Malaysian Journal Of ELT Research
ISSN: 1511-8002
Vol. 6, 2010

The Vocabulary Performance of Native and Non-Native Speakers and Its Relationship to Learning Style

Paul Booth

Kingston University
UK

Abstract

One of the ways in which non-native speakers differ from native speakers is that the former tend to be more unpredictable in their use of vocabulary. This study explores how the learning style of second language learners may be associated with this variability. The participants were first year university students of engineering. Vocabulary from their report writing was analysed for the amount of lexical recycling which occurs in their texts. This was done by using a measure of lexical diversity which calculates a mathematical curve fitting procedure to model the fall of the type-token ratio in order to give a value, parameter *D*. Learning style was based on Skehan's (1998) memory-analysis framework. The participants were tested for associative memory of unfamiliar words and grammatical sensitivity of an unknown language. The results suggest that L2 learners who show extreme lexical diversity (either high recycling or low recycling of vocabulary), are related to low grammatical awareness and high associative memory. The implication for teaching and learning is that although non-native speakers of English may use technical vocabulary in a quantitatively similar manner to native speakers, some may need to notice the complexity and coherence in L2.

Introduction

This paper reports on a study which looks at whether the vocabulary used by non-native speakers is similar to native speakers in relation to technical writing. Specifically, it looks at how learning style can determine, to some extent, the vocabulary profiles of second language learners. This raises implications for teaching in that learners have diverse predispositions in their adoption of particular learning strategies. The paper goes on to discuss how different learning style predispositions can have different payoffs in terms of the complexity of how words are put together and shows how teachers can foster restructuring in second language (L2) English by learners who may not be inclined to do so.

Productive vocabulary can be highly heterogeneous for both native (NS) and non-native speakers (NNS) in which there can be more variability within groups than between groups so learning style will be examined to help understand this variability. One reason for this variability is that it is clear from the literature that there is interplay between grammatical and lexical complexity (Bell 2009, p. 126). As this study will show, although L2 texts may contain similar academic or technical vocabulary as L1 texts, the similarity ends there. The way less frequent words are used seems to be highly variable from learner to learner. It could be that learners may have internalized more complex technical vocabulary but do not have the language resources to use the lexis in a coherent manner. Skehan (2009) found that, in the majority of cases, the more learners recycle vocabulary in speaking the greater the complexity as measured through an index of subordination (p. 117). The cost of technical vocabulary may mean a greater demand placed on the learner to recycle other lexis to achieve coherence in discourse.

Lexical diversity

One way of studying lexical recycling is to measure the lexical diversity of texts. Lexical diversity is a measure of the type-token[i] relationship within a text. It gives an indication of the amount of lexical recycling in a set of words. The greater the recycling of tokens of the same type the lower the diversity within a text. Jarvis (2002, p. 73) found that although the learner groups with the least formal instruction in English showed the lowest levels of lexical diversity and the learner group with the most formal instruction showed the highest level of diversity, differences between the learners and native speakers were not at all straightforward. For example, only some non-native speaker groups showed differences to native speaker groups. Within the native speaker groups, US fifth grade native speakers attained higher levels of diversity than seventh graders. Interestingly, Jarvis found a curvilinear relationship between lexical diversity and holistic quality ratings of texts. In other words, the highest diversity exhibited the lowest correlation to writing quality. A lack of repetition, i.e. high lexical diversity, precludes the amount of lexical repetition which is necessary for coherence in discourse. Therefore a certain amount of lexical recycling is necessary to retain quality in discourse and this may be particularly pertinent in technical writing.

To measure the lexical diversity, i.e. type-token ratio (TTR) of the texts, Malvern, Richards, Chipere, & Durán (2004) devised a method of measurement called parameter *D.* Basically, parameter *D* is a measure which overcomes one of the serious problems recently signalled in TTR, i.e. the falling type-token ratio that occurs. The longer the text is the more words are used, and so the greater the chance of repetition of the same word. The *D* statistic is a single parameter of a mathematical curve that models the falling TTR and allows researchers to compare texts of different word lengths. Moreover, this measure of lexical sophistication is not influenced by the repetition of a relatively few rare words which is an important factor because the texts used in this study contain technical jargon that would skew the results if the measure was based on lexical rarity alone.

Learning style

One way of understanding the individual differences in lexical diversity is to examine the learning styles of native and non-native speakers. Although the term is discussed in the literature there is a confusing array of learning style taxonomies that can leave the reader confused as to what this term actually means. Das (1988, p. 101) defines learning style as a predisposition to adopt a particular learning strategy. However, there seems to be a scarcity of information about the reliability and validity of the various taxonomies. The learning style framework used in this study is grounded in the field of linguistics and the theory behind the framework is based on Skehan's cognitive theory of L2 learning and processing. Skehan (1998) proposes a dual-coding approach to language learning and performance, which is made up of a memory-based system and a rule-based system. According to this framework, high memory learners store a wide range of lexicalised exemplars which are mobilised for communication in real time. High analysis learners develop an organised, rule-based representation of language which is more open to restructuring, and so sensitive to feedback, and who would value precision in their meanings. This dual-coding emerged from L2 learning profiles taken from cluster analysis of aptitude test scores. These two approaches to L2 processing, exemplar-based (memory) and rule-based (analysis), are particularly relevant for this study because learners who are able to analyse language may have an advantage over memory orientated learners since technical vocabulary with its relatively complex morphological structure may demand more analysis than simply learning by rote memory.

The aim of this study then is to examine lexical diversity profiles in relation to learning style to determine whether learning style is similarly associated with any patterns of variability in native or non-native speaker profiles. If learning style is associated with variability in lexical production it could give us insights into the learning of L2 lexis and ultimately the teaching of lexis.

Research questions

To examine lexical diversity in relation to learning style and to determine whether learning style is associated with patterns of variability of native speakers and learners lexical profiles, the following questions need to be answered:

1. Do native and non-native speakers matched for learning style profile produce similar patterns of lexical diversity?
2. Is greater variability in lexical profiles associated more with non-native speakers than native speakers?

This research concerns itself with whether native speakers' lexical profiles as measured by the amount of lexical recycling (diversity) mirror non-native speakers' profiles when both groups are each sub-divided into learning style strengths. In doing so it will examine whether the variability (i.e. the unpredictability) is more inherent with non-native speakers than native speakers. The results will then be discussed in relation to teaching and learning and will explore the possibility of whether learners who do not analyse language may use quantitatively similar levels of technical vocabulary to those who do but at the expense of complexity and coherence in their text.

The Study

The participants

The participants were first year university students of Civil Engineering. The group was comprised of 20 native speakers and 16 non-native speakers from a wide variety of first language backgrounds (see Appendix 1). Males (N = 28) outnumbered the females (N = 8) by about three to one. The average age was 22 years old (the ages ranged from 18 to 25 years old). An on-line test was used (Oxford University Language Center Placement Test) so as to get a rough idea of NS and NNS language proficiency. The test is a 50-question, multiple-choice format designed to test English grammar by recognition of the grammatically correct words or appropriate phrases to complete a sentence. Participants were required to select the correct answer out of five choices to obtain a score out of 50. Although this test has no external validity, it was intended simply to compare native and non-native speakers' English proficiency in this study. The mean scores[ii] of the native and non-native speakers are shown in Table 1 below.

Table 1

Level test

	N	Mean	Std. Deviation	T-Test	Sig. (2-tailed)
Native speaker	20	42.90	7.70	1.00	.321
Non-native speaker	15	40.67	4.35		

These scores indicate that there is little difference in proficiency as measured by this grammar test between native and non-native speakers. An independent samples t-test was used to compare the means which indicated that they were not significantly different. Interestingly, there is a greater standard deviation for the native-speaker group. Within the native and non-native groups there was a range of proficiency levels with some L2 learners scoring higher on the test than native speakers.

Data collection

Two pieces of writing from each participant were collected over a period of one month to help ensure reliability. The texts used were part of the first year civil engineering students' assignment for a module in electrical engineering, electronics and computing. The reports were written to describe laboratory demonstrations. The first report covered a selection of industrial heating demonstrations; the second covered different types of heating applications. The total number of participants dropped from 36 to 33 for the second piece of writing. Longer texts are the source of lexical data for this study because a single short piece of writing may be atypical of what the learner knows and can do in other contexts (see Daller *et al*. 2007, p. 12). The participants were informed that marks would be given for neat layout, coherent presentation of technical information and selection of material. Because the reports were written in the students' own time, the reports were hand-written to encourage the participants to use their own words and so avoid cutting and pasting from various on-line sources.

As part of the background questionnaire, participants were asked how much they relied on external sources (i.e. dictionaries, translators and thesauruses). Figure 1 shows the number of students and the frequency of external source use in the form of a bar chart.

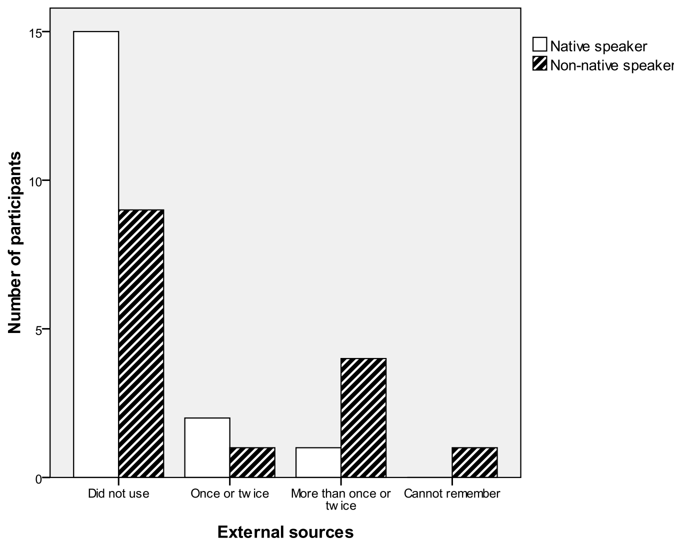

Figure 1: Use of external sources

Figure 1 shows, predictably, native speakers relied less on external sources than non-native speakers. About half the non-native speakers, however, reported not using any type of reference source which is surprising when we consider the highly specific subject matter of these technical reports and the complex language needed to explain the processes.

The reports were then transcribed for the software to analyse the lexis. So as not to skew the findings, the following were not inputted for analysis: formulae, symbols, tables, lists and direct quotes. Hyphenated words (e.g. nickel-chromium) and phrasal verbs (e.g. made up) were treated as one word. Non-words (e.g. lossiness) were deleted but spelling mistakes were corrected. In this study, Meara and Miralpeix's *D-Tools* (2007) was used to calculate the *D* value. The value of the *D* statistic is labelled as *D*1 for the first set of reports, and the second set is labelled *D*2.

All participants were tested for learning style by using two tests of language aptitude (Meara *et al.* 2001). The associative memory test (LAT B) is a measure of the ability to remember pairs of words shown visually whereas the analysis test (LAT C) is a measure of the ability to infer rules from examples of a fictitious language. Both tests use a fictional language so as not to discriminate against non-native speakers.

Results

As can be seen in Table 2 below, the differences in mean number of tokens of the native and non-native speakers is relatively small even over two different texts; however, the standard deviation for the native speakers is greater than for the non-native speakers suggesting greater variability for this group.

Table 2

Tokens NS and NNS

Tokens	Native and non-native speakers	N	Mean	Std. Deviation	Std. Error Mean
Report 1	Native speaker	20	716.95	232.71	52.04
	Non-native speaker	16	621.06	168.96	42.24
Report 2	Native speaker	18	645.89	188.58	44.45
	Non-native speaker	15	622.00	130.17	33.61

Table 3 shows the differences between the two native and non-native speakers' mean lexical diversity (*D*) scores for the first and second texts. An independent samples t-test was carried out and found that none of the differences were significant but the standard deviation for the non-native speakers is higher in both texts.

Table 3

Mean lexical diversity (D) NS and NNS

Diversity (*D*)	Native and non-native speakers	N	Mean (*D*)	Std. Deviation	Std. Error Mean	T-test	Sig (2-tailed)
*D*1	Native speaker	20	78.76	12.89	2.88	.217	.829
	Non-native speaker	16	77.65	17.86	4.46		
*D*2	Native speaker	18	68.88	7.53	1.78	-1.66	.106
	Non-native speaker	15	77.93	21.58	5.57		

The native speakers' second text showed a lower mean *D* statistic (M = 68.88; SD = 7.53) than non-native speakers (M = 77.93; SD = 21.58). Notice too that the standard deviation for the second text is nearly three times higher for the non-native speakers.

Within text 1 and 2 Memory and Analysis

The data were then analysed using "box-and-whisker plots". This approach was used because it shows variation both within one set of data and across different sets of data. Recall that the aim of this research is to find any patterns, i.e. whether different groups of learners matched for learning style strengths and weaknesses show consistency or variation in lexical diversity. The box-and-whisker plots are similar to a vertical bar chart in that the vertical axis shows the measured quantity (i.e. *D* values). The upper edge of a particular box is the 75th percentile[iii] of the distribution for a single group and the lower edge is the 25th, so 50% of the measured *D* values for that group lie between the 25th and 75th percentiles. The vertical lines (whiskers) outside the box show the largest and smallest *D* values measured for that group. The line across the middle of the box represents the median for that group (Dunleavy 2003, p. 188).

The participants were grouped firstly according to their scores in LAT B (Memory) and then in LAT C (Analysis). The grouping, bottom, middle and top is based on Meara *et al.*'s (2001) normative data for the percentage of learners who fall into the bottom 30%, middle 40%, and top 30% (see Table 4 below).

Table 4

LAT B Memory and Analysis scores and their interpretation

LAT B (Memory)		LAT C (Analysis)	
74-100	Top 30% of all scores	70-100	Top 30% of all scores
43-73	Middle 40% of all scores	60-69	Middle 40% of all scores
0-42	Bottom 30% of all scores	0-59	Bottom 30% of all scores

It is clear that Memory and Analysis are factors which relate to lexical profiles in combination and not as separate factors because people have both. In this set of data though, the participants are not grouped according to combined strengths and weaknesses in Memory and Analysis because it would make the data sets too small if they were then further sub-divided into native and non-native speakers. Instead the participants were grouped according to Memory or Analysis strengths, and then native or non-native speaker background.

Figures 2 and 3 show the box-plots of the parameter *D* statistic for two sets of texts from native and non-native speakers. With non-native speakers we can see a fairly wide distribution of scores (*D*). In the second set of data we see that the higher the Memory (LAT B) score the greater the variability in lexical diversity (*D*). It is difficult to see any patterns in the median *D* values across the sub-groups from each data set. Intuitively, one would expect a higher *D* statistic for the high Memory group but this is not the case in both sets of data. Higher Memory scores tend to be associated with greater lexical diversity for non-native speakers but this is not linearly related. In other words, there seems to be no direct relationship between higher Memory scores and greater diversity.

With the native-speaker lexical output, there seems to be no coherent pattern emerging. A greater spread of *D* values does not seem to be associated with higher Memory scores across both data sets. The native speaker box-plots show considerably less variation than non-native speaker box-plots across the different sub-sets of Memory. Interestingly, the second data set (*D*2) shows much less variation for the native-speakers than the first data set. In the second data set there are two native speaker outliers who are more than 1.5 box-lengths above the box.

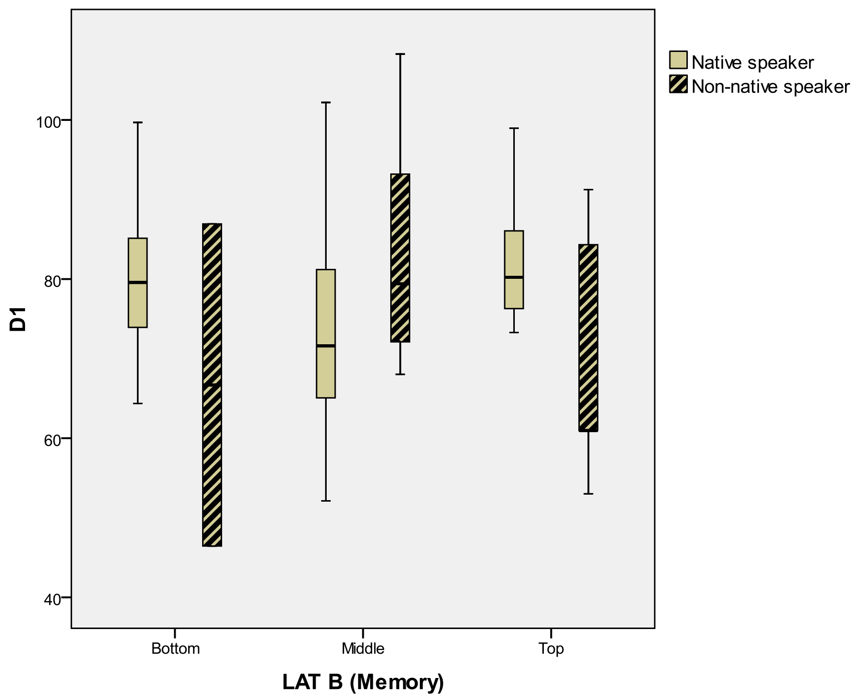

Figure 2: Lexical diversity (*D*1) and LAT B native and non-native speakers

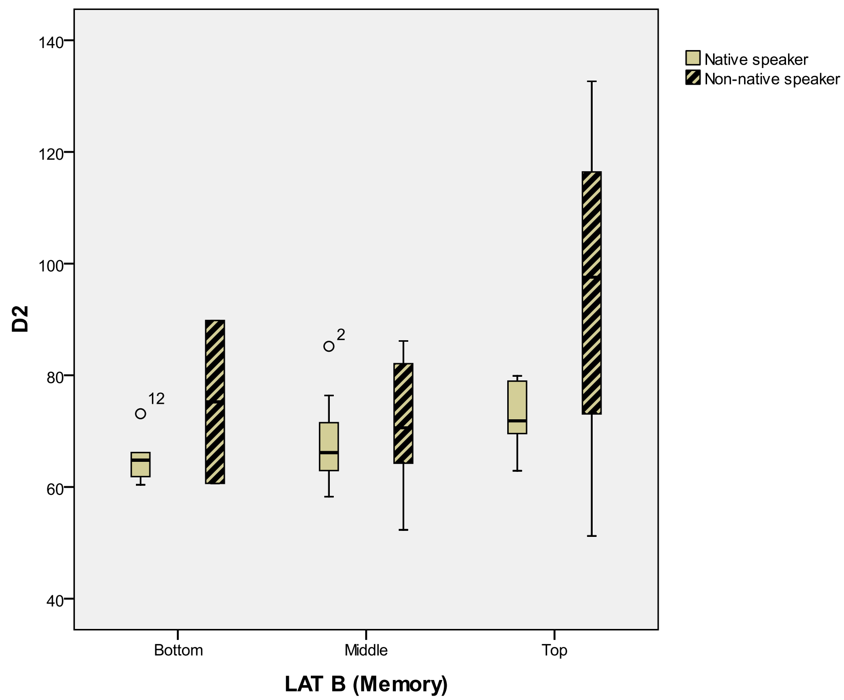

Figure 3: Lexical diversity (*D*2) and LAT B native and non-native speakers

Table 5 below displays the number of participants in each sub-group.

Table 5

LAT B group numbers: *D*1 and *D*2

Memory		*D*1 N	*D*2 N
Bottom	Native speaker	7	5
	Non-native speaker	2	2

Middle	Native speaker	7	7
	Non-native speaker	9	9
Top	Native speaker	6	6
	Non-native speaker	5	4

The next set of data is from learners and native speakers grouped according to their Analysis scores. Figures 4 and 5 display the box-plots of the participants grouped according to their Analysis scores (LAT C). The pattern for the non-native speakers in Figure 4 suggests that the greater the Analysis score (LAT C) the more tightly clustered the *D* value. The second set of data in Figure 5 suggests that the greater the Analysis score the lower the median *D* (diversity) value. There are extreme high and low *D* scores for the learners in the bottom LAT C in both sets of data. Taking these two datasets together for non-native speakers, higher Analysis scores tend to be associated with tighter clustering and lower D values. Native speakers, on the other hand, do not mirror this pattern. In fact, as in the previous Memory data set, it is difficult to see any coherent pattern for native-speakers grouped according to their LAT C score. In the first native speaker data set, there is an outlier and an extreme case, which is more than 3 box-lengths above the box, which could be because of the tight clustering of participants in the bottom Analysis group. In this dataset both native and non-native speakers show a wide distribution of lexical diversity scores when classified as bottom in Analysis.

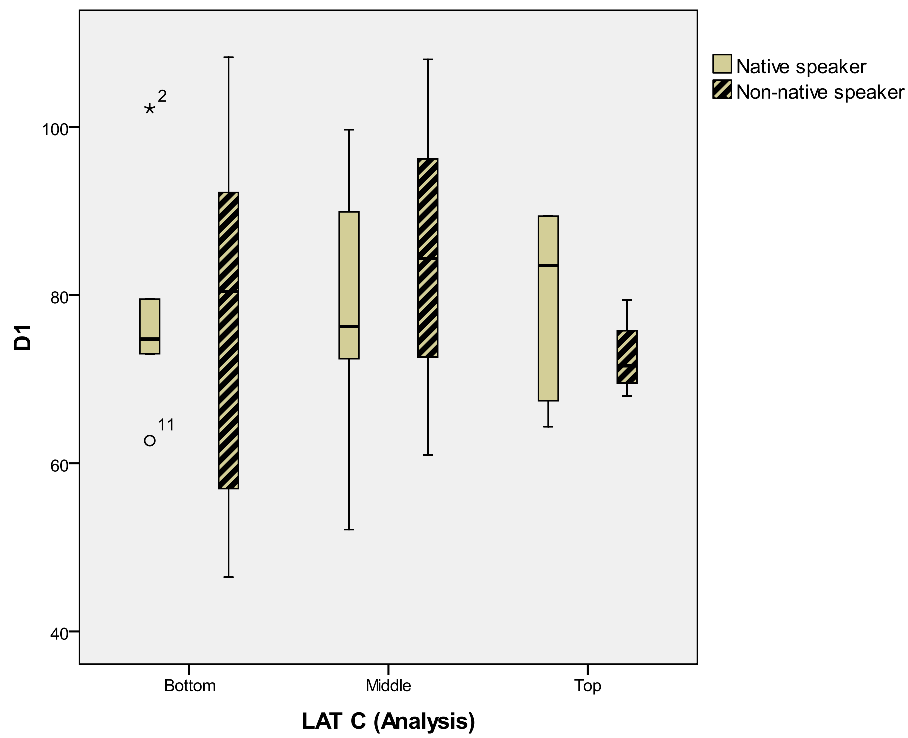

Figure 4: Lexical diversity (*D*1) and LAT C groups: native and non-native speakers

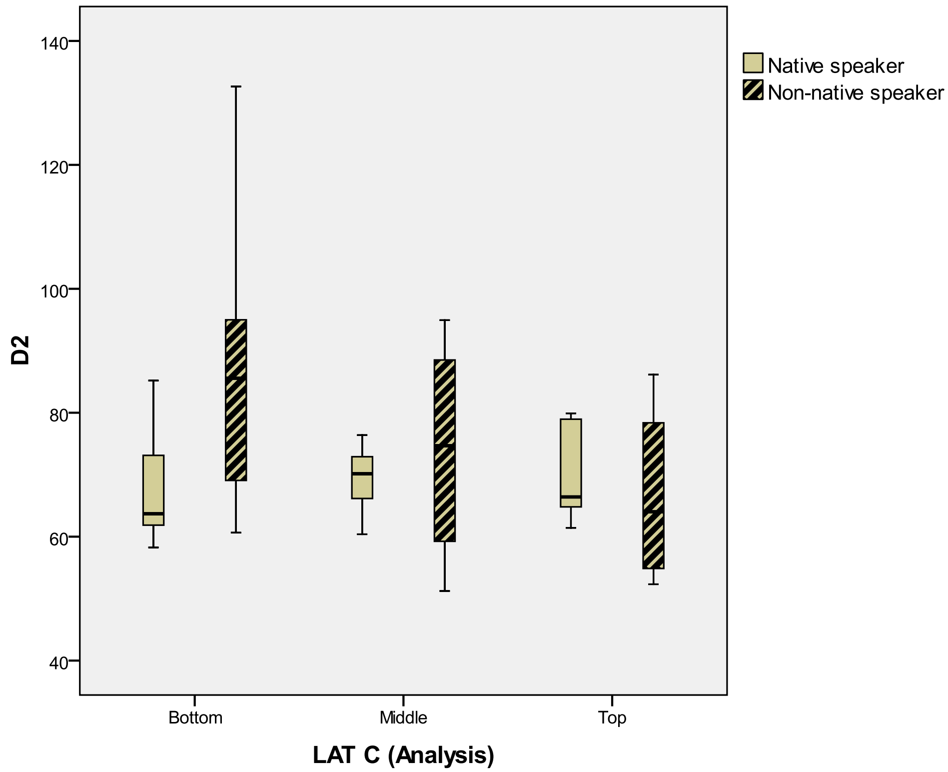

Figure 5: Lexical diversity (*D*2) and LAT C groups: native and non-native speakers

Table 6

LAT C group numbers: *D*1 and *D*2

Analysis	Native and non-native English	*D*1 N	*D*2 N
Bottom	Native speaker	7	6
	Non-native speaker	8	7

Middle	Native speaker	7	6
	Non-native speaker	4	4
Top	Native speaker	6	6
	Non-native speaker	4	4

Although the numbers of participants in the sub-groups are very small (see Tables 5 and 6), the same patterns appear across both sets of data, i.e. with reports 1 and 2. Memory tends to be associated with variability in lexical diversity, whereas Analysis tends to be associated with much less variability and lower lexical diversity. These patterns are evident with non-native speakers but not native speakers. Nevertheless native speaker lexical diversity is variable. This section of the results has examined native and non-native speakers' lexis from each of the texts (Reports 1 and 2) in isolation. The next section examines the learners' diversity (*D*) and learning style but this time across the two texts (Reports 1 and 2) to find any intra-variability and stability.

In Figure 6 below, when non-native speakers are classified according to their Memory scores, their lexical diversity tends to be erratic from Text 1 to Text 2. The jump in diversity from Text 1 to 2 for the top Memory group is relatively high compared to the other two sub groups.

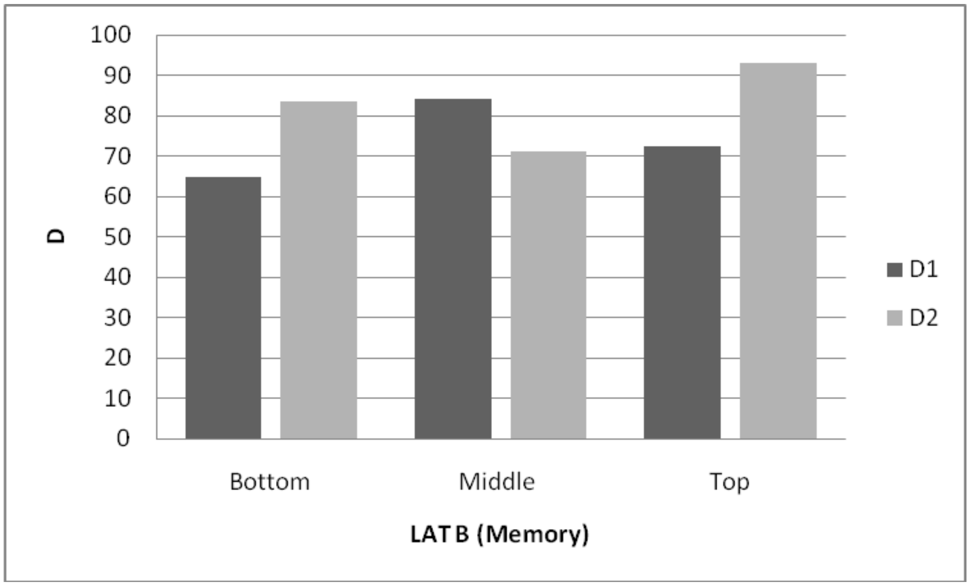

Figure 6: Non-native speakers: Mean diversity (*D*1 & *D*2) and LAT B (Memory)

In Figure 7 below, when we look at learners classified by Analysis scores then we can see some pattern emerging; the higher the Analysis, the lower the diversity. The middle and top Analysis groups both show a lower diversity for the second text compared to the first, whilst the bottom group does not. It could be that the relationship between Analysis and diversity is nonlinear: the lower the Analysis, the less the relationship with diversity.

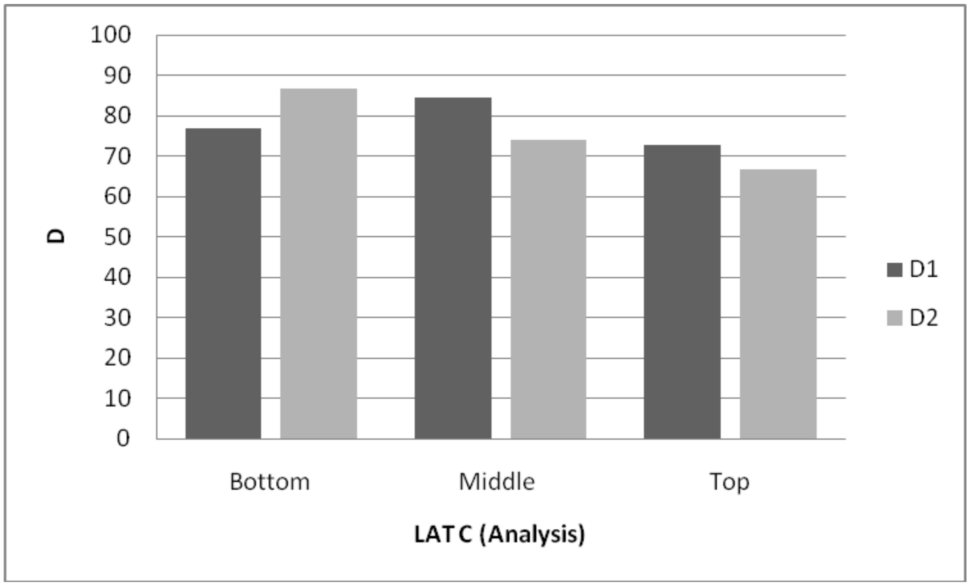

Figure 7: Non-native speakers: Mean diversity (*D*1 & *D*2) and LAT C (Analysis)

Discussion

The aim of this study was to examine the lexical diversity of native and non-native speakers and the relationship with learning style. For the non-natives the distribution of lexical diversity (*D*) values appears to be tightly clustered for analysis-orientated learners. Memory-orientated learners, on the other hand, display a wider distribution of *D* values. It should be noted that lexical repetition in texts is necessary for coherence and that extremely high lexical diversity would preclude the amount of lexical recycling necessary to obtain coherence. Therefore extreme diversity scores would indicate a lack of lexical complexity in a text. Native speakers, in contrast, show little relationship between the distributions of lexical diversity categorised according to Memory and Analysis strengths.

Non-native speaker texts show a greater standard deviation of the mean *D* scores, and so this tells us that there is greater variability within this group. Within this group, learners with a predisposition towards grammatical sensitivity (Analysis) seem not to vary so much in lexical diversity. Verspoor, Lowie, & van Dijk (2008, p. 215) argue that by looking at different degrees and patterns of variability in developmental data we can start to understand how and when sub-systems change and develop. L2 lexis then could be interpreted as a system in a more developmental phase than L1 lexis and is arguably less stable than L1 lexis. Although we should not forget that native speakers do not stop acquiring lexis, they are perhaps more advanced in their lexical knowledge than, in comparison, non-native speakers (Davies 2003, p. 41). It is also possible that L2 lexis, which is more variable than L1 lexis, might be more susceptible to the associations with learning style because of its more developmental nature.

Stability and variability

What the results seem to also suggest is that L2 learners who are more able to analyse language, i.e. analysis-orientated learners, tend to produce texts of which the lexis is less susceptible to wide variations in lexical diversity. Skehan suggests that a grammatical ability may provide an organizing framework in which a developing interlanguage system can grow (Skehan 1982, p. 312). One possible explanation is that a deeper analysis of lexis encourages grammatical associations which encourage more stability in lexical diversity. In the LAT C test (Analysis), learners are required to notice not only word order but also the grammatical function of words which requires considerable cognitive effort. These learners may tend to complexify their language and use function words more often to express precise meanings. Analytic learners may be more consistent in the use of semantically opaque function words and phrases which could encourage lower diversity. Another reason why analytic learners are less likely to show wide swings in diversity may be because they are less likely to write in a "telegraphic" style (i.e. a register which is lacking in grammar words) which could encourage repetition of certain key content words or the use of lists.

Memory-orientated learners, on the other hand, may have wider swings in diversity because of two factors. A memory orientated learning style may encourage greater repetition of lexis which would decrease the diversity score. In addition, a high reliance on memory could also discourage the use of function or grammatical words with low semantic meaning but which are necessary to give precision. Instead these learners may tend to rely more on lexis with high semantic meaning, which could increase the diversity of lexis. In the LAT B test (Memory) learners are only required to associate the stimulus word to its translation without any demonstration of understanding how the test word is integrated into a coherent language system. This may explain the greater variability in diversity scores: high diversity through lexis (rather than through low semantic function words) but also low diversity through repetition. The upshot of learning lexis as unitary wholes through rote or analyzing lexis for their grammatical content could account for the wider spread of diversity scores with memory-orientated learners, and the narrower spread of lower diversity scores with analysis-orientated learners. To illustrate this point, compare participant 32, who scores high on LAT B (Memory) but low on LAT C (Analysis) with participant 3 whose learning style profile is the opposite. Both participants wrote about dielectric loss but the second appears more coherent than the first.

Participant 32: top Memory, bottom Analysis

Dielectric loss can be caused by two processes electrical conduction current flow in an oscillating electric field allows material to absorb energy as heat. Dipole rotation this loss occurs due to movement of atoms or molecules in an alternating electric field. Dissipated heat is a form or kinetic energy of atoms and molecules. The most widely used application of this type dielectric loss is microwave ovens alternating current generates electromagnetic forces thus food and liquids getting hot.

Participant 3: middle Memory, top Analysis

Dielectric loss is the dissipation of power in a dielectric under alternating electric stress. Dielectric loss factor is a measure of the amount of dissipation of energy in the dielectric medium and is equal to the product of the dielectric constant and the loss is tangent. In the other hand dielectric loss is due to the effects of finite loss tangent in which the losses rise proportional over the operating frequency. Dielectric loss tangent is the imaginary part of the dielectric

constant and determines the [word removed] of the medium.

Participant 32 writes in a “telegraphic” manner whereas participant 3 writes in a more coherent and complex manner. Interestingly, participant 32 scored 44 (out of 50) on the grammar test whereas participant 3 scored 36. This may indicate that a multiple-choice grammar test may not give a reliable indication of a learner’s ability to construct complex meaning in a text. What this type of test may draw on is memory of language rather than analysis. Table 7 below shows the correlation between the grammar test and LAT B and C.

Table 7

Correlation between grammar test and learning style

		LAT B (Memory)	LAT C (Analysis)
Multiple-choice grammar test	Pearson Correlation	.485	.065
	Sig. (2-tailed)	.067	.819
	N	15	15

Although the correlation does not quite reach statistical significance, the strength of correlation is greater with Memory than Analysis. It could be that to express meaning precisely and cohesively an analytic learning style is required (Booth 2009). It is possible that these learners have a more efficient organizing framework for new lexis to develop or to become active. A more organized framework could favour not only stability but also encourage development. In other words, new lexical items may integrate more effectively into a more organized framework. This is supported by Mandler (1980, p. 253) who found that when participants sorted sets of unrelated words into a number of categories, the number of categories (of their own choosing) used in the sorting correlated highly with the number of items subsequently recalled. Therefore, analysis, or in this case categorization, relates to recall.

Precision and diversity

Another issue which arose out of the results is that the second reports written by the native and non-native speakers were lower in lexical diversity. This could indicate that these participants are repeating more technical words in their second reports and so do not have to use alternative vocabulary to explain or describe. Moreover native speakers may have a wider repertoire of English technical words because they would have studied technical subjects in an English speaking environment. This may mean that the learners are at a disadvantage because they could have started their degree course with a lower store of technical vocabulary available for use and so may have to employ higher frequency words with a broader meaning and repetition in their writing to compensate. A study by Foster (2009) found that non-native speakers taught in an English language speaking environment (London) had native speaker levels of lexical diversity whereas non-native speakers taught English abroad (Tehran) were found to have significantly lower lexical diversity in a speaking task. Non-native speakers (London) were found be able to select more precise lexis e.g. *notice* or *realize* instead of words with a broader meaning e.g. *understand* (Foster 2009, p. 103).

Therefore to investigate whether more technical vocabulary is employed by native speakers, who arguably have been exposed to more technical words in English, I undertook a post hoc analysis of the data, collating all the native and all the non-native speaker texts to obtain lexical frequency profiles of these two groups. The texts were then analysed by using Cobb's (2002) *Web VocabProfile* software which was developed by Laufer & Nation (1995). The *VocabProfile* software calculates the proportions of tokens (running words) in a text that can be found in different frequency bands. It uses word lists based on the British National Corpus (BNC) and Coxhead's (2000) Academic Word List (word families that are frequent in academic writing) to categorize words by frequency bands of up to 20k. In version 3.2, words beyond the 2k frequency band and not on the academic word list (AWL) are classified as 'technical'. Tables 8 and 9 below show the results.

Table 8

All NS and NNS (report 1)

Word Families (%)	NS	NNS
K1 Words (1 to 1000)	71.93	70.52
Function	41.01	39.88
Content	30.92	30.64
K2 Words (1001 to 2000)	6.51	7.43
Academic Word List	8.39	8.71
Technical Words (beyond 2000 but not in AWL)	**11.03**	**11.48**
Off-List Words	2.13	1.87

Table 9

All NS and NNS (report 2)

Word Families (%)	NS	NNS
K1 Words (1 to 1000)	72.07	70.55
Function	44.39	42.41

Content	27.68	28.14
K2 Words (1001 to 2000)	5.17	5.92
Academic Word List	7.06	7.75
Technical Words (beyond 2000 but not in AWL)	**11.79**	**11.73**
Off-List Words	3.91	4.05

What is striking about the different profiles is that they are all remarkably similar. There does not seem to be a greater percentage of technical words in the second reports. Moreover non-native speakers do not seem to be disadvantaged by a lower percentage of technical vocabulary use. It also seems that technical vocabulary is not used significantly more frequently after one month of studying for a degree in Engineering by either NS or NNS. It could be that studying specific modules in engineering draws upon the depth of processing (Craik and Lockhart, 1972) needed for NNS to acquire the domain specific vocabulary to use in their laboratory reports.

Implications for teaching and learning

It would be reasonable to assume that to teach technical vocabulary one could simply determine what the words are and then teach them. Whilst this group of words is obviously important for students, it is equally important not to ignore how these words are actually used in coherent texts. This study has shown that although technical vocabulary is prevalent in L2 texts, learners who do not analyse language may lack coherence and complexity in their texts. Learners who use a memory-orientated approach to language learning may use highly technical vocabulary but they may not be aware of, or use, the semantically opaque grammar words, such as prepositions and conjunctions, or phrasal verbs which need to be recycled to help make the expression of complex ideas coherent. Skehan (2009, p. 116) underlines this notion of complexity and rarity. He found that errors were provoked by less frequent words for non-native speakers. From a pedagogical point of view, low frequency technical vocabulary needs to be taught in context since isolating these words in a list format might not be sufficient for integration into coherent sentences.

One of the ways to foster the integration of technical lexis into coherent text is to select technical vocabulary which is embedded in sentences. This in itself is not enough to encourage the learners restructure and complexify their language. They need to restructure the sentences themselves; the outcome of this process is that they notice how technical words are used in combination with the surrounding words which are usually less concrete and more semantically opaque. The British National Corpus (http://corpus.byu.edu/bnc) can be consulted and most technical or low frequency lexical items can be found embedded within authentic text. The sentences can then be manipulated by the teacher in order for the students to supply or reconstruct the missing words.

1. Parallel structure: Learners correct the phrases to make the sentence parallel.

Faulty parallel structure	**Reworked sentence**
Among other reported techniques the most useful include differential thermal analyse, dielectric loss measurements, ray absorbing, and gas permeability studies.	Among other reported techniques the most useful include differential thermal analysis, dielectric loss measurements, ray absorption, and gas permeability studies.

2. Restructuring: Learners are given the content words. They then have to supply the missing words in between so they can restructure the sentence into a coherent whole.

Sentence to be restructured	**Reworked sentence**
Interestingly, / dielectric loss appears / match the loss modulus more / than the loss compliance / data / analysed.	Interestingly, the dielectric loss appears to match the loss modulus more closely than the loss compliance when data is analysed.

3. Run on sentence: Learners reformulate the sentence with link words with the purpose of showing the relationship between ideas in a sentence.

Run on sentence	**Reworked sentence**

X is to be expected for a given process and this results in a broadening of the dielectric loss peak, the more mobile a dipolar group, the easier it…	X is to be expected for a given process and this results in a broadening of the dielectric loss peak. Thus, the more mobile a dipolar group, the easier it…

What learners tend to confuse is not necessarily the technical vocabulary but how it is actually used in conjunction with other words. The target for teaching should not only be the technical words but also, just as importantly, the words which fit in between. If learners do not analyse low frequency lexis the danger is that they may not notice how these words are actually used in coherent and complex sentences. The problem for them is not so much what the technical lexis refers to but how this type of lexis fits together with semantically opaque words in a sentence. Understanding which learners are more disposed to analyse vocabulary and which learners are more disposed to learning vocabulary as unitary wholes can help a teacher understand the strengths and weaknesses of learners within in a class. It would seem likely that learners who avoid analysis and learn by rote memory will need greater support in understanding how words fit together to develop their language for coherence and complexity.

Conclusion

This study has sought to find similarities and differences between native and non-native speakers of English. This is because native-like linguistic ability is often held up as the goal for language learners and so it makes sense to see whether native speaker data could provide a basis for this comparison. In particular, the aim was to find any patterns of diversity profiles from the participants' written output grouped according to strengths and weaknesses in Memory and Analysis. Clear cut differences in lexical diversity between native and non-native speakers were not established. However, when the data was analysed in terms of Memory and Analysis strengths, patterns in diversity (*D*) were found for non-native speakers but these patterns were much weaker or even non-existent for native speakers. In the non-native speaker group, a memory-orientated learning style was associated with variability in their lexical diversity. In contrast, analysis-orientated learners are more stable in their diversity. When all native speaker texts were analysed together and compared with non-native speaker texts the two groups displayed a remarkably similar percentage of technical vocabulary.

The effects of previous learning may affect the susceptibility of lexical profiles to learning styles. In other words, native speakers may over-learn lexis which then cancels out the subtle effects of learning style whereas non-native speakers through a lack of time and use may have acquired lexis which is less predictable. Skehan's (1982, p. 337) hypothesis that 'previous learning affects the nature of long-term memory which, in turn, affects current processing' may be relevant here. The implication for teaching and learning is that differences in lexical diversity may point to differences in learning style. Evidence in this study suggests that a predisposition to analyse language may encourage a lower lexical diversity and less variability than learners who adopt a memory based approach in lexical diversity. Learners who do not analyse language can miss the important syntactical information attached to words. This may be particularly acute when learners need to use low frequency technical vocabulary. To help learners who are not predisposed to analyse language, teachers can foster noticing and restructuring through problem solving tasks which are particularly beneficial to memory-orientated learners.

Notes

---

[i] Different words are 'types' whereas every running word form is a 'token'.

[ii] One participant did not complete this test.

[iii] The x^{th} percentile is the value below which x% of the distribution lies.

Acknowledgements

I would like to thank Dr Clarissa Wilks, Dr Marina Lambrou and Karen Hurley for their helpful comments on a previous version of this paper, as well as Pauline Stephenson and the engineering students who made this study possible.

References

Bell, H. (2009). The messy little details: a longitudinal case study of the emerging lexicon. In T. Fitzpatrick & A. Barfield (Eds.), *Lexical processing in second language learners* (pp. 111-127), Bristol: Multilingual Matters.

Booth, P. (2009). The development of vocabulary proficiency in relation to learning style. In A. G. Benati (Ed), *Issues in second language proficiency* (pp. 95-115), London: Continuum.

Cobb, T. (2002) *Web VocabProfile/BNC-20* Retrieved September 18, 2009, from: http://www.lextutor.ca/vp/

Coxhead, A. (2000) A new academic word list. *TESOL Quarterly*, 34, (2) 213-238.

Craik, F. I. M. & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behaviour*, 11, 671-684.

Daller, H., Milton, J. and Treffers-Daller, J. *Modelling and assessing vocabulary knowledge*. Cambridge: Cambridge University Press.

Das, J. P. (1988). Simultaneous-successive processing and planning. In R. Schmeck (Ed.) *Learning strategies and learning styles*. (pp. 101-129). New York: Plenum Press.

Davies, A. (2003). *The native speaker: myth and reality*. Clevedon: Multilingual Matters.

Dunleavy, P. (2003). *Authoring a PhD*. Basingstoke: Palgrave Macmillan.

Foster, P. (2009) Lexical diversity and native-like selection: the bonus of studying abroad. In B. Richards, H. M. Daller, D. Malvern, P. Meara, J. Milton, J. Treffers-Daller (Eds.) *Vocabulary studies in first and second language acquisition: the interface between theory and application* (pp.91-106). Basingstoke: Palgrave Macmillan.

Jarvis, S. (2002). Short texts, best-fitting curves and new measures of lexical diversity. *Language Testing* 19 (1), 57-84.

Laufer, B. & Nation, P. (1995). Vocabulary size and use: lexical richness in L2 written production. *Applied Linguistics*, 16 (3), 307-322.

Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development*. Basingstoke: Palgrave Macmillan.

Mandler, G. (1980). Recognizing: the judgment of previous occurrence. *Psychological Review*. 87(3), 252-271.

Meara, P. & Miralpeix, I. (2007). *D_Tools*. (version 2.0) Swansea: Lognostics.

Meara, P., Milton, J., & Lorenzo-Duz, N. (2001) *Language aptitude tests*. Newbury: Express Publishing.

Oxford University Language Center Placement Test. Retrieved 8 November 2007 from http://www.lang.ox.ac.uk/courses/tst_english_placement.html

Skehan, P. (1982). *Memory and motivation in language aptitude testing*. Unpublished PhD thesis, University of London.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Skehan, P. (2009). Lexical performance by native and non-native speakers on language learning tasks. In B. Richards, H. M. Daller, D. Malvern, P. Meara, J. Milton, J. Treffers-Daller (Eds.) *Vocabulary studies in first and second language acquisition: the* *interface between theory and application* (pp.107-124). Basingstoke: Palgrave Macmillan.

Verspoor, M., Lowie, W., & van Dijk, M. (2008). Variability in second language development from a dynamic systems perspective. *The Modern Language Journal*. 92 (2), 214-231.

Appendix 1

First language backgrounds

L1	Frequency	Percent
English	20	55.6
Farsi	2	5.6
Hindi	2	5.6
Russian	2	5.6
Dari	1	2.8
Dutch	1	2.8
Guajarati	1	2.8
Malayalam	1	2.8
Norwegian	1	2.8
Persian	1	2.8
Sinhala	1	2.8

Somali	1	2.8
Urdu	1	2.8
Yoruba	1	2.8
Total	36	100.0