

## **English Grammar Profile: a corpus-based inventory of English grammar competency using the Cambridge Learner Corpus**

Anne O’Keeffe (Mary Immaculate College, Limerick) and Geraldine Mark  
(independent scholar)

The English Grammar Profile (EGP) Project was a four-year quasi-longitudinal study investigating learner grammar from the Cambridge Learner Corpus (a 55 million-word learner corpus containing over 200,000 exam scripts, from over 200 countries, by speakers of over 140 first languages). The focus of the project was to scrutinize the grammar used by learners in the Cambridge English Language Assessment exams across the levels of the Common European Framework of Reference (CEFR), tracking the evolution of the use of forms and features, across different learners, contexts, L2 backgrounds, levels of proficiency. The main output of the research is the EGP, a free educational online database, which provides a profile of over 1,200 corpus-based grammar competency statements about real learner grammar use across the six CEFR levels.

The main focus of this paper will be on the findings of this extensive corpus study in relation to the acquisition of grammar across different levels of competency, as well as interesting insights into our misconceptions about learner grammar and its development across competency levels. These findings parallel and/or complement existing corpus and non-corpus SLA work.

We note that there are some important departures in this study from hitherto pioneering contrastive learner corpus research on interlanguage:

By using the CEFR as the calibration, we are looking at learner language across six staging points A1-C2, which have been intuitively drawn within the CEFR (Council of Europe 2001).

1. By using the CEFR as the calibration, we are looking at learner language across all six staging proficiency points A1-C2, which have been drawn within the CEFR (Council of Europe 2001). Hitherto studies have looked at only some levels.
2. We are investigating a summative aggregation of grammatical competences of learner English (from those who have taken Cambridge English exams). Hitherto studies usually examine one or a set of grammatical features.
3. We do not take a systematic contrastive focus between learners’ L1 and the target language.
4. We are not setting out to examine the learner English of one L1 cohort; we are exploring data from a multitude of L1 backgrounds across 55 million words of learner data.
5. We are not tracking the evolution of any one interlanguage feature.
6. We are investigating the quasi-longitudinal development of grammar competence, with data spanning 17 years, across more than 260,000 learner exam documents.

7. We are not setting out to chart learner development in relation to error decline, plateau or regression. We are tracking and describing the development of learner grammar competence towards an 'idealized L1 norm'.

The authors acknowledge the potential circularity in the use of exam data from learners who will have been preparing for the exams using targeted exam practice material. We anticipate that learners will have been exposed to a range of input, pedagogical or otherwise, other than traditional ELT textbooks and exam-focussed materials.

As noted by Ortega & Byrnes (2008), accuracy development at the higher levels of proficiency has been under-researched and this is echoed by Thewissen (2013) who makes her own substantial empirical contribution by looking at learner error data developmentally, within the CEFR framework. Interestingly, there is complementarity between Thewissen's (2013) findings on grammar errors across levels of the CEFR and the patterns of EGP grammatical competencies. Thewissen (2013), in referring to instances where there is lack of significant progress in accuracy between B2 and C2 levels (e.g. lexical phrases), cites the accuracy-complexity trade off effect whereby learners at higher levels take more risks with language and use it in more complex ways but in doing so their risk of making an error increases and so this can hinder improvement in accuracy. This stabilization, depending on where it occurs, according to Thewissen, is related to the "ceiling effect" (after Milton & Meara 1995), which could be at play between B2 and C2. She advises that although errors still remain, a significant amount of learning has already taken place and she stresses that the stabilization tendency should not be narrowly interpreted as an absence of any development whatsoever, or in Larsen-Freeman's words "linguistic rigor mortis" (2006: 597).

The present study of competence, not error, brings some insight to the ceiling effect. We often see a 'syntactic ceiling' where a form has stabilised and reached its syntactic 'developmental endpoint' (after Thewissen 2013) at A and B levels but as learners go beyond these levels, they deploy the stabilised forms with a greater complexity of meaning and with greater dexterity of use. As they get to higher levels, this is constantly borne out in how a learner can use a syntactic form pragmatically and skilfully play with a form for greater subtlety of meaning or focus.

The past simple tense offers one of many examples of this developing functional and pragmatic competence. At A1, the affirmative form is evident and stable, but only with a very limited range of regular and irregular verbs, leading to the can-do statement '*Can use the affirmative form with a limited range of regular and irregular verbs*'. The only function of the form at A1 that meets the criteria for competent use (based on the methodology for this research) is to talk about everyday events or states, such as *Every time I went to Hendon Park, I saw so many people there.* (A1, 2007, Polish).

This study illustrates that two developments happen as learners go up levels of the CEFR:

- 1) they learn more verbs and so their lexical range of verbs expands, and
- 2) the past simple affirmative form is deployed in more complex ways, illustrating growth in functions and often pragmatic competence.

Were we to look at the pattern of errors, as others have done, we would not see this picture of growing competence beyond point of stabilisation of form. We would rather see a ceiling effect, a stabilisation or developmental end point at A1 in the correct use of the affirmative form and miss a rich seam of competence from A1 upwards. Put simply, our insight is that by looking at competence rather than error stabilisation end points, we can see how learners develop beyond the acquisition of a form to how they can *use* that form with greater functional and pragmatic effect and complexity.

Thewissen (2013) rightly points to the dearth of attention to the importance of investigating lexical development, apart from work such as Ife et al (2000) and Milton & Meara (1995). Our findings offer complementary insight into the role of lexis. As illustrated by past simple affirmative form examples at A1 level above, while the syntactic form can be used competently at A1, it can only be used with a limited range of verbs. This is because a learner at A1 has not yet acquired many verbs. In the present study, we see an increase in lexical repertoire going hand-in-hand with an expanding repertoire of grammatical uses of the same form, level by level. Certain verbs are more typically used for particular functions and so as a learner's lexical range increases so does the potential range of usage contexts increase.

Within the ELT community there is widespread agreement on what grammar structures are taught and when. Other insights from this study include some mismatches between what is taught at certain levels in a traditional ELT syllabi and when it is actually competently acquired. In some cases, we found more frequent use of structures in the learner data than in native speaker corpora. In other cases, we even found that learners showed competency in the use of grammatical items before they appear formally on traditional syllabi. This may be due to a number of factors, some of which are related to individual differences and variation, or indeed, it may link to the findings from usage-based theories of learning (Ellis et al. 2015, Gablasova et al. 2017).

Learner use of conditional clauses illustrates an example of competence of structures not formally on the syllabus. Students are traditionally taught four conditional clauses, described as zero, first, second and third conditionals, whose forms are limited to four prescribed patterns. The four conditional clauses are typically introduced one by one from A2 level to B2. Two surprises emerge from the learner data:

- 1) evidence of all four traditional clauses in stable use by B1, and
- 2) evidence of 27 different forms and uses expressing conditionality from A1 to C2 (supporting the findings of Gabrielatos 2010).

It appears that learners are showing implicit knowledge of forms that have been both explicitly and not explicitly taught in the classroom. Here we find resonance with interest in explicit and implicit learning and theories of incidental vs intentional language learning.

In summary, by looking at a 55 million word corpus of learner English exam writing, from a global range of L1 backgrounds, we have been able to explore largescale grammar patterns of use and their development across the CEFR. The

insights from this research add to the existing body of work on learner corpora but do so from the perspective of what learners can do rather than what they cannot.

## References

- Ellis, N. C., O'Donnell, M. and Römer, U. (2015). Usage-based language learning. In B. MacWhinney and W. O'Grady (Eds.), *The Handbook of Language Emergence* (pp. 163-180). Hoboken, NJ: Wiley.
- Gablasova, D., Brezina, V. and McEnery, T. (2017). Exploring learner language through corpora: comparing and interpreting corpus frequency information. *Language Learning*, 67(1): <http://onlinelibrary.wiley.com/doi/10.1111/lang.12226/full>
- Gabrielatos, C. (2010). *A corpus-based examination of English if-conditionals through the lens of modality: Nature and types*. Unpublished PhD Thesis. Lancaster University.
- Ife, A., Vives Boix, G., & Meara, P. (2000). The impact of study abroad on the vocabulary development of different proficiency groups. *Spanish Applied Linguistics*, 4, 55–84.
- Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics*, 27, 590–619.
- Milton, J., & Meara, P. (1995). How periods abroad affect vocabulary growth in a foreign language. *ITL Review of Applied Linguistics*, 107/108, 17–34.
- Ortega, L., & Byrnes, H. (Eds.) (2008). *The longitudinal study of advanced L2 capacities*. New York: Routledge.
- Thewissen, J. (2013). Capturing L2 accuracy developmental patterns: insights from an error tagged learner corpus. *The Modern Language Journal*, 97, 77–101.