Sharing an Open Methodology for Building Domain-specific Corpora for EAP

Martin Barge, William Tweddle, Saima Sherazi, Alannah Fitzgerald

http://creativecommons.org/weblog/entry/35165/
Outline

• FLAX Language Project at Waikato University
• Developing an EAP Resource Interface between Traditional EAP and Massive Open Online Courses
• Developing ESAP Collections in FLAX (Academic English for Law at QMUL)
  – What’s in the Demo Collection and What’s to Come!
  – Formatting Open Access Articles for FLAX Corpora
• Fully Open Texts
  – Beyond Parsing with Text Augmentation & Linked Data
  – Lexical Bundles, Collocations, Wordlists, Cherry Picking Functions
  – Building in Interactivity
• Design-based Research with FLAX, Queen Mary and the OER Research Hub
  – Research & Development Cycles with Design-based Research for Iterating Collections Development
  – Rapid Prototyping of Online Demo Collections to Evaluate the Design Process and to Share with Stakeholders
FLAX Language at Waikato University

http://flax.nzdl.org  FLAX image by permission of non-commercial reuse by Jane Galloway
FLAX Language Project at the Greenstone Digital Library Lab, Waikato University NZ

Professor Ian Witten
FLAX Project Lead

Dr Shaoqun Wu
FLAX Project Lead Researcher & Developer
QM’s Critical Thinking & Writing in Law

• Queen Mary’s Critical Thinking and Writing in Law (CTWL) Programme has been running successfully for over 7 years.
• It is delivered by QM Language Centre’s EAP/ESAP team as part of the Insessional provision.
• Over 600-800 LLM students enroll on it every year.
• A team of 6-7 EAP tutors teach on it, and are under constant pressure to develop better and new materials for their high calibre students.
The FLAX System for Subject-Specific Corpus Development


An empirical method of linguistic enquiry
• Used to discover the lexico-grammatical properties of genre or text-type
• Used to discover the key terminology given field or discipline – English for Specific Academic Purposes (ESAP)
• Used for exploring collocations:

“You shall know a word by the company it keeps.” (Frith, 1957:11)
Collaboration with Subject Specialists

“In the emerging academic literacies approach involving cooperation between subject specialists and writing teachers, the aim is to help the students develop metacognitive awareness of the roles and functions of writing in that discipline, to enable them to stand back from it and observe how it functions, and then to help them gradually participate in the genres, where genre is understood as a constellation of actions rather than a list of formal features.” (Breeze, 2012)
Benefits

- Inductive – promotes critical thinking
- Promotes learner autonomy
- Based on evidence, not instinct
- Especially relevant for ESP and ESAP

Limitations

- Need for Ts and Sts to have technical skills to use corpora and concordancers
- Need for access to corpora and software programmes
- Large amount of data can be overwhelming

“Every student is Sherlock Holmes.” (Johns, 2002:108)
Interfacing Traditional EAP & MOOCs
## ESAP Law Collections in FLAX

<table>
<thead>
<tr>
<th>Type of media in the FLAX Law Collections</th>
<th>Number and source of items in the FLAX Law Collections</th>
</tr>
</thead>
<tbody>
<tr>
<td>Podcast audio files &amp; transcripts (OpenSpires)</td>
<td>10-15 Lectures (Oxford Law Faculty &amp; the Centre for Socio-Legal Studies)</td>
</tr>
<tr>
<td>MOOC lecture transcripts &amp; videos (streamed via YouTube &amp; Vimeo)</td>
<td>4 MOOC Collections: Copyright Law (Harvard/edX), English Common Law (Uni. of London/Coursera), Age of Globalization (Texas at Austin/edX), Environmental Law &amp; Politics (OpenYale)</td>
</tr>
<tr>
<td>Student PhD thesis writing and Pre-sessional for Law ESAP essay writing</td>
<td>70 QMUL EThoS Theses at the British Library (Open Access but not licensed with Creative Commons – will need permission to develop for Non-Commercial Educational &amp; Research purposes); 20+ Essays from QMUL Law Pre-sessional</td>
</tr>
<tr>
<td>Open Access research articles (relevant to QMUL Law and EAP for Law and Globalisation)</td>
<td>40 Articles (DOAJ - Directory of Open Access Journals)</td>
</tr>
</tbody>
</table>
Formatting OA Articles for FLAX
### Lessons from Technology and Intellectual Property in the Oil and Gas Industry in Scotland: a Scholarly Journey and an Empirical Review

- 1. Introduction and overview
- 2. The project landscape: three strands - Strand 2.1
- 2. The project landscape: three strands - Strand 2.2
- 2. The project landscape: three strands - Strand 2.3
- 3. Combining the strands: a new approach to encouraging the development of technology?
- 4. Empirical Pilot
- 5. Key themes from pilot interviews
  - 5.1. There is no such field as ‘energy’
  - 5.2. Technological innovation is not the focus of (most) of the oil and gas industry
  - 5.3. IP is important but not essential to innovation and rarely a significant obstacle to activity
- 5.4. There can be collaboration and sharing
- 5.5. There should be different regulation
- 7. Conclusions
1. Introduction

The EU is unique in its transboundary laws, institutions, and cooperation in the field of environmental matters. With its quasi-federal structure, the EU has managed to go beyond the rigid distinction between international and national law that prevails in intergovernmental cooperation [2].

Historically, legislation concerning water has been some of the most developed and progressive in European Community law. This trend persists, and the most significant and momentous developments are taking place with regard to aquatic species and ecosystems, and therefore, their principles and assessment methods may eventually be applied to other sectors of environmental law, and to non-aquatic biodiversity [3]. Therefore, aquatic ecosystems, and water in general, may be perceived as testing grounds for contemporary regulations [3].

In the EU, the current ecological objective for surface water is 'good ecological status', established by the Water Framework Directive [46]. Through this institutionalization of ecosystem-based objectives, community water policy is supposed to become functionally oriented towards sustainable development (e.g. Articles 1 and 4 emphasize sustainability) [7,8]. To reach this objective, suitable legal instruments are needed, and quality standards have been discussed and emphasized as one set of legal instruments with the potential to bring about an improved environmental status [3,911].

For example, quality standards have led to successful reductions of concentrations of toxic substances in bodies of water, and may be useful for addressing environmental problems related to air and water quality, which are linked to human health [10,1214]. The establishment of quality standards signified an important change in environmental regulations; for example, they take their point of departure in the conditions of a body of water, are based on the precautionary principle, and are legally manageable. Nonetheless, when trying to manage coupled
2.1 Taxonomy of the lex specialis principles of international climate law

Article 3 UNFCCC principles are indicative of ubiquitous legal values that give expression to the object and purpose of the UNFCCC. As a starting point, attention turns to these principles in order to confound which is...

The United Nations Framework Convention on Climate Change (UNFCCC or FCCC) is an international environmental treaty produced at the United Nations Conference on Environment and Development (UNCED), informally known as the Earth Summit, held in Rio de Janeiro from June 3 to 14, 1992.

Related topics in Wikipedia
- Montreal Protocol
- Kyoto Protocol
- United Nations Convention to Combat Desertification
- Clean Development Mechanism
- Intergovernmental Panel on Climate Change
- 2009 United Nations Climate Change Conference
- Carbon sink
- Joint Implementation
Wordlist from OA Articles

<table>
<thead>
<tr>
<th>Word</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>principle</td>
<td>380</td>
</tr>
<tr>
<td>create</td>
<td>263</td>
</tr>
<tr>
<td>contract</td>
<td>148</td>
</tr>
<tr>
<td>creditor</td>
<td>128</td>
</tr>
<tr>
<td>bond</td>
<td>107</td>
</tr>
<tr>
<td>approach</td>
<td>82</td>
</tr>
<tr>
<td>area</td>
<td>86</td>
</tr>
<tr>
<td>security</td>
<td>85</td>
</tr>
<tr>
<td>investor</td>
<td>84</td>
</tr>
<tr>
<td>framework</td>
<td>45</td>
</tr>
<tr>
<td>formal</td>
<td>40</td>
</tr>
<tr>
<td>financial</td>
<td>40</td>
</tr>
<tr>
<td>federal</td>
<td>40</td>
</tr>
<tr>
<td>focus</td>
<td>36</td>
</tr>
<tr>
<td>regime</td>
<td>35</td>
</tr>
<tr>
<td>relevant</td>
<td>32</td>
</tr>
<tr>
<td>process</td>
<td>31</td>
</tr>
<tr>
<td>variable</td>
<td>31</td>
</tr>
<tr>
<td>subsidy</td>
<td>30</td>
</tr>
<tr>
<td>welfare</td>
<td>29</td>
</tr>
<tr>
<td>achieve</td>
<td>28</td>
</tr>
<tr>
<td>culture</td>
<td>28</td>
</tr>
<tr>
<td>occur</td>
<td>27</td>
</tr>
<tr>
<td>community</td>
<td>26</td>
</tr>
<tr>
<td>code</td>
<td>25</td>
</tr>
<tr>
<td>sector</td>
<td>25</td>
</tr>
<tr>
<td>acknowledge</td>
<td>24</td>
</tr>
<tr>
<td>option</td>
<td>22</td>
</tr>
<tr>
<td>legislation</td>
<td>22</td>
</tr>
<tr>
<td>parallel</td>
<td>22</td>
</tr>
<tr>
<td>domestic</td>
<td>22</td>
</tr>
<tr>
<td>evidence</td>
<td>22</td>
</tr>
<tr>
<td>corporate</td>
<td>22</td>
</tr>
<tr>
<td>option</td>
<td>22</td>
</tr>
<tr>
<td>legislation</td>
<td>22</td>
</tr>
<tr>
<td>parallel</td>
<td>22</td>
</tr>
<tr>
<td>domestic</td>
<td>22</td>
</tr>
<tr>
<td>evidence</td>
<td>22</td>
</tr>
<tr>
<td>corporate</td>
<td>22</td>
</tr>
<tr>
<td>option</td>
<td>22</td>
</tr>
<tr>
<td>legislation</td>
<td>22</td>
</tr>
</tbody>
</table>
Collocations from Law Lectures

125 collocation(s) associated with the word globalization

<table>
<thead>
<tr>
<th>Noun + of (63)</th>
<th>Verb (24)</th>
<th>Noun (20)</th>
<th>Adjective (18)</th>
</tr>
</thead>
<tbody>
<tr>
<td>age of globalization (10)</td>
<td>aspect of globalization (5)</td>
<td>phase of globalization (5)</td>
<td>globalization of sport (4)</td>
</tr>
<tr>
<td>context of globalization (3)</td>
<td>study of globalization (3)</td>
<td>effects of globalization (3)</td>
<td>boosters of globalization (2)</td>
</tr>
<tr>
<td>symptom of globalization (2)</td>
<td>characteristics of globalization (2)</td>
<td>underbelly of globalization (2)</td>
<td>feature of globalization (2)</td>
</tr>
<tr>
<td>logic of globalization (2)</td>
<td>history of globalization (2)</td>
<td>phase of the globalization (2)</td>
<td></td>
</tr>
</tbody>
</table>
Linking Collocations in Law-Specific Corpus to Reference Collections in FLAX (BNC, BAWE, Wikipedia)
Lexical Bundles from Law Lectures

So I'm going to have to shoot ahead here.
So I'm going to conclude there and just remind you that the electronic exam will be emailed out to you today at five o'clock, that you have forty-eight hours to return it, and return it both to the Classes*v2 server drop box, but also please return it to your teaching assistants.
So I'm going to scoot ahead here with one final aspect of this story.
So I'm going to just pause with reflections on the work of Tyrone Hayes, who is a former chemical company employee, a scientist who worked for Syngenta.
So I'm going to run out of time in a few minutes.
So I'm going to skip really quickly now to a variety of different cases.
So I'm going to move on to discuss several other statutes now.
So I'm going to jump ahead here to a discussion about evacuation.
So I'm going to jump ahead here because I'm going to run out of time.
So I'm going to pause there and just take a couple of minutes for any questions that you've got.
So I'm going to stop there.
So I'm going to jump ahead here.
So I'm going to do my own little survey here.
So I'm going to try to show a couple film clips here, hopefully this will pick up, on the effects of radiation and fallout.
So I'm going to close here, and I'm going to just take a little bit of license, and I apologize for this, because you may think of me as being paternal.
So I'm going to run through just a few summary of findings here before I close today that should give you a sense of what's working and what's not working too well with the LEED program.
### Building Interactivity into FLAX

<table>
<thead>
<tr>
<th>Language activities</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Scrambled Sentences</strong></td>
</tr>
</tbody>
</table>
| The words of sentences are scrambled and students must sort them into their original order.  
This activity helps you study sentence structure by providing you with genuine text and allowing you to select suitable materials to practice on.  
**Type:** Individual  
**Instructions:** Put the bold words in the correct order by drag-drop them into the correct position. |
| **2. Word Guessing** | exercises | create an exercise |
| Students enter words in the gaps, based on the context within a given article, individually or collaboratively.  
This activity helps improve your communication skills and vocabulary.  
**Type:** Individual or Group collaboration  
**Instructions:** Click on the gap and type in a word. |
| **3. Completing Collocations** | exercises | create an exercise |
| Learners fill in a blank to form a valid word combination.  
**Type:** Individual  
**Instructions:** |
| **4. Scrambled Paragraphs** | exercises | create an exercise |
| The paragraphs of a document are scrambled and students must sort them into their original order.  
This activity helps you study paragraph structure by providing you with genuine text and allowing you to select suitable materials to practice on.  
**Type:** Individual  
**Instructions:** Put the paragraphs in the correct order by drag-drop them into the correct position. |
| **5. Punctuation and Capitalization** | exercises | create an exercise |
| The punctuation of a paragraph is removed and students must insert the missing punctuation.  
This activity helps you study the role of punctuation in language.  
**Type:** Individual  
**Instructions:** Drag and drop punctuations on the top and put them in the right position. |
This activity helps you study sentence structure.

**Type:** Individual

**Instructions:** Click a box on the left and then a box on the right to make the correct sentence.

7. **Collocation Matching**

Select a set of collocations, normally from the same collocation type; split them into left and right parts; and shuffle each set of parts. For example, the secretary of state, course of action, hundreds of dollars might be presented as: the secretary of action hundreds of state course of dollars.

Learners must rematch them.

**Type:** Individual

**Instructions:**

8. **Collocation Dominoes**

This activity mimics the dominoes game: the last word of the previous collocation becomes the first word of the next. Here is an example: bank cheque -- cheque book -- book club -- club sandwich -- sandwich board -- board room ...

Learners are asked to complete collocation dominoes.

**Type:** Individual

**Instructions:**

9. **Related Words**

Pick several related words and some of their associated collocations; remove the words from the collocation text; and shuffle what is left. For example:

pay make

_____ bill, _____ effort, _____ debt, _____ difference

Learners are asked to choose the right word to complete a collocation, e.g. pay bill, make effort, pay debt and make difference.

**Type:** Individual

**Instructions:**

10. **Collocation Guessing**

Choose a target word and a number of associated collocations; remove the target word from the collocation text; and then reveal the remaining text gradually to a learner who must guess the target word. For example, given the following words:

plain, dark, white, bitter, milk, bar of,
FLAX Do-It-Yourself Podcast Corpora with Oxford OER

http://www.youtube.com/watch?v=Si24d3Z-8nQ
FLAX Do-It-Yourself Podcast Corpora 2: Building interactivity into your collections

http://www.youtube.com/watch?v=fysDzYjbhh0
Developing Podcast Activities in FLAX

The Global Recession - How Did It Happen? [Part One]

Description: Linda Yueh talks to Jonathan Michie about the credit crunch and looming global recession, the effect it has had on both banks and businesses, and how a Green New Deal may provide a solution to the crisis.
Close Exercises in FLAX

How Did It Happen? [Part One]

Jonathan Michie about the credit 1 crunch and looming global 2 _____, the effect it has had on both banks and 3 _____, and how a 4 _____ No

lrow in 6 _____ at St. Edmund Hall, University of 7 _____, I am an 8 _____ and actually also a lawyer by 9 _____, but I teach economics at t 2 _____ Economics', an 13 _____ textbook, which will soon need to be updated given the 14 _____ of the past year. And another book which _____' to look at the 16 _____ in which the economy has been changed, in both economic and legal 17 _____, over the last few years.

Michie, I am also an economist by 18 _____, although I am now 19 _____ of Oxford University's department for continuing 20 _____, which does ______, and in fact our most popular online course is on the global 23 _____ by Doctor Linda Yueh. I am also President of Kellogg 24 _____.
Scrambled Sentences in FLAX

Split Sentences
Exercise name: Credit Crunch - part one
Select a category: none  Add a category

Exercise parameters
Choose article: The Global Recession - How Did It Happen? (Part 1)
Number of words in sentence: Min 3 Max 15
Max sentences per page: 10

Number of sentences available: 17

Review: 17 sentences selected
I am also President of Kellogg College.
My most recent book is a handbook on globalisation.
Linda: I think 2008 has been a tremendous year in terms of economic developments.
Since then of course, things have only gone from bad to worse.
That roughly brings us to where we are today in terms of the financial crisis.
Drag ‘n’ Drop exercises in FLAX

**It Sentences**

- I am also President to be true, it probably is.
- My most recent book doesn't mean no regulation.
- Linda: I think 2008 has been a tremendous year of Kellogg College.
- Since then of course, things have only gone from bad to worse.
- That roughly brings us to where we are today in terms of economic developments.
- And thirdly the government has extended to them guarantees of lending originating now.
- Light touch regulation is a handbook on globalisation.
- It means that if something looks slightly too good a financial sector problem becomes a real economy problem.
- That is the lesson in terms of the financial crisis.
- By doing policy too late, of the great depression.

*Total score: 0  Question No. 1  out  Check Answer  Previous question  Next que.*
Learning Collocations in FLAX
Automated Collocations Guessing in FLAX (drawing on the British National Corpus)

**Ilocation Guessing**

<table>
<thead>
<tr>
<th>feature of an arbitration</th>
<th>How to p</th>
</tr>
</thead>
<tbody>
<tr>
<td>annulment of the arbitration proceedings</td>
<td></td>
</tr>
<tr>
<td>merits of arbitration</td>
<td></td>
</tr>
<tr>
<td>conduct of arbitration</td>
<td></td>
</tr>
<tr>
<td>competence of the arbitration court</td>
<td></td>
</tr>
<tr>
<td>presidency of the arbitration authority</td>
<td></td>
</tr>
<tr>
<td>right of arbitration</td>
<td></td>
</tr>
<tr>
<td>concept of arbitration</td>
<td></td>
</tr>
<tr>
<td>example of an arbitration</td>
<td></td>
</tr>
<tr>
<td>number of recognised arbitration agencies</td>
<td></td>
</tr>
<tr>
<td>courts of arbitration</td>
<td></td>
</tr>
<tr>
<td>offer of arbitration</td>
<td></td>
</tr>
<tr>
<td>use of arbitration</td>
<td></td>
</tr>
<tr>
<td>arbitration of rents</td>
<td></td>
</tr>
<tr>
<td>arbitration of proper action</td>
<td></td>
</tr>
<tr>
<td>guarantor of an arbitration</td>
<td></td>
</tr>
<tr>
<td>types of arbitration</td>
<td></td>
</tr>
<tr>
<td>costs of the arbitration</td>
<td></td>
</tr>
<tr>
<td>arbitration of disputes</td>
<td></td>
</tr>
<tr>
<td>method of arbitration</td>
<td></td>
</tr>
</tbody>
</table>

- The Community wants to create bureaucratic machinery for joint **arbitration of disputes**.
- Central to all the procedures, indeed to the very mandate of the United Nations is peaceful **arbitration of disputes**.
- The 1990 legislation authorizes the establishment of a stock exchange, sets out the preconditions for its statutes and rules of procedure, provides for its membership and operation, and establishes a system for **arbitration of disputes** between securities traders and between issuers and traders.

Bricks left: 19
Attempts: 1
Design-Based Research Cycles with FLAX, the OER Research Hub & Queen Mary

- Practitioners/Researchers involved in iterative development of ESAP language collections
  - Interfacing with open Law resources
    - Open Access articles, Open Government research reports with contributions from QMUL Law professors, Case Law, Open lectures, Openly-licensed student writing
  - Developing expertise with open tools and resources
  - Developing interaction within the corpus and derivatives from the corpus
  - Documenting the collections development process for sharing across the EAP and Open Education sectors
Free to Do Whatever You Want

- Open Resources for EAP Soup Dragons:
  - Building ESAP Corpora
  - Developing Interactivity into ESAP Corpora
  - Developing ESAP Course Book and Lesson Plan Derivatives
  - Researching and Developing ESAP Corpora & Derivatives
  - Researching and Developing Corpus Tools e.g. Interfaces, Text Augmentation and Linked Data Approaches

http://en.wikipedia.org/wiki/The_Soup_Dragons
Thank You

Shaoqun Wu: shaoqun@waikato.ac.nz / Ian Witten: ihw@cs.waikato.ac.nz

OER Research Hub http://oerresearchhub.org/
Alannah Fitzgerald: fitzgerald@education.concordia.ca; @AlannahFitz;
www.alannahfitzgerald.org TOETOE Blog; Slideshare:
http://www.slideshare.net/AlannahOpenEd/

The Language Centre – Queen Mary University of London
http://language-centre.sllf.qmul.ac.uk/
Martin Barge m.i.barge@qmul.ac.uk
William Tweddle w.tweddle@qmul.ac.uk
Saima Sherazi s.n.sherazi@qmul.ac.uk