular proper noun), VV0 (base form of lexical verb), JJ (general adjective), ND1 (singular noun of direction), IF (preposition *for*), and NPM1 (singular month noun).

5) Semantic Comparison between the KHC and BNC Sampler Written

According to Figure 11, Z2 which is 'Geographical names' is overused in the KHC compared to the BNC Sampler Written and it shows the unique characteristics of English headline in Korea. That is, the KHC has a high frequency of the semantic fields of 'law and order', 'business and money', 'warfare, defence and army', and 'education'. This indicates the current issue or situation of Korea, for example the 'warfare, defence and army' accounts for divided Korean peninsula and 'education' shows high attention for being educated and cultivated in Korean society.

Figure 11. Key Concepts compared to BNC Sampler Written

Item	01	%1	02	%2	LL	
Z2	47	4.29	14502	1.50 +	37.71	Geographical names
G2.1	18	1.64	2418	0.25 +	37.21	Law and order
Z99	60	5.48	22165	2.29 +	34.82	Unmatched
I1.1	18	1.64	2654	0.27 +	34.41	Money and pay
Y1	7	0.64	778	0.08 +	16.75	Science and technology in general
P1	15	1.37	3691	0.38 +	16.69	Education in general
I2.1	12	1.10	2634	0.27 +	15.37	Business: Generally
I1.1++	2	0.18	15	0.00 +	14.86	Money: Affluence
I2.2	12	1.10	2738	0.28 +	14.68	Business: Selling
A13	1	0.09	0	0.00 +	13.57	Degree
N3.7---	2	0.18	24	0.00 +	13.10	Short and narrow
I3.1-	4	0.37	296	0.03 +	12.47	Unemployed
G3	12	1.10	3152	0.33 +	12.24	Warfare, defence and the army; weapons
T2+	11	1.00	2757	0.28 +	11.95	Time: Beginning
G2.1-	5	0.46	570	0.06 +	11.74	Crime
S5+	17	1.55	5811	0.60 +	11.44	Belonging to a group
I1.3	7	0.64	1254	0.13 +	11.16	Money: Cost and price
M6	24	2.19	9859	1.02 +	11.08	Location and direction

As we have seen in the POS and USAS tagging results above, the study on *The Korea Herald* headlines leads us not only to analyze the linguistic characteristics of the headlines, but to look into the social and cultural aspects. Therefore it is much more interesting than just analyzing the POS of headlines and gives learners some opportunities to consider social or cultural facts when reading headlines.

IV. Conclusion

The findings of the KHC semantic tagging include the following. First, the KHC showed a high frequency of 'names and grammar'. Especially, geographical names are overused in the KHC compared to the BNC Sampler Written. Second, the frequency of 'Grammatical bin' which consists of prepositions, genitive markers, or articles was high. Third, it also showed the high frequency of 'Unmatched' words. It means that the semantic tagset could not classify some words such as Korean family names, city names, and abbreviations. This is related to limitation of the semantic tagging.

However, the present study only focused on *The Korea Herald* headlines and referred to the NYT headlines, BNC, BNC Sampler Written when required. Thus, it is difficult to see whether there is any other different cases in the characteristics of the headline. It would be necessary in the future to collect more headlines from different newspapers, for example a newspaper in U.K or Japan, China, India, etc. The results of such studies will serve as important references for studying on the generalizations and characteristics of headlines and even for better understanding the unique style of each country's English.

Although all these results cannot be generalized to other headlines, they will serve as a good model for the further study on the characteristics of the headlines. In particular, the analysis of semantic tagging of headlines, despite its limitation, could lead learners to be interested in social and cultural aspects since they are interesting and relevant to them. Therefore, once learners are interested in reading headlines, it is possible for them to infer a complete sentence of a news story, predict the whole news story, and rewrite their own headlines. In this way, English headlines could be utilized as useful learning materials.

REFERENCES

- 김성식. (2007). 코퍼스와의 영어교육, *영어교육*, 62(2), 281-307.
- 김용균. (2005). 인터넷 영자신문을 활용한 영어 읽기 교육 연구. 석사학위논문. 중앙대학교 교육대학원.
- 김준기. (2007). 세계 영어의 지역적 차이에 대한 말뭉치 분석연구, *새한영어영문학*,

- 49(2), 143-155.
- 김현배. (2003). 영자신문을 고등학교 영어학습 동기유발 자료로 활용하는 방안. *영미어문학 연구*, 19(2). 43-59.
- 박재현. (1999). *Teaching reading comprehension through English newspapers*. 석사학위논문. 서강대학교 교육대학원.
- 신영순. (2003). 영자신문을 활용한 영어교육. 석사학위논문. 경희대학교 교육대학원.
- 이영옥. (2002). 한영간 신문 표제어 표현의 차이 -동음이의어와 인용문. *커뮤니케이션학 연구*. 10(1), 150-170.
- 이정원 & 이태옥. (2007). 영자신문을 활용한 고등학생의 영어 독해능력 향상에 관한 사례 연구. *영어어문교육*. 13(2). 159-180.
- 최상희. (1999). *한국 NIE의 성공전략*. 대일.
- 허병두. (1997) *신문활용교육이란 무엇인가*, 서울: 중앙M&B
- Bell, A. (1991), *The language of news media*. Oxford: Blackwell.
- Choi, Yong-Son. (2002). The applications of on-line English newspapers into classroom instruction. *영어어문교육*, 8(2), 1-22.
- Chung, Daeho (2006). A quantitative study on the English headline grammar: Focused on the NYT headlines. *Studies in Modern Grammar* 46, 33-68.
- Conrad, S. (2000). *Will corpus linguistics revolutionize grammar teaching in the 21st century?* *TESOL Quarterly*, 34, 47-63.
- Grundy, P. (1998). *Newspapers: Resource books for teachers*. (4th). Hong Kong: Oxford University Press.
- Kim, Y.-J (2003). A Sociolinguistic analysis of newspaper editorial titles. *The Sociolinguistic Journal of Korea*, 11(1), 31-54.
- Kitao, K. (1996). Teaching the English newspaper effectively. *The Internet TESL Journal*, 2(3), Retrieved March 7, 2002, from the World Wide Web:
<http://www.aitech.ac.jp/~iteslj/Lessons/Kitao-Newspaper.html>
- McEnery, T., & Wilson, A. (2001). *Corpus linguistics* (2nd ed.). Edinburgh, U.K.: Edinburgh University Press.
- Olivares, R. (1993). Using the newspaper to teach ESL learners. *International Reading Association*. Newark, Delaware. USA.
- Reah, D. (1998). *The language of newspapers*. London: Routledge.
- Saxena, S. (2006). *Headline writing*. New Delhi: Sage Publications.

Web Sites Referred to

- <http://pie.usna.edu/explorepg.html>
<http://pie.usna.edu/POScodes.html>
<http://info.ox.ac.uk/bnc/getting/sampler.html>
<http://www.natcorp.ox.ac.uk>

<http://www.edict.com.hk>

<http://ucrel.lancs.ac.uk/wmatrix>

<http://ucrel.lancs.ac.uk/wmatrix2.html>

<http://www.usingenglish.com/resources/text-statistics.php>

<http://www.answers.google.com/answers/threadview?id=470042>

http://titania.cobuild.collins.co.uk/boe_info.html

<http://clwww.essex.ac.uk>

<http://uk.cambridge.org/elt/reference/cic.htm>