Lexical diversity in speaking and writing performances
Guoxing Yu
Graduate School of Education, University of Bristol, 35 Berkeley Square, Bristol, BS8 1JA
Guoxing.Yu@bristol.ac.uk

Lexical diversity has been considered an illuminative predictor of learners’ general language proficiency (e.g. Zareva et al. 2005) and an essential indicator of the quality of their writing (e.g. Laufer and Nation 1995) and speaking (e.g. Jarvis 2002; Malvern and Richards 2002; O’Loughlin 1995) task performances. Such a positive relationship is also claimed explicitly in the rating scales of major international language tests (e.g. International English Language Testing System, Michigan English Language Assessment Battery, Test of English as a Foreign Language), as well as in the development and implementation of automated writing evaluation systems (e.g. e-rater).

For example, IELTS writing and speaking responses are rated according to their “lexical resource” which refers to “the range of vocabulary the candidate has used” (IELTS Handbook 2007). According to the MELAB rating scales (MELAB Technical Manual 2003), test takers need to demonstrate “a wide range of appropriately used vocabulary” in order to achieve a high score for written compositions, and similarly an “excellent speaker” should demonstrate that his use of “idiomatic, general, and specific vocabulary range is extensive”; and for a “marginal speaker”, his “vocabulary is limited”.

In TOEFL iBT, one of the criteria to score the responses to the independent and the integrated writing tasks is “appropriate and precise use of grammar and vocabulary” (TOEFL iBT Scores 2005). Lexical diversity has also been employed as one of the most important parameters in automated writing (e.g. Chodorow and Burstein 2004) and speaking (e.g. Zechner et al. 2007) evaluation systems. It seems that lexical diversity as an important quality indicator of test performance has been widely assumed in both research and practice in the field of language teaching and testing. However, the precise nature of such relationships remains largely asserted and sometimes elusive.

Given this widespread assumption and the high stakes of international language tests and the influences of these tests on language education worldwide, empirical posterior validation studies are much needed to establish whether and to what extent such an explicit link between lexical diversity of written and spoken discourses and test scores assigned to the writing and speaking task performances, as claimed by the rating scales of the international language tests and by the parameters of automated evaluation systems, exist in the actual test performance data.

Furthermore, several other questions remain unanswered such as: to what extent test takers’ lexical diversity in writing and speaking task performances are correlated, and to what extent lexical diversity of a written or spoken discourse is attributable to test takers’ overall language proficiency, and what task characteristics may affect lexical diversity of performance and how.

This paper reports an empirical posterior validation study that analysed a sample of an international language test programme’s archived data to investigate to what extent such relationships exist, in particular: the relationships between lexical diversity and the holistic quality of test takers’ writing and speaking task performances as well as their general language proficiency, the differences in lexical diversity between writing and speaking task performances, and to what extent the prompts of the writing tasks affected lexical diversity of written discourses.

Using D as a measure of lexical diversity (Malvern and Richards 1997, 2002; McKee et al. 2000), it was found that D had statistically significant and positive correlation with the overall quality ratings of both writing and speaking performances as well as test takers’ general language proficiency. Nevertheless, the significant relationships did not bear out across the sub-groups of the sample in terms of gender, first language background, purpose of taking the test and topics of the writing prompts. The different topics and topic types of the writing prompts also had significant effects on lexical diversity – especially when test takers were highly familiar with the topics – even after controlling their writing abilities and overall language proficiency. The lexical diversity of test takers’ writing and speaking performances were approximately at the same level; further, D was found to be more able to predict test takers’ speaking than their writing performances.

The implications of these findings are discussed with specific reference to the use of lexical diversity measures to inform language test validation and the development of lexical diversity parameters in automated evaluation systems.

References


