

## **You know Korean English? Lexical priming in short strings of Korean Spoken English**

Glenn S. HADIKIN, University of Portsmouth, UK

This paper introduces the emerging new variety of Korean English to what may be a relatively new audience. Shim (1999) suggested that Korean English was codified in the form of school textbooks by the end of the 1980s; now, more than 20 years later it would be reasonable to expect a number of Koreanised forms to be detectable in contemporary Korean spoken English but, to my knowledge, there has never been a corpus-driven study that explores this variety. With a theoretical backdrop based on a new theory of language, Hoey's theory of Lexical Priming (Hoey, 2005), I explore three high frequency lexical strings, '*do you know*', '*but you know*' and '*and you know*' and their variation in two corpora of Korean English and two of British English. The results suggest that Korean speakers use certain strings as extended connectives to 'buy' extra processing time and the study raises interesting questions about the relationship between string form and meaning.

**Keywords:** Lexical Priming; Spoken Englishes; Corpus Linguistics; Konglish; Self-Depreciation; Necessitation; Externalization

### **1. Introduction**

This research described in this paper started out as a series of thoughts and observations as I taught English in South Korea. As I was travelling around several cities and provinces I noticed that people would speak English to me in a distinctive way that was surprisingly consistent across the country but not simply because of accent features or their choice of vocabulary—it was more of a 'phraseological accent'.

Shim (1999) tells us that there had been 116 years of formal English education when her paper was published but, as Hadikin (2006) points out one is more likely to hear negative references to *Konglish*: a disparaging term often used to refer to unusual vocabulary choices or a strong accent rather than *Korean English* which would imply a more standardised version. Standardised is, indeed, the correct word but few people are aware of the

standardisation in Korean high schools that allow usages such as *valleys that are covered with fresh green growths, we go to school day by day* and *she studied very hard. After all, she passed the examination* (Shim, 1999). The author explains that *growths* is taught as being a count form equivalent to *trees and plants*, *day by day* is closely synonymous with *every day* and that *after all* has the same function as *finally* (but reports nothing that would affect the focus of this paper: the short string *you know*). For the purposes of this paper it will be assumed that Shim's claim of codification marks the beginning of Korean English as a new variety. If such forms have been taught for over 20 years (the examples were taken from a textbook (Chang, Kim & Chang, 1989)) they would have some effect on the English spoken in Korea on the whole but how significant would the effect really be?

To begin to explore this issue, or, more accurately, to capture a snapshot of Korean English as it is actually used rather than simply looking for peculiar forms, I collected recordings of Korean adults speaking to me in informal interview-like exchanges discussing topics such as the volunteers' hobbies and interests and any culture differences they may know about between the UK and Korea. Data was collected from two groups—one in Seoul, Korea and one in Liverpool, UK, so that I could explore the effects of *Lexical Priming* which will be described in more detail in the following section. The two word string (in this paper I will use the terms *chunks* or *strings* to refer to all strings of words that may or may not be stored as wholes by the language user) *you know* is the main focus of this paper (the data was collected specifically for an earlier PhD project) and as such the overriding research questions are as follows:

1. What are the highest frequency strings containing *you know* in the four corpora?
2. What are the key sites for variation within each string and what can this tell us about the use of English in South Korea?

The short string *you know* was chosen at first because it is the most high frequency two word structure in both the Seoul data and the British data but, mostly, because it provides a convenient starting point for an interesting discussion about this new, emerging variety. Its limited lexical content may make it appear an unusual choice at first but in section 2.2 I will show how this, in fact, makes it an excellent choice for a discussion of priming effects.

## 2. Background

Since Shim (1999) first reported the codification of Korean English there has, to my knowledge, been no published corpus-driven study of the English

spoken in South Korea. Studies have tended to focus on ideological issues surrounding the use of English or have been very much pedagogy oriented with little descriptive value. Park (2009, cited in Porter 2011), for example, tells us that English is ‘a language that drives Koreans into strange and irrational obsessions which unduly burden every Korean, both emotionally and financially’ (2009, p. 2); the author goes on to highlight three principle components of their language ideology:

1. *necessitation* is the belief that English is now essential for professional success in a global community;
2. *externalization* is the still commonly held belief that English is the language of the other and thus is a source for identity conflicts; and
3. *self-depreciation* is a shared idea that Korean people can never be ‘good’ at English no matter how hard they try.

I mark the word *good* as potentially problematic here because it relates to the perception that Korean English users should try to sound exactly like native English speakers; for the purposes of this paper, however I argue that Korean English users can begin to develop their own norms in the spirit of Achebe:

I feel that the English language will be able to carry the weight of my African experience . . . but it will have to be a new English, still in communion with its ancestral home but altered to suit its new African surroundings (Achebe, 1975, p. 62, cited in Widdowson, 1994, p. 384)

For this reason I will be avoiding references to Konglish—the negative term that is used in a similar way to Spanglish and Chinglish—though the term is sometimes associated with useful academic studies. Doms (2003), for example, uses an article about Konglish for a text analysis and reminds us that it is a popular and sometimes controversial topic in Korean classrooms. *Korean learner English* such as that described in Lee (2001) is a useful construct for teachers and learners in the sense that it highlights common Interlanguage forms and problems such as a tendency to answer yes/no questions with the inappropriate answer; Lee gives the example of a proficient speaker asking ‘don’t you like mushrooms?’ followed by a Korean learner answering ‘yes’ but meaning to indicate that they do not. This sort of publication still carries the implication that all forms unique to Asian speakers are problematic without any information about which forms cause difficulties in communication and which forms are, in fact, quite clear. (Note that Lee 2001 also says nothing about Korean learners’ use of *you know*).

The decision to regard Korean English and, indeed, other Asian Englishes as language varieties with no negative implications is partially inspired by work in the field of *English as a Lingua Franca* (ELF). Barbara Seidlhofer is a key figure in this area and in Seidlhofer (2005) she summarises the language's role as a means of International communication with non-native speakers now playing a very significant part. Two striking results are mentioned from recent research into ELF: the first is that a number of standard English sounds are not needed in terms of intelligibility ('th' sounds /θ/ and /ð/ as well as the dark l [ɫ] found at the end of the word pool). The second is that a speaker omitting the third person marker -s from verbs appears to cause no communication problems. Such findings highlight the fact that English can be suited to new surroundings without necessarily hindering communication.

Despite a lack of corpus-driven studies in Korean English such studies have been useful for research into other Asian varieties of English (Cava, 2011 discusses the difference between *corpus-driven* and *corpus-based* studies). Robertson (2000), for example, discusses article usage in Chinese English and suggests a 'lexical transfer principle' whereby some learners have a tendency to use *this* and *one* as alternatives to articles. Nelson (2006) uses corpora of six World Englishes including Singapore, Hong Kong and the Philippines to develop a list of core and peripheral lexical items; his analysis shows that shorter words tend to reside in the core and confirms that scientific and technical language terms are spread across both the core and the periphery (i.e. words that are common to all six corpora and words that are unique to certain varieties respectively). Izumi, Uchimoto, Saiga, Supnithi and Isahara (2003) show a number of 'errors' in a Japanese learner corpus but it appears that the research team simply used their own judgment for detecting errors based on what appears to be a traditional prescriptive approach.

### *Lexical priming*

The work reported in this paper is most closely aligned with Hoey's (2005) *Lexical Priming* which will be introduced in the following section. As I look through a pile of student essays on my desk, I am struck by the sentence *he is a very kind guy*. It is not grammatically incorrect (in the traditional sense) and the register is appropriate for the informal description that the student was writing but, nevertheless it strikes me as unusual. It appears that there is only a single occurrence of *kind guy* in the British National Corpus: a 100 million word database that is available for language research; this compares with 83 examples of *nice guy* and 70 of *good guy*. *Very kind*, it seems, is not often used to modify people nouns with just three occurrences of *very kind man* and three of *very kind person*. The fact that we can intuitively spot such phrases in our students' work suggests that we have a mental store of collocational

behaviour that is fundamental to Hoey's new theory of language: *Lexical Priming* (Hoey, 2005). Hoey argues that:

As a word is acquired through encounters with it in speech and writing, it becomes cumulatively loaded with the contexts and cotexts in which it is encountered, and our knowledge of it includes the fact that it co-occurs with certain other words in certain kinds of context. (Hoey, 2005, p. 8)

Sorbi (2010) echoes this view by stating 'it now seems likely that language acquisition begins with word learning rather than syntax triggering, with words gradually "grammaticalized" through experience on a largely associative basis' (2010, p. 1). Hoey's work brings the phenomenon of collocation from the backwaters of an EFL syllabus to the very centre of human language acquisition and it was with this theoretical background that I began the data collection. This paper shares two key points with Hoey (2005)—lemmatisation is avoided because one cannot assume that a community's primings for a string such as *you know*, for example, will be similar to those for *you knew* and concordance data such as that shown in concordances 1 and 2 (below) is indicative of the speakers' primings though it is acknowledged that no text can provide direct evidence of such an abstract, psychological construct. Note that priming effects can be crudely divided into two divisions—ones we are consciously aware of on the basis of our formal education (for example) and ones we are not (in many cases a native English speaking child will produce all their utterances on the basis of subconscious primings). A string such as *you know* will be affected by both conscious primings and subconscious primings when used by an adult who learned English as a second language but it is hoped that the volunteers in this study will have paid little attention to such a string while focussing on the more salient lexical content of their utterances.

If our recent conversations, education, the films we watch and the music we listen to prime us and change our knowledge of English every moment then it is reasonable to suggest two communities of Korean L1 speakers separated by living in two countries will show slight variation in the forms they select though it is not known which forms of language are vulnerable to change in a short space of time. Two corpora of British spoken English were used for comparison with the Korean data.

### 3. Method

The two spoken Korean English corpora were collected in Liverpool and Seoul in 2008 and are named SK (for Seoul Koreans) and LK (for Liverpool Koreans). The recording context consisted of the Korean informant and myself in a small room and resembled an interview setting as I asked

questions about their reasons for studying English, hobbies and career ambition (these were some common starter topics but note that they were not structured interviews); I aimed to keep the conversations as informal as possible and was keen to find a subject that would ‘get them talking’ freely without concern for ‘grammatical correctness’. My own utterances were removed from the main Korean corpora and not used in subsequent frequency counts but all audio files and complete transcripts were kept for reference and checked in some cases for intonation or reference to the full discourse exchange. The total number of words (or tokens) in each of the four corpora used in this study is shown in table 1.

Table 1  
*Size of Four Corpora*

Corpus Name	Abbreviated Name	Size
Liverpool Korean corpus	(LK)	83 446
Seoul Korean corpus	(SK)	112 621
Scouse corpus	(SCO)	106 562
Demographic section of spoken BNC	(BNC)	3 945 881

Basic demographic details for the Korean corpora are shown in table 2.

Table 2  
*Korean Informant Data*

		SK	LK
Number of respondents		39	28
Average age		25	27
Average years learning English		9.7	12.2
Gender	F	29 (78)%	16 (57)%
	M	8 (22%) <sup>1</sup>	12 (43)%

LK volunteers had spent an average of two years living in the UK. Note that SK has more informants than LK and is more notably biased towards female volunteers—this is a limitation in terms of direct comparability but it was not deemed a major concern for this particular study. A corpus of native Liverpool spoken English or ‘scouse’ (SCO) was prepared by a colleague between 2001 and 2004 (Pace-Sigge, 2010), so I used this as a comparator corpus because of its similar size and to allow me to account for any possible influence of the local primings on the Liverpool (LK) volunteers. SCO consists of 51 speakers, a larger proportion of males at 54% and slightly older informants with an average age of 33. Finally the much larger demographic spoken section of the British National Corpus (BNC) was used. This is a well known reference corpus collected by 124 volunteers in 38 different UK

<sup>1</sup> Two respondents in Seoul chose not to complete demographic data sheet

locations (What is the BNC? 2012) but with notably older sound recordings collected in the early nineties one has to be cautious about any language structures that may be changing in this timescale; the difference between the 'interview' type data collection in SK and LK and the freer recording used in SCO and the BNC (which included recording of groups rather than strictly one-to-one conversations) must also be noted as a limitation of this study.

In line with other corpus-driven studies such as Hoey (2005) and Barbieri (2008), frequency data is considered in the first instance with the researcher turning to other aspects of language such as phrase structure or pragmatics only when an issue with the data demands it. All analysis was carried out using corpus analysis software *WordSmith tools version five* (Scott, 2011). For all chi-squared statistics a standard value of  $p = 0.05$  was taken as the cut-off point for statistical significance though results with  $p$  value above that are not necessarily without value (all values for chi-square are based on two-tailed calculations, and Yates' correction for 2x2 contingency tables has been applied).

#### 4. Results and discussion

In Hadikin (2011), individual word variation was the main focus and the apparent priming effects that surround each word so for this paper the logical step from single words to larger texts was taken—the most frequent two word strings in each of the corpora.

Table 3

##### *High Frequency Two Word Strings in the Four Corpora*

	SK Items	SK	LK	SCO	BNC		SCO Items	SK	LK	SCO	BNC
1	BU yeah	1649	1518	780	44771*		BU yeah	1649	1518	780	44771*
2	yeah EOU	651	758	585	31090*		you know	602	205	619	15523
3	BU er	615	541	n/a	n/a		yeah EOU	651	758	585	31090*
4	you know	602	205	619	15523		in the	277	174	272	11061
5	I think	575	402	161	8908		it was	366	296	233	8009
6	BU and	542	356	162	16298*		I don't	495	372	218	11763
7	so I	529	370	58	2902		I mean	123	114	206	9912
8	I was	521	274	162	5202		I was	521	274	162	5202
9	I don't	495	372	218	11763		I think	575	402	161	8908
10	and I	447	221	149	6900		and then	208	76	160	5337

  

	LK Items	SK	LK	SCO	BNC		BNC Items	SK	LK	SCO	BNC
1	BU yeah	1649	1518	780	44771*		you know	602	205	619	15523
2	yeah EOU	651	758	585	31090*		I don't	495	372	218	11763
3	BU er	615	541	n/a	n/a		in the	277	174	272	11061
4	I think	575	402	161	8908		I mean	123	114	206	9912
5	I don't	495	372	218	11763		I think	575	402	161	8908
6	so I	529	370	58	2902		do you	121	84	122	8169
7	BU and	542	356	162	16298*		it was	366	296	233	8009
8	it was	366	296	233	8009		on the	60	41	161	7266
9	I was	521	274	162	5202		and I	447	221	149	6900
10	when I	346	271	71	2378		I know	60	52	113	6493

I will briefly explain the layout of table three because it is rather unusual. The table consists of four quarters—the top left quarter contains the highest frequency two word strings in the Seoul corpus (SK) and then shows frequency figures for SK, LK, SCO and the BNC in four columns. The top right quarter shows corresponding data for SCO; some of the information is repeated such as the frequency figures for *you know* (highlighted) but I have kept the data in this format so that one can easily see that this string is a very high frequency string in SK, SCO and the BNC but does not rank in the top 10 for LK and thus is missing from the lower left quarter. This format is particularly advantageous as it does not privilege any corpus over the others—we can see the high frequency strings in each one for comparison.

*BU yeah* is not technically a two word string but represents utterances where *yeah* and *er* were used at the beginning of an utterance (BU stands for *beginning utterance*) and, similarly, *yeah EOU* shows that the word occurred at the end of the utterance. As there are almost 59 000 occurrences of *yeah* in the BNC and almost 91 000 occurrences of *and* I estimated the frequency of values marked with an asterisk based on manually checking sets of 300 occurrences.

The level of variation of *you know* struck me as particularly interesting; of course, the corpora vary in size but when normalised to words per million the frequencies for *you know* are 5345 per million (pm) in SK, 2457pm in LK, 5809pm in SCO and 3934pm in the BNC. Chi-squared with Yates' correction (YC) is 58.33,  $df=1$ ,  $p < 0.0001$  when the raw frequencies of 602 occurrences of *you know* in SK and 205 occurrences of *you know* in LK are compared with 423 *OTHER know* strings (SK) and 331 *OTHER know* strings in LK so this is clearly statistically significant (recall that the cut-off point is  $p = 0.05$ ). In the following sections I will explore this string, the role it plays within larger strings and attempt to shed light on why there is such variation and what it may tell us about the speakers involved.

#### 4.1. 'you know' strings

Table four shows frequency figures for *you know* and the three most frequent three word chunks that appear to be influencing usage in the Korean corpora (based only on items to the left of *you know*): '*do you know*', '*but you know*' and '*and you know*'. The left side of *you know* was chosen as a focus area because of the consistency shown between LK and SK whereas SCO has *do you know*, *like you know* and *that you know* as its top three strings and the BNC has *do you know*, *yeah you know* and *it you know*. Items to the right of *you know* would be equally important but were not selected for the scope of this study. One can see from table four that SK's profile is not especially similar to the British profiles – the high frequency of *you know* is superficially similar to



that of SCO but 16% of all *you know* strings take the form of either *do you know* or *but you know*. This is notably closer to the 20% seen in LK than the 9% in the BNC and the 3.5% in SCO. This suggests that Korean speakers are primed to use these forms but at the expense of the wider range of alternatives that British speakers use.

Table 4

*Frequency, Normalised Frequency and Percentage Figure for 'You Know' Strings in Four Corpora*

	SK			BNC		
	f	PM	%	f	PM	%
you know	602	5345	100	15523	3934	100
do you know	48	426	8	1047	265	7
but you know	46	408	8	271	69	2
and you know	24	213	4	401	102	3

  

	LK			SCO		
	f	PM	%	f	PM	%
you know	205	2457	100	619	5809	100
do you know	30	360	15	16	150	3
but you know	11	132	5	3	28	0.5
and you know	10	120	5	6	56	1

The following subsections will focus on each of these three strings to explore the variation that occurs.

#### 4.2. '*do you know*'

Table five shows frequency and percentage information for the string *do you know* and its three constituent parts (based on Biber 2009). For example, the first column with the heading *do* represents *\*you know* as a search term and tells us that 48 out of 602 occurrences of *you know* take the form *do you know* in SK. The lower part of the chart shows that this corresponds to 8%. The central column represents the search term *do \*know* and tells us that 48 out of 52 occurrences of *do \*know* take the form *do you know* in SK and that this is 92%. The \* symbol is called a wildcard and brings up all words in this position so, for example, the other 8% of *do \*know* in SK is made up of one case of *do I know* and three occurrences of *do not know*.

This system of breaking down the string allows internal variation to be explored as well as how the string as a whole behaves in the corpora. It is clear from figure three that LK stands out as having a very high reliance on the form *do you know* whereas the other three corpora show more flexibility in the first column. Chi-squared with YC is 7.03,  $df=1$ ,  $p = 0.008$  when the *do you know* figures for SK and LK are compared with *OTHER you know* figures

so this indicates a statistically significant difference between the two corpora. SK, for example, has ten occurrences of *actually you know* compared with just one in LK; there are no cases of this string in SCO and just eight in the BNC (which is, you may recall, thirty-five times larger than SK) so it is possible that the LK speakers are primed to avoid such rare structures in Britain and prefer prototypical \* *you know* strings that they know are used by proficient speakers around them. The difference between SK and LK in this regard is not statistically significant with chi-squared with  $YC = 0.815$ ,  $df=1$ ,  $p=0.37$  but note that the value for LK clearly tends towards the values seen in the British corpora and the difference between SK and the BNC is clearly statistically significant with chi-squared = 120.6,  $df=1$ ,  $p < 0.0001$  when the data for *actually you know* is compared with *OTHER you know* strings.

Table 5  
*Frequency and Percentage Details for the String 'Do You Know'*

	do	you	know
SK	48/602	48/52	48/121
LK	30/205	30/31	30/84
SCO	16/619	16/18	16/122
BNC	1047/15522	1047/1144	1047/8155
SK	8	92	40
LK	15	97	36
SCO	3	89	13
BNC	7	92	13

In the central column the LK data stand out as marginally higher than the other corpora as 97% of *do \* know* strings take the form *do you know*. A single case of *do not know* makes up the remaining 3%. SK contains a single occurrence of *do I know* and three cases of *do not know*. This variety should not be seen as necessarily British-like – in fact, this is a rather high frequency of *do not know* for such a small corpus (27pm) and the LK normalised value of 12pm is closer to the very low BNC value of 3pm. There are no occurrences in SCO.

Finally, the frequency of *do you know* as a proportion of all *do you \** strings marks SK as being the extreme case though there is a clear distinction between the Korean corpora using the form in 36% and 40% of occurrences (in LK and SK respectively) and both British corpora showing a value of 13%. This difference between *do you know* string frequency and other *do you \** strings in SK and the BNC is statistically significant with chi-square = 72.4,  $df=1$ ,  $p < 0.0001$ . The difference appears to be partially associated with *do you know* being used more frequently to raise topics/confirm shared knowledge before moving on to say something in the Korean corpora (Concordance one, for example, shows a sample of SK texts and a sample of BNC texts for

comparison) and this could be usefully confirmed by further studies and/or compared with corpora of other Asian Englishes.

One must be careful, however, not to assume that this is the only, or indeed, the most important difference. The string *do you want* is actually the most frequent *do you* \* string in both SCO and the BNC and makes up 16% and 24% of all *do you* \* strings respectively. It is often used for invitations in the British corpora such as *do you want to listen to some of this* and *do you want to bring your friends* as well as *do you want me to be specific* and *do you want me carry on talking*; the difference in data collection method where the British speakers are in more naturalistic settings would likely be a very significant factor in this area – one might reasonably expect the frequency of such offers to drop in interview-like settings and move the percentages closer to those seen in the Korean data. A second string that appears to be affecting the data is *do you do* in strings such as *what do you do* and *how do you do*, particularly in SCO. 48% of the occurrences were produced by the researcher himself and may be a form of idiosyncratic preference but, nevertheless, with his usages removed there are still nine occurrences which would place it as the fourth most frequent string and affect the overall proportions reported in table five.

ific hobby for me J: sex and the city do you know that J: I really love that drama and I learn f  
 us for travelling korea and especially do you know Jungfrau my pronunciation is not good so  
 ia before just I study in the chungno do you know y b m V: y b yeah yeah hogwan educatior  
 eading books I like do you know aga do you know Agatha Christie V: I love her novels V: real  
 beer and they said no do you have a do you know there is a a very nice invented human inver  
 ching American drama animation A: do you know family guy A: and friends A: this is season  
 og in the well V: frog frog in the well do you know that idiom V: like frog in the well V: never l  
 oy it yeah V: medical research V: [] do you know the fmr V: yes yes my major's about the fr  
 o you know ah sorry I forgot V: yeah do you know the woman who worked with chimpanzee :  
 glish movie atonement K: atonement do you know atonement K: yeah atonement love actualy  
 ry much is well I lived in woojong bo do you know woojong bo V: it's not far from here about i  
 know K: I don't know everything like do you know david david K: yeah K: so I have been in F  
 id I'm currently working company for do you know Epson J: yes J: yeah printer and scanner  
 ayer plays for Manchester united H: do you know who H: and you know er I like Christian ro  
 how much size the painting and like do you know crate a box J: so whenever we wonder the  
 ed I graduated from my university J: do you know ah maybe you don't know my university J:  
 ll leaving all that outside! Ooh ooh! Do you know Margaret I go back and it looks so dirty  
 pleased to meet you . Oh, no ooh! Do you know I can't understand that! You can't see it  
 ight. David'll learn when he's older. Do you know any more dad? I thought you were gonn  
 it then! Ask her! Ask her! Hang on! Do you know how much these are a bag? No. Really  
 g but kind of now, he's kind of old, do you know what I mean. Exactly. There's no decen  
 point seven percent. He's an old . Do you know anybody that sort of you know happy O  
 potato round you Can I have one? Do you know how fast? of those Do you want a cake  
 whether you can answer this one, do you know if Angela is still at Metropolitan Housing  
 her one. No then I'll have that one. Do you know what she said Fred What? What did I s  
 n. Oh and again. And another one. Do you know what you're going up against? Whose b  
 tgs in a funny I just found it. place! Do you know I've Yeah but co emptied this cupboard  
 ary's Dye dyes, they're on plastic, do you know what she does, she gathers all the plan  
 at Kelly does it to loads of people, do you know, cos she was Geraldine's best friend, K:  
 that they would like to be picked . Do you know what I mean? yeah yeah, could be. Err  
 No you don't. for twenty five pound. Do you know how much it is? Yes I know how much  
 mean, it was, it wasn't predictable, do you know what I mean? And it was, I mean, right.

*Concordance 1: Sample concordance of do you know in SK (top) and the BNC (bottom)*

The complex interplay between these strings suggests that this is one area that would be particularly interesting to look at in a repeated study with more directly comparable corpora and other, nearby varieties of English. The apparent narrower range of functions of *do you know* seen in the Korean corpora could have implications for teaching materials if students are not being exposed to (or not perceiving) other functions of the string.

#### 4.3. '*but you know*'

Table six shows similar information about the string *but you know*. Many of the overall patterns are similar to the *do you know* data with the Korean corpora showing less flexibility than the British corpora. In the first column it is SK that has the strongest tendency to use *but you know* out of all possible \**you know* strings. The difference between SK and LK is not statistically significant when *but you know* is compared with *OTHER you know* (chi-squared with  $YC = 0.88$ ,  $df=1$ ,  $p=0.35$ ) but the difference between SK and the BNC values is significant with chi-squared with  $YC = 102.5$ ,  $df=1$ ,  $p < 0.0001$  suggesting a potential difference between Korean English and UK English.

Table 6  
*Frequency and Percentage Details for the String 'But You Know'*

	but	you	know
SK	46/602	46/51	46/53
LK	11/205	11/12	11/18
SCO	3/619	3/4	3/23
BNC	271/15522	271/396	271/1454
SK	8	90	87
LK	5	92	61
SCO	0.5	75	13
BNC	2	68	19

The most striking factor influencing this relatively high percentage appears to be the frequency of the string itself (rather than a lack of variety in related forms). When normalised it comes to 408 occurrences per million compared with 132pm in LK and just 69pm and 28pm in the BNC and SCO. A notable four word string (or 4-gram) in the SK data is *yeah but you know* which occurs 10 times – there is only one occurrence in LK, none in SCO and 10 in the BNC (a sample concordance from SK is shown in concordance two); the SK speakers appear to be primed to use this particular structure in many cases to make a point that is in some way in contrast with the previous one whereas the speakers in the other corpora make use of a wider variety of forms.

jering and sometimes I burn my skin but **yes but you know** the cellular phone in korea is  
 Zealand J: er it depends on academy J: **yeah but you know** at that time the teacher was  
 ears V: ah around twenty three yeah V: **yeah but you know** actually I think about that yo  
 : and then he was training J: yeah yeah **yeah but you know** after I came back to korea I l  
 just a very short time just two years V: **yeah but you know** that I don't know about that I  
 ouse tower bridge V: koalas and I think **yeah but you know** the England or swiss uh Swi  
 vel with me for few days inside korea K: **yeah but you know** I can't afford you know K: ye  
 mething like that so V: I think that is V: **yeah but you know** the baseball is just three or f  
 : so you're English K: [laughs] sorry K: **yeah but you know** even I'm not English like briti  
 irean and china V: uh just apartment V: **yeah but you know** that V: when when when I m  
 er for five hours V: and then she said V: **yeah but you know** and er actually when I starte  
 ent up to the mountain it's just V: oh yeah V: **but you know** what getting on a plane by h  
 d er live in wadding near the rest stop area V: **but you know** there was some kind miscor  
 n right V: yeah but actually weve broke up V: **but you know** it's kinda complicated V: like  
 j the top like kinda Australian thing V: yes V: **but you know** even though I was in Australi

*Concordance 2: Sample concordance of but you know from SK (selected to highlight use of yeah but you know)*

The middle column of table six, by contrast, shows that all four corpora are rather restricted when it comes to the central component of *but \* know* – the frame tends to be completed by the item *you* in at least 68% of occurrences (68% is the value for the BNC). There is room, however, for strings such as *but I know*, *but we know*, *but they know* and *but don't know* in the BNC. SCO shows a comparable level of variety on a smaller scale with one occurrence each of *but you know*, *but they know*, *but didn't know* and *but I know* in its total of four *but \* know* strings. SK and LK have unusually similar percentage values of 90 and 92 respectively (the difference between SK and the BNC is statistically significant with chi-square = 9.35,  $df = 1$ ,  $p = 0.002$  suggesting an overall difference between Korean usage and British usage); SK's percentage is, again, influenced by a large number of *but you know* occurrences and, of course, the *yeah but you know* strings shown in figure six. The other strings used in SK are four occurrences of *but I know* and a single line of *but didn't know* while LK only makes use of a single occurrence of *but I know* alongside its 11 occurrences of *but you know*. The only four word string (4-gram) used more than once in this part of the LK data is two cases of *but you know I*. One may suggest that the speakers are breaking down such large chunks as part of their learning experience in the UK but it would take further studies to confirm this.

The right column of table six shows an even more notable division between the Korean and British corpora than the equivalent column figure three; the Korean speakers appear to be strongly primed to form *but you know* at the expense of rival *but you \* strings* and there are no notable invites or offers involved (cf. the lack of *do you want* in section 4.2).

a mia V: oh it's very nice I want to live here **but you** know in australia there aren't many black pe  
er never V: yeah I know it's very interesting **but you** know what when I look at her I could tell you  
h huh L: oh L: yeah yeah L: yeah L: I know **but you** know Korean people are pretty shy L: yeah y  
h asian people you cannot play the cricket **but you** know sometimes we have some problem we  
ne expression in my face on my face V: no **but you** know that time I lost some my nickname my  
y guys wear like shorts like that not shorts **but you** know tennis shorts V: er [] how do you call t  
like different it's up to like er every situation **but you** know seoul national university they like most  
: very difficult change to V: I tried to do that **but you** know sometimes if we study soccer V: Aust  
computer electronic or something like that **but you** know actually er when I joined the university  
erature V: er I really don't like the literature **but you** know my score was not very good enough go  
he top like kinda Australian thing V: yes V: **but you** know even though I was in Australia I there is  
h yeah yeah uh uh uh V: I don't understand **but you** know that weird peo person make that show  
at getting on a plane by herself it's hard V: **but you** live with your mother V: I do I do too V: uh hu  
V: uh V: er I don't think so V: yeah V: yeah **but you** wanna V: er hi I'm Helga what's your name o  
it up to the mountain it's just V: oh yeah V: **but you** know what getting on a plane by herself it's h  
ight V: yeah but actually we've broke up V: **but you** know it's kinda complicated V: like we can m  
she really didn't she didn't want to go there **but you** know like she applied for it V: yeah compute  
rs V: ah around twenty three yeah V: yeah **but you** know actually I think about that you know I th  
er live in wadding near the rest stop area V: **but you** know there was some kind miscommunicati  
**ur personal privacy. Fucking right it is. But. You know it's not, no it's erm for the Longi**  
ool. . No, no, muggins, muggins does, **but. You don't** need, he, that sort of thing that t  
Idn't have to be told! It should be there. **But you** don't have to be told to put your socks  
actually simulate but then again there, **but you**, you fire off the blank, that's what trigge  
vell you were hungry a way down there **but you** would make good child. . Oh oh oh! Oh  
arket and then And then another there **but you** don't go in and buy five cases of and yo  
Where? You got the staircase in there **but you** will have got ornamental wrought iron le  
r you can go for like a nice meal there, **but you** can like also eat, you know dips and Y  
! It's a bit sort of stilted. Ah! Well there **But you are!** er apart from that it all went very w  
bag next week! Oh dear! actually there **but you** know what I mean. Yeah. Right. I'll go  
ouldn't think of a salad that isn't there. **But you** can't, and there's, then there's loads of  
ey like? Same as that one down there. **But you** go from the stairs and in the er You've  
ng else Well I was gonna change them **but you** might as well . cos there the most Ooh  
und for you. You don't like to ask them **But you** don't like to because of the fact you kn  
rd they'd get something for themselves **but you** heard they'd get the town something ar  
ah, they were we really enjoyed them, **but you** see if we'd of got four they wouldn't of b

Concordance 3: Sample concordance of *but you* in SK (top) and the BNC (bottom)

Concordance three shows the striking contrast between *but you* strings in samples of SK and the BNC; it is clearly influenced by a difference in reliance on the string *but you know* and suggests that the SK speakers have quite a restricted use of the string that may be influenced by *you* having no or very limited use in its sense of *people in general*. The LK data makes much less frequent use of *but you know* – 11 occurrences compared to 46 in SK (132pm in LK compared to 408pm in SK) but is also interesting because two other *but you* strings appear more than once whereas SK only has *but you know* occurring multiple times. These two strings are *but you don't* and *but you can* in contexts such as *but you don't have to prepare for it* and *but you can work anyway like take some payment* and suggest that the Korean speakers living in Liverpool are slowly being primed to use *but you* with this *people in general* sense and that appropriate collocations are developing to reflect this use (collocation in lexical priming studies could include strings such as *but you*

*can* if they occur sufficiently frequently as well as more ‘psychologically salient’ collocations like *fish and chips*); in SCO the strings *but you don’t*, *but you do*, *but you can* and *but you can’t* are used more than once but, actually, *but you know* is not present at all in 23 lines of *but you* data.

#### 4.4. ‘and you know’

The final set of data that this study is based on is the percentage table for *and you know* shown in table seven. It is the third most frequent *you know* string in both Korean corpora but is actually the second most frequent in both British corpora. The first column shows the most consistency out of all the chart sections discussed in this paper and shows that *\* you know* is a very open ‘frame’ with between 1% and 5% making use of the form *and you know*. One interesting point is that SK has the highest normalised frequency of *and you know* with 213pm compared with 120pm in LK, 101pm in the BNC and 56pm in SCO (the raw numbers of 24 in SK and 397 in the BNC give chi-squared = 4.11,  $df=1$ ,  $p=0.04$  which is statistically significant; this suggests that the SK speakers may be relying on this as a ‘lexical teddy bear’ (term from Hasselgren 1994) – a fixed form that they know they can rely on – to allow more time for language processing during speech). Related to this suggestion is the presence of a four word string *yeah and you know* which occurs three times in SK, not at all in LK or SCO and just seven times in the much larger BNC.

Table 7

*Frequency and Percentage Details for the String ‘And You Know’*

	and	you	know
SK	24/602	24/27	24/40
LK	10/205	10/11	10/17
SCO	6/619	6/8	6/55
BNC	397/15522	397/541	397/1084
SK	4	89	60
LK	5	91	59
SCO	1	75	11
BNC	3	73	37

The middle column shows a similar pattern to the middle column of figure five (the corresponding data for *but you know*) with the Korean data presenting *and you know* in approximately 90% of *and \* know* occurrences and the British data showing more flexibility with percentage figures in the low to mid 70s. Unlike the *but you know* data, however, the remaining 11% of SK *and \* know* occurrences is not largely made up of multiple occurrences of a single string but one each of *and I know*, *and they know* and *and we know*. The speakers then are not dependent on a collocation but rather a colligation

(discussed in some detail in Hoey, 2005:42): the relationship between the frame *and \* know* and, in this case, a set of pronouns. This is similar to SCO which only makes use of *and I know* out of all possible alternative strings (in fact, with a higher frequency than *and you know*) and the BNC which also only has *and you know* and *and I know* occurring at normalised frequencies above the 10 per million mark that one might expect to see in all four corpora. In this respect LK is unusual with a single occurrence of *and don't know* in the string *don't like UK accent and don't know well UK accent*; this appears to be a single speaker with a tendency to elide the subject of clauses. Note that in a sense the Korean corpora have more variety in this area than the British corpora but the large number of *and you know* strings affects the percentage values.

The right column of table seven showing the percentage of *and you know* strings out of all possible *and you \** strings shows the most variety across the four corpora with SCO at 11%, the BNC at 37% and LK and SK at 59% and 60% respectively (this difference between SK and the BNC, for example, is statistically significant with chi-squared with  $YC = 8.03$ ,  $df=1$ ,  $p = 0.004$ ). As mentioned earlier in this section SK makes the most use of the string *and you know* in the normalised data with a value of 213pm compared with 120pm in LK, 56 pm in SCO and 101pm in the BNC so much of the difference comes in the variety of *and you \** forms. Alongside several forms that occur only once SK has two forms (other than *and you know*) that occur more than once and these are two each of *and you have* and *and you can*. With contexts such as *you have to take a test and you have to get a score* (to get into business school) it shows clearer use of *you* to mean 'people in general' than was seen in section 4.3 – it would be useful to test the hypothesis that this aspect of meaning is influenced by the collocational environment in further work. Similarly LK has two each of *and you have* and *and you can* in contexts such as *and you have er extra subjects* and *and you can find many Korean people and you can interview them*. The difference in variety becomes clear when one compares the numbers of strings that occur more than once in the similar sized SCO:

- 4 x *and you go*
- 3 x *and you turn*
- 3 x *and you have*
- 2 x *and you say*
- 2 x *and you can't*
- 2 x *and you can*
- 2 x *and you be*

The high frequency of *and you go* is most striking and appears to be used to introduce reported speech or actions as in the following exchange from SCO:



Whack them on the head  
If they don't say nothing  
Whack  
They'll be like - ahh  
B And you go - manners  
(laughter)

The Korean speakers may not have been exposed to this informal form in taught materials so could listen out for it after their consciousness has been raised by studies such as this. The BNC has a comparable range of forms that occur more than 18 times per million (the value that two occurrences would come to in SK) including:

and you can  
and you don't  
and you get  
and you go  
and you have  
and you just

This suggests that the use of *and you go* is not restricted to Liverpool English and could be useful for learners (and many experienced L2 English users) in other parts of Britain. With a normalised frequency of 29pm it is striking that the BNC's third most frequent *and you* \* form *and you don't* has not appeared in this discussion. It appears only once in SCO, once in SK and not at all in LK; the alternative form *and you do not* does not occur in the Korean corpora so does not appear to play a role. It appears that the function of *you* meaning 'people in general' has not fully developed to include this string.

## 5. Conclusion

In this paper I have begun to explore select features of English as it is spoken by 21<sup>st</sup> century Korean adults and, with Hoey's theory of Lexical Priming as a theoretical backdrop (Hoey, 2005), I compare similarities and differences between a corpus of Korean English collected in the UK and a comparable corpus collected in Korea –to my knowledge, this is the first published study that compares two 'non-native' corpora with a shared L1 and shared cultural background so as to allow potential recent primings to show through. In the spirit of Lexical Priming I have tried to explore areas where language change may be happening and being driven at the level of collocations and note that even my choice of words has reflected certain working assumptions such as a community's preference for a particular string affects its meaning rather than vice-versa (the directionality is almost certainly not quite this simple).

The results focused on three high frequency strings, *do you know*, *but you*

*know* and *and you know* and I introduced a procedure inspired by Biber (2009) that breaks the strings into their constituent parts thus allowing one to discuss the way priming effects influence, for example, the central component as well as just the positions to the left and right of a chosen node. The *do you know* study showed that this string is getting used in Korean English to introduce topics and/or confirm shared knowledge with occurrences such as *do you know Family Guy* and *do you know Agatha Christie* but certain fixed forms such as *do you know what I mean* and rhetorical devices such as *do you what she does* are lacking compared with British data; differences with the way the Korean data and the British data were collected prevent me from making strong claims but this would be a useful area for further study.

*But you know* is used with very high normalised frequency in the Korean corpora and I argue that it is used as an extended connective that 'buys more time' for online speech processing. The presence of a very high proportion of *yeah but you know* only in the Seoul Korean corpus (SK) supports this and suggests that Korean speakers living in the UK are breaking up the string and exploring alternative conversational forms. In the final section the *and you know* data shows an interesting range of variety in the way its constituent parts differ between the corpora and the work raises questions about the meanings of *you* and how they may be influenced by slight differences in the forms used by a community. To my knowledge this is the first published study that applies Hoey's Theory of Lexical Priming to spoken English as well as the first to look at a variety of English other than British and the approach of looking at short strings and considering the effects of the combination of conscious and subconscious primings provides a unique opportunity to explain why certain strings may be getting used while others are being avoided (though with such small corpora all results must be treated with some caution). For students wanting to sound more like native speakers many of the details included in this paper would provide useful areas of focus and a similar technique could be used to find further areas of variation between Korean and UK English (of course, readers of IJLS are especially encouraged to compare findings with small corpora of spoken American English or other World Englishes). If further studies support my argument that certain senses of words like *you* are not being learned from standard pedagogic materials this study would have clear implications for that field.

Throughout this work I have been cautious not to assume that Korean adults *want* to sound exactly like British speakers – that would be a choice for the individual – but by exploring how the use of a small set of lexical strings varies across four different communities we can generate questions and begin working towards answers about how individuals learn and how languages change on a larger scale.

### ***Acknowledgements:***

I would like to thank Professor Michael Hoey and all the staff and students at the University of Liverpool for supporting this work when it was part of my PhD and the University of Portsmouth's School of Languages and Area Studies and the Centre for European and International Studies Research (CEISR) for further support and funding.

### ***The Author***

Dr Glenn Hadikin (Email: glenn.hadikin[at]port.ac.uk) has worked as a lecturer in English and Applied linguistics for 12 years in both South Korea and the UK. He completed his MA TESOL at London's Institute of Education, obtained a PhD in English from the University of Liverpool and is currently based at the University of Portsmouth on the English south coast. He is interested in a range of topics from World Englishes and pragmatics to corpus studies and EAP.

### **References**

- Achebe, C. (1975). The African writer and the English Language. In C. Achebe (Ed.), *Morning Yet on Creation Day* (pp. 55-62) London: Heinemann.
- Barbieri, F. (2008). Patterns of age-based linguistic variation in American English. *Journal of sociolinguistics*, 12(1), 58-88.
- Biber, D. (2009). *A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing*. Presentation given at Corpus Linguistics 2009 on 23<sup>rd</sup> July 2009.
- Cava, A.M. (2011). Abstracting science: A corpus-based approach to research article abstracts. *International Journal of Language Studies*, 5(3), 75-98.
- Chang, W., Park, Y., Kim, C., & Chang, D. (1989). *High school English I: Teacher's guide*. Seoul: Dong-a Publishing.
- Doms, D. (2003). *A text analysis of a newspaper article about Konglish taken from 'The Korea Herald'*. Unpublished university report retrieved from <http://www.bhamlive3.bham.ac.uk/Documents/college-artslaw/cels/essays/writtendiscourse/Doms5.pdf>.
- Hadikin, G. S. (2006). *World Englishes and Konglish: What is Konglish and what are local attitudes to it?* Available at [www.oocities.org/glennh69/EDWCmar11d.doc](http://www.oocities.org/glennh69/EDWCmar11d.doc).

- Hadikin, G. S. (2011). *Corpus, Concordance, Koreans: A corpus-driven study of an emerging New English*. Manuscript submitted for publication.
- Hasselgren, A. (1994). Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics*, 4 (2), 237-258.
- Hoey, M. (2005). *Lexical priming: A new theory of words and language*. London: Routledge.
- Izumi, E., Uchimoto, K., Saiga, T., Supnithi, T., & Isahara, H. (2003). Automatic error detection in the Japanese learners' English spoken data. In *Proceedings of the 41<sup>st</sup> annual meeting of the association for computational linguistics* retrieved from <http://aclweb.org/anthology-new/P/P03/>.
- Lee, J. (2001). 'Korean speakers' in Swan, M. and Smith, B. (eds.) *Learner English: a teacher's guide to interference and other problems*. Cambridge: Cambridge University Press.
- Park, J. S. (2009). *The local construction of a global language: Ideologies of English in South Korea*. Berlin: Mouton de Gruyter.
- Pace-Sigge, M. T. L. (2010). *Evidence of lexical priming in spoken Liverpool English*. Unpublished PhD thesis, University of Liverpool.
- Porter, C. (2011). Review of *The local construction of a global language: Ideologies of English in South Korea*. *TESL-EJ*, 14(4), 1-3.
- Scott, M. (2011). *WordSmith tools 5.0*. Available at <http://www.lexically.net/wordsmith>.
- Seidlhofer, B. (2005). *English as a lingua franca*. Retrieved from [http://people.ufpr.br/~clarissa/pdfs/ELF\\_Seidlhofer2005.pdf](http://people.ufpr.br/~clarissa/pdfs/ELF_Seidlhofer2005.pdf).
- Shim, R. (1999). Codified Korean English. *World Englishes*, 18(2), 247-259.
- Sorbi, A. (2010). The relationship among word knowledge, word prediction and reading comprehension. *International Journal of Language Studies* 4 (1), pp. 1-28.
- What is the BNC?* (2012). Retrieved from <http://www.natcorp.ox.ac.uk/corpus/index.xml?ID=intro>.
- Widdowson, H. (1994). *Defining issues in English language teaching*. Oxford: Oxford University Press.