The Lexical Frequency Profile: Problems and Uses

Richard Smith
International University of Japan

This paper examines three claims that Laufer and Nation make for the Lexical Frequency Profile (LFP), a tool that measures lexical richness in learner writing. The paper argues that the claim that the LFP is a valid measurement tool needs to be revised because the LFP is not a comprehensive measure of lexical richness and that the claim that the LFP is a sensitive measurement tool is questionable. Evidence is presented that casts doubt on the stability of the LFP when text extracts exceed 200 words. Some suggestions about ways to design studies using the LFP to mitigate the effects of its weaknesses are forwarded. The paper concludes by identifying areas of research that may yield improvements in the validity, sensitivity, and stability of lexical richness measurement tools.

　この論文は、ラウファー及びネイションが the Lexical Frequency Profile（以下LFP）、すなわち英語学習者の書く文面に見られる語彙力を測る手段について述べている3つの論点を吟味するものである。まず、ここではLFPが有効な測定手段であるという議論には修正が加えられるべきであるとの見方をしている。なぜならLFPが包括的な尺度ではないからだ。次に、LFPが優れた測定手段であるとする議論には問題があるとしている。そして本文中の抜粋部分が200語を超える場合LFPには変動がない、としている点については完全に事実であるとはいえないということを明らかにするものである。これらの考察は、LFPが最も問題になりにくいリサーチ目的に対して示唆を与えることになる。この論文では、語彙力を測る手段の妥当性、感受性、安定性を高めるであろう研究分野を見極めることによって結論を導き出す。

Introduction

Before the arrival of the Lexical Frequency Profile (LFP) in 1995, the measurement of lexical richness in learner writing represented a research domain which was fragmented by the existence of several incompatible measurement tools, none of which achieved dominance or inspired confidence (Laufer & Nation, 1995). Laufer and Nation make bold claims for the LFP in their 1995 article even though the empirical database that supports these claims is small, far from representative of the entire learner population, and ambiguous in some respects (see Meara, forthcoming). The strength of these claims and the fact that the LFP is available as a freely downloadable software program, VocabProfile, (Nation & Heatley, 1996) have encouraged a steady flow of research which utilizes the LFP, much of it unpublished or in in-house publications, but some of it reaching general audiences through widely disseminated journals (e.g., Coniam, 1999; Meara, Lightbown & Halter, 1997; Morris & Cobb, 2004; Muncie, 2002). With the lone and recent exception of Meara, the authors of this body of work accept uncritically the large claims that Laufer and Nation make for the LFP. Now that the body of published work that utilizes the LFP is reaching a critical mass, it is time to reconsider some of Laufer and Nation's claims.

After providing a brief summary of the measurement structure of the LFP, of the type of data it yields, and why it represents an advance on previous measures of lexical richness, this paper examines three of the claims which Laufer and Nation make for the LFP in their 1995 article: *validity*, *sensitivity* and *stability*. The examination of the first claim takes the form of an observation and brief discussion of the measurement structure of the LFP. The examination of the second claim comprises a short survey and assessment of typical LFP research outcomes and a reference to a forthcoming paper by Meara, who has reviewed this claim, as well some other LFP claims, by using computerized simulations of LFP analyses. The examination of the third claim involves a new empirical investigation. The paper then draws some conclusions about the sorts of research designs for which the drawbacks of the LFP appear to be non-critical, identifies directions for future research into the development of good lexical richness measurement tools, and points out the importance of such tools for both research and teaching.

The Lexical Frequency Profile

The LFP is a measure of lexical richness in writing which counts the number of word tokens in a text and distributes these word tokens among four frequency levels which are derived from standardized word frequency lists. The four frequency lists which Laufer and Nation stipulate are:

> **Band 1**   the most common 1,000 words in English (1-1,000)

Band 2   the next most common 1,000 words in English (1,001-2,000)

AWL   the Academic Word List, an updated version (Coxhead, 2000) of the original "University Word List," which contains 3,100 words belonging to 570 word families which appear frequently in a wide range of academic textbooks

Not in the lists   less frequent words

The results of this frequency analysis are expressed in terms of the percentage of word *families* (as opposed to word *types*) that belongs to each frequency band. In every case, the percentages add up to one hundred percent. The final computation of a Profile for a text might look like this:

Band 1: 86.5%   **Band 2**: 7.0%   **AWL**: 3.5%   **Nil**: 3%

Some merits of the LFP

Laufer and Nation make several claims for the LFP, a majority of which are ambitious and deserve to be treated with some skepticism. It is important, however, to recognize as valid one claim in their 1995 article, that the LFP represents an advance over older measures of lexical richness in writing such as Lexical Density, Lexical Sophistication, and Lexical Variation. The incorporation of real-world frequency lists in the measurement tool mean that the LFP may be a more discrete, objective and sensitive measure of lexical richness than its predecessors. It is more discrete than tools like Lexical Density because its results are not affected directly by non-lexical features of text. It is more objective than tools like Lexical Sophistication because

its results are independent of subjective interpretations of sophistication. It is more sensitive than tools like Lexical Variation because it can segment the variation according to four frequency levels.

LFP claims

Not content to rest their case, Laufer and Nation (1995) take the notion of the LFP's superiority over existing measures to a higher level when they argue that LFP results might serve as proxies for non-native speakers' productive vocabulary knowledge, arguing that Profiles change in systematic ways in step with changes in the size of learners' vocabularies. Laufer and Nation make six claims about the LFP in order to support their larger claim. The six claims are that the LFP (Laufer & Nation, 1995):

- is a reliable and valid measure of lexical use in writing

- provides similar stable results for two pieces of writing by one person

- discriminates between learners at different proficiency levels

- correlates with an independent measure of vocabulary knowledge

- is a useful diagnostic test

- is a sensitive research tool (p. 319)

This paper limits its attention to two of these claims - that the LFP:

- is a reliable and valid measure of lexical use in writing

- is a sensitive research tool

The examination of the first of these two claims focuses on its second proposition, that the LFP is a *valid* measure. This paper also examines a third claim. This is a specific technical claim advanced by Laufer and Nation in an earlier part of their article that "Profiles over 200 words were found to be stable …" (1995, p. 314). Laufer and Nation do not distinguish between the meanings of "reliable" and "stable," so the author assumes that Laufer and Nation believe that their claim in support of the "stability" of the Profiles contributes to their claim in support of the "reliability" of the LFP.

Claim 1: the LFP is a valid measure of lexical use in writing

The claim that the LFP is a valid measure of lexical use in writing depends in part on its capacity to measure productive vocabulary knowledge in a comprehensive way. An important component of productive vocabulary knowledge is knowledge of "multi-word expressions." According to Calzolari et. Al. (2002), multi-word expressions (MWEs) are:

> used to describe different but related phenomena, including fixed or semi-fixed phrases, compounds, support verbs, idioms, phrasal verbs, collocations, etc. At the level of greatest generality, all of these phenomena can be described as *a sequence of words that acts as a single unit at some level of linguistic analysis*. (p. 1934)

In his own definition of productive vocabulary knowledge, Nation (2001: 27) acknowledges that MWEs represent an important part of total productive word knowledge. He identifies nine dimensions of knowledge that are involved in "knowing" a word for productive purposes (see Table 1).

Table 1. Nation's (2001) taxonomy of productive word knowledge (abridged)

Form	spoken	
	written	
	word parts	
Meaning	form and meaning	
	concept and references	
	associations	
Use	grammatical functions	In what patterns must we use this word?
	collocations	What words or types of words must we use with this one?
	constraints on use (register, frequency…)	

Two of these dimensions, grammatical functions and collocations, involve knowledge of MWEs. In the case of collocations, MWEs represent the entire knowledge

dimension; in the case of grammatical functions, MWEs represent some areas of the knowledge dimension, particularly the area occupied by lexico-grammar. The LFP, however, is similar to all other existing lexical richness measurement tools in that it has no capacity to recognize and count MWEs.

Determining the extent to which the LFP's blindness to MWEs weakens its validity is not easy since there is considerable disagreement about the precise definition of MWEs and, thus, disagreement about how many exist in the English lexicon. What is undeniable is that even the more conservative research studies in this area have yielded significant frequency figures for MWEs. Wiktorrson (2001), who prefers to use the term, *prefabricated phrases*, in place of MWEs, compared essays on argumentative topics written by native speaker university students with essays written by non-native speaker university English students and hand counted the number of "verb phrases," "noun phrases," and pragmatic expressions within the two sets of texts. Although Wiktorrson had no interest in repetitions in the texts of these sorts of prefabricated phrases and thus only counted the numbers of "types," the results showed that both native speaker writers and learner writers used similar numbers of prefabricated phrases (829 for the native speaker writers and 915 for the non-native speaker writers) extracted from texts which totaled 10,907 and 10,876 words in length. This means that, even shorn of repetitions and even assuming that prefabricated phrases comprise only two words, the number of words in MWEs accounted for at least 16.8% (2 x 8.4%) of the total number of words in the non-native speaker texts. Other studies by Erman and Warren (2000) and

Foster (2001) lend direct or indirect support to this finding. This conservative estimate of 16.8% frequency of MWEs bears significance for the claim that the LFP, incapable of measuring MWEs, is a valid measure of lexical richness in learners' texts.

Restricting the definition of MWEs only to those MWEs that currently find dictionary entries (such as phrasal verbs, prepositional verbs, and some compound noun phrases) reduces but does not eliminate the LFP's validity problems. Many phrasal verbs and some prepositional verbs consist of words that, separately, count as highly frequent prepositions such as *in*, *on*, *at*, *by* and *off* and as highly frequent verbs such as *bring*, *come*, *go* and *take*, but, together, the word compounds such as *bring in*, *come by*, *take off,* and *go on* would count as much less frequent lexemes. Counting or not counting these MWEs could in many cases result in discrepancies in LFP counts of perhaps a few percentage points. Although such discrepancies seem small, they are in fact quite significant in relation to the restricted variances that reported LFP values typically exhibit (see the discussion of Claim 2 below).

Claim 2: the LFP is a sensitive research tool

The results of published studies, including those in Laufer and Nation's 1995 article, demonstrate that the LFP has some sensitivity. The question under consideration here is whether the LFP is as sensitive as its creators claim. Laufer and Nation claim that, based on the analysis of single short pieces of learner writing, the LFP "…discriminates between learners of different proficiency levels …" and "… is thus a useful diagnostic tool as well as a sensitive

research tool" (1995, p. 319). Research usually focuses on differences among groups, but diagnostic tools should serve to identify differences among individuals. Any examination of this claim should therefore consider the LFP's capacity to discriminate among individuals as well as groups. This examination consists of a mini survey of published research in which there are correlation data for LFP values and general proficiency, a brief discussion of some data from the author's own research, and a reference to a finding from Meara's (forthcoming) computerized simulation of the LFP.

The author's survey of the published research finds that there are two published LFP studies that focus directly on the relation between LFP values and general proficiency. Laufer and Nation's original 1995 study shows conclusively that the LFP can discriminate significantly between two widely separated proficiency *groups*, one of which consists of "low intermediate" learners and the other of which comprises learners whose level is equivalent to the Cambridge First Certificate in English (CFCE). The reported standard deviations also show that the LFP is able to discriminate in most cases between *individuals* from the two widely separated groups. This same study, however, shows that the LFP is much less successful in discriminating between the two CFCE groups separated by just under two semesters of university study; the only significant difference between them lies in the use of "not-in-the-lists" vocabulary. Furthermore, the reported standard deviations show a very large overlap between the two sets of *individuals* in the CFCE groups (see Table 2).

Morris and Cobb's (2004) study compares LFP values obtained from two groups studying in the same university degree course: a group of native speakers and a group of

JALT 2004 NARA — Language Learning for Life

Table 2. Laufer & Nation's (1995, p. 316) reported LFP values for three groups of learners

	1st 1,000		2nd 1,000		UWL		Not in lists	
	Comp 1	Comp 2	Comp 1	Comp 2	Comp 1	Comp 2	Comp 1	Comp 2
Group 1	86.5	87.5	7.1	7.0	3.2	4.1	3.3	2.8
SD	3.8	5.3	2.0	2.3	1.8	2.5	2.3	1.8
Group 2	79.7	79.4	6.7	6.8	8.1	7.8	5.6	6.6
SD	5.3	4.5	1.7	2.2	2.3	2.3	3.5	3.3
Group 3	77.0	74.0	6.6	5.6	8.1	10.1	7.5	8.7
SD	6.1	5.9	2.6	2.5	3.2	2.9	2.9	3.5
F-test	19.35	33.1	0.29	1.89	24.86	27.40	10.46	22.74
p value	.0001	.0001	0.29	.16	.0001	.0001	.0001	.0001

Note: Group 1 = "low intermediate" learners;
Group 2 = CFCE level;
Group 3 = CFCE level + two semesters study

"proficient" non-native speakers. This study does not present any raw LFP data. It finds statistically significant differences between the two *groups* for all four frequency bands, but also finds that there is considerable overlap between the two sets of *individuals* in terms of achieving an "ideal" Profile in which Band 1 values are lower than 85% and AWL values exceed 5%.

Table 4 shows the LFP values from the author's own study obtained at different essay extract lengths for three proficiency groups: Group1, TOEFL 450-577; Group 2, TOEFL 590-627; Group 3, TOEFL 630-650 and native speakers. The discussion of "Claim 3" in the next section describes the data collection for this study. These results suggest that the LFP is unable to distinguish consistently among the three proficiency levels. A one-way ANOVA of

the LFP values at essay extract lengths of 220 words and 300 words and t-tests of the LFP values at essay extract lengths of 450 words show only one statistically significant difference among the three groups: for "not-in-the lists" values at extract lengths of 300 words ($F = 4.8$, $p < 0.05$). A *post hoc* analysis (using Fisher's PLSD) shows significant differences between Group 1 and Group 3 for "1st 2000" and "Beyond 2000" values at both 220 and 300 word extract lengths. Groups 2 and 3, however, are statistically indistinguishable. The standard deviations show some overlap between values for *individuals* belonging to Group 1 and *individuals* belonging to Group 3.

Meara (forthcoming) observes similar patterns of sensitivity in computerized simulations of the LFP. He shares

this paper's conclusion that the LFP probably discriminates quite well between lower level learners and more advanced learners, but does not distinguish well between groups that are closer to each other. He adds that, if we assume productive lexical growth of about 500 words each year, it is not likely that the LFP will be able to discriminate between two groups of learners separated by one year of study.

One likely factor that contributes to this observed pattern of reduced sensitivity is the restricted variance ranges which LFP values exhibit. Studies published in the sources cited above, two other studies (Laufer, 1998 & 1994), and the empirical study included in this paper suggest that the full range of variance for the "1st 1000" values lies somewhere between the low 70s and the high 80s, the range for the "2nd 1000" values lies somewhere between 1% and 9%, and the ranges for "AWL" and "Nil" are somewhere around 3-14% and 2-10% respectively. These figures are approximate since chance error and the use of different essay extraction lengths and of different essay genres mean the figures cannot be equally compared. Although approximate, the variance ranges do indicate the inherent difficulty the LFP faces in separating groups of learners whose vocabulary sizes differ by less than an order of magnitude. An interesting feature of the variance ranges is the tendency of lower proficiency learners to produce Profiles in which "1st 1000" values range between the low 80s and the high 80s and of learners from intermediate level upwards to yield equivalent values in the low 70s to high 70s range. This suggests that the "1st 1000" values, which exhibit the widest variance, might be reasonably good at capturing the relatively small increases in vocabulary size which occur at lower proficiency levels, but much less

effective at capturing the much larger increases in vocabulary size which occur from intermediate levels upwards.

In terms of the criteria which Laufer and Nation set, existing evidence suggests that the LFP is only a *partially sensitive* tool. Diagnostic applications of the LFP would most likely work best in separating individual learners into rather *broadly* defined proficiency levels.

Claim 3: Profiles over 200 words are stable

Laufer and Nation simply report that "Profiles over 200 words were found to be stable …," (1995, p. 314). Identification of significant criteria by which to assess "stability" is less obvious than it is for assessing sensitivity because Laufer and Nation neither provide sample stability data nor make any explicit reference to stability criteria. A common sense definition of "Profile stability" would be that Profiles over 200 words return the same numbers regardless of the text lengths from which they are calculated. This common sense definition may be unreasonably strict since very small discrepancies in Profile number distributions could have insignificant effects on the overall results of an LFP study. A more reasonable definition of stability is one that measures these discrepancies in relation to their effects on the importance of study results. Discrepancies that are significant in relation to variances in observed values would count as having significant effects on a study's results. Thus, a reasonable definition of Profile stability could be formalized as how the size of observed standard deviations in mean LFP values is related to the tendency of LFP values to change with changes in essay extract lengths. If the changes in standard deviations are much larger than changes induced through

manipulation of essay extract length, then the LFP could be considered stable. If the opposite is true, and manipulation of essay extract length induces changes in values nearly as large as standard deviations, then the LFP can be considered unstable. The standard deviations reported by Laufer and Nation in their foundation study (1995) range from as low as 1.7 to as high as 6.1, with 3.5 being a typical middle value (see Table 2). In order to check whether LFP values have any tendency to change with changes in essay extract lengths, the author conducted his own empirical investigation.

Subjects and data collection/processing

The subjects were 47 candidates for admission to the English-medium MBA school within the author's host institution. Nine of the subjects were native speakers and the other 38 were native speakers of a variety of Asian and European languages. This GMAT subject population was chosen because of its likely capacity to write essays that would extend up to 400 words and beyond even under standardized test conditions.

Two sets of data were obtained from the subjects. The first set of data comprises argumentative or analytical essays written for the *Analytical Writing Assessment* (AWA) portion of the GMAT, which is a standardized screening tool used for MBA admission at most English-medium MBA programs around the world. The second set of data comprises TOEFL scores obtained for the 38 non-native speakers from TOEFL tests administered within four months of the GMAT.

In accordance with standard LFP procedures (Laufer & Nation, 1995), the essays were edited to eliminate all proper nouns and to correct all non-significant errors such

as wrong derivatives and minor spelling errors. The essays were then entered into the computer for data analysis at three composition lengths: the first 220 words, the first 300 words and the first 450 words. Some of the essays written by lower proficiency candidates did not much exceed 300 words, which is reflected in the values for Group 1 in Table 4.

Results and discussion

Table 3 shows the differences between LFP values for *individual* essays at different essay cut-off points. Positive and negative changes are represented as neutral values since both are equally relevant to the stability of a single subject's Profile. Table 4 shows the mean average LFP values for the 47 essays when sorted into three proficiency *groups*. This group analysis allows us to see whether the individual differences (Table 3) neutralize each other when the differences are aggregated and whether any aggregated differences display a direction of LFP variation. All values shown in Table 4 are *positive* values.

The results suggest that both individual and aggregated Profiles suffer essay length-related instability. In extreme individual cases, the change in observed Profile values is equal to or greater than the observed standard deviations for the group, and the mean average of individual changes is about 30-40% the size of group standard deviations. The aggregated Profile values in Table 4 show that these individual differences do not neutralize each other. Instead, there is a consistent tendency for aggregated "1st 2000" values to diminish with increasing extract length and for their mirror image, the "Beyond 2000" values, to grow. This tendency is marked for Groups 2 and 3, but subdued for

Group 1, the lowest proficiency group.

The results in Table 4 appear to be consistent with the intuitive expectation that, the more text a more proficient writer writes, the more opportunity he has to deploy his lexical knowledge. If this is a correct interpretation, it follows that the comparative stability of the values for the lowest proficiency group is a function of its shorter essay lengths and its dependence on a smaller lexicon. In other words, it appears that the LFP is more stable when there is less to measure.

Table 3. Mean average of differences between the LFP values for <u>individual</u> essays at different essay lengths

3A: Differences (+/-) in LFP values for 220 word and 300 word Lengths

	1st 2000	Beyond 2000	AWL
Mean Average	1.5%	1.5%	1.0%
Maximum	4.8%	4.8%	3.6%

3B: Differences (+/-) in LFP values for 220 word and 400 word Lengths

	1st 2000	Beyond 2000	AWL
Mean Average	2.0%	2.0%	1.4%
Maximum	6.7%	6.7%	4.9%

Table 4. Mean average LFP values for <u>groups</u> of essays at different essay lengths

Three proficiency groups:

Group 1: TOEFL 450-577   (13 subjects)
Group 2: TOEFL 590-627   (16 subjects)
Group 3: TOEFL 630-650   (9 subjects)   + 9 native speakers

220 word essay length

Mean Avs.	1st 1000	2nd 1000	AWL	Nil	1st 2000	Beyond 2000
Group 1	76.8	7.6	11.3	4.3	84.4	15.6
SD	5.1	2.6	3.4	1.9	4.1	4.1
Group 2	73.7	8.0	11.9	6.4	81.7	18.2
SD	5.1	2.2	4.0	2.5	4.5	4.5
Group 3	73.8	7.1	12.6	6.5	80.9	19.1
SD	6.1	2.0	3.2	3.3	5.4	5.4

300 word essay length

Mean Avs.	1st 1000	2nd 1000	AWL	Nil	1st 2000	Beyond 2000
Group 1	76.3	7.6	11.6	4.5	84.0	16.0
SD	4.7	2.7	3.5	1.8	4.3	4.4
Group 2	72.7	8.5	11.9	7.0	81.1	18.9
SD	4.6	1.7	3.9	2.5	4.6	4.6
Group 3	72.7	6.9	12.9	7.4	79.6	20.3
SD	5.5	1.6	3.3	3.4	5.3	5.5

450 word essay length

Mean Avs.	1st 1000	2nd 1000	AWL	Nil	1st 2000	Beyond 2000
Group 1	-	-	-	-	-	-
Group 2	72.1	8.5	12.4	7.1	80.6	19.4
SD	4.5	1.6	4.1	2.1	4.5	4.5
Group 3	71.2	7.3	13.8	7.7	78.5	21.5
SD	4.5	1.5	2.8	2.9	4.5	4.5

Uses for the LFP

The conclusions reached about the validity, sensitivity, and stability of the LFP lead to suggestions about the sorts of research design for which the LFP can be used with some confidence. The LFP may be valid with research designs that:

1. focus on group differences rather than on individual differences unless the latter are large

2. select groups which have significantly different vocabulary sizes; low proficiency groups may be a partial exception to this rule

3. select low proficiency groups because their unsophisticated and small vocabularies suit the LFP

4. use a consistent essay extract length (300 words appears to be the most common choice)

Meara (forthcoming) makes much the same suggestion as #2 above but more forcefully by stating that the LFP might work reasonably well when group differences are very large.

Directions for future research

Future research into the development of effective measures of lexical richness in written text should aim to improve validity, sensitivity, and stability. Current research suggests that there are two main directions these efforts at improvement may take: (1) enhancements of the wordlist/frequency approach represented by the LFP or (2) enhancements of older measurement tools, most notably the type-token ratio (TTR) measurement tool. As

already pointed out, enabling LFP-type tools to include multi-word expressions in their analyses of texts is a precondition for the improvement of their validity. This is a major project in computational linguistics. Calzolari, Fillmore, Grishman and their colleagues have mapped out the necessary project stages (Calzolari et.al, 2002), but it will likely take several years before usable results materialize. Applied linguists are taking the lead in addressing the sensitivity and stability problems associated with existing measures of lexical richness. Malvern and Richards (1997) have suggested that sampling methods might reduce the severity of these and other problems. Meara and his colleagues at Swansea University have followed this lead by developing a pilot variant of the LFP called Plex, which uses the same basic wordlists as those found in the LFP but uses sampling techniques instead of cumulative counts to yield richness results. Meara claims that Plex values are stable to the extent that variations in text length appear not to affect the results significantly. Another benefit is a single measurement scale that appears to allow for large real variances (Meara, 2001). Malvern and Richards themselves have applied a modern sampling approach to the traditional TTR measurement tool. The new tool, which is called D, has shown that it is more stable in respect to variations in text length than was the case for the old TTR (Malvern & Richards, 2002). There is plenty of scope for second language researchers to adopt and adapt these experimental tools, which have so far been used mostly for studies concerning children's L1 development.

Conclusion

Researchers, syllabus designers, and teachers in English as an L2 stand to benefit substantially from the efforts which are

being made to improve lexical richness measurement tools. Syllabus designers will have access to increasingly precise tools which can measure the size and the variances in the productive vocabulary knowledge of groups of students. Teachers will be able to use these tools both for diagnostic purposes and to test the extent to which vocabulary instruction input is reflected in the learners' free written output. Probably the most significant benefit will be for researchers who are trying to develop theories of vocabulary development in L2 learners. The theories currently under development hinge on the modeling of the relationship between passive and active vocabulary knowledge, the latter of which is best represented in samples of subjects' free writing. Without tools that can accurately measure lexical richness in free writing, it will not be possible to assess the competing theories about the ways in which the passive-active knowledge relationship develops.

This paper has evaluated three claims regarding the validity, sensitivity and stability of a well-known measure of lexical richness in free writing, the Lexical Frequency Profile, which Laufer and Nation make and has concluded that each claim needs to be qualified. In 1995, the time had come for an advance in the measurement of lexical richness because the research community was asking questions about the lexical richness of learners' written output to which the existing measurement tools provided answers that were often irrelevant or unsatisfactory. Although there are good reasons to believe that the LFP is not equal to several of the claims made for it, its innovative design marks the beginning of advances in the measurement of lexical richness. Future progress will require the collaborative efforts of many researchers, all of whom have benefited and will benefit from the insights and the interest generated by the LFP.

A surprise encounter with Paul Meara at JALT 2004

Paul Meara has gained a reputation within the field of L2 lexical research for outstanding breadth and technical depth of knowledge and for a willingness to apply theory to practice in the form of a wide range of vocabulary tests and vocabulary measurement tools, developed with the help of colleagues at Swansea University's Centre for Applied Language Studies. I was curious about his views of the Lexical Frequency Profile (LFP), the subject of my JALT 2004 presentation, but had not summoned the confidence to contact him directly. I discovered one clue to his views just before the JALT 2004 conference. It was a hard-to-find page on the CALS web site comprising abstracts of papers delivered at a small conference held at Swansea University in 2003. One of these abstracts announced a paper by Paul Meara which would comprise a critique of the LFP. My heart immediately sank at the thought I had been beaten to the draw and, moreover, had been beaten by a leading authority on L2 vocabulary research. I then revived almost as quickly as my spirits had sunk when I read that his critique was to be based on computerized *simulations* of the LFP rather than on actual samples of the LFP in action. This raised the prospect that my paper might be a fortuitously timed complementary effort rather than a redundant one, and, in order to impress on my JALT audience that I was not alone in my views, I resolved to make a point of referring to Paul Meara as my "ally" during my JALT presentation.

About five minutes into my presentation, I made the reference to Paul Meara and added in an unscripted aside that I would like to meet him. As soon as these words were uttered, someone in the audience yelled, "He's over there!"

Not knowing what Paul looked like, I scanned the room and there in the corner was a man with his arm raised in a confident manner and with a big and very reassuring (for me) smile on his face. To my relief, he listened closely to my presentation and at the end encouraged me to continue with my line of research, promising to send me a proof copy of his forthcoming article, *Lexical Frequency Profiles: A Monte Carlo Analysis*. He kept this promise and the email connection I should have established months before was established by him at JALT 2004.

References

Calzolari, N., Fillmore, C., Grishman, R., Ide, N., Lenci, A., MacLeod, C., Zampolli, A. (2002). Towards best practice for multiword expressions in computational lexicons. In *Proceedings of the 3rd international conference on language resources and evaluation*, *Las Palmas, Canary Islands, Spain. 29-31 May, 2002* (pp. 1934-1940). Paris: European Language Resources Association.

Coniam, D. (1999). Second language proficiency and word frequency in English. *Asian Journal of English Language Teaching 9*, 59-74.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*, 213-238.

Erman, B. & Warren, B. (2000). The idiom principle and the open-choice principle. *Text 20*, 29-62.

Foster, P. (2001). Rules and routines: A consideration of their role in the task-based language production of native and non-native speakers. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching Pedagogic Tasks: Second Language Learning, Teaching, and Testing* (pp. 75-93). Harlow: Longman.

Laufer, B. (1994). The lexical profile of second language writing: does it change over time? *RELC Journal, 25*, 21-33.

Laufer, B. (1998). The development of passive and active vocabulary in a second language: Same or different? *Applied Linguistics, 19*, 255-271.

Laufer, B. & Nation, P. (1995). Vocabulary size and use: lexical richness in L2 written production. *Applied Linguistics, 19*, 255-271.

Malvern, D. & Richards,B. (1997). A new measure of lexical diversity. In A.Ryan & A.Wray (Eds.) *Evolving Models of Language: Papers from the Annual Meeting of tile BAAL held at the University of Wales, Swansea, September 1996* (pp. 58-71). Clevedon: Multilingual Matters.

Malvern, D. & Richards, B. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing, 19*(1), 85-104.

Meara, P. (2001). *Plex*. Textfile and software program downloaded September 21, 2004, from: http://www.swan.ac.uk/cals/calsres/free/free.htm

Meara, P. (forthcoming in 2005). Lexical Frequency Profiles: a Monte Carlo analysis. *Applied Linguistics, 26*, 1-16.

Meara, P., Lightbown, P. & Halter, R. (1997). Classrooms as lexical environments. *Language Teaching Research*, *1*, 32-37.

Morris, L & Cobb, T. (2004). Vocabulary Profiles as predictors of the academic performance of Teaching English as a Second Language trainees. *System, 32*, 75-87.

JALT 2004 NARA — Language Learning for Life

Muncie, J. (2002). Process writing and vocabulary development: comparing Lexical Frequency Profiles across drafts. *System, 30*, 225-235.

Nation, P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, P. & Heatley, A. (1996). *VocabProfile*, *Word and Range: programs for processing text*. LALS, Victoria University of Wellington. Available online at: http://www.vuw.ac.nz/lals/staff/paul_nation/index.html

Wiktorsson, M. (2001). Register Differences between Prefabs in Native and EFL English. In S. Manninen & C. Paradis (Eds.) *The Department of English in Lund: Working Papers in Linguistics* (vol 1) (pp. 85-94). Lund: Lund University. Retrieved June 24, 2004, from: http://www.englund.lu.se/research/workingpapers/pdf/Maria.pdf.