Assessing the impact of an intensive EAP programme: multidimensional analysis of an authentic learner corpus

Steve Issitt (University of Birmingham, UK)

Multidimensional analysis (MDA) forms a central component of the paper. MDA uses computers to count the frequencies of linguistic features in tagged corpora, and then to carry out a factor analysis on these feature counts in order to identify 'dimensions', that is, sets of meaningful associations among individual variables.

The multidimensional approach can be used to identify individual features which can then be quantified and compared so that it is possible to examine a particular text, identify a feature, for example the incidence of agentless passives or average word length and compare it with another text. This enables an analysis of pre and post-writing samples which the study presents as a key methodological technique.

The current study applies multidimensional analysis to a learner corpus, namely the EAPCORP which stands for English for Academic Purposes Corpus and consists of 526 individual scripts and 263 matched pairs (pre and post course) representing a complete set of essays for two separate EAP summer programmes. All the scripts were written under exam conditions by hand and copied in their entirety with no corrections. Students were given a thirty minute writing task at the beginning of their respective EAP programme which were of twenty, fifteen, ten or six weeks duration. They were given the same writing task at the end of their programmes. The task was " What are the differences between your country and the UK?". The question was chosen because of its simplicity, relevance and the fact that it offers considerable opportunity for explanation, comparison and exposition. The aim was to standardise the prompt question without creating too much boredom and to prevent students from preparing an answer in advance.

The corpus itself presents data for two EAP programmes with two separate student cohorts, one from 2010 and one from 2012, offering an opportunity to examine developmental features. The EAPCORP itself may have some advantages over some other learner corpora in that it is designed to measure linguistic feature movement for a specific programme and can be considered in this sense to be purpose-built. It also covers two separate years and has a relatively large number of matched pair samples.

The data was analysed by using the multidimensional analysis tagger (MAT) which "replicates Biber's (1998) tagger for the multidimensional functional analysis of English texts … and generates a grammatically-annotated version of the corpus or text selected [and] the statistics needed to perform a text-type or genre analysis" (Nini, A., 2014, p.1). The MAT programme operates by employing the algorithms used by Biber (1988) and employs an adaption of the Stanford tagger (Toutanova, K. et al., 2003) to analyse the text. The programme offers a range of analytical categories of which two are used in the present study, the individual variables and the dimensions.

The results of the study suggest that the students' written production is marked by movement along a talk to writing strand suggested by Shaw and Liu (1998) and that one characteristic of developing student second language writing, is

text of a more recognisably "written" and to a certain degree more complex character. This is evidenced by the MAT programme analysis which has shown that in terms of individual features there has been an increase in use of longer words, passives, nominalisations, verb participles, sentence relatives, demonstratives, prepositional phrases and attributive adjectives. There has been a decrease in first and second person pronouns, private verbs, use of the present tense, use of the past tense and fewer emphatics. In terms of dimensional characteristics, this movement from talk to writing is further evidenced by the observed shift from nonabstract to abstract information which characterises all eight programmes (2010 and 2012) and the change from involved to informational production, from high overt to low overt persuasion and from low to high information elaboration which characterise seven out of eight of the programmes.

References

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Nini, A. (2014). *Multidimensional Analysis Tagger 1.2 - Manual*. Available at: https://sites.google.com/site/multidimensionaltagger/versions

Shaw, P. and Liu, E. (1998). What Develops in the Development of Second-language Writing. *Applied Linguistics*, 19 (2), pp. 225-254.

Toutanova K., Klein, D., Manning, C. and Singer, Y. (2003). Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. In *Proceedings of HLTNAACL 2003*, pp. 252-259.