

Lexical Patterns in the Conversational English of L2 Speakers: Is There Evidence of Individual Variation?

Glenn Stewart Hadikin¹

1 Introduction

There is little doubt that the study of large bodies of natural text is providing the research community with new and exciting ways of exploring language. Halliday (1994) suggests that linguistics today is ‘more or less where physics was in the fifteenth century’, and Sinclair (1991) states that until recently ‘the situation was similar to that of the physical sciences 250 years ago.’ Amongst this wave of excitement, though, one may wonder how it all affects the humble language learner sitting in a classroom wondering how to express the thoughts in her head in English.

My personal experience teaching in the Republic of Korea first sparked my interest in phraseology and lexical patterns. Students used certain phrases and ways of expressing themselves that appeared consistent from class to class yet were significantly different from both the target variety, standard American English and the L1. Indeed, Thorkelson (2005) showed that many Korean students will edit an English article differently when informed that the target audience is Korean. Hadikin (2006) showed a high frequency of non-standard collocations such as *lose my weight, *play bowling and *make a girlfriend that were written in a gap-fill exercise and I have heard students ‘correcting’ themselves when accidentally using the standard equivalent in class.

In 1989 Carlos A. Yorio described comparable forms used by a Korean student known as ‘K’ at City University of New York. Figure one shows a few examples:

1. Sunday was lousiest day
2. just two of us
3. for short time
4. I have lots money
5. at the morning

Figure 1: Non-standard forms used by ‘K’. (Yorio 1989: 61)

¹ School of English, University of Liverpool

e-mail: glenn.hadikin@liv.ac.uk

Arguably, the forms discussed by Yorio could have been one off production errors that were never repeated. Hoey's Theory of Lexical Priming (2005), however, suggests that every time a piece of language is used all participants are 'primed' to hear and use the language in a similar way. The more times one hears 'at the morning', Hoey suggests, the stronger one becomes primed to use such an expression in the future. Could Lexical Priming explain the consistent 'errors' being used in everyday Korean English?

This paper describes a methodology designed to begin to answer such a question. It is a work in progress that begins with the search for a suitable method of capturing a 'snapshot' of an individual's phraseology and a valid means to look at how phrase usage may vary between individual users. Part two describes some of the key works in L2 phraseology research and definitions of lexical patterns. In part three I will describe the methodology used to compare phrase usage between individual speakers and part four summarises the main findings of a short pilot study.

2. Second language learner phraseology

A key paper that asks questions about L2 phraseology is Pawley and Syder (1983). This paper describes two 'puzzles'; 'the ability of the native speaker to convey his meaning by an expression that is not only grammatical but also nativelike' (*ibid.*:189) and the puzzle of how a speaker can produce 'fluent stretches of spontaneous connected discourse' that is believed to be far beyond the natural human capacity to produce utterances one word at a time. The authors argue that fluent language use rests on a body of *lexicalised sentence stems* defined as 'a unit of clause length or longer whose grammatical form and lexical content is wholly or largely fixed'.

Pawley and Syder (*ibid.*) illustrate the puzzle of nativelike selection with a text consisting of a seventy year old New Zealand man describing his memories of the First World War. Utterances such as 'I had four uncles' and 'one Christmas th't I'll always remember' were paraphrased with 'the brothers of my parents were four' and 'there is not a time when my remembering that Christmas did not take place' to show how grammatically acceptable replacements can sound quite unnatural. The two puzzles are not, perhaps, so puzzling when one considers that the man is simply using the lexicalised sentence stems of 'I had number type of family member' and 'one event (that) I'll always remember'.

One claim that Pawley and Syder make, however, is that a 'distinction must be made between a morpheme sequence that is memorized and one that is lexicalized' (*ibid.*:208). They go on to explain that memorisation belongs to the domain of

‘performance’ in the Chomskyan sense while lexicalisation belongs to the domain of ‘competence’. This is a step too far towards suggesting a fundamental difference between the memorised phrases of the individual and perhaps a few friends and family members and the lexicalised phrases of the wider community.

Are there meaningful phraseological units below clause length however? Bolander (1989) adopts a broad definition of formulaic speech stating that the only criterion for qualification as a chunk or prefab is ‘either by a more frequent than average correct production of a certain structure, or by errors in structures that are otherwise correct’ (*ibid.*:73). Bolander also refers to the variety of terms used to describe formulaic speech such as ‘ready-made utterances’ and ‘schemata’ (Lyons 1968), ‘prefabricated routines’ (e.g. Brown 1973) and ‘prefabricated patterns’ (e.g. Hakuta 1974, all cited in Bolander 1989). Bolander’s study is based on data from sixty adult learners of Swedish and it suggests that the positioning of the negative particle *inte* supports a hypothesis on chunk learning.

Taking a closer look at Bolander’s definition one may be surprised to notice that a certain structure being produced with average frequency would *not* be judged as a chunk. Pawley and Syder’s (1983) division between an individual user’s *memorised sentence stems* and the community’s *lexicalised sentence stems* seems to share a problem with other authors’ divisions between chunks and non-chunks in that the line is imagined; largely unsupported by empirical evidence. Concerns such as these led me to consider whether a method could be found that held no such assumptions.

Months of pondering Sinclair’s (1991) Two Models of Interpretation: The open-choice principle and the idiom principle led me to a problem. What if there is no real division? Hunston and Francis (1999) offer the example of the words *I must confess* to illustrate the idiom principle and suggest that the words can be interpreted two ways. The authors tell us that the first interpretation, using the idiom principle, means ‘I am going to tell you something you may find unpleasant, or something I find embarrassing’ (*ibid.*:22) and the second, interpreted according to the open-choice principle, means ‘I am under an obligation (possibly self-imposed) to admit a wrong-doing’ (*ibid.*:22). I, personally, find it very hard to imagine a situation where anybody would say *I must confess* in a literal sense. The fifty concordance lines offered randomly by the BNC sampler show no cases that appear to show such a usage:

I must confess I find being a complete freelance is a strain and I love the

I must confess, my dear Fanny (his sister), that I found your judgement of

I must confess I was deeply depressed,‘; he said later.

I must confess that I find this deeply attractive, and I want to draw

I must confess,'; she said cheerily.

I must confess I hadn't participated in any giant financial coups.

who really believe, must confess to an unbridgeable gulf between

Moreover, I must confess that my quip was directed towards deflecting

though I must confess, I remain rather unsure as to how I should respond.

I must confess I didn't enjoy others at all.

Figure 2: First ten concordance lines of *must confess* taken from BNC sampler

Of course no computer corpus can provide conclusive evidence that a form is never used but such an example can generate questions such as 'When does one ever interpret language according to the open-choice principle?' Assuming there were a case where a person chose to confess to a crime this way the accused would likely cause confusion and participants in the dialogue may expect a joke.

Hoey's (2005) *Drinking Problem Hypothesis* suggests a reason why people may be primed to actively avoid the form *I must confess* when confessing to a crime or a serious incident. The hypothesis is named after a scene in the film *Airplane* that describes a pilot with a 'drinking problem'. The joke is that the pilot has a physical problem and cannot drink anything without spilling it. One part of Hoey's hypothesis is that users will avoid the rarer sense of a word or phrase because of their awareness of the more common form. In normal (serious) conversation a person wanting to describe a physical problem that results in difficulty drinking, say, water would be primed to avoid the phrase *drinking problem* just as a person wanting to confess to a crime would avoid the phrase *I must confess*.

It could be argued that it is exactly these kinds of situation and the sudden 'switch' to open-choice interpretation that can cause confusion and helps us make jokes but the wording 'open-choice' suggest that listeners suddenly begin to process each word individually. The phrase *I must confess* may be processed holistically, however, in both cases. Assume, for a moment, that every advanced speaker does have two primings for this phrase. One, the more common sense, that a person is about to give an opinion or admit to something slightly embarrassing and the second, the rarer sense, that a person wishes to admit to a serious wrongdoing. I have seen no evidence to suggest that

the second sense can only be understood by breaking the phrase down into individual words.

Essentially then, it is possible that every utterance influences the surrounding text and an investigation of what phrases are stored holistically and which ones are stored individually may ultimately be futile. This concern led me away from the problem of defining phraseological language and towards the problem of how to look at a piece of text and explore what mechanisms are at work when a speaker ‘chooses’ lexical items.

Hunston and Francis’ (1999:37) definition of *the patterns of a word* seemed like a good starting point:

The patterns of a word can be defined as all the words and structures which are regularly associated with the word and which contribute to its meaning.

This definition has two advantages. It allows one to focus on a particular word but it does not carry any assumptions about how that word affects the surrounding text nor where the effect ends if, indeed, it ever does actually end. The notation used by Hunston and Francis (*ibid.*) will be familiar to users of COBUILD dictionaries and includes the following examples:

V n verb group followed by a noun group such as *explain the fact*

V wh verb group followed by a wh-clause such as *explain how that happened*

V n to n lexical items can be included as part of a pattern, an example of this would be *explain it to me*

Many readers would agree that the greatest success of Hunston and Francis’ work is their observation that patterns can be categorised according to meaning. The V n group, for example, includes a BE group with verbs such as *be, compose, comprise and make*. This is not intended to imply that these verbs necessarily have the same meaning. It simply means that these verbs all share the form V n so that one can say *I am a teacher, the brain is composed of billions of nerve cells, the information pack comprises fifteen single sheets and he’d make a good president*. (examples based on Francis *et al.* 1996: 15).

As I am ultimately planning to look at the lexical patterns in the everyday English of a Korean speaker I felt it was important to note that the Pattern Grammar of

Hunston and Francis (1999) (PG) is based on native speaker interactions and it cannot be assumed that Korean English speakers will share the primings of the informants used in the development of PG. Furthermore, PG is not intended to be a full description of the English language so it seems prudent to allow for the fact that new patterns may be found or that the patterns described in PG are not the most appropriate for describing spoken Korean English. For the purposes of this paper I will be using the term *lexical pattern* to specify any string of words that could potentially be stored by the speaker as a chunk rather than members of a semantically related group. The phrase *go home*, for example is counted as a verb noun pattern (V n) without implying that other members of this group necessarily share a similar meaning.

3. Methodology

3.1 Collecting the data

I advertised for informants from the University of Liverpool's English Language Unit, a division of the English department that specialises in teaching EAP to students from various faculties and the pilot study began with interviews with the only three respondents. The three informants were from China, Mexico and Poland. At this stage I specified only that they should be L2 English learners in order to get a feel for the way they use lexical patterns and the practicalities of thirty minutes interviews with relative strangers.

My initial plan was to select every verb pattern (V) from each informant along with a list of all the Vs from the three texts consisting of only my utterances to see how much variation there was between the four speakers and also to see how consistent my own usage was on the three separate days data was collected. Verb patterns found included V, V n, V amount, V adj, V that, V so/not and V wh but questions regarding definitions of the categories and potential overlap began to concern me and I decided that a more detailed study of one category would be preferable. As verb noun patterns (V n) made up between 26 and 38 percent of all verb patterns at this stage of the study this category, the largest group of verb patterns, was chosen.

3.2 Verb-noun patterns

Three conversations were recorded and transcribed based on thirty minute interviews. A sample of one text is reproduced in figure three:

Glenn: did you apply for lots of universities in England

E: just three er I applied for Imperial college Cambridge and University of Liverpool

Glenn: what attracted you to Liverpool

E: scholarship

Glenn: okay they had similar deals the three colleges had similar programmes

E: yeah here in Liverpool the programme has more erm erm electrical and electronic and in Cambridge and Imperial college there is more like computer science

Glenn: okay what was your bachelor's degree

E: electrical and electronic engineering

Figure 3: Sample text used in study

My own turns were separated from those of the informants resulting in six texts. The three texts consisting of utterances from the three informants were labelled INF A, INF B, INF C and the corresponding texts consisting of my own utterances were labelled GLENN A, GLENN B and GLENN C. Figure four shows a sample of V n patterns from INF A.

My father sent me here
 I'm my only child in my house
 My father has brothers and sisters
 Chinese has policy
 The government only allows yeah one couple per child
 People have first child
 First child is girl
 Have a chance for boy

Figure 4: Sample of Vn patterns used in INF A

There is a multitude of ways of breaking down such a list to look for patterns that may be specific to the individual. I began to separate Vn patterns according to the three structures listed in Francis *et al.* (1996:14), i.e.: verb with complement, verb with object and verb with adjunct but it quickly became obvious that I would then need to include V amount, V about n, V across n and various forms of modal verbs. My definition of verb noun pattern is significantly broader than Francis *et al.*'s V n category therefore I

decided to simplify the study by breaking down the V n patterns into three new categories: those containing a BE verb, those containing a HAVE verb and OTHER verbs. The BE category was chosen because of its similarity to *verb with complement* and the HAVE category due to this verb's unusual properties. The verb HAVE as a main verb has no passive form. One can say *I have a car* but not **A car is had by me*.

Each of the six V n lists were divided up into the three categories BE, HAVE and OTHER. Figure five shows an example of each section taken from GLENN A.

V n patterns containing BE

1. You're mainland Chinese
2. It's a very big country
3. It's just one country
4. That's IELTS
5. That's writing
6. That's traditional Chinese
7. This is your final year
8. It was an easy subject for me

V n patterns containing HAVE

1. Have a chat
2. I had a sausage toastie in there
3. I have had Chinese food
4. It doesn't have some other meaning
5. Korea has wow real culture differences
6. They have little shops
7. If you have enough money
8. No-one has manual gears

V n patterns containing OTHER verbs

1. We need a hook
2. We'll try that
3. Yeah I get that
4. You got other members of your family brothers sisters
5. We just say China
6. We kind of forget the size

7. I learnt French
8. I did it twice a week for five years

Figure 5: Examples of V n patterns divided into three categories

As the ratio of BE, HAVE and OTHER V n patterns in the GLENN files showed some consistency across three different interviews in contrast to the rather different charts representing the V n patterns of the three NNS informants I felt it was important to collect multiple data sets from a single NNS informant. It is possible that the similarities in the three sets of GLENN data were coincidental, that they were somehow affected by my status as a native speaker or even that my status as the interviewer in each case affected the results so I recruited a fourth informant (my wife) and three further texts were produced STACY A, STACY B and STACY C to test the hypothesis that the ratio of a second language learner's V n patterns would remain consistent on three different interview days.

4. Results

All the data is taken from three thirty minute interviews about everyday topics such as life in Liverpool, the informant's hobbies and pastimes and culture differences. I was the interviewer in each case.

4.1 Participant's data

	Nationality	Age	Studies	English usage
GLENN	British	33	PhD English	Native speaker
Inf A	Chinese	19	Undergrad economics	IELTS 5.5 2 years ago, rarely speaks English
Inf B	Mexican	25	MSc Signal Processing	IELTS 6.5 often speaks English with friends
Inf C	Polish	24	PhD Biology	IELTS not taken, speaks English with friends

4.2 Ratio of various V patterns

Initially the results were broken down into the eight categories listed in figure six to see how usage may vary between individuals. Francis *et al.*'s (1996) original Vn category was subdivided into *verb with object*, *verb with complement* and *verb with adjunct* although none of the latter were found in informant A's data.

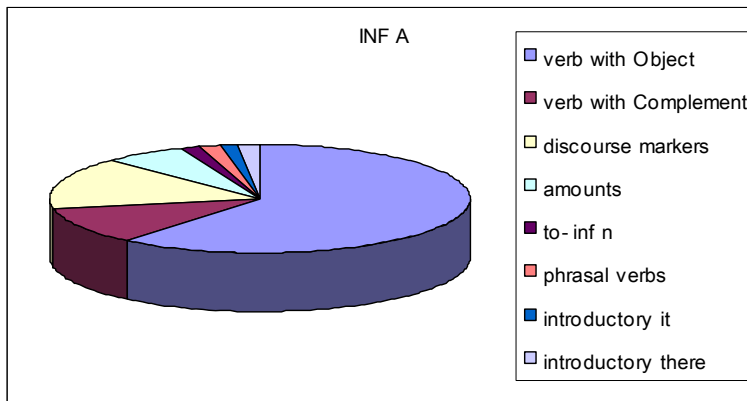
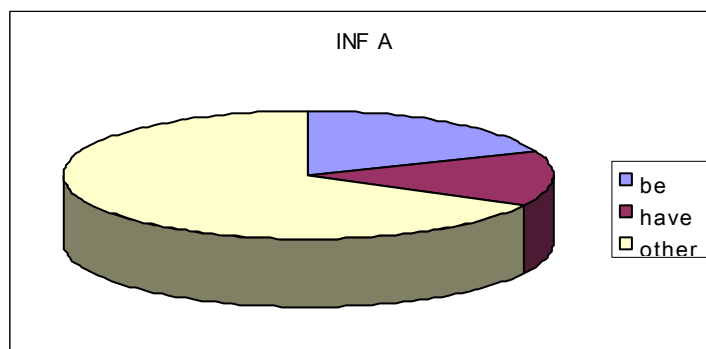


Figure 6: Initial breakdown of V n patterns used by informant A

Verb with amount, *phrasal verbs* and *introductory it* were included at this stage, despite being classed separately from Vn patterns by Francis *et al.* (*ibid.*) and *discourse markers*, *to-inf n* and *introductory there* were added as they appeared to make up a significant proportion of the data. The job of defining each these categories and accounting for utterances that could arguably go in more than one category began to seem overly complicated and I re-categorised the data into Vn containing BE verbs, Vn containing HAVE verbs and OTHER Vn categories. The Vn phrases used by the three informants are shown in figure seven.



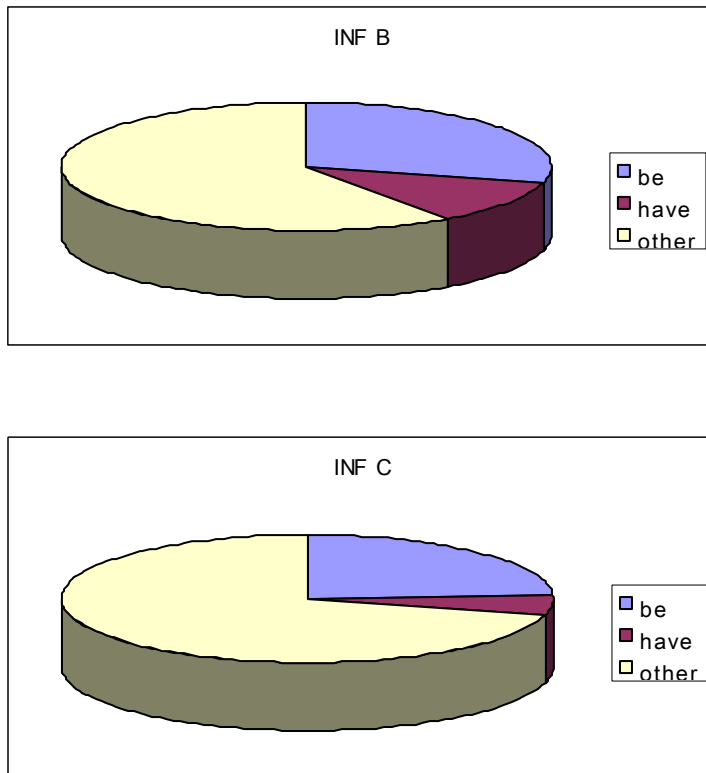


Figure 7: NNS texts broken down into BE, HAVE and OTHER V n patterns

The difference in the ratios is clear from the charts although, obviously, any cause is far from apparent. Vn containing BE verbs, for example range from only 19% in the case of informant A to 29% for informant B. Informant C appears to be using HAVE Vn patterns very differently from the other informants as seen by the mere 5% shown in the chart (compare 14% and 11% respectively for informants A and B) but one must remember that details of the BE patterns and different usage of certain OTHER Vn patterns will have an important effect. It should also be noted that the NNS informants were interviewees in the texts and I was the interviewer in every GLENN text. This important difference and the fact that only one text from each NNS informant exists limits the usefulness of these results on their own.

Figure eight shows the three GLENN texts broken down into the new categories to illustrate the consistent use of V n patterns.

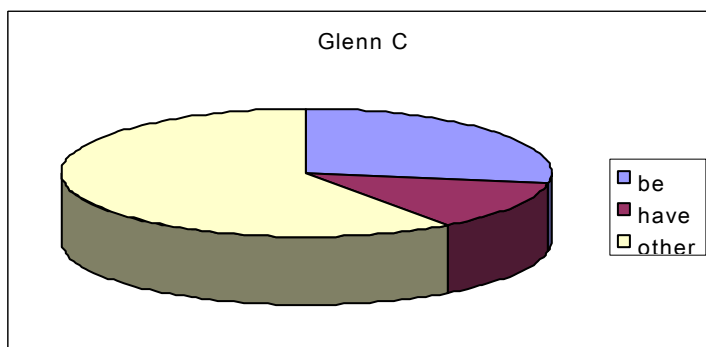
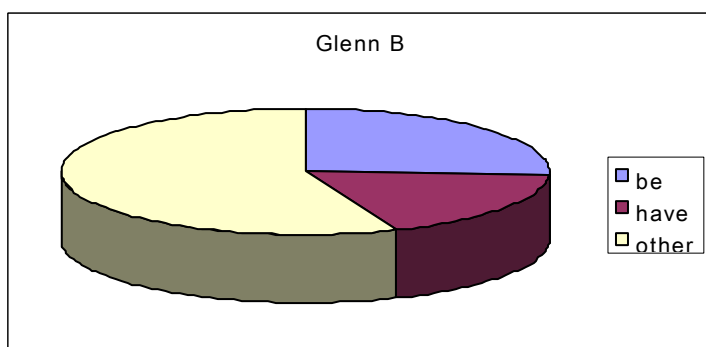
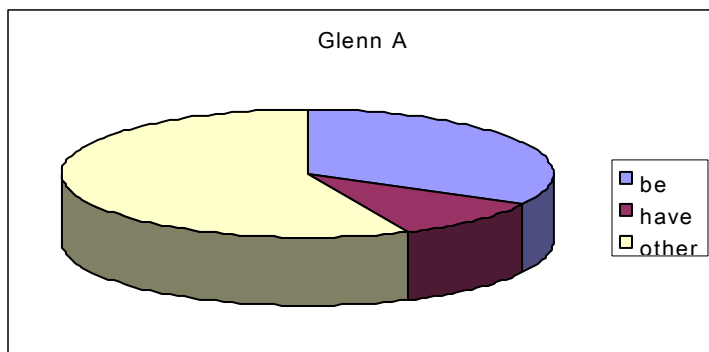


Figure 8: GLENN texts broken down into BE, HAVE and OTHER V n patterns

The Vn patterns containing BE verbs, for example, are seen to be consistent at 26-33% for these three texts. Of course there is still a range of 7% but this was deemed acceptable pending further statistical tests. The Vn patterns containing HAVE verbs range from 10-18% and the Vn patterns containing neither *be* nor *have* marked as OTHER are in the range 56-60%. Note that each of the three recordings was made more than a week apart.

At this stage of the study it was clear that at least three texts were required from a single NNS informant to check for consistency in the sense that can be seen in the GLENN data. As I will soon be looking at the data in more detail to investigate how (and perhaps why) certain forms are used differently from native speakers I felt it was

time to focus on an informant whose first language and culture I am quite familiar with: Korean. My wife is a Korean L1 speaker and stepped in as the final informant STACY.

The breakdown of the STACY data is shown in figure nine.

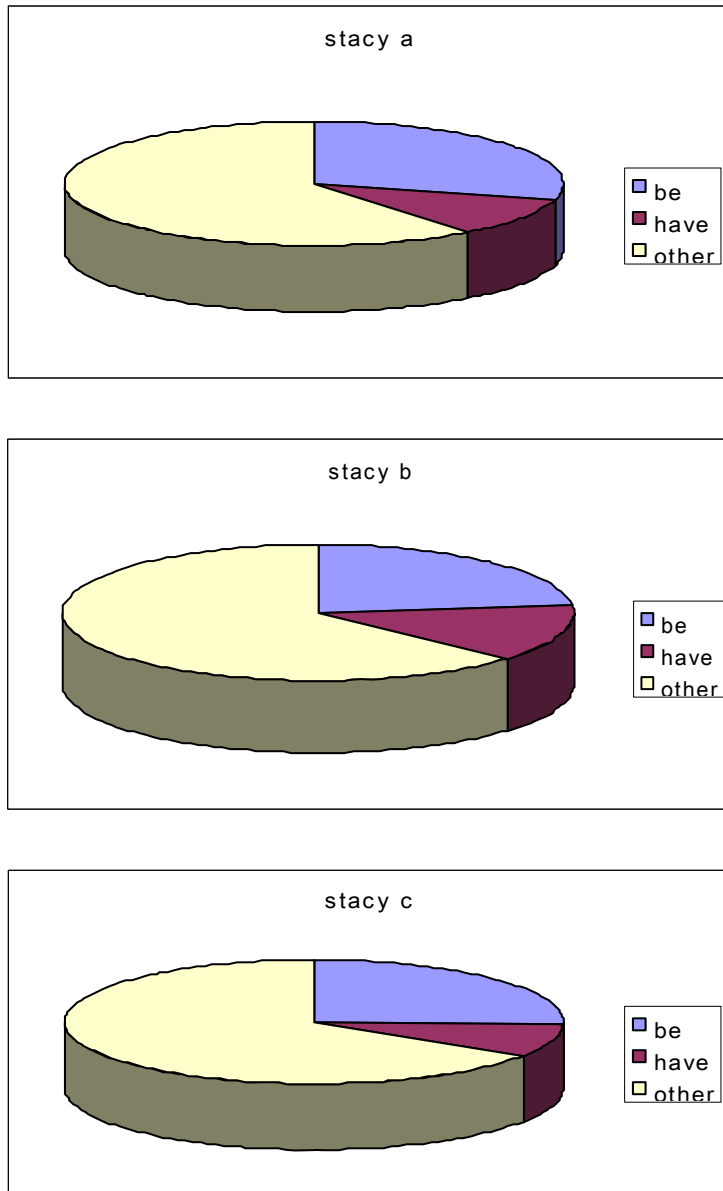


Figure 9: STACY texts broken down into BE, HAVE and OTHER V n patterns

The BE V n patterns in the STACY data occur at 29%, 23%, and 26% respectively. The HAVE patterns are 10%, 14% and 8% and the OTHER Vn patterns occur at levels of 61%, 63% and 66%. The largest variation in all the STACY data is 6% for the HAVE patterns and these data actually shows more consistency than the GLENN data. The results provide some indication that the kinds of Vn patterns used by a single speaker remain constant over the time period investigated i.e. approximately one month. Further

work is needed to investigate which specific patterns are used by each user and what reasons there might be for this. One informant, for example, relied heavily on the structure *you know n* when introducing new topics and I have not yet had the opportunity to investigate how this might have affected the presence of other patterns.

Similarities between the charts for GLENN, STACY and the informant with the highest given IELTS score might make one wonder if there is simply an ‘accepted range’ for native speakers that the learner gradually moves towards as they progress in their studies. There is, however, a real chance that my wife and I prime each other in our everyday interactions and these results may not reflect differences between any given native speaker and an advanced Korean English speaker. If there is a nativelike range what can be said of learners that are not in this range? Would a study of their Vn patterns simply reflect their individual idiolects or are there real geographical and cultural variations that could be taken as evidence of Lexical Priming? Is there significant variation amongst native speakers? Is there variation between members of the same learner community when they speak English and does this type of variation correlate with other means of judging the proficiency of a language learner? The results shown here are merely suggestive and I hope that future work will shed some light on these and the many other questions that have come up during this study.

5. Conclusions

In order to test theories such as Hoey’s Lexical Priming (2005) against L2 data a method of ‘measuring’ phraseological activity was sought. If one hypothesises that all language is primed and is essentially phraseological in nature it was judged necessary to choose one category of phrase and investigate how it varies between users in a given community. Verb-noun phrases (Vn) were chosen based on the category described in COBUILD dictionaries and Hunston and Francis (1999) but expanded to include any verb followed by a noun group including verbs followed by amounts, modals followed by nouns and verbs followed by prepositions and nouns. Nine sets of data were collected: three comprising every Vn used by three informants in separate interviews (INF), three comprising every Vn used by myself in those interviews (GLENN), and three comprising every Vn used by my wife, an L1 Korean speaker, in three different interviews (STACY).

The data from the first three informants showed three different patterns of phrase usage but the lack of further data from each informant made it impossible to tell whether a language learner/interviewee’s patterns might remain constant on three different days. The GLENN data indicated that the speaker’s use of Vn patterns in three

categories remained consistent on three different days although it is noteworthy that I was the only native speaker in the study and I was always the interviewer. My wife was recruited as a fourth informant and three data sets were prepared capturing her usage of Vn patterns on three different days across a period of three weeks. This STACY data did, indeed, stay approximately constant on three different days (see figure 9). Slight differences in definitions make direct comparisons between the GLENN data and the STACY data very difficult but these data sets offer support to the hypothesis that a speaker's phraseological usage is constant for a relatively short time period and can be recorded this way. These results are just the beginning but they do suggest that the methodology described produces valid, repeatable results that can be used to compare the phraseological usage of both individuals and communities in further studies

References

- Bolander, M. (1989) Prefabs, patterns and rules in interaction? Formulaic speech in adult learners' L2 Swedish, in Hyltenstam, K. and Obler, L.K. (eds.) *Bilingualism across the lifespan: Aspects of acquisition, maturity, and loss* (Cambridge: Cambridge University Press), 73–86.
- Brown, R. (1973) Meaning and Memory. *Forum Linguisticum* 1: 1-14. Cited in Bolander, M. (1989) Prefabs, patterns and rules in interaction? Formulaic speech in adult learners' L2 Swedish, in Hyltenstam, K. and Obler, L.K. (eds.) *Bilingualism across the lifespan: Aspects of acquisition, maturity, and loss* (Cambridge: Cambridge University Press), 73–86.
- Francis, G., Hunston, S. and Manning, E. (eds.) (1996) *Grammar Patterns: Verbs* (London: Harper Collins).
- Hadikin, G. S. (2006) *What is phraseology and does the imageability of verb-noun collocations affect a student's performance?* Unpublished MA dissertation, Institute of Education, University of London.
- Hakuta, K. (1974) Prefabricated patterns and the emergence of structure in second language acquisition. *Language Learning*, 24. 287–97. Cited in Bolander, M. (1989) Prefabs, patterns and rules in interaction? Formulaic speech in adult learners' L2 Swedish, in Hyltenstam, K. and Obler, L.K. (eds.) *Bilingualism across the lifespan: Aspects of acquisition, maturity, and loss* (Cambridge: Cambridge University Press), 73–86.
- Halliday, M.A.K. (1994) *An Introduction to Functional Grammar*. (London: Arnold).
- Hoey, M. (2005) *Lexical Priming: A new theory of words and language*. (London: Routledge).

- Hunston, S. and Francis, G. (1999) *Pattern Grammar: A corpus-driven approach to the lexical grammar of English*. (Amsterdam/Philadelphia: John Benjamins).
- Lyons, J. (1968) *Introduction to Theoretical Linguistics*. (Cambridge: Cambridge University Press) cited in Bolander, M. (1989) Prefabs, patterns and rules in interaction? Formulaic speech in adult learners' L2 Swedish, in Hyltenstam, K. and Obler, L.K. (eds.) *Bilingualism across the lifespan: Aspects of acquisition, maturity, and loss* (Cambridge: Cambridge University Press), 73–86.
- Pawley, A. and Syder, F.H. (1983) Two puzzles for linguistic theory: Native-like selection and native-like fluency. In Richards, J. and R. Schmidt (eds.) *Language and Communication*. (New York: Longman).
- Sinclair, J.M. (1991) *Corpus, Concordance, Collocation*. (Oxford: Oxford University Press).
- Thorkelson, T.S. (2005) *Konglish in the classroom: The Teacher's Back Door*. Presented at KOTESOL Daegu workshop 4/6/2005.
- Yorio, C.A. (1989) Idiomaticity as an indicator of second language proficiency. In Hyltenstam, K. and Obler, L.K. (eds.) *Bilingualism across the lifespan: Aspects of acquisition, maturity, and loss* (Cambridge: Cambridge University Press), 55–72.