How I Learned to Stop Empiricising and Love my Intuitions

Dougal Graham
King Mongkut's University of Technology Thonburi

Abstract
Corpus linguistics' approach to the analysis of language data is often highly focused on the use of statistics to generate research results. This is exceptionally useful in terms of replicability of research, however results are not always completely useful for the target audience. Many corpus linguistics resources discuss the importance of intuition in the interpretation of findings. Despite this, research is still often highly focused on results which appear to be purely empirically driven, when in fact the process of analysis and interpretation may be highly intuitive but be written to appear straightforwardly empirical. This article will discuss the author's process of moving from a purely statistical approach to corpus linguistics to a more blended intuitive-empirical approach throughout the analysis and interpretation of corpus data. While this article will focus on research performed in an ESP context, the content should be generalizable to much of corpus research.

1. Introduction

In 2013, soon after beginning my first academic teaching position since the completion of my masters degree, I read the Academic Formulas List (Academic Formulas List) article in Applied Linguistics, by Simpson-Vlach and Ellis (2010). As a teacher in post-secondary education teaching English as a foreign language for specific academic purposes (at an engineering university in Thailand), I became interested in the use of formulaic sequences and their potential implications for my students. Since Pawley and Syder's (1983) work on formulaic language as a component of native speaker fluency, interest and research in this area has blossomed. Many researchers have shown that formulaic language mirrors native and second-language learner learning strategies (Nattinger & DeCarrico, 2009, pp. 24–29), is a tool used to increase both productive and receptive fluency (Conklin & Schmitt, 2007), and allows the speaker greater comprehension and retention of information (Tremblay, Derwing, Libben, & Westbury, 2011).

Coincidentally, at this time one of my colleagues had just completed the compilation of the Engineering English Corpus (see Graham, 2014). This corpus, compiled from student English-language Engineering textbooks represents the authentic language that my students use throughout their undergraduate engineering careers. It seemed like the perfect opportunity to take my interest from the Academic Formulas List and try to apply some of the concepts that they described there to my own context. Furthermore, as a part-time programmer, and coming from a partially computational linguistics background, the highly rigorous, computational, and empirical approach taken in the generation of the Academic Formulas List and much of corpus-linguistics appealed to me. Finally, as a novice researcher, the possibility of conducting a piece of replication research to introduce me to a new field seemed like a reasonable way to both create something which my students could immediately use, and also to produce a useful piece of research for students and teachers of engineering English.

This article will begin with a discussion of my methods and my results, and finally move to a discussion of the implications for other researchers considering conducting similar research.

It is hoped that this in depth discussion will prove especially useful for other novice researchers, considering conducting corpus research.

2. Methodology

2.1 Corpora

Following the methods used in the development of the Academic Formulas List, two corpora were used to identify which phrases were more frequent in engineering English than in general English. The target corpus (from which I derived the phrases) was the Engineering English Corpus, or EEC. This corpus is comprised of representative samples of the textbooks used in the English language (international) programme at the Thai university at which I was employed at the time. The use of this corpus allowed me to both meet practical requirements by generating materials which would be specifically useful to my students, while also generating materials which could well be used by other English as a Foreign Language engineering students elsewhere. For more specific details on the construction of the corpus see (Graham, 2014). The comparison corpus representing general English in this case was the British National Corpus, or BNC. The British National Corpus, despite being somewhat dated is extremely comprehensive (over 100 million words and over 50 genres of spoken and written language), and still used in much research (eg: Heatley & Nation, 1994; Martinez & Schmitt, 2012). One issue with the use of the British National Corpus was that it represents British English, whereas the textbooks are representative of American English (at least in terms of spelling). This meant I needed to normalize the data for spelling (eg: color vs. colour).

2.2 Locating Frequent Phrases

The next step was that I applied very similar statistical analyses to those used in the creation of the Academic Formulas List. For the purposes of this research, formulaic language was operationalized in statistical terms. Formulaic language is defined to be sequences of three, four, or five words which occur together with surprising frequency, and occur significantly more frequently in the engineering data than in general English. The reason for the first criterion was that I wished to examine chunks, not words, and not collocations, so the sequences should be at least three words in length. The criterion of surprising frequency of word co-occurrence is to aid in the finding of formulaic language. As formulaic language often tends to be somewhat frozen and highly cohesive (Nattinger & DeCarrico, 2009, pp. 31–35), a statistical measure of this property would be useful (Schmitt, 2010; Simpson-Vlach & Ellis, 2010). The final criterion was to make sure that I found phrases which would be useful to the students and for this I would use the Formula Teaching Worth score described below.

First, I created a frequency list of all three, four, and five-word sequences from both corpora. The list from the engineering corpus contained approximately 2.4 million items. As in the Academic Formulas List I then cut off all phrases below a given frequency threshold. The reasoning here is that as the goal of the research is pedagogical, it makes sense to locate language to which the learners will be exposed frequently. Similarly, I wanted to make sure that the phrases I chose were representative of engineering English as a whole and not a single sub-discipline which might happen to use a single phrase extremely frequently. As such, I also made sure to remove phrases which occurred in fewer than 10% of the texts, and did not occur in a set of core texts used by all students.

2.3 Locating Engineering Phrases

Using the resulting frequency list I calculated the log-likelihood score for each sequence. Log-likelihood is a statistic which allows the researcher to measure if a term is more frequent in one corpus or another (Dunning, 1993; Schmitt, 2010; Simpson-Vlach & Ellis, 2010). It can be applied equally well to words (keyword analysis), or to sequences of words treated as a single chunk as in my research and in the Academic Formulas List. Once this log-likelihood score was calculated, I set a minimum log-likelihood cutoff. This cutoff made sure that I would pick language that was significantly more frequent in my corpus, so that I would not advocate teaching students language that they might already have learned in their previous English courses, and was in fact representative of engineering English as opposed to general English.

2.4 Locating Useful Phrases for Teaching – Statistical Methods

Now that I had a list of the phrases which were surprisingly frequent in engineering English, I was free to move on and try to determine which ones constituted useful formulaic language for my engineering students. To do so, I followed the methods of the Academic Formulas List, calculating the mutual information (MI) score, and using their Formula Teaching Worth statistic to rank the phrases. Mutual information is a statistic which is used to see if two (or more) items occur together with surprising frequency. One characteristic of formulaic language is that it is often "frozen", or "less variable" in the terms of Nattinger and DeCarrico (2009). Compared with the analytic construction of language in which words are put together using grammar rules, formulaic language is often characterized as being stored in the brain as a single lexical unit and retrieved as such with little to no modification possible. If this is the case, we can easily measure how often words occur together. If we see that they occur together very frequently, but rarely or never with other words (high mutual information), then it is likely that this means that we are witnessing less variable word choice, or frozen word organizations. This characteristic is what mutual information seeks to measure.

Finally, with frequency and mutual information calculated, and the various cut-offs applied, I was ready to calculate the "Formula Teaching Worth" score and then use it rank my list of 3,000 phrases by usefulness for teaching. The Formula Teaching Worth score created by Simpson-Vlach and Ellis (2010) for the Academic Formulas List is an attempt to create a statistical representation of teacher intuitions into usefulness for teaching. By taking a stratified random sampling of their data (based on phrase frequency rank and mutual information scores) and having teachers rate them for usefulness for teaching, they were able to determine a statistical correlation between teachers' judgments and the frequency and mutual information scores. Using that correlation, if it is valid, it should be possible to rank the phrases by applying the amount of correlation to the frequency and mutual information to determine a composite score that more closely resembles teacher intuitions. In this way, they have taken an intuitive judgment and quantified it in terms that are easily measurable.

3. Results

After I applied the calculation of and ranking by the Formula Teaching Worth score I was then able to compile a list of phrases. To give an idea of what the phrases were like Table 1 shows the top 15 phrases ordered by Formula Teaching Worth from my corpus.

Table 1 Engineering English Phrases ranked by Formula Teaching Worth

Rank	*Three Words*	*Four Words*	*Five Words*
1	what is the	can be used to	at a rate of #
2	the number of	as a function of	you should be able to
3	as shown in	the magnitude of the	beyond the scope of this
4	# and #	as shown in figure	how long will it take
5	can be used	with respect to the	the first law of thermodynamics
6	shown in figure	in this chapter we	in such a way that
7	the value of	the value of the	the rate of change of
8	in terms of	the sum of the	the external forces acting on
9	be used to	newton 's second law	recall from chapter # that
10	# to #	in terms of the	in this section we will
11	in which the	the centre of the	what is the difference between
12	to determine the	is equal to the	that can be used to
13	the centre of	the length of the	determine the magnitude of the
14	is to be	the weight of the	in this chapter we will
15	with respect to	at a rate of	it is not possible to

Now that my results were complete, all that was left was to interpret and discuss them, and see what there might be to say. However, a problem emerged. I didn't like these results. It was a feeling, and I couldn't shake it. I turned to introspection, and asked myself why I didn't like these results, and, after considering them I decided I didn't feel they were useful for teaching. Sure, some of them might be useful for teaching, but others certainly were not. When I showed my results to colleagues, they agreed with me. Some of the phrases were too simple (e.g., *what is the*) and the students would understand them already from their prior education in general English. Some of the phrases, while formulaic, were highly technical terms that students were already explicitly learning as part of their engineering studies (e.g., *the first law of thermodynamics*). Others still just seemed *uninteresting* somehow (e.g., *the value of (the)*, *the sum of the*). This final category was the most troubling to me. Of course some students have trouble with articles, and prepositions. But is teaching them the phrase "*the value of the*" as an unanalyzed chunk going to address the issue? I came to the conclusion that the answer to that question was "no".

4. Re-evaluation

At this point, I needed to step back and take a second look at my methods, and my assumptions. Given that my goal was to create a list of phrases which were "useful for teaching" and that many of these phrases clearly did not fit that criterion I began to look for alternative criteria which could be added to my existing ones or alternate methods of ranking the phrases which might be useful for teachers and students. First, I calculated Z-score for my collocations. The z-score statistic is another measurement of probability of words occurring together. Where MI tends to prioritize phrases containing more frequent words z-score prioritizes those with less frequent words (Schmitt, 2010). I had thought that possibly by using this statistic and applying an additional cut-off point for my data, I could remove some of the phrases with highly technical

terms which only occur occasionally. Unfortunately, this did not help. Seemingly useful and less useful phrases were interspersed within the data equally.

Beginning to feel frustrated, I decided to look for frames in my data. A frame is a variable type of formulaic language in which there is at least one open slot which can be filled by words of certain types. In fact I found several such frames, and based on an analysis of the words which filled the slots using Roget's Thesaurus (Roget & Kirkpatrick, 1987), I found several clear patterns. However, this was not really what I had set out to look for, and the frames, while they did tend to be filled with certain types of words, and some might have proved useful to students, they primarily seemed to show use of prepositions and determiners, and were often characterized by those phrases which I earlier mentioned as "feeling uninteresting" or which just contained a technical term. One of the potentially more useful frames was "at a __ of #" which could be filled with "measurable quantities" (per Roget & Kirkpatrick's classification) such as "rate", "speed", or "pressure". It is unlikely that students would be familiar with such a use of "at" to denote something other than a physical location. However, most of the other results such as "to the __ of" ("right", "left", "sum", "axis", "direction", "surface") did not appear to be interesting for teaching purposes using very typical syntactic patterns which learners would be familiar with.

4.1 Comfort in the literature

Eventually, when discussing my problems with a senior colleague, he asked me why I had not tried using intuitive measures to locate the most useful phrases for teaching. This shocked me. I felt that intuitive measures would be unscientific and leave my work open to extensive criticism, and potential unpublishability. However, I decided to return to my literature review to see what other researchers said specifically about the use of intuition. First, I re-read my original source article, which I was replicating, the Academic Formulas List. Their approach to intuition, as I mentioned above, was an empirical one, attempting to link intuitive judgments with empirical measurements via the Formula Teaching Worth score. They explained that empirical judgments were subject to claims of subjectivity, mirroring statements by other researchers. Laviosa (2011), particularly worries that hand-picked example sentences from the data could easily result in cherry-picking data which matches the researcher's beliefs rather than being representative of the beliefs of the speech community being studied. This seemed like exactly what I might be doing if I were to try to intuit which phrases were useful for learners. By applying subjective criteria, I would be locating phrases which seemed useful to me, but would they in fact be useful? Similarly, Louw (2011) has broad complaints about the use of intuition in corpus linguistics in general, complaining that most researchers fail to trust the text, and rather use intuition as a cover for speculation and armchair linguistics. These are not labels which I want applied to my work.

As I continued to re-read work which I had read previously, I realized that in fact many corpus researchers, doing work in the field of formulaic language were using intuitive methods to conduct their research. For example, Martinez and Schmitt (2012) first applied statistical corpus techniques to obtain a list of common phrases, and then they applied a set criteria to that list of common phrases to attempt to determine which were the most useful for learners of general English. However, when reading that article before I had ignored most of the discussions of intuitions as they related to formulaic language. As I had conceptualized my work as being a partial replication of the work of Simpson-Vlach and Ellis (2010), just in a new context, I had internalized the concept that my research would be empirical, and that I did not need to consider intuitive methods, or their justifications. This had been aided and abetted by my background in programming, and computational approaches to linguistics problems. I had always felt that if

something was not quantifiable, or explicitly defined, it should not be included in research. As I continued re-reading, now with a focus on intuition, I realized that many of the classical characteristics of formulaic language were intuitive ones, which I had prior to this ignored in favor of statistical definitions. The following general definition of formulaic language emerged. It is a multi-word unit which generally has at least one of these characteristics: (1) semantically opaque, (2) grammatically unusual, (3) unitary, (4) situationally dependent, (5) somewhat frozen in form (Myles, Hooper, & Mitchell, 1998; Nattinger & DeCarrico, 2009; Wray & Namba, 2003).

Some of these criteria were already met, using statistical measures in my existing list. I used log-likelihood to measure situational dependency in terms of genre, and MI to attempt to measure language which is frozen in form. However, the other three characteristics were more difficult to conceptualize in empirical terms. Potentially, grammatical unusualness could be measured statistically if the sentences were coded for part-of-speech, syntactic categories, and so forth, it could be compared with a general English corpus. In fact, extensive discussion of this approach is given by Villada Moirón (2005) who looking in great detail at using statistical and computational methods to locate formulaic language in a fully tagged and parsed corpus was able to greatly increase the effectiveness of computerized, automated models for the retrieval of formulaic language. However she notes that currently there may not be sufficiently advanced automatic thesauri to effectively measure semantic opacity. She similarly notes as a deficiency of her model that it does not take into account "lexical, morphological, or syntactic idiosyncrasies" of formulaic language. As my data was neither tagged nor parsed, such approaches were not available to me, and given the caveats it seems likely to be a case of diminishing returns to attempt comprehensive tagging and parsing for relatively minor gains in ability to computationally determine formulaic language in my corpus.

4.2 Embracing the Intuitive

At this point, I decided that I would need to use some sort of intuitive criteria for the determination of a list of useful phrases. In discussion with a senior colleague I settled on the use of markedness criteria. These were useful because they allowed me represent characteristics such as semantic and grammatical issues which I felt were most likely to be important to learners in the context of how different they are from the most common usage. This is explicitly useful for my students because they come from a background of studying English for general purposes. Any language that deviates from that norm is likely to be more difficult for them, and the further the deviation, the more difficult and therefore likely useful it becomes. I also added a second intuitive criteria: that if a phrase were only interesting only due to a single technical term, but otherwise following normal grammatical and semantic patterns, then it would be excluded. This allowed me to exclude phrases which the students would likely be able to understand clearly only by learning their vocabulary from engineering courses, and focus on phrases which they were unlikely to be able to understand purely based on that knowledge. In the end I was left with the following seven criteria (Graham, 2014):

1) **marked part of speech**: any of the words in the phrase do not have their usual part of speech;
2) **marked word form**: any word in the phrase does not occur in the most common form of that word;
3) **non-prototypical word meaning**: any word in the phrase does not occur with its most prototypical meaning;

4) **marked collocations**: the phrase contains any collocations or co-occurrence patterns that differ from general English patterns;
5) **non-literal phrase meaning**;
6) **specialized syntax**: the phrase contains or is connected with complex or unclear syntax.
7) **Non-technical**: the phrase is not interesting only due to containing a technical term

As in work by Martinez and Schmitt (2012), and Wray and Namba (2003), these categories are not criteria to exclude formulaic language, but instead to include them (criterion 7 is a special case in this regard as it excludes phrases which only fit one of the above criteria because of the use of technical terms). In this way I was able to include much language which matched only one of these criteria allowing me to find a wide range of results, reflecting the diversity of formulaic language. Finally, after applying these criteria to my list of phrases, I determined a list of about 40 phrases which I felt would be useful for teachers and students studying engineering English.

5. Lessons Learned

I learned several lessons in the conducting of this research and through the process of re-evaluating my methods, assumptions, and beliefs. These may be useful, especially to other novice researchers. One primary lesson I learned in conducting this research is that an intuitive judgment does not necessarily have to be unclear or vague. My initial goal was to find phrases that would be useful for teachers and students, which is of course far too vague. I was able to narrow this down to phrases which were likely difficult for students. This was somewhat more precise, but it still did not give me the level of granularity and clarity which I sought. By asking myself what specifically might make a phrase difficult for a student I was able to determine my seven criteria which were much more focused. By having such explicit criteria, I was much more comfortable working with my data.

Furthermore, I realized that many of my criteria were easily empirically verifiable by doing quick checks against the data. By checking the use of a word in a phrase that I felt intuitively to be marked against its use in the BNC and my corpus, I was able to quickly determine if it was used with a common part of speech in my data, or if the collocations were indeed unusual. While generally I found my intuitions to be quite well representative of the data, in several cases I was able to exclude phrases which I had initially felt might fit the criteria, but in fact did not. This procedure worked well for categories (1-4) but was more difficult (or impossible) to apply with categories (5-7). I believe that this type of highly systematic approach combines many of the benefits of both intuitive and empirical approaches. By attempting to determine which specific features are likely to be at the root of our intuitive judgments we may be able to more easily conduct intuitive research. Furthermore, by making these intuitive judgments as explicit and detailed as possible, I hoped to avoid the kinds of criticisms made by Laviosa (2011) and Louw (2011), that this work might represent purely my own beliefs about which language is useful for teaching, and instead represent my beliefs about what type of language is useful for teaching and then locate all language of that type from within my data.

There are many issues to be aware of when working in corpus linguistics, and if one is particularly biased towards either empirical approaches or intuitive ones, then it is easy to overlook other methods which may be appropriate to one's work. Just as I had ignored potentially useful intuitive methods, others may also do the reverse, and ignore potentially useful empirical methods in favor of intuitive approaches. For example, Martinez and Schmitt (2012), used a mixed-method approach, using a frequency-based approach to locate 15,000 n-grams which they believed could potentially be useful to learners, and then applying their six

qualitative criteria to each one. The play-off here between empirical and intuitive approaches seems to be that given the state of current metrics for analyzing formulaic language, that an intuitive approach is still more valid, resulting in higher accuracy, and potentially higher recall, than an empirical approach. However, this comes at the cost of time. While categorizing 15,000 items by hand certainly shows dedication and commitment, I wonder why the authors of the Phrasal Expressions List did not attempt any of the statistical measures which attempt to determine highly cohesive and unitary expressions such as mutual information. This could well have been used as a soft cut-off to remove all phrases with extremely low scores, potentially saving them the time and effort of categorizing several thousand additional phrases.

In terms of precision and recall, the statistical methods used by Martinez and Schmitt (2012) were somewhat less precise than those I used. Precision refers to whether our statistics only retrieve the data they are meant to retrieve. If they retrieve many phrases which are not useful for us, then they have low precision. Normally, this is empirically measured, however if our goal is pedagogical usefulness, then perhaps a comparison against intuitive measures may be in order. From their empirical list of 15,000 items, they located 515 phrases, or approximately 1 phrase in 29 was deemed useful. However, for the creation of my list of phrases, I found 40 useful phrases out of a list of 700, giving me a precision of about 1 in 18. This seems to show that there may well have been room to use additional statistics to refine their search before applying their intuitive criteria. However, another issue to consider is that of recall. While precision is loosely defined as getting only what you want (and no other data), recall can be considered to be getting all of what you want. In this sense, using a less empirical method will likely produce better results. There certainly are phrases which might be useful for teaching, but which had to be excluded from my Engineering Phrases List on the basis of statistical characteristics. Nevertheless due to the low threshold set by Martinez and Schmitt (2013), they were able to retrieve a much larger number of phrases overall. It should be mentioned that the goal of the low threshold was not to maximize quantity of results, but to ensure that phrases through the range of frequency determined to be likely of most use to students be retrieved.

On the other hand Simpson-Vlach and Ellis (2010), in the creation of the Academic Formulas List only used statistical measures, resulting in formulaic language such as "*the University of Michigan*" which is unlikely to be useful for many learners. One of the underlying assumptions of the development of the Formula Teaching Worth score is that either frequency or MI should be well related to intuitive judgments of the importance of formulaic language. As we have seen above, frequency is useful to find language which is often used and therefore important for learners and MI may reasonably be related to frozenness of form, or unitarity. However, these are just two of many criteria for the identification of formulaic language, and it is not clear that they are even the most psychologically important ones. It is certainly possible that unusual syntax or semantics may hold higher relevance to teacher intuitions, and therefore a more complex statistical measure that seeks to take these concepts into account may prove more useful than the current Formula Teaching Worth statistic. The formula teaching worth score therefore is a useful first step towards this goal, but more research to determine what other empirical measures might well reflect the other aspects in which formulaic language is intuitively judged is warranted.

It appears that in the field of corpus linguistics there is certainly room for both empirical and intuitive methods, although attempting to apply one at the expense of the other may result in problems. When only applying statistical methods, my findings were significantly less precise than was ideal. This lack of precision was displayed in the very small number of useful phrases

that were located using only my empirical methods, and was greatly increased by applying intuitive criteria. As well, when applying intuitive criteria, the higher the level of specificity, the better. This will ensure clarity for the researcher, and reduce the chance of researcher bias becoming a possible source of judgments. Finally, it was extremely difficult for me to come to terms with the validity of an intuitive approach. As a novice researcher whose background was in highly empirical methods, I had ignored the amount of intuition present in other researchers' work, and instead focused on their empirical methods. This blindness led me to significant difficulties and frustration once my results did not match my goals, and I believe that this is a lesson that may be useful to other novice researchers, both empirically and intuitively minded. The intuitive-minded should keep in mind the benefits in time and clarity that can be gained from taking an empirical approach, and the empirically-minded should keep in mind the benefits in accuracy that can be gained from an intuitive approach which are crucial to useful pedagogical tools.

References

Conklin, K., & Schmitt, N. (2007) Formulaic Sequences: Are They Processed More Quickly than Nonformulaic Language by Native and Nonnative Speakers? *Applied Linguistics*, *29*(1), 72–89. doi:10.1093/applin/amm022

Dunning, T. (1993) Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, *19*(1), 61–74.

Graham, D. (2014) The Engineering Phrases List - Towards Teachable ESP Phrases. Presented at the The 34th Annual ThaiTESOL International Conference.

Heatley, A., & Nation, P. (1994) Range. *Victoria University of Wellington, NZ.[Computer Program, Available at Http://www. Vuw. Ac. Nz/lals/.]*.

Laviosa, S. (2011) Corpus linguistics and translation studies. In V. Viana, S. Zyngier, & G. Barnbrook (Eds.), *Perspectives on corpus linguistics* (Vol. 48, pp. 131–154). John Benjamins Publishing.

Louw, B. (2011) Philosophical and Literary Concerns in Corpus Linguistics. In V. Viana, S. Zyngier, & G. Barnbrook (Eds.), *Perspectives on corpus linguistics* (Vol. 48, pp. 171–196). John Benjamins Publishing.

Martinez, R., & Schmitt, N. (2012) A Phrasal Expressions List. *Applied Linguistics*, *33*(3), 299–320.

Myles, F., Hooper, J., & Mitchell, R. (1998) Rote or Rule? Exploring the Role of Formulaic Language in Classroom Foreign Language Learning. *Language Learning*, *48*(3), 323–364.

Nattinger, J., R., & DeCarrico, J., S. (2009) *Lexical Phrases and Language Teaching*. Oxford University Press.

Pawley, A., & Syder, F. H. (1983) Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. *Language and Communication*, *191*, 225.

Roget, P. M., & Kirkpatrick, B. (1987) Roget's thesaurus of English words and phrases.

Schmitt, N. (2010) *Researching Vocabulary: A Vocabulary Research Manual*. Palgrave MacMillan.

Simpson-Vlach, R., & Ellis, N. C. (2010) An Academic Formulas List: New Methods in Phraseology Research. *Applied Linguistics*, *31*(4), 487–512. doi:10.1093/applin/amp058

Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011) Processing Advantages of Lexical Bundles: Evidence from Self-paced Reading and Sentence Recall Tasks. *Language Learning*, *61*(2), 569–613.

Villada Moirón, M. B. (2005) *Data-driven identification of fixed expressions and their modifiability*. GRODIL, Secretary Department of General Linguistics. Retrieved from http://www.villada-moiron.nl/thesis.pdf

Wray, A., & Namba, K. (2003) Use of formulaic language by a Japanese-English bilingual child: a practical approach to data analysis. *Japan Journal for Multilingualism and Multiculturalism*, *9*(1), 24–51.