Using Vocabulary Profiling Assessment Software to Promote Independent Process Writing

Arthur McNeill

The Hong Kong University of Science and Technology

Introduction

The extent to which assessment tools can make a positive contribution to student learning is likely to remain a controversial topic in education in spite of the recent interest in exploring the learning potential of some approaches to assessment and the pedagogical uses of assessment data. In the minds of many educators, assessment and teaching are activities that need to be distinguished clearly from each other. For example, in examination-dominated education systems, the widespread use of examination papers as teaching materials has been criticized on the grounds that the practice promotes test-taking skills at the expense of developing students' understanding of the subject matter. However, the current interest in Assessment *as* Learning has encouraged new perspectives in language assessment, including a reappraisal of some established language tests. This paper explores the potential of a customized version of the Lexical Frequency Profile (LFP) assessment software (Laufer & Nation 1995) as a learning tool.

The original purpose of LFP was to measure the lexical richness of texts. The classic version of the software provides a profile of a text's lexical content by grouping the vocabulary items into four categories: (a) words from the first 1000 frequency level, (b) words from the second 1000 frequency level, (c) words included in the Academic Wordlist (AWL) (Coxhead 2000), and (d) words not included in the previous three categories, i.e. "off-list". The words included in the first and second 1000 frequency levels are basically the items in the General Service List (GSL) (West 1953). A later version of LFP divides a text's lexis into twenty levels based on the British National Corpus (BNC)

(Leech et al. 2002), plus "off-list". LFP has typically been used by teachers to assess texts written by students. It has also been used to scrutinize texts when considering their suitability for use in examinations and textbooks. The LFP performs a quick analysis of a text's lexical content by showing the frequency level of the words used and providing some basic descriptive statistics about the text based on word frequency. The profile shows the extent to which a text consists of low frequency and academic words. Although LFP has proved to be a popular tool among researchers, other approaches to measuring lexical richness have been adopted (e.g. Meara 2001, Meara & Bell 2001, Miralpeix & Celaya 2002, Bell 2003, Šišková 2012).

Lexical richness has been identified as one of the most important features of second language writing and can determine the quality of a text (Koda 1993). Many advanced learners of English, in particular undergraduate and postgraduate students studying through the medium of English, also benefit from LFP and use it to check the vocabulary content of course assignments and academic papers before submitting them. Unfortunately, however, this use of the software for checking drafts is currently less accessible to intermediate level learners because the wordlists used in the original LFP are based on the frequency of occurrence of words within the language in general, rather than a particular L2 vocabulary syllabus where the prescribed lexical content may vary considerably from frequency lists such as GSL and BNC, which comprise the databases of LFP.

This paper describes an attempt to produce a version of LFP that is of particular relevance to learners in Hong Kong, by replacing the frequency-based wordlists (GSL and BNC) with the four wordlists (Key Stages 1 – 4) that were developed for Hong Kong schools (Curriculum Development Council 2009, 2012; McNeill 2011). The paper also shows how institutional wordlists, such as the prescribed words of a university English course can be included within a customized LFP. It is argued that students will benefit more from using LFP as a learning tool if its lexical database corresponds to the vocabulary curriculum they are expected to study. In its original form, LFP is driven by wordlists that consist of lexical items selected because of their frequency of occurrence in English generally.

Hong Kong English Wordlists and a Customized LFP

Increasing evidence has emerged to suggest that many Hong Kong students leave school with an inadequate English vocabulary. For example, a study conducted with first-year undergraduates at the Chinese University of Hong Kong concluded that the majority of school leavers entering university know fewer than 3000 English words (Chui 2005). International research into the English language proficiency of students studying degree courses through the medium of English has suggested that a vocabulary size of 5000 words is necessary for students to cope with the demands of reading academic texts in English (Laufer 1989, 1992). In order to promote higher English vocabulary targets for Hong Kong school leavers, the Education Bureau commissioned a study of the vocabulary needs of Hong Kong primary and secondary students, with a view to developing an English vocabulary curriculum for primary and secondary education. The first stage of this curriculum project involved the creation of four wordlists, consisting of the vocabulary items that students could be expected to know at different stages of their education.

It was agreed at the outset of the curriculum project that a sensible starting point in selecting words for the lists would be corpus data showing the frequency of occurrence of English words. It was assumed that frequency of occurrence would provide a useful provisional ranking of the potential words. This initial ranking would then be subjected to scrutiny by a number of stakeholders, including teachers familiar with the student population for whom each list was intended, before arriving at a final selection of target lexis. However, when analyzing the patterns of response from teachers, it soon became apparent that the number of words rejected was far greater than expected (McNeill 2011). The use of the BNC corpus was, at first sight, attractive to the research team because of its spoken component. The Hong Kong word lists were intended to reflect vocabulary used in spoken as well as written English. However, because BNC is representative of contemporary usage within UK, a large number of the words were judged to be inappropriate in the S.E. Asian context. Many of the items were considered to be restricted to users of colloquial British English or relied on familiarity with contemporary British culture.

Examples of high-frequency BNC words judged to be restricted to the UK context were *bobby, dodgy, dole, heck, lass, plonk, posh, innit, shilling, tuppence, kiddy, owt.* BNC words considered too colloquial for a L2 school curriculum included *baffle, boo, blob, buck, clobber, fiddle, eve, grumble, grotty, hassle, fuss, tumble, wobble, potter, dodgy, swap, poke, cop, buzz, whack.* The high number of rejected items from the 4000 level of BNC illustrated the enormous difference that appears to exist between the high-frequency vocabulary in the English used by British people for everyday communication among themselves and the kind of vocabulary which ESL learners might be expected to learn for the purpose of education and employment in their own countries.

Probably because of their very high frequency within English, the words in GSL presented fewer cultural barriers to being accepted into the Hong Kong lists than the comparatively 'newer' vocabulary of BNC. In spite of the development of enormous corpora in recent years, there remains little controversy over the words that constitute the first 2000 words of English. However, when GSL words were rejected in the teacher decision-making tasks of KS1 and KS2, it was usually because the items were judged to be inappropriate for young learners. The majority of the items were eventually included in the lists for KS3 and KS4. Therefore, unlike the original LFP, the customized Hong Kong version does not assume that the first 2000 words of English, by themselves, should be the starting point of an L2 vocabulary curriculum for school children.

Customizing the LFP Software

At first sight, the task of replacing one word list with another in the LFP software looks straightforward. However, one of the challenges for LFP, particularly when serving as a learning tool, is to assign a text's lexis to the prescribed lists (in our case the vocabulary of the four Key Stages) as accurately as possible. Basically, the software needs to anticipate and accommodate all morphological variations of the words in each list. While the vocabulary target for Hong Kong students by the end of secondary education is 5000 English

words, the number of word forms which students actually know is much larger than 5000. The lists contain 5000 word families (Bauer & Nation 1993), which means all the derived forms of each headword need to be included in the database so that the LFP will recognize them and place them in the correct category when profiling students' texts. If a word form is not recognized, LFP places it in the "Off-List Words" category. Since all 570 of the AWL words are included across the four KS lists, there is no separate AWL list. The basic LFP report format is shown in Figure 1.

	Tokens	Percentage
KS1 Words		
KS2 Words		
KS3 Words		
KS4 Words		
KS1+KS2+KS3+KS4		
Off-List Words		
Total		

Figure 1. Categories used to profile lexis based on Hong Kong's local wordlists

Post-Secondary Customization

The main appeal of producing a customized version of an assessment software program such as LFP is the opportunity to focus learners' attention on the language they are expected to know according to their local curriculum, rather than being judged according to knowledge of words that may not be relevant to their own situation and needs. Preparing a version of LFP based on the Hong Kong primary and secondary school vocabulary lists can make the tool available and accessible to the thousands of students who study English up to Diploma of Secondary Education (DSE) level. However, for students who proceed to post-secondary education, it is possible to include additional lists, such as the prescribed words of particular English courses and lists of discipline-specific lexis for subjects such as business, engineering, science, medicine, law, architecture, etc.

This section provides an illustration of an additional wordlist based on a first-year university English course. The course, "English for University Studies", is taken by the majority of undergraduates at the Hong Kong University of Science and Technology (HKUST) and focuses on improving overall English proficiency and developing academic literacy. The course has a prescribed wordlist, derived mainly from the course reading materials, which students are expected to master.

A wordlist consisting of the prescribed words for "English for University Studies" (e-Core Words) was added to the Hong Kong version of LFP and students were encouraged to make use of the tool by submitting drafts of their course assignments into LFP and making revisions to their texts in the light of the lexical profile. Many university writing courses adopt a process approach, whereby a student submits a draft of a writing assignment to the instructor, who then provides formative feedback on the written text so that the student can re-draft and improve it. The pedagogical advantages of process writing are undisputed and it has long been established that re-drafting a text in the light of feedback results in improved writing quality (e.g. Raimes 1983, White & Arndt 1991). However, as every writing instructor knows, giving feedback on drafts is enormously time-consuming on the instructor's part. In some respects, LFP allows students to learn to write in a process mode by providing them with an individual lexical profile of a draft. Students can then decide whether they need to make changes to the content in order to demonstrate that they are able to use the words that have been prescribed for study on a particular course.

Experience of using the LFP with course-prescribed words at HKUST reveals that students typically produce first drafts with an over-reliance on high-frequency items, then make a serious effort to monitor their lexical choices and improve the sophistication of the vocabulary in the light of their LFP profile. Fears that students might insert and substitute words in a reckless manner just to improve their profile were largely unfounded. However, to test the system, a text was prepared which deliberately used as many of the course-prescribed words as possible. The results are shown in Figure 2.

Vocab-Profile:

	Tokens	Percent
KS1 Words:	58	40.28%
KS2 Words:	10	6.94%
KS3 Words:	7	4.86%
KS4 Words:	13	9.03%
KS1 + KS2 + KS3 + KS4	(88)	(61.11%)
eCore Words:	44	30.56%
Off-List Words:	12	8.33%
Total:	144	100%

intuition is the antithesis of rational thinking and can be radically differentiated from coherent models of cognitive deduction intuitive thinkers may be individualistic spontaneous self-reliant sophisticated and prone to self-indulgence paradoxically however they are not immune to criteria such as randomness and consensus their deviation from socioeconomic and ethnic trends can imply a transitional immersion in dogma yet we should be reluctant to generalize about intuition on the basis of anecdotes and superficial brainstorming although we may sympathize with and even idolize the hallmarks of unique engagement in diverse ventures texts such as this assert that lexical richness is a relevant variable of tech-savvy writing for resourceful learners however some instructors may be skeptical about its stark contribution to cohesive anything

Figure 2. LFP analysis of a text written to display mastery of a prescribed university wordlist

Although the above text represents an extreme lexically-oriented approach to writing in which the author deliberately included words from the course prescribed wordlist, the resulting text suggests that high-quality academic writing is heavily dependent on lexical choice. In the above example, more than 30% of the words were taken from the course-prescribed wordlist! Nevertheless,

the practice of encouraging students to reflect on the vocabulary content of their draft assignments and allowing them the opportunity to re-draft their texts through lexical substitution is clearly worthwhile educationally.

The use of different colours to represent the words of the various lists provides students with an immediate visual analysis of their text. Once students learn which colours are highly valued in terms of the profiling, they generally find the lexical editing of their work engaging and enjoyable.

Conclusion

This paper has discussed the potential of the LFP software as a learning tool to promote second language writing. As explained above, LFP was developed as an assessment instrument to measure the lexical richness of written texts and has, in fact, been used by learners, usually advanced level adult learners, to check drafts as part of a revision process. Therefore, LFP has always been an assessment tool that can help language learners to improve the quality of their written texts. However, as explained in the paper, LFP's potential as a learning tool can be greatly enhanced when used in contexts where there is a local vocabulary curriculum whose wordlists can replace the frequency-based lists of the original LFP. Such customization of the tool can allow it to be used by students of all proficiency levels, including young learners, since the vocabulary databases used correspond completely to the students' local English curriculum.

References

Bauer, L. and Nation, I.S.P (1993). Word Families. *International Journal of Lexicography*, Vol. 6, No. 3, 253-279.

Bell, H. (2003). Using Frequency Lists to Assess L2 Texts. University of Wales Swansea: Unpublished Thesis.

Chui, S.Y. (2005). *An Investigation of the English Vocabulary Knowledge of University Students in Hong Kong*. Unpublished M.Phil. thesis. Hong Kong: The Chinese University of Hong Kong.

Coxhead, A. (2000). A new Academic Word List. *TESOL Quarterly*, Vol. 34, No. 2, 213-238.

Curriculum Development Council (2009). *Enhancing English Vocabulary Learning and Teaching at Primary Level*. Hong Kong: Government Logistics Department.

Curriculum Development Council (2012). *Enhancing English Vocabulary Learning and Teaching at Secondary Level*. Hong Kong: Government Logistics Department.

Koda, K. (1993). Task-induced variability in FL compositions: Language-specific perspectives. *Foreign Language Annals*, 26 (3), 332-346.

Laufer, B. (1989). What percentage of text-lexis is essential for comprehension? In Lauren, C. and Nordman, M. (eds.) *Special Language: From Humans Thinking to Thinking Machines*. Clevedon: Multilingual Matters.

Laufer, B. (1992). How much lexis is necessary for reading comprehension? In Arnaud, P.J.L. and Béjoint, H. (eds.) *Vocabulary and Applied Linguistics*. London: Macmillan 126-132.

Laufer, B. and Nation, I.S.P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, Vol. 16, No. 3, 307-322.

Leech, G., Rayson, P., and Wilson, A. (2002). Companion website for word frequencies in written and spoken English: based on the British National Corpus. Available at http://ucrel.lancs.ac.uk/bncfreq

McNeill, A. (2011). From corpus to language curriculum: a data-based or data-driven exercise? *Proceedings of EDULEARN11*, International Association for Technology, Education and Development. ISBN: 978-84-615-0441-1.

Meara, P. (2001). PLEX v1.1. University of Wales Swansea. Available at http://www.swan.ac.uk/cals/calsres

Meara, P. and Bell, H. (2001). "P_Lex: A simple and effective way of describing the lexical characteristics of short L2 texts", in *Prospect*, 16, 3: 323-337.

Miralpeix, I. and Celaya, M.L. (2002). "The Use of P_Lex to Assess Lexical Richness in Compositions Written by Learners of English as an L3", in I. Palacios (ed.), *Actas del XXVI Congreso de AEDEA*. Santiago de Compostela/ University of Santiago de Compostela Press, 399- 406.

Raimes, A. (1983). *Techniques in Teaching Writing*. Oxford: Oxford University Press.

Šišková, Z. (2012). Lexical richness in EFL students' narratives, *University of Reading Language Studies Working Papers*, Vol. 4, 26-36.

West, M. (1953). *A General Service List of English Words*. London: Longman.

White, R. and Arndt, V. (1991). *Process Writing*. London: Longman.

Author's e-mail:　lcamcneill@ust.hk