Bridging Informal Massive Open Online Courses and Formal English for Academic Purposes Programmes with Language Corpora

Alannah Fitzgerald  Shaoqun Wu, Ian H. Witten
Concordia University & The Open Educational Resources Research Hub University of Waikato (shaoqun, ihw) @cs.waikato.ac.nz alannahfitzgerald@gmail.com

Martin Barge, William Tweddle, Saima Sherazi
Queen Mary University of London {m.i.barge, w.tweddle, s.m.sherazi} @qmul.ac.uk

1 Introduction

Massive Open Online Courses (MOOCs) provide a compelling opportunity for domain-specific language learning. They supply a large corpus of interesting linguistic material relevant to a particular subject area, including text, supplementary images (slides), audio and video. It follows then that these domain-specific corpora can also be used in formal English for Academic Purposes (EAP) programmes as well. Such corpora can be automatically analysed, enriched, and transformed into a resource that learners can browse and query in order to extend their ability to understand the language used, and help them express themselves more fluently and eloquently in that domain.

To illustrate this idea, an existing online corpus-based language-learning tool (FLAX) is applied to Open Educational Resources (OER), including openly-licensed Coursera and edX MOOC content, and Open Access (OA) research content for the development of domain-specific language collections for uptake by informal MOOC learners and formal EAP students. Open education acts as a bridge to formal education, and is complementary, not competitive, with it. This is one of a cluster of research hypotheses currently under investigation at the OER Research Hub for the development of open language corpora in FLAX.

2 Domain-specific Corpora

The use of domain-specific corpora is a growing trend in language teaching and learning (e.g. Gabrielatos, 2005). Most corpora are based on particular domains, genres, or collections of certain types of document from which recurrent phrases and grammatical patterns can easily be retrieved (Stubbs and Barth, 2003). Among other aspects of language, the domain-specific corpora we are developing in FLAX will provide an excellent context in which to study lexical bundles (Biber & Barbieri, 2007; Biber et al, 2003, 2004) and collocations, a notoriously challenging aspect of productive language use even for quite advanced learners (Bishop, 2008; Nesselhauf, 2003). The academic language collections in FLAX use an automated scheme that extracts salient linguistic features from academic text and presents them in an augmented text interface, designed for the non-expert corpus user (Wu & Witten, In Press). Rather than relying on complex search commands to query corpora within involved concordancer interfaces (which have been designed by and for the corpus linguist) FLAX links relevant tools and resources into streamlined online interfaces for the language learner. For example, FLAX connects academic collections in the language learning system to the Wikipedia Miner tool to extract key academic concepts and their definitions from Wikipedia articles (Milne & Witten, 2013) to assist with reading and vocabulary.

3 Open Educational Resources

OER were selected for re-use to demonstrate how the FLAX corpus tools can linguistically enhance MOOC and domain-specific content. Two demonstration academic English language collections in FLAX are currently under development.

One collection is based on virology courses and resources developed by Professor Vincent Racaniello of Columbia University. His lectures were already popular across a range of web channels, including iTunesU and YouTube, before being imported into the Coursera MOOC platform. These lectures, along with Racaniello’s weekly podcast *This Week in Virology*, his academic *Virology* blog, and OA articles relevant to his virology courses, are all published under a Creative Commons Attribution licence for easy processing as a FLAX language support collection for the virology MOOC learners.

The other domain-specific collection in FLAX is
centred on the re-use of OER for academic English for law. This collection is being developed for EAP students who will be following the Law Pathway Pre-sessional and the Critical Thinking and Writing in Law In-Sessional programmes at Queen Mary University of London this 2014-15 academic year.

Lecture transcripts and videos (streamed via YouTube and Vimeo) will be featured from four different MOOCs: Copyright Law at Harvard (edX), English Common Law at the University of London (Coursera), Age of Globalization from Texas at Austin (edX), and Environmental Law and Politics at Yale (OpenYale). Podcast audio files and transcripts from the University of Oxford’s Law Faculty and the Centre for Socio-Legal Studies (OpenSpires) are also being added to the collection.

The spoken language sub-corpus will be paired with a written language sub-corpus made up of OA research articles, samples of student writing from previous Law Pathway Pre-sensionals and sections of EThoS law theses held at the British Library and written by Queen Mary law students.

4 Open Systems Design for the Uptake of Educational Language Corpora

Open corpus-based systems and resources like the academic English collections in FLAX have unique characteristics and challenges with regards to diffusion, adoption and integration. Insights from EAP teachers involved in the FLAX collections building process will be presented with respects to how they perceived and interacted with these open educational systems, as they exist and as we are designing them in this research. In addition to this, a deeper understanding of how to design, iterate, integrate and evaluate open technological systems in support of advanced approaches to language learning and instruction within the specific context of open educational resource initiatives will be shared for discussion with TaLC conference participants.

References


