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

Jayanti Banerjee, Florencia Franceschina, Anne Margaret Smith
Lancaster University, UK
4th EALTA Conference, 15-17 June 2007, Sitges, Spain

Aims

- Identify the defining characteristics of written language production at different points on a rating scale.
- Explore how these language features interact as test-takers progress up the rating scale.
- Find out whether different language features are salient at different levels on a rating scale.
- Investigate the effect of L1 and task type on the language features that are relevant to a performance.

Why these aims?

- Validation of rating scales (providing an empirical basis for descriptions of competence at each performance level)
- Placing the results of SLA research on a common scale so that it can be related and compared.
- Providing teachers with a description of language development that can be used for diagnosis, agenda setting and curriculum planning.

The main study: overview

- Approach

  Analyse writing performances that have been placed at different band scores through normal rating procedures

- Language features explored
  - Cohesive devices
  - Vocabulary richness
  - Syntactic complexity
  - Grammatical accuracy

The main study: language features explored

Why these features?

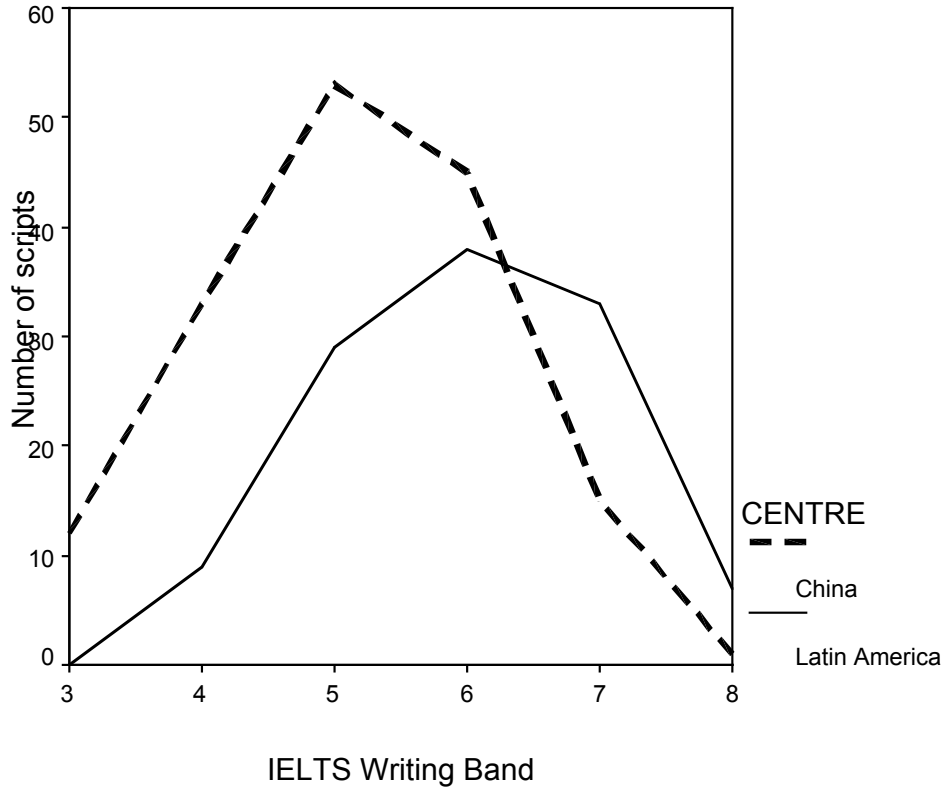
- We wished to cover a range of key areas of language performance
- These features represented key aspects of most language assessment scales
- Previous research (e.g. Wolfe-Quintero et al., 1998) has shown that these features are productive developmental measures.

The main study: the data

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

The main study: the data

- Distribution of scripts



The main study: analyses

- Bands 3 – 8
- Analyses for:
  - Cohesive devices
  - Vocabulary richness
  - Syntactic complexity
  - Grammatical accuracy
- Trends by measure

Interaction between measures: analyses

- Ordinal regression analysis - to find out how these measures interact in predicting band level (i.e. outcome variable is band level)
  - Because we cannot assume linearity or interval measurement

Procedure for obtaining ordinal regression

- Only bands 4-7 (255 test takers) included.
- PCA/EFA to reduce measures in each language area so as to end up with measures that tap different aspects of proficiency in each of the 4 language areas.
- Step-wise backward regression.
- Procedure done separately for each task.

PCA/EFA

- 8 measures reduced to 6
  - Cohesive devices:
    - Ratio of all demonstratives/tokens
  - Vocabulary richness:
    - Lexical variation/density
    - Weighed lexical density
    - Lexical sophistication
  - Syntactic complexity:
    - Dependent clauses per clause
  - Grammatical accuracy:
    - Target-like use

Ordinal regression results

Task 1

		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.

Ordinal regression results
Task 2

		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
	Gra mmatical 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.

Ordinal regression results
Summary

- Best predictors of band level regardless of task:
  - Syntactic complexity
  - Grammatical accuracy
- Best predictors of band level for Task 1:
  - Lexical density
  - Syntactic complexity
  - Grammatical accuracy
- Best predictors of band level for Task 2:
  - Lexical diversity
  - Lexical sophistication
  - Syntactic complexity
  - Grammatical accuracy

Conclusions
Task 1

- The model explains bands 4 – 6 but not band 7
  - Task does not give higher level test-takers latitude to show what they can do while simultaneously supporting lower level test-takers.
  - For this task the measures that we have investigated are not relevant for higher level test-takers.
- The model accounts for a relatively modest proportion of the variability in the band levels.

Conclusions
Task 2

- The model explains all 4 bands
- The model accounts for a higher proportion of the variability in the band levels (compared to Task 1).
  - Because more of the measures contribute to the model

Conclusions

- The cohesion measure used here does not seem to contribute to the model. However this is likely to be an artefact of the measures we conducted.
- Vocabulary measures contribute to both Task 1 and Task 2 BUT the measures are different.
  - Lexical density might be more relevant in Task 1 because of the lexical support provided by the input material.

Final thoughts

- Plenty that remains to do:
  - More robust measures of coherence/discourse features.
  - Explore the interactions band by band i.e. what defines a Band 4?

Final thoughts

- We are *edging* closer to understanding the features of language that underlie the developmental stages implied by the IELTS scales:
  - Vocabulary richness (measured by lexical diversity, lexical density and lexical sophistication)
  - Syntactic complexity (measured by the number of dependent clauses per clause)
  - Grammatical accuracy (measured by the extent to which test-takers correctly supply specific early and late-acquired grammatical features)

Questions, comments, suggestions?

If you would like to contact us:

Jayanti Banerjee
j.banerjee@lancaster.ac.uk

Florencia Franceschina
f.franceschina@lancaster.ac.uk

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