Accuracy across proficiency levels: A learner corpus approach. Jennifer Thewissen. **Presses Universitaires de Louvain, Lougain-la-Neuve, Belgium (2015). 342pp.**

Another great book has come out of a PhD study. As a researcher in the field of language testing, I take great interest in identifying the criterial features of learner language that distinguish between different levels of proficiency, and in developing level descriptors that are empirically-based as well as user-friendly. When the call for a review for this book came out, I jumped to the opportunity because I had read the journal article published by the same author several years ago (i.e. Thewissen (2013)), and had long wanted to know more. I am glad that I did.

This book is based on Thewissen's PhD thesis, which used the International Corpus of Learner English (ICLE) (Granger et al., 2009) for two main purposes. The first purpose was to capture the development of linguistic accuracy of the argumentative essays written by learners of English at intermediate to advanced levels, namely B1 to C2 in the Common European Framework of Reference (CEFR) (Council of Europe, 2001). The second purpose was to create a set of L1-specific CEFR descriptors related to the linguistic accuracy by building on the results from the essays written by learners who have French as their L1.

The methodological rigor and the usability of the results make this book a great addition to knowledge in the field of language testing as well as language learning and teaching. The organisation of the book is as follows; Chapter 1 reviews the early error analysis studies in order to feed into the methodology of this study, in terms of a) the methods of collecting the language samples, highlighting the need for specifying and controlling for the length of texts, proficiency levels and the learners' L1s, b) types of errors (e.g. overt vs. covert errors), c) classification of errors (e.g. different categories of verb phrase errors), d) error counting (what the denominators are, e.g. total number of words, T-units, obligatory occasions) and e) error explanation (where the error might be coming from, such as L1 transfer). The literature review is very thorough and informative that I would recommend anyone who is interested in investigating learner language accuracy to have a read. Thewissen eventually decides on using potential occasions as the denominator for error counting for this book, i.e. by calculating errors of a particular category out of the corresponding part-of-speech (POS) category. For example, the number of verb tense errors out of the total number of verbs used in the performance, as each use of verbs presents an opportunity for tense error.

Building on the review and discussion of early error analysis studies in Chapter 1, Thewissen moves on to the next chapter, discussing the use of learner corpora in error analysis and presents how errors have been detected and tagged, and presents findings from key studies using learner corpora on the accuracy trends across different proficiency levels in errors involving, for example, articles, tenses and lexical choice.

The rigorous methodology of this book is presented in detail in Chapters 3 and 4. From the ICLE, a total of 223 argumentative essays in English (of 500-900 words) were selected based on the writers' L1 (French, German or Spanish), and were manually error-tagged using Louvan

error taxonomy (Dagneaux et al., 2008) with 45 error categories. Also, each of the 223 essays were assigned an overall CEFR level for linguistic accuracy by two professionally-trained raters (who used to work for Cambridge and/or IELTS exams) based on their initial analytic assessment in the relevant CEFR subscales of *Grammatical Accuracy*, *Vocabulary Control*, *Vocabulary Range*, *Orthographic Control* and *Coherence and Cohesion*. The 223 essays were POS-tagged using CLAWS system (Garside & Smith, 1997) in order to produce a figure which represents the rate of erroneous uses in a particular POS category.

The answer to the research question, ***How do the 40-plus error types in the ICLE learner corpus develop from proficiency levels B1 to C2?***, was explored by comparing the mean error percentages of each error category based on the potential occasion analysis across B1 to C2 level groups using ANOVA. If ANOVA results revealed any statistical significance, then Ryan post-hoc tests were used to determine exactly where (between which levels) the significant difference(s) lay.

While the choice of the statistical methods seems the best for the research design, it was not clear whether the distribution of data satisfied the assumptions for using parametric tests. Moreover, it would have been useful to clearly justify not adjusting the p-values, as some might argue that running more than 40 ANOVAs on the dataset from the same learners would require it. I would assume, however, that this is defensible as adjusting the p-values would have led to a higher probability of committing Type II errors (i.e. judging a difference as non-significant where it should really be significant), especially for less frequent errors with small percentages across proficiency levels.

It is not surprising to find that different error types exhibited different patterns as the proficiency levels increased, but it is extremely useful to know how and which error types distinguished (or did not distinguish) between different levels of proficiency. There were 5 'strong' developmental patterns where significant differences were found between at least one pair of adjacent levels. Specifically, 22 error categories showed a 'strong' pattern:

1) Between all adjacent levels (B1>B2>C1>C2) (e.g. single lexical error, such as using *technique* instead of *technology*)
2) Between B1 and all other levels (B1>B2, C1, C2) (e.g. uncountable nouns, noun-dependent prepositions, verb morphology)
3) Between B1 and B2, and all non-adjacent levels (B1>B2, B1>C1, B1>C2, B2>C2) (e.g. articles)
4) Between B and C levels (B2>C1) (e.g. adverb order)
5) Between B2 and B1 (B2>B1) (e.g. punctuation)

Moreover, there were 8 error categories that showed 'weak' developmental patterns in which significant differences were found only between non-adjacent levels (which will not be detailed here due to the limited space). Even more intriguing was that 14 error categories did not show any significant progress across 4 proficiency levels, including errors of missing punctuation, finite/non-finite verbs, and conjunction coordination.

In summary, out of the 46 error categories, only 4% showed steady improvement, 59% exhibited an initial improvement then reached a 'plateau', 35% showed no marked progress across levels, and 2% followed partly regressive patterns. This finding emphasises that the error profiles do not necessarily progress linearly. Furthermore, the development in the linguistic accuracy of the ICLE samples was found primarily between B1 and B2, which "suggests that the construct of accuracy may be a useful tool in discriminating B1 vs. B2 performance, but that additional constructs would need to be relied on to better identify B2 to C2 performance (p.211)." This echoes the findings from recent studies which examines the criterial features of learner output, such as by Tavakoli et al. (2016), and has important implications in the field of language testing where level descriptors and CEFR levels are extensively (and increasingly) used.

In fact, the results presented in Chapter 5 have already been published (in a much shorter article) in Thewissen (2013), but readers will be glad to know that there is another chapter in this book that was not included in the article. Chapter 6 attempts to create a set of L1-specific CEFR descriptors related to the linguistic accuracy by selecting the data of L1 French learners' essays from those analysed in Chapter 5. The CEFR has been criticised for its under-specificity due to its non-language specific nature, and the need for more detailed descriptors is surging. In order to address this under-researched area, Thewissen produced L1- and L2-specific descriptors based on the linguistic accuracy trends found in L1 French learners' English essays. The L1 French-specific descriptors at B2/C1 levels in *Grammatical Accuracy* are presented on pages 248-249, starting as follows:

> Good grammatical control overall but some errors remain, both in more subtle but also sometimes more basic grammatical areas. There is still noticeable mother tongue influence from French. Errors are likely to be found […] in the following areas: […] adverb placement with recurrent use of the erroneous verb+adverb+noun phrase pattern […]. (p.248)

As a language tester, learner and a former language teacher, I myself would find such descriptors extremely helpful in conceptualising what B2/C1 for a specific L1 would look like. In addition to these descriptors in *Grammatical Accuracy*, a number of language samples are presented throughout Chapter 6 in the areas of *Vocabulary Control*, *Orthographic Control* and *Punctuation*. I personally feel that it is worth purchasing the book just to read Chapter 6.

To conclude, it is my firm view that this book presents a solid research project which was executed with great attention to detail, and it offers great deal of insight and practical implications not only from the findings and L1-specific descriptors, but also from various aspects of research methodology. Readers who are not very familiar with corpus analysis should not be put off (I am by no means a corpus specialist), as the book does not go into nitty-gritty of technical procedures but provides thorough explanations with helpful summaries and examples. Highly recommended.

References

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

Dagneaux, E., Denness, S., Granger, S., Meunier, F., Neff, J., & Thewissen, J. (2008). *Error Tagging Manual Version 1.3.* Centre for English Corpus Linguistics. Universite catholique de Louvain, Louvain-la-Neuve.

Granger, S., Dagneaux, E., Meunier, F., & Paquot, M. (2009). *The International Corpus of Learner English. Handbook and CD-ROM* (2nd ed.). Louvain-la-Neuve: Presses Universitaires de Louvain.

Garside, R., and Smith, N. (1997). A hybrid grammatical tagger: CLAWS4. In R. Garside, G. Leech & A. McEnery, A. (eds.) *Corpus annotation: Linguistic information from computer text corpora*. Longman, London, pp. 102-121.

Tavakoli, P., Nakatsuhara, F. & Hunter, A. (2016). *Fluency constructs measured in speaking tests: Implications for semi-direct EAP speaking tests*. Paper presented at the Language Testing Forum. Reading, UK.

Thewissen, J. (2013). Capturing L2 accuracy developmental patterns: Insights from an error-tagged EFL learner corpus. *Modern Language Journal, 97*(1), 77-101.