ReCALL **14** (1): 133–145. © 2002 Cambridge University Press
DOI: 10.1017/S0958344002001118 Printed in the United Kingdom

Assessing learners' writing using lexical frequency

ROBIN GOODFELLOW
Institute of Educational Technology, Open University, Milton Keynes MK7 6AA, UK
(*email: r.goodfellow@open.ac.uk*)

MARIE-NOËLLE LAMY
Faculty of Education and Language Studies, Open University, Milton Keynes MK7 6AA, UK
(*email: m.n.lamy@open.ac.uk*)

GLYN JONES
City and Guilds International, 1 Giltspur Street, London EC1 9DD
(*email: glynj@city-and-guilds.co.uk*)

Abstract

In this work we set out to investigate the feasibility of applying measures of lexical frequency to the assessment of the writing of learners of French. A system developed for analysing the lexical knowledge of learners, according to their productive use of high and low frequency words (Laufer and Nation 1995), was adapted for French and used to analyse learners' texts from an Open University French course. Whilst we found that this analysis could not be said to reflect the state of the learners' vocabulary knowledge in the same way that Laufer and Nation's study did, elements of the system's output did correlate significantly with scores awarded by human markers for vocabulary use in these texts. This suggests that the approach could be used for self-assessment. However, the feedback that can be given to learners on the basis of the current analysis is very limited. Nevertheless, the approach has the potential for refinement and when enhanced with information derived from successive cohorts of learners performing similar writing tasks, could be a first step in the development of a viable aid for learners evaluating their own writing.

1 Automatic text analysis

Technologies for giving automatic feedback to learners are of particular interest to providers of large scale language courses at a distance, such as the Open University. It is seen a way to support students' motivation and the development of study skills, without increasing the workload on tutors or the financial cost to the institution. Unfortunately, the quality of currently available automatic feedback on written language use is rather low, due to the technical difficulty of text analysis, and a lack of conviction in the CALL

https://doi.org/10.1017/S0958344002001118 Published online by Cambridge University Press

community of its usefulness! (Ironically, the level of work going on in the technically even more difficult area of speech analysis, is much higher (see Ehsani and Knodt, 1998).

The most commonly-used applications of text analysis are spell-checking, grammar-checking, and style and usage checking. These are now widely available and are useful tools. They are, however, very limited in their suitability for automatic feedback, because they focus on short (word or phrase) segments of language, which are analysed in isolation. Pennington (1992) has criticised the feedback such tools give as 'out of context', and 'arbitrary' in their decisions about style and readability. As a way of distinguishing between levels of language knowledge or competency they are even less appropriate. Similar criticisms can be levelled at syntactical parsers, although they have interested CALL researchers for some time. Whilst many interesting and ingenious prototypes have been developed, mainly in languages with a high degree of regularity, such as German, their application has remained focused on the analysis of individual errors, and their use restricted to the research lab (e.g. Vandeventer, 2001). Parsing free text for meaningful feedback is still a very hard problem, and more recent developments in automatic text analysis have tended to look at more statistical approaches. An example is Latent Semantic Analysis (see Folz *et al.*, 1999). This is a method of comparing texts with 'models' representing the genre they belong to. Whilst it has proved a feasible way of automatically spotting non-standard student writing in subject areas such as psychology, it has not yet been applied to language learners' texts, nor to the provision of feedback to the students themselves, rather than to the people who are marking the texts. In any case it is computationally quite complex and would demand a greater level of resource to investigate than is available to most university language departments.

It seems clear that technologies for providing feedback on accuracy or meaning in learners' texts are complex, and still fall far short of what is needed if those learners are to make use of the feedback to improve their writing. Our approach in the work described here has been to focus our attention instead on the area of vocabulary – the individual lexical items that learners use – and to take a relatively simple process of automatic analysis which has been shown to be a reliable measure of knowledge in one context, and try to adapt it to the requirements of assessment in another. The process is known as the Lexical Frequency Profile (LFP).

2 Lexical Frequency Profile (LFP)

79.9% of written English uses only the first 2000 most-frequent words in the language (Laufer, 1999). Knowledge of these 2000 most-frequent words plus the 570 most-frequent 'academic' words is considered 'critical for academic success' (Beglar, 1999). Because most L2 words are learned incidentally (i.e. through reading and listening rather than through specific vocabulary-learning exercises) we can assume that a learner's vocabulary builds up in layers made up of words having similar frequencies. We could expect vocabulary knowledge at an early stage of development to consist mainly of high frequency words, and at a later stage to have a higher proportion of low frequency words.

The lexical frequency profile method of assessing vocabulary knowledge by analysing learners' texts was developed by Laufer and Nation (1995). They developed a procedure

https://doi.org/10.1017/S0958344002001118 Published online by Cambridge University Press

which categorises the words in a learner's text according to which frequency band each word belongs to: first 1000 most-frequent, second 1000 most-frequent, 570 most-frequent 'academic' words. Academic words are defined in the University Word List – 836 word families containing vocabulary that is not in the first 2000 words of English, but which is frequent and has a wide range across a variety of written academic texts from a variety of disciplines (Xue and Nation, 1984). They called this analysis the lexical frequency profile (LFP) of the text. The LFP analyser program, (now renamed RANGE), can be downloaded from Paul Nation's web site: http://www.vuw.ac.nz/lals/staff/paul_nation/index.html. The program shows the numbers and percentages of words and word families in a target English text coming from each of the three word lists, plus those which are not recognised (see Table 1).

In Table 1, all the words in a sample text have been classified into categories of frequency (word list one is the first 1000 most-frequent words in English, column B row 2 shows the number and percentage of words in the text that come from that list etc.). The program has also performed a type and token analysis. A token is any occurrence of a word form in the text, regardless of whether it is occurring for the first or the nth time. A type is any word form which occurs once, regardless of how many more times it might occur. Both numbers and percentages of occurrences are given. A word family is the base form of a word, such as might appear as a headword in a dictionary, plus the derived and inflected forms of it. For the LFP word families Laufer and Nation included all the derivations and inflections listed at 'level 3' of the scale devised in Bauer and Nation (1993), that is: *-able, -er, -ish, -less, -ly, -ness, -th, -y, non-*, and *un-*. The LFP program is not able to classify any words that do not appear in the three frequency lists into their word families (hence the question marks in the fifth row of column D).

Laufer and Nation showed that the LFP measure of learners' texts can be compared with scores that the same learners achieve on standard vocabulary tests. They found that there is a correlation between performance on vocabulary tests and the proportions of low and high-frequency words in the free-written texts. They give the following results for correlation between the use that their English learners at the University of Haifa made of high and low frequency word families, and their scores in a vocabulary-based 'levels' test (see Table 2).

In Table 2, the negative correlations at the bottom of column B show that learners who used higher proportions of high-frequency words in their texts scored lower in the vocabulary test, and vice versa. The positive correlations in column D show that learners who used higher proportions of academic words in their text also scored higher in the vocabulary test. Similarly for column E, which deals with words that were not in the first three

Table 1. *Sample output from the LFP program*

A. Word List	B. Types/%	C. Tokens/%	D. Families
one	54/72.0	34/69.4	33
two	2/ 2.7	2/ 4.1	2
three	14/18.7	9/18.4	9
not in the lists	5/ 6.7	4/ 8.2	?????
Total	75	49	44

https://doi.org/10.1017/S0958344002001118 Published online by Cambridge University Press

Table 2. *% of word families from each frequency band correlated against level test scores (N = 65). (Data re-presented from Laufer and Nation, 1995, op cit., p.317)*

A	B		C		D		E	
	% 1st 1000 (high frequency) word families		% 2nd 1000 (medium frequency) word families		% Academic (low frequency) word families		% word families not in the other 3 lists (low frequency)	
	Text1	Text2	Text1	Text2	Text1	Text2	Text1	Text2
Levels Test/LFP	–0.7	–0.7	0.01	0.2	0.7	0.6	0.6	0.8
Significance	0.0001	0.0001	0.9	0.3	0.0001	0.001	0.0002	0.0001

lists and are therefore by definition low frequency. Laufer and Nation conclude that use of low frequency words is an indicator of richness in a learner's vocabulary, and recommend this procedure as a stable and reliable measure of lexical use in writing.

Laufer and Nation's test procedures were carefully controlled and their subjects were mainly students who had a common educational background in the Israeli school system. They were therefore able to make a strong case for the usefulness of the LFP measure for curriculum-design purposes. Our interest in it, on the other hand, was less experimental and more opportunistic, as we saw in it a potential source of automatic feedback to distance learners on the quality of the texts they submit for assessment. Whilst the LFP focuses only on vocabulary, we assumed that the learner's use of vocabulary would be an important determinant of the overall quality of their text (Laufer and Nation report in the same paper on two studies which found correlations between lexical measures and more holistic measures of quality in written text). If the LFP was capable of providing a reliable measure of the learner's lexical knowledge as reflected in a text, in the way that Laufer and Nation's study suggested, then we hypothesised that its analysis might bear some relation to the scores that human markers gave the same text, especially where they were marking specifically for vocabulary use. We saw in this hypothethised relation the potential to give a learner some indication of the kind of mark they might get for a free-writing assignment, before it was marked. Feedback of this type, we believed, would be useful in a formative way, giving the learner a focus for reflection on their work as well as an opportunity to improve it before submission. A study was set up to determine whether an LFP measure did in fact correspond to tutor marks for a group of assignments on one of the OU's French courses.

3 Comparing the LFP with tutor marks

The testbed we chose for the study was the OU Level 1 French Course L120. The reasons for choosing this course, given that the OU does not have an English programme which would have enabled us to use the Laufer and Nation system more or less exactly as they did, were:

- The level is appropriate (low intermediate)

https://doi.org/10.1017/S0958344002001118 Published online by Cambridge University Press

- The course has at least one tutor-marked assignment which is graded under four criteria, one of which is explicitly vocabulary-related.
- Because of the amount of work going on in French lexicography it is feasible that word frequency lists could be found or developed for this language.

Once the French LFP system was built we proposed to test it in two ways. Firstly by comparing its analysis of a number of L120 tutor-marked assignments (TMAs) with the marks given by the course tutors under the vocabulary criterion. Secondly, the system would be evaluated qualitatively by learners and teachers, to establish the optimal form in which feedback on a text should be given, in order to help a learner to benefit from it. In the event, the second part of the evaluation has not yet been carried out, and this paper focuses only on the results of the first.

3.1 Creating the word lists

In adapting the LFP program for French texts it was found necessary to create the French word-frequency lists from scratch, as no suitable equivalent already existed. The general lists (first 1000 and second 1000 most frequent words) were extracted from word lists developed and lemmatised (categorised into word families) at the Catholic University of Leuven, as part of the Eurolex project (Verlinde and Selva, 2001) from a corpus of texts from *Le Monde* and *Le Soir*. The academic list was extracted from the ELRA Parole French corpus (available for purchase from the European Language Resources Association at http://www.elda.fr), and lemmatised by one of the current authors. His report on some of the feasibility considerations relating to this work is available at http://iet.open.ac.uk/pp/r.goodfellow/ltic/report1.htm

3.2 The study - procedure

For the comparison we transcribed 36 student essays which had been submitted and marked during a recent presentation of the L120 course, and submitted them to the French LFP program for analysis. We then searched for correlations between key aspects of the lexical profile for each text, and the marks awarded for grammatical accuracy and vocabulary range The main differences between our procedure and Laufer and Nation's were as follows:

- They used specially-written texts – they were on 'essay/discussion' topics such as "Should a government be allowed to limit the number of children a family can have?" or "A person cannot be poor and happy..". Our project selected from texts submitted for assignment by learners on L120 – all were on the same topic, a 'journalistic' account of the life of fire-fighters in Quebec.
- Laufer and Nation had all their texts 'corrected' by hand prior to processing. Obviously incorrect words were deleted, misspelled words were corrected, proper nouns were deleted. We wanted to limit human intervention as far as possible, but on the assumption that learners able to use any feedback system based on this analysis would also be able to use a French spellchecker, all texts were spellchecked and where obvious corrections were suggested these were accepted,

https://doi.org/10.1017/S0958344002001118 Published online by Cambridge University Press

but where appropriate corrections were not obvious or not suggested the word was deleted. Two proper nouns that occurred in most of the texts were deleted.

- Laufer and Nation may have done some manual post-processing of the LFP output. This is not acknowledged in the 1995 paper, but can be inferred from the fact that they report figures which include word 'families' not found in the frequency lists. As the analyser is not able to categorise words which do not appear in the lists it is assumed that they had the 'not-in-a-list' words assigned to families manually. Our analysis does not use the category word 'family' for these unrecognised words, but instead uses word 'type'.
- Where they compared their LFP analysis of students' texts with results in vocabulary tests, we compared LFP analysis of the L120 student texts with the marks the tutors had given. Each tutor had given a mark out of 25 for each of four criteria: two 'content-related' criteria, one 'accuracy' criterion and one 'vocabulary range' criterion.

3.3 Discussion of first results

The initial comparison (see Table 3) did not produce the same kinds of correlation between the LFP analysis and the tutors' marks as Laufer and Nation found between LFP and vocabulary test scores.

In Table 3, the correlations are neither as strong as Laufer and Nation found, nor do they occur in the same areas of the data. Weak negative correlations ($p = 0.05$) exist between the use of high frequency word families and marks for range and accuracy (column B), where Laufer and Nation found strong ones, and there is no correlation at all between use of academic words (column D) or 'not-in-a-list' word types (column E) and the tutor marks. On the other hand, medium strength correlations ($p = 0.01$) were found between use of medium frequency word families (column C) and the range and accuracy marks, whereas Laufer and Nation found no correlation at this level of frequency.

These differences in strength between Laufer and Nation's correlations and ours might be explained by the less-controlled conditions of our study. The L120 adult distance learners are likely to have been more varied individually in age and background (43 of Laufer and Nation's subjects were recent graduates from the Israeli school system and had passed the same entrance exam). The tutors' marks against which the L120 LFP scores were correlated were produced by four different tutors and had not been standard-

Table 3. *% of word families from 3 frequency bands, and % of word types not in any list correlated against marks for vocabulary range and accuracy (N = 36)*

A	B	C	D	E
	% 1st 1000 word families	% 2nd 1000 word families	% Academic word families	% 'Not-in-list' word types
Range mark/LFP	–0.35	0.45	0.05	–0.06
Accuracy mark/LFP	–0.35	0.42	0.004	0.02

https://doi.org/10.1017/S0958344002001118 Published online by Cambridge University Press

ised (except implicitly via the fact that all tutors were experienced at marking assignments for that course). The contrast between Laufer and Nation's association between vocabulary knowledge and use of academic and low frequency words in the texts, and our finding that tutor marks correlated instead with learners' use of medium frequency words is a more important discrepancy, however, as it does not accord either with the theory that low frequency words are a product of a richer personal vocabulary, or the assumption that the tutor marks are equivalent to a standardised vocabulary test as an assessment of this knowledge. There are at least four possible ways to account for this mismatch.

First, the way we defined a word 'family' gives the first 1000 most-frequent words smaller coverage in our French lists than it has in Laufer and Nation's English ones. For example, in our lists the French equivalents for agree, agrees, agreed and agreeing all belong to the same family because they are parts of the same verb, but agreeable belongs to another family because it has a different meaning. Laufer and Nation applied a broader scheme in which these words would all be in the same family, together with agreement, disagreeable and several more. One result of this may be that a group of words which all appear under the same family in the first 1000 most-frequent list in Laufer and Nation's classification are actually split, in the French version, between the first 1000 and the second 1000 most-frequent lists. Thus some of the words our learners used that were classed as medium frequency might have been classed as belonging to high frequency families in Laufer and Nation's study.

Secondly, and following on from the above, it is possible that the L120 students were at a lower level in French than Laufer and Nation's subjects were in English, and that medium frequency was for them what low frequency was for the Haifa students. However, this begs the question why the use of 'not-in-a-list' low frequency words failed to discriminate amongst our learners in any way, as it did so decisively amongst Laufer and Nation's students. An analysis of two cases from the L120 group, the ones who got the 'top' and 'bottom' marks for vocabulary range, shows that they had similar numbers of words classified as not-in-list (24 for the top student, 20 for the bottom). If we take out from this list the words that were common, those that are English/French cognates, and those that are proper nouns or hyphenated words (which the LFP system does not recognise), the numbers are reduced to 14 and 10, a quantitative difference which is not enough to account for these learners' relative positions at top and bottom of the marks for vocabulary range. The most likely explanation for this failure of the 'low-frequency' words to discriminate between learners is that the first part of the assignment these students were writing was effectively a comprehension test from written and audio input. Students typically reproduce some of the vocabulary that is in the input when they are writing their answers. It is possible that this happened here, reducing the number of 'not-in-a-list' words which could be expected to discriminate between those who really knew the vocabulary and those who were reproducing what they had recently heard or read.

Thirdly, whilst the L120 tutors were allocating a mark for 'vocabulary range', they were in fact marking according to a criterion-referenced rubric that did not allow them to give credit for use of vocabulary which was not assumed by, or introduced in, the course itself. It is not possible to know how far they adhered to this stricture, however. Nevertheless, if the tutors were withholding recognition of low frequency words which were not introduced in the course then this would account for our failure to find a

https://doi.org/10.1017/S0958344002001118 Published online by Cambridge University Press

correlation between the use of these words and the marks awarded. This analysis would also account for the correlation we found between tutor marks and medium frequency words in the texts, if it were found to be the case that many of the lexical items introduced in the L120 course were in fact drawn from the medium frequency range.

The fourth reason for the difference between our results and Laufer and Nation's may be put down to the failure of the Academic word list to discriminate amongst the L120 learners, due to the genre of the writing task. The account of the life of a Québecois fire-fighter is essentially a non-discursive task, capable of a journalistic, almost conversational realisation. Such a text may have very little in common with forms of academic writing in content, intention or degree of abstraction, and therefore may draw on a very small subset of the academic word list. Certainly the overall percentages of words used from this list differs considerably between the English and French learners, with the Haifa students using up to 10% of academic words and the L120 learners only averaging 2.3%.

Whilst our initial results, therefore, do not reproduce Laufer and Nation's, they still suggest that a lexical frequency analysis can discriminate amongst L120 learners of French as assessed by their tutors, as it does between learners of English at Haifa University assessed through a vocabulary test, albeit in a different way, i.e. through their use of words classed as medium frequency. As the object of the research was to establish whether there are grounds for basing an assessment of the quality of a learner's text on the lexical frequency profile, it was considered worthwhile to see if there were other correlations between the LFP output and the human markers' scores. Having accepted that the different system we had used for classifying word families in French was at least partly responsible for the discrepancy between our study and Laufer and Nation's, we focused on looking for patterns in other areas of the data, such as the total numbers of word types and families used, rather than relative percentages.

3.4 Discussion of further results

A second round of analysis showed that the most significant measures of difference

Table 4. *Number of word types from three frequency bands, and types not in any list correlated against marks for range and accuracy*

A	B	C	D	E	F	G
	No. 1st 1000 word types	No. 2nd 1000 word types	No. Academic word types	No. 'Not-in-list' word types	Total word types	Total word families
	Text1	Text1	Text1	Text1		
Range mark/LFP	0.38	0.56	0.05	0.19	0.46	0.45
Accuracy mark/LFP	0.38	0.53	0.1	0.25	0.48	0.49
significance	p = 0.05	p = <0.001			p = <0.004	p = <0.005

https://doi.org/10.1017/S0958344002001118 Published online by Cambridge University Press

amongst the L120 learners are in fact those which relate to overall quantities of word types and families used in the texts. Table 4 shows that the learners who received the best marks for range and accuracy tended to be the ones who used the most word types from the medium frequency list (column C), and the most word types (column F) and word families (column G) overall.

In this table, the most significant correlations with tutor marks here occur over numbers of word types from the medium frequency list. This is a result that might be expected if, as we have supposed, the LFP measure does in principle reflect some aspect of text quality in terms of lexical richness, but, as we have suggested, our smaller word families and constraints on the use of low frequency and academic words have shifted the differentiating point from low frequency words to those in the medium frequency range. We believe, in other words, that both tutor marks and lexical frequency profile are identifying the learners who have written prolifically and with the most lexical diversity. Again, if it were possible to show that the L120 'course vocabulary' coincided to some extent with the list of the second 1000 most frequent words in general French, we might be able to make the claim that both tutors and lexical frequency profile are rating the learners' actual vocabulary knowledge, but we have not carried the investigation this far. Suffice it to say that, for the purpose of generating some form of feedback to enable students to reflect on the quality of what they have written, the French LFP analysis we have developed would seem to have something to offer. In the following section we briefly describe the way such feedback might work.

4 Feeding the LFP results back to the learner

The feedback system we envisaged was to be used by remote learners, automatically, so it was necessary to incorporate the LFP program into an application which could be accessed through a web browser. A prototype has been developed and is available for demonstration at http://iet.open.ac.uk/cgi-bin/vat/vat.html – as indicated earlier, it has yet to be evaluated with actual learners. An example of the prototype's feedback on a student text is given in appendix A.

The messages this prototype returns to the user are intended to give them an idea of the kind of mark they might get for a text if they submitted it for a given assignment, and also indicate where there is room for improvement in their vocabulary use. The LFP output is capable of indicating whereabouts in a ranking of student essays a particular text might fall, provided the system has knowledge of the way students have been ranked on previous similar tasks. L120 enrols approximately 1000 students each year and the course materials do not change substantially from year to year, with the assignments modified as little as possible, consistent with regulations for awarding credit. This means that an LFP analysis from a set of assignments in one year should still retain relevance to the assignments for the following year. The main factor likely to interfere with this reusability is the genre of the text required for the assignment, as the extent of the LFP analysis's genre-sensitivity is not clear. Whilst the analysis could be expected to differ substantially between an academic essay and a letter home, it is yet to be established whether, for example, presenting essentially the same information as an article in a popular magazine, or as a letter to a friend, would also produce a different lexical frequency profile. This prototype feedback has been based on the assumption that it would

https://doi.org/10.1017/S0958344002001118 Published online by Cambridge University Press

Table 5. *Predicting marks from the LFP analysis and sample averages*

Feedback principle (derived from the correlations)	Feedback to student (generated from the LFP output for the particular text, plus generalisations made from the whole set of marks)
High use of medium frequency words associates with high mark	*You used X word types from this frequency range...* *The average for this task =14* *If you used less than 9 then the maximum score you're likely to get is 57* *If you used less than 15 then the maximum score you're likely to get is 75* *If you used more than 20 then the minimum score you're likely to get is 79*
High proportion of high frequency word families associates with low mark	*You used X% word families from this frequency range...* *The average for this task =85%* *If you had more than 89% then your maximum score is likely to be 76* *If you had less than 81% then your maximum score is likely to be 79* *High total word families associates with high mark. You used X word families overall* *The average for this task= 112* *If you used less than 85 then your maximum score is likely to be 70*

not, and that the LFP-based feedback for a given assignment may be enhanced with information about student rankings and marks awarded for previous, similar assignments.

Table 5 shows how it is possible to put feedback principles derived from our correlations of LFP output with tutor marks, into the context of average numbers of word types used, and marks awarded, to produce a prediction of likely score for any given text that conforms to the 'Quebec fire-fighter' type of writing task.

These predictions may be integrated into a more contextualised discussion of general performance at the assignment, using, for example, figures such as type-token ratio (number of word types as a percentage of all word tokens) to indicate levels of repetition. To indicate directions of improvement of the text some more qualitative information may be included, such as particular words from the medium frequency list used by learners who have scored highly.

Whilst this example represents a way to give students meaningful feedback on free text automatically, it is clearly not yet of sufficient quality to engage the average learner for very long, nor does it offer them a clear model of how to improve their writing. Giving students feedback of a kind which will encourage them to revise a text is a matter not just of having something pedagogically valid to say about that text, but also of convincing the student that it is worth their time doing the revision. Whilst an estimation of a likely mark for an assignment is easy to understand, motivating, and can be used iteratively, the system as it stands does not address key aspects of writing quality, such as style, creativity, grammatical invention, etc. It is merely the first step in the development of a relatively low-tech approach to these larger issues, utilising a focus on vocabulary, a customised analysis, and attempting to exploit the information that

https://doi.org/10.1017/S0958344002001118 Published online by Cambridge University Press

successive cohorts of students generate as they tackle essentially similar tasks. These limited aims may eventually be developed to provide a more pedagogically rich level of feedback, integrating other sources of advice on writing, and the creative use of vocabulary.

5 Summary and conclusions

We have seen that a measure by frequency of a learner's use of words in a free-written text may correlate in some cases with both their performance on standard vocabulary tests, and the marks awarded by their tutors for vocabulary range. This has given us grounds to think that such a measure, which can be generated automatically, may serve as a basis for a CALL system which helps distance language students to assess their own work, prior to submitting it for human marking. The strength of the approach is that it introduces into the measurement of vocabulary richness the notion of word frequency. Effectively, learners take credit for using words that are relatively unusual, as determined by the objective measure of their ranking in a reliable frequency count for the language as a whole. However, we have noted that, for the purpose of giving meaningful feedback to learners, the way in which the LFP uses word frequency data is rather limited. The words in a learner's text are sorted into just four categories, and the importance accorded to academic vocabulary is problematic on two counts. Firstly, it raises the issue as to what constitutes academic vocabulary, and secondly, it limits the validity of the measure to assignments which are supposed to be academic, or at least compromises its validity for other kinds of written work. Whilst it is feasible, as long as we have data from successive cohorts of students performing essentially similar tasks, to contextualise the lexical frequency output in terms of the performance of the whole group, and thus enhance the feedback, it may be that the only way to make the analysis itself more meaningful is to adapt it to reflect the actual frequencies of individual words the student uses, rather than simply the broad band (first 1000, second 1000, etc.) into which they fall. A development of the analysis, for example, which measured the total numbers of words in a student's text as a proportion of the sum of frequency indices for each of those words, might have the following advantages:

- It would be highly discriminating amongst the words used. Each word has a discrete and finely graded effect on the overall result, rather than as a member of one of three or four broad bands. This would enable feedback to pick out individual items for comment.
- By assigning frequency values to individual word types it might capture the learner's range of grammatical as well as purely lexical resources. In French, difficult grammatical forms tend to be associated with low frequency graphical forms. For example, parlez, as in *Parlez-vous français*? is likely to have a much higher frequency index than *parlâtes* ("you spoke", formal past). Feedback could give the learner credit for using the latter form on account of its rarity.
- It would be possible to create a frequency profile curve showing how words from each of several narrow frequency bands are represented in a learner's writing.
- It would be possible to implement for other languages, as it avoids all issues of how to lemmatise word lists

https://doi.org/10.1017/S0958344002001118 Published online by Cambridge University Press

In speculating about such a development it is important to note that the focus has moved away from the original purpose of Laufer and Nation's research, which was to characterise lexical knowledge for the purposes of curriculum development, towards the more contingent goal of helping individual students to notice and develop features of their vocabulary use in writing. The next stage of our work must clearly be to expose the prototype system to learners and tutors in order to determine how the information given in lexical feedback could be integrated into meaningful activities intended to offer real possibilities for improvement of the texts in question. Only after that would we be justified in experimenting with the kinds of development of the lexical profiling approach that we have just discussed.

References

Bauer, L. and Nation, P. (1993) Word familes. *International Journal of Lexicography* **6**/3:253–279

Beglar, D. and Hunt, A (1999) Revising and validating the 2000 word level and university word level vocabulary tests. *Language Testing* **16**/2.

Ehsani, F. and Knodt, E. (1998) Speech Technology In Computer-Aided Language Learning: Strengths And Limitations Of A New Call Paradigm. *Language Learning and Technology* **2**/1, http://llt.msu.edu/vol2num1/article3/

Folz, F., Laham, D. and Landauer, T. (1999) The Intelligent Essay Assessor: Applications to Educational Technology. http://imej.wfu.edu/articles/1999/2/04/index.asp

Laufer, B. and Nation, P. (1995) Vocabulary size and use: lexical richness in L2 written production. *Applied Linguistics* **16**/3.

Laufer, B. and Nation, P. (1999) A vocabulary-size test of controlled productive ability. *Language Testing* **16**/1.

Pennington, M. (1992) Beyond off-the-shelf computer remedies for student writers. *System* **20**/4.

Vandeventer, A. (2001) Creating a grammar checker for CALL by constraint relaxation: a feasibility study. *ReCALL* **13**/1:110–120

Verlinde, S. and Selva, T. (2001): Corpus-based vs intuition-based lexicography: defining a word list for a French learners' dictionary. *Proceedings Corpus Linguistics 2001 Conference*, Lancaster, UK, pp. 594–598.

Xue Guoyi and Nation, P. (1984). A university word list. *Language Learning and Communication* **3**(2):215–229

Appendix A

Sample feedback from prototype (the numbers in bold are those which have been calculated from the lexical frequency profile output for the given text, the others are 'canned' text derived from overall averages for this task)

Profile Results

Your data has been accepted. The transaction number is 60; please quote this number in the event of a query.

Type:Token Ratio

According to the analysis there are **335** words used in this text, and the ratio of unique words (word types) to total-words-in-the-text (the type-token ratio) is **56**%.

The average length of text submitted for this assignment is usually about **385** words. Average

https://doi.org/10.1017/S0958344002001118 Published online by Cambridge University Press

type-token ratio is about **48**%. A lower-than-average type-token ratio for an average length text might indicate a lot of word repetition, which could get marked down. Type-token ratio usually falls as the text gets longer, the average for texts over 400 words long was 22%.

The top mark gained for this assignment was 96. The bottom was 41. Average was 77.

Word Types

In this text there are **190** separate word types used. Of these, **23** are word types which are found among the second one-thousand-most-frequent words in written French. These are considered medium-frequency words (the first one-thousand-most-frequent words are considered high-frequency – they include very frequent words such as 'le' and 'la', 'pour', 'avec', etc.) The average number of medium-frequency word types for this assignment is 14.

Knowledge of medium-frequency words could indicate a higher level of French vocabulary. Students whose texts for this assignment contained more than the average number of medium-frequency words generally scored much better. Here is an indication of the likely relation between the numbers of these words and the mark given for the TMA:

- Less than 9 medium-frequency words = maximum mark likely to be around 57
- Less than 15 medium-frequency words = maximum mark likely to be around 75
- More than 20 medium-frequency words = minimum mark likely to be around 79

Click here for an indication of some of the medium frequency words used by students who scored well in this assignment.

Word Families

The total number of word 'families' used in this text is **127**. A word family is the group of word types which are all derived from the same 'root' word. For example 'actuel', 'actuels', 'actuelle', and 'actuelles' all belong to the same word family 'actuel'.

Marks for this assignment tend to be higher the more different word families are used. The average number of families is around 112. If a text has less than about 85 word families, the mark it will get is unlikely to be more than 70.

The number of word families may be an indication of the richness of vocabulary in the text - a criterion which is often used by markers. High-frequency word families are the least rich, the average for this assignment is 85%. Texts having more than 89% high-frequency word families are unlikely to score more than 76. Texts having less than 81% high-frequency word families will be likely to get a mark of at least 79. Here is the distribution of high and medium frequency and 'academic' word families in this text:

- Percentage of high frequency word families: 78 %
- Percentage of medium frequency word families: 17 %
- Percentage of low frequency 'academic' word families: 3 %

If you think there may be scope to improve your text before submitting it for marking, we suggest you begin by finding ways to increase the variety and number of different word families you use, starting with those which are found in the medium-frequency list.

https://doi.org/10.1017/S0958344002001118 Published online by Cambridge University Press