

ELT



Ireland bulletin

IN THIS EDITION

Could Ireland's internationalisation policy reshape future ELT curricula?

by Touria Jouilla McKee (page 5)

Stimulating Autonomy in Speaking Through Self and Peer Correction

by Gabriela Pozza (page 10)

Thank you for helping with my Big Idea!

by Marianne Jordan (page 12)

Corpora, Language Learning and Teaching

by Gerard O'Hanlon (page 16)

Teaching the Foundations of ESL to Adults: Challenges and Techniques

by Maximiliano E. Orlando (page 21)

Getting Unstuck

by Marjorie Rosenberg - ELT Ireland 5th Annual Conference Plenary Speaker (page 24)

No. 6

21st February 2020

ELT Ireland features (Pages 2-4)

Welcome to our Sixth Bulletin

by Peter Lahiff

Keeping you up to Date by Ben Dobbs

ELT Hackathon by Peter Lahiff

Letter from the Editor by Laura O'Grady

Email us to write for the next ELT Ireland Bulletin!

info.eltireland@gmail.com @ELTIreland

elt-ireland.com

Corpora, language learning and teaching.

By Gerard O'Hanlon

What is a corpus?

A corpus is an electronically stored, searchable collection of texts. Such banks of texts can be of written or (transcribed) spoken language. Corpora vary enormously in size, from a few thousand words to the billions of words found online at sites such as Sketch Engine.

How can we use corpora?

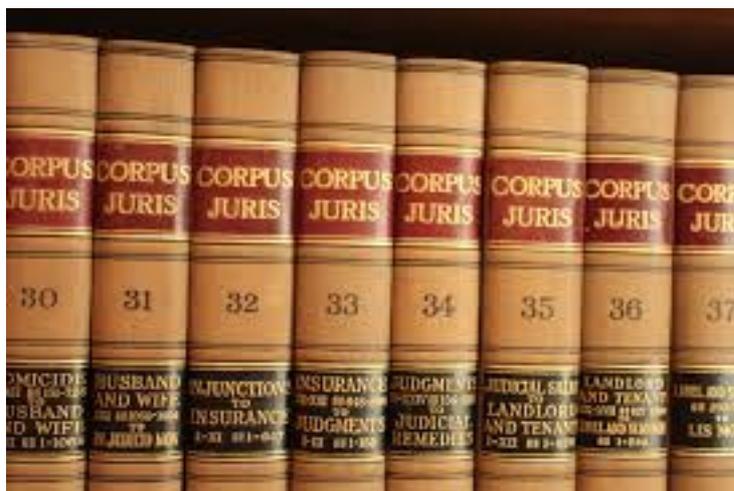
There are various strands as to how corpora can be applied to language education. The most common distinction is that of direct and indirect routes (Römer 2010). Teacher training also plays an important factor (O'Keeffe and Farr 2003; Flowerdew 2012) while learner corpora is vital to many facets of language learning and teaching. Let us now explore each of these strands in turn.

An indirect route

The indirect route defines the role of corpora in reference works (such as dictionaries and grammars), class materials and course books, all of which are informed by corpus-informed descriptions of language. In such circumstances, the use of corpora allows teachers to present language within varying degrees of context. This usually starts with collocations, patterning and concordance lines (i.e. the search item in a sentence). Further information about these concordances is available via links. The user can view the source text of the search item alongside other important information such as when it was initially published or recorded, or the genre it belongs to (academic writing, a radio interview etc). Being genre specific allows the user to see how words behave in different contexts.

Data-driven Learning

The direct route, meanwhile, involves students exploiting corpus data themselves. This permits the learner to perform linguistic analyses to aid language acquisition. Developed by Tim Johns in the early 1990s, this methodology is known as Data-Driven Learning (DDL). The idea behind DDL is to encourage learners to access authentic language, which helps refine L2 knowledge as learners observe how the target language behaves (Lenko-Szymanska and Boulton 2015). Such searches often help learners during phases when reference, clarification or accuracy is needed. This might occur during written work or after corrective feedback, for example. DDL is based on the premise that learners formulate queries and do the detective work themselves. This process of 'making sense' of corpus results contributes to the development of linguistic ability, learning skills and autonomy (ibid).



A legal corpus in need of a database

Pic By: Pexels

“This process of ‘making sense’ of corpus results contributes to the development of linguistic ability, learning skills and autonomy”

DDL – Issues and solutions

Yet DDL and the use of corpora in language education face certain problems. Firstly, there is a lack of awareness of corpora and their pedagogical potential, particularly outside of third-level education and research programmes - they are rarely used on mainstream ELT teacher training courses. Two studies cited further below show how efforts at integration had varying degrees of success.

This leads to a second issue, that of a reluctance on the part of teachers and learners to use corpora, given that they are perceived to be demanding and challenging. Moreover, corpora are often believed to be best suited to learners at advanced or proficient levels and off-limits for lower level students.

DDL is by nature an inductive approach, which means that one learns by doing. Learner agency is important in DDL and students are encouraged to take an active role in their own learning by interpreting corpus results autonomously. However, such autonomy may not sit well with certain institutions or teachers and may well clash with their learners' preferred way of learning. Furthermore, problems of access to technological or paid online content also present obstacles.

Notwithstanding such issues, the teacher can mediate and scaffold to help students with their learning, as we shall see in Wilcher (2019) below. As part of their planning, a teacher can use corpus results to highlight examples of words, phrases or textual features as patterns in contexts. These can be edited, cleaned up, printed out and underlined/highlighted with colours to facilitate class work. Such interventions would also be particularly effective for lower level learners. Editing, printing or photocopying these enhanced versions would also help bridge any technological divides.

Teacher training

What place do corpora have in teacher training? O'Keeffe and Farr (2003) state how corpora offer 'great potential in developing language awareness and research skills within teacher education'. Flowerdew (2012) further details its potential by outlining three routes of corpora in teacher training. These include technological awareness (e.g. online resources, corpus software), pedagogic awareness (e.g. DDL) and linguistic awareness (e.g. chunks, frequencies, genres). Cobb and Boulton (2015) emphasise the potential of linguistic awareness, stating that the use of corpora can sensitise teachers and learners to important issues such as frequency, chunking, collocations and genre.

“Editing, printing or photocopying these enhanced versions would also help bridge any technological divides.”

Corpora and CELTA

Naismith (2017) details how corpus-based tools were introduced into a CELTA programme in IH Vancouver. The sites were Google N-Grams and Just-The-Word, the first a chronological frequency tracker, the second a concordancer. Both are basic, open-access platforms well suited to those taking their first steps in this area. An input session was delivered midway through the course, focusing on aspects of language awareness which helped deepen awareness of corpus-related themes, such as frequency and vocabulary chunking.

Naismith shows that trainees found corpora useful for their own language research and analyses but were apprehensive about using them during their own lessons. The time constraints of such a short training programme limited further exploration of the benefits of corpus-informed teaching. Naismith recommends changes to the CELTA syllabus to reflect areas of language study (such as frequency and chunking through concordancers) while recognising the limitation that all centres may not be equipped with the technology to provide access to computers, internet and online corpora.

Teacher engagement

Lenko-Szymanska (2017) designed a semester-long course which involved trainee English Teachers at the University of Warsaw compiling small ESP corpora and preparing corpus-based lessons. The aim was to exploit corpora for the development of courses, materials and class activities and consisted of three thematic modules: types of corpora, exploration of corpora for teaching language systems and skills and designing an ESP corpus for syllabus and classroom activities.

The implementation of such a course required three competencies –“practically handling corpus tools, interpreting corpus results and incorporating corpus research into instructional contexts. The course usually ran over one semester and the study covers 2013-2016.

Lenko-Szymanska found that teachers lacked autonomy in terms of the pedagogical skills necessary for successful exploitation of corpora in language education. The paper calls for the development and empirical validation of a model of effective teacher training in corpus informed instruction.

For those considering the incorporation of corpora into teacher training courses, Lenko-Szymanska recommends that teacher trainees be exposed to corpora in their own language classes and introduced to the tenets of DDL alongside other instructional procedures and techniques. If teacher trainees can discover the potential of corpora in their own learning and development, she concludes, this may motivate them to make better use of corpora in their professional lives.

“...practically handling corpus tools, interpreting corpus results and incorporating corpus research into instructional contexts.”

Corpora and lesson structure

Wicher (2019) asks how best to integrate corpus-driven activities (DDL) into the foreign language classroom. In particular, his study focused on both a ‘PPP’ and a ‘Task-Based Language Teaching’ lesson.

Where do corpus tasks best fit in with these lesson phase structures? Regarding the PPP model, concordances don’t suit the Presentation or Production phrases of such lessons, Wilcher argues, given that Presentation will need the target language situated in a broader context (which corpus search results may lack) while the Production phrase requires spoken or written production, neither of which will be ‘best achieved utilising corpus search results’ (ibid).

This leaves the Practice phase, where corpus search results can be useful as they allow for the exploration of co-text, i.e. patterns and structures of the language immediately surrounding keyword searches. Such concordances can serve as a model exemplifying typical usage

patterns from the Presentation phase or provide models looking forward to the Production phase. This target language can be highlighted with different colours, fonts or underlining to help learners observe and follow language content. Wilcher (2019) emphasises that PPP is not without its critics but does emphasise that an integrated DDL/PPP approach can move such lessons towards a more learner-centred, inductive direction.

In terms of Task-Based Language Teaching, the post-task language focus or feedback phase encourages reflection and feedback. The paper suggests that corpus-related tasks as highlighted in the PPP section (modelling, exemplification, correction) can provide an excellent basis to demonstrate and explore genre specific features or provide reactive error correction, for example. (*A link to a sample TBLT lesson employing corpora can be found at the end of this article.)

Practical Approaches

For teachers who employ a range of approaches to their lessons, Wilcher (2019) highlights three practical ways that corpora can be integrated into a lesson: input enhancement, variable assignments and the manipulation of concordance lines.

Input enhancement draws learner attention to specific structures e.g. with coloured or underlined examples (as detailed above); variable assignments permit the teacher to differentiate tasks for their classes by presenting the same corpus search results but asking differentiated learners to focus on different degrees of directionality through narrowly or openly formulated questions; thirdly, while the manipulation of concordance lines may go against the DDL tenet of ‘textual authenticity at all costs’, some practitioners recommend a ‘softer’ approach. In essence, this means tidying up, reformatting and/or editing concordance results for easier reading and comprehension. A frequent complaint of learners when using corpora is the messy nature of transcribed language or cut-off utterances.

“...this means tidying up, reformatting and/or editing concordance results for easier reading and comprehension.”

Learner corpora

One further aspect of corpus research is that of collecting large banks of learner language. Two important open access sites are of importance to teachers here. These are the English Profile (a 55 million word corpus of written learner language from Cambridge) and TLC Hub (a spoken learner corpus of some 3.5 million words). Although both collect corpus data from learners of various L1 backgrounds, other learner corpora can have a specific L1 focus for local markets and research.

The English Profile (EP) consists of the Vocabulary Profile and the Grammar Profile. EP searches are ‘filtered’, meaning that many aspects of the searches have already been categorised for the user. This allows us to view search items in a contextual sentence (taken from past Cambridge exam student scripts) and observe the CEFR levels at which language features occur. The Vocabulary Profile allows an exploration of categories by theme or by lexical features (such as suffixes). The Grammar Profile charts over 1,200 grammar features with items detailed by level, can-do statements and learner examples.

The Trinity Lancaster learner corpus (TLC Hub) is a spoken corpus with data collected from Trinity’s GESE test. A central focus is to study spoken interaction and meaning, based on learner

identity. Nonetheless, a large range of spoken learner language features can be garnered from a corpus such as this. Much like the EP, this corpus is made up of language from a variety of learner L1 backgrounds at various levels, all of which can be observed via graphics and contextual data.

What can we do with learner corpora?

Features of learner corpora can be compared with target language structures and genres, allowing teachers to see what learners are doing correctly as well as incorrectly. By comparing learner language with target corpora (such as Academic English), teachers can observe at what levels learners start producing target features, how competently they are using them and if these features are being overused or underused.

Learner corpus research helps provide frequency information, common errors, improved register/style description, advice on overuse/underuse, common collocates for target language and lists of key linguistic features for specific written genres, among other features (Xu 2016). This information has informed the compilation of dictionary and grammar books, course materials and the development of language tests and assessment criteria.

Learner corpora are vital given that they describe when learner language emerges in spoken or written production, thus allowing us to see at what CEFR level the word or grammar feature first starts being produced by a broader cohort of learners. The learner corpus will not tell us when a language feature was acquired or at what level comprehension of it first occurred. Nor should it dictate at what level we need to employ such a language feature in our classes. The needs of the group, the differentiation of the class and the experience of the teacher and learners in situ add an important 'pedagogic filter' to how we employ learner (or any kind of) corpora in language learning and acquisition.

*TBLT lesson incorporating corpus tools by Neil McMillan on Freed:

<https://www.freed.com/articles/911/tblt-on-the-fly-the-vendor-client-relationship>

References

- Cobb, T., and Boulton, A., (2015) 'Classroom Applications of Corpus analysis' in Biber, D., and Reppen, R., eds., *English Corpus Linguistics*, CUP: Cambridge, 478-498.
- Flowerdew, L., (2012), *Corpora and Language Education*, London: Palgrave Macmillan.
- Lenko-Szymanska, A. and Boulton, A. eds. (2015) *Multiple Affordances of Language Corpora for Data-Driven Learning*, John Benjamins Publishing Company: Amsterdam/Philadelphia
- Lenko-Szymanska, A., (2017) Training Teachers in data-driven learning: Tackling the Challenges, *Language Learning and Technology*, 21 (3) 217-241
- Naismith, B., (2017) Integrating corpus tools on intensive CELTA courses, *ELT Journal*, 71 (3) 273-283
- O'Keeffe, A. and Farr, F. (2003) Using language corpora in initial teacher education: pedagogic issues and practical applications, *TESOL Quarterly*, 37 (3): 389-418.
- Römer, U. (2010) 'Using general and specialised corpora in English language teaching: past, present and future'. IN M. Campoy-Cubillo et al (eds), *Corpus-based approaches to English Language Teaching*, 18-38
- Wilcher, O., (2019) Data-driven Learning in the Secondary Classroom: A critical Evaluation from the Perspective of Foreign Language Learning Didactics. In Crosthwaite, Peter (ed.) *Data-Driven Learning for the Next Generation: Corpora and DDL for Pre-tertiary Learners* (pages pending), London: Routledge
- Xu, Q., (2016) Application of Learner Corpora to Second Language Learning and Teaching: An Overview, *English Language Teaching*, 9 (8) 46-52

ELT 
Ireland
bulletin

