How do features of written language production interact at different performance levels?

Paper presented at 4th EALTA Conference, 15-17 June 2007, Sitges, Spain

Jayanti Banerjee
j.banerjee@lancaster.ac.uk

Florencia Franceschina
f.franceschina@lancaster.ac.uk

Anne Margaret Smith
a.m.smith3@lancaster.ac.uk

Lancaster University, UK

A. Outline

1. The research context
   - Relating described competence levels and L2 developmental stages
2. The main study
   - Documenting features of written language production typical at different performance levels
3. Analyses
   - Principal Components Analyses
   - Ordinal Regression Analysis
4. Conclusions and reflections

B. Background to the study

1. Collaboration between Language Testers and SLA researchers (see Bachman & Cohen, 1998; Laufer, 2001; Ellis, 2001 and Douglas, 2001)
2. Areas of benefit:
   - Insights about the acquisition process
   - Close examination of language produced by language learners
   - Increased emphasis on reliability and validity of measures

C. Aims

1. Identify the defining characteristics of written language performance at each band score
2. Explore how these features of written language change from one band score to another
3. Explain the effects of L1 and writing task type on the different features of written language production.

D. Language features explored

Cohesive devices	
• Ratio of frequency of use of demonstratives	We used Atlas-ti code the scripts for the occurrences of the demonstratives 'this', 'that', 'these' and 'those'. The pattern of use of individual demonstratives was not clear and there was a lot of variability between bands so we explored the patterns for the overall use of demonstratives. The eventual figure was divided by the number of tokens per script.
Vocabulary richness	
• Lexical variation/ diversity	We used the Type-Token Ratio (TTR) measure where the number of types is divided by the number of tokens and multiplied by 100. This approach has been criticised because it is affected by text length. A more robust method of measuring lexical variation has been developed by Malvern and Richards (2002) (see also, Duran et al., 2004). Based on mathematical modelling, the method samples increasingly larger chunks from the text, calculating the TTR for each sample. It then plots all the TTR values obtained on a graph and calculates the fit of the curve obtained to mathematical models. The result of this process is a D-value which represents the best fit between the two curves. However, we did not have access to this method at the time of the study.
Vocabulary richness cont/…	
• 'Weighted' Lexical density	Measures of lexical density calculate the proportion of lexical words to grammatical words in the text. In this study we have adopted O'Loughlin's (2001) definition of grammatical and lexical items as well as his division of lexical items into high-frequency and low-frequency items. O'Loughlin (2001) classified all the lexical items found in the COBUILD (Willis and Willis, 1988) list of 700 most frequently occurring words in English as high-frequency items. These words account for 70% of English text. As suggested by Halliday (1985: 64-5, cited in O'Loughlin, 2001: 102), O'Loughlin weighted these at ½ the value of the low-frequency items because this is likely to provide a truer estimate of lexical density. He then calculated the ratio of this weighted set of lexical items to the grammatical items in order to arrive at a measure of lexical density. We did not have access to the COBULD list so we developed a high-frequency list using the British National Corpus (BNC). In all other respects our procedure was like O'Loughlin's.
• Lexical sophistication	We adopted the approach developed by Nation and Heatley (1996), using the Range program. The program classifies the words in a text into four categories. The first two categories are the first and second thousand most frequently occurring words in English (West, 1953). The third category is the Academic Word List (Coxhead, 2000) and contains 570 word families. The final category is an open category for all words that are not contained in the first three lists.
Syntactic complexity	
• Clause per T-unit	The 'T-unit' (minimal terminal unit) was developed by Hunt and is defined as the unit generated when text is divided into the smallest possible independent segments, without leaving sentence fragments behind. Each T-unit consists of a main clause, and all the subordinate clauses that belong to it. This preserves both the subordination and the co-ordination of a written text, without yielding an unrealistically high word count per sentence. We calculated the number of clauses per T-unit.
• Dependent clause per clause	We counted the number of clauses in each text and the number of dependent clauses. We then divided the number of dependent clauses by the number of clauses.
Grammatical accuracy Grammatical features analysed for accuracy: number on demonstratives, copula in the present and past tense, subject-verb agreement on main verbs in the present (3PS 's') and passives.	
• SOC	SOC stands for 'Suppliance in Obligatory Context'. It is calculated as follows: $\frac{\text{number of correct suppliance in obligatory contexts x2} + \text{number of misformations in OCs}}{\text{total OCs x2}}$
• TLU	TLU stands for 'Target-Like Use'. It is calculated as follows: $\frac{\text{number of correct suppliance in obligatory contexts}}{\text{number of obligatory contexts} + \text{number of suppliance in non-OCs}}$

E. The data

1. 275 test takers
2. L1 Chinese (159) + L1 Spanish (116)
3. Male (112) + Female (106) – evenly distributed between the two L1 groups
4. IELTS band scores 3 – 8
5. Tasks 1 and 2

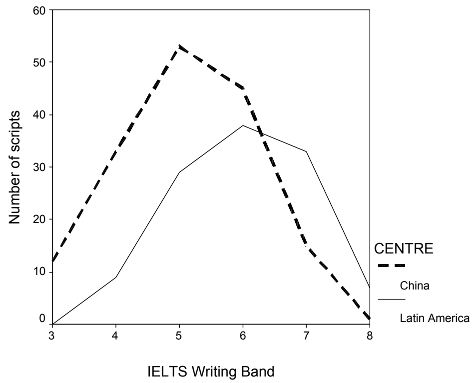

Figure 1: Distribution of scripts by L1 and IELTS band score

F. Results: Regression Analysis

Task 1:

Parameter Estimates

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[band = 4]	2.413	.847	8.122	1	.004	.754	4.073
	[band = 5]	4.335	.878	24.382	1	.000	2.614	6.055
	[band = 6]	6.116	.911	45.031	1	.000	4.329	7.902
Location	Lexical density	2.246	.948	5.611	1	.018	.388	4.105
	Syntactic complexity	2.914	.754	14.948	1	.000	1.437	4.392
	Grammatical accuracy	2.947	.684	18.581	1	.000	1.607	4.287
	[L1=1]	-.915	.257	12.729	1	.000	-1.418	-.413
	[L1=2]	0(a)	.	.	0	.	.	.

Link function: Logit.
a This parameter is set to zero because it is redundant.

Pseudo R-Square

Cox and Snell	.238
Nagelkerke	.255

Task 2:

Parameter Estimates

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[band = 4]	-14.370	3.190	20.294	1	.000	-20.622	-8.118
	[band = 5]	-12.091	3.160	14.636	1	.000	-18.285	-5.897
	[band = 6]	-9.954	3.125	10.149	1	.001	-16.078	-3.830
Location	Lexical diversity	-.051	.021	5.697	1	.017	-.093	-.009
	Lexical sophistication	-.189	.029	42.677	1	.000	-.246	-.132
	Syntactic complexity	3.709	1.247	8.840	1	.003	1.264	6.154
	Grammatical accuracy	5.996	1.024	34.280	1	.000	3.989	8.003
	[L1=1]	.014	.290	.002	1	.960	-.554	.583
	[L1=2]	0(a)	.	.	0	.	.	.

Link function: Logit.
a This parameter is set to zero because it is redundant.

Pseudo R-Square

Cox and Snell	.409
Nagelkerke	.439

G. Selected References

- Bachman, L.F. and Cohen, A.D. (eds.) (1998) Interfaces between second language acquisition and language testing research, Cambridge: Cambridge University Press.
- Botley, S., and McEnery, A.M. (2000) Discourse anaphora: the need for synthesis, in Botley, S. and McEnery, A.M. (eds.) Corpus-based and computational approaches to discourse anaphora, Amsterdam: John Benjamins Publishing Company, 1 – 41.
- Douglas, D. (2001) Performance and consistency in second language acquisition and language testing research: a conceptual gap, Second Language Research, 17(4), 442 – 456.
- Halliday, M.A.K. and Hasan, R. (1976) Cohesion in English, London: Longman Group Ltd.
- Hawkey, R., and Barker, F. (2004) Developing a common scale for the assessment of writing. Assessing Writing. 9, 122 – 159.
- International English Language Testing System (IELTS), http://www.ielts.org (accessed 09 January 2006)
- Laufer, B. and Nation, P. (1995) Vocabulary size and use: lexical richness in L2 written production, Applied Linguistics, 16, 307-322.
- Malvern, D. and Richards, B. (2002) Investigating accommodation in language proficiency interviews using a new measure or lexical diversity, Language Testing, 19(1), 85 – 104.
- McNamara, T. (1996) Measuring second language performance, London: Longman.
- Meara, P. and Miralpeix, I. (2004) D Tools, Swansea: Lognostics (Centre for Applied Language Studies, University of Wales Swansea).
- Nation, P. and Heatley, A. (1996) Range, Wellington: School of Linguistics and Applied Language Studies, Victoria University of Wellington.
- O'Loughlin, K. (2001) The equivalence of semi-direct speaking tests, Cambridge: University of Cambridge Local Examinations Syndicate & Cambridge University Press.
- Odlin, T. (2003). Cross-linguistic influence. Handbook of Second Language Acquisition. C. J. Doughty and M. H. Long. Malden, MA, Blackwell: 436-486.
- Read, J. (2005) Applying lexical statistics to the IELTS speaking test, Cambridge Research Notes, 20, 12 – 16.
- Shohamy, E. (1998) How can language testing and SLA benefit from each other? The case of discourse, in Bachman, L.F. and Cohen, A.D. (eds.) (1998) Interfaces between second language acquisition and language testing research, Cambridge: Cambridge University Press, 156 – 176.
- Wolfe Quintero, K., Inagaki, S., and Kim, H.-Y. (1998) Second language development in writing: measures of fluency, accuracy and complexity, In Technical Report 17. Honolulu: University of Hawai'i at Manoa, Second Language Teaching and Curriculum Centre.