Wow! The FLAX Language System
So Much Open Data

Shaoqun Wu
Alannah Fitzgerald
Ian Witten
Vici LinkedUp Challenge @ISWC

• Open Data-Driven Language Learning with FLAX
  – Open-Source Tools, Open Resources & Practices
  – Design-Based Research with the FLAX Project

• Research with Corpus-Based Approaches
  – Critical Reflection by Language Teachers with Open Do-It-Yourself Language Collections in FLAX
FLAX Wins the Vici LinkedUp Challenge for Open Data in Education

http://alannahfitzgerald.org/2014/10/06/vici-competition/
WHO ARE WE IN THIS FLAX RESEARCH & DEVELOPMENT COLLABORATION?
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
Research on Open FLAX Collections

OER Research Hub
Join us in building understanding of open education

Home  About  Research  News and Events  OER  Community  Contact Us

Alannah Fitzgerald
Open Fellow with OERRH
FLAX Language & Open Education Researcher

Recent Work: Evidence Hub / Hal Plotkin Interview / OER in 10 videos / Reports

http://oerresearchhub.org/
OER Research Hypotheses

Hypothesis A – Performance  Use of OER leads to improvement in student performance and satisfaction.

Hypothesis E – Reflection  Use of OER leads to critical reflection by (language) educators, with improvement in their practice.

Hypothesis G – Indicators  Informal (MOOC) learners use a variety of indicators when selecting OER.

Hypothesis H – Support  Informal (MOOC) learners adopt a variety of techniques to compensate for the lack of formal support.

Hypothesis I – Transition  Open education acts as a bridge to formal education, and is complementary, not competitive, with it.

Hypothesis K – Assessment  Informal means of assessment (in MOOCs) are motivators to learning with OER.

http://oerresearchhub.org/collaborative-research/hypotheses/
OPEN SOURCE LANGUAGE TOOLS DEVELOPMENT
FLAX Digital Library Collections

- Collocations database
- Open Educational Resources
- Glossary
- British National Corpus
- HTTP://www
The FLAX (Flexible Language Acquisition) project aims to automate the production and delivery of practice exercises for overseas students who are learning English. Exercise material comes from digital libraries, which can supply a virtually endless supply of collaborative and competitive language activities. (All software produced by this project is open source, issued under the GNU General Public License.)
Google-esque Interface Designs

Designed for the non-expert corpus user, namely:

learners, teachers, subject academics, instructional designers and language resource developers.
Introducing the Wikipedia Miner Toolkit
(Milne & Witten, 2013)
# Building Interactivity into FLAX Language Collections

## Language activities

<table>
<thead>
<tr>
<th>Activity</th>
<th>Exercises</th>
<th>Create an exercise</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Word Guessing</strong></td>
<td>exercises</td>
<td></td>
</tr>
<tr>
<td>Students enter words in the gaps, based on the context within a given article, individually or collaboratively. This activity helps improve your vocabulary and sentence structure and your communication skills. Type: Individual or Group collaboration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructions: Click on the gap and type in a word. Click on the light bulb icon (if any) for help.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>2. Completing Collocations</strong></td>
<td>exercises</td>
<td></td>
</tr>
<tr>
<td>Learners enter words in the blanks to complete accurate multi-word combinations, also known as collocations. This activity helps to improve your understanding of how native speakers and writers of the language you are studying combine words, also known as collocations. Type: Individual</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructions: Enter the missing words into the blanks.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>3. Scrambled Paragraphs</strong></td>
<td>exercises</td>
<td></td>
</tr>
<tr>
<td>The paragraphs of a document are scrambled and students must sort them into their original order. This activity helps to improve your understanding of paragraph structure, using authentic texts for you to practice on. Type: Individual</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructions: Put the paragraphs in the correct order by dragging and dropping them into the correct position. Click on the ‘Check Answer’ button at any time during the activity to see which paragraphs are in the correct order.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>4. Split Sentences</strong></td>
<td>exercises</td>
<td></td>
</tr>
<tr>
<td>A sentence is split in half and students must match the halves together. This activity helps you study sentence structure. Type: Individual</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructions: Reconnect the sentence halves by dragging and dropping them into the correct position. Click on the ‘Check Answer’ button at any time during the activity to see which sentences have been correctly reconnected.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
FLAX Activities Continued

5. Collocation Dominoes

This activity mimics the dominoes game and is populated by language from the British National Corpus (BNC) of 100 million words. The last word of the previous collocation becomes the first word of the next collocation. Here is an example: bank cheque -- cheque book -- book club -- club sandwich -- sandwich board -- board room ...

This activity helps to improve your understanding of how native speakers and writers of English in the BNC corpus combine words, also known as collocations.

Type: Individual

Instructions: Drag the words onto the domino ends to complete the sequence. Click on the ‘Check Answer’ button at any time during the activity to see which collocations have been correctly made into collocations dominoes.

6. Related Words

Two or three related words are presented in groups to learners to choose the correct word to complete collocations. For example:

pay make

_____ bill, _____ effort, _____ debt, _____ difference

This activity helps you to distinguish how related words are used and combined (collocated) by native speakers and writers of English in the BNC corpus of 100 million words.

Type: Individual

Instructions: Drag and drop each related word into their correct (collocational) positions to make complete phrases. Click on the ‘Check Answer’ button at any time during the activity to see which related words have been correctly made into collocations.

7. Collocation Guessing

In this game-based activity words and phrases with a missing blank will appear on the screen and start moving downwards while learners enter guessed words to try and complete the collocations before they reach the bottom of the screen. Here is an example of the types of moving target words and phrases: plain _____, dark _____, white _____, bitter _____, milk _____, bar of _____.

Learners must guess one keyword that collocates with all of them (The answer is obvious to chocoolics).

This fun activity helps you to interact with a wide range of collocations in English using the same keywords and is populated by the BNC corpus of 100 million words.

Type: Individual

Instructions: Select the ‘Start Question’ button to begin the collocations guessing game. As words and phrases with blanks start moving down your screen enter words into the ‘Guessed Word’ box to try and complete the collocations. When you have entered the correct word the collocations will stop moving down your screen and a ‘Show Remaining Collocations’ button will appear for you to learn more collocations using the same word in context from the BNC.
FLAX Across Platforms

- FLAX Website flax.nzdl.org for hosting open online language collections
  - Building directly onto the Web with OER
- FLAX multilingual open-source software for download
  - Set up your own FLAX server online or;
  - Build collections offline for use on your PC
- FLAX Android app for download
  - Interact with game-based FLAX collections while on the go
- FLAX for MOODLE plug-in for download
- FLAX for MOOC Platforms?
- FLAX in conjunction with translation technologies?
Training Videos for FLAX on YouTube

https://www.youtube.com/user/bananakiwiful/videos
DOMAIN-SPECIFIC OPEN LANGUAGE COLLECTIONS BUILDING
Demo Law Collections in FLAX

**British Law Report Corpus (BLaRC)**

**About this collection**

The British Law Report Corpus (BLaRC) is an 8.85 million-word legal corpus of 1,228 judicial decisions issued between 2008 and 2010 by British courts and tribunals. It was compiled and classified by Dr. Maria Jose Marin, a legal English lecturer with the LACELL research group at the University of Murcia, Spain.

The BLaRC is structured into five main sections reflecting the different jurisdictions of the British judicial system, that is, the geographical scope of its courts and tribunals: a) Commonwealth countries; b) United Kingdom; c) England and Wales; d) Northern Ireland; e) Scotland. Additionally, each corpus section is divided into different sub-sections coinciding with the hierarchical structure of the courts and tribunals comprised therein. By maintaining this structure, the texts are grouped according to the field of law they belong to (but for the Supreme Court, most courts and tribunals are organised according to the branch of law they pertain to, i.e. criminal law, family law, commercial law, intellectual property, right law, etc.), hence the similarity of their lexicon. Therefore, comparing results by studying the sections separately could prove useful and responsive to thematic criteria, which is fundamental as far as the identification and study of the specialised vocabulary of this legal English genre is concerned.

All United Kingdom Crown Copyright content re-used for educational and research purposes in this FLAX BLaRC language collection was derived from freely available judicial reports on the British and Irish Legal Information Institute (BAILII) website, and is available under the Open Government Licence v2.0 except where otherwise stated.


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)
# 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 - OER)</td>
<td>15 Lectures (Oxford Law Faculty and the Centre for Socio-Legal Studies)</td>
</tr>
<tr>
<td>MOOC lecture transcripts &amp; videos (streamed via YouTube &amp; Vimeo - OER)</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>PhD Law theses (Open Access)</td>
<td>50 EThoS Theses at the British Library (Abstracts, Introductions, Conclusions)</td>
</tr>
<tr>
<td>British Law Report Corpus (BLaRC) by Marin, 2012 (Open Access)</td>
<td>8.85 million-word corpus derived from freely available legal content on the BAILII website</td>
</tr>
<tr>
<td>Research Articles (Open Access)</td>
<td>40 Articles (DOAJ - Directory of Open Access Journals)</td>
</tr>
</tbody>
</table>
Working with Full Texts
The Powers of Parliament

What I want to do now is talk in much more detail about parliamentary sovereignty. What I'm going to do in order to elaborate this argument is make reference to one of the major writers on the British Constitution, Alber. Parliamentary sovereignty (also called the sovereignty of parliament, parliamentary supremacy, or legislative supremacy) is a concept in the constitutional law of some parliamentary democracies.

Related topics in Wikipedia
- Factortame litigation
- European Communities Act 1972 (UK)

Successor or bind a successor. Thirdly, no second, the British constitution.
Domain-specific Collocations

We focus on lexical collocations with noun-based structures because they are the most salient and important patterns in domain-specific text.

Collocations from the English Common Law MOOC:

• verb + noun   e.g. abolish judicial review
• noun + noun   e.g. precedent case
• adjective + noun   e.g. common law
• noun + of + noun   e.g. court of appeal
Lexical Bundles

“Lexical bundles” are multi-word sequences with distinctive syntactic patterns and discourse functions that are commonly used in academic prose (Biber & Barbieri, 2007; Biber et al, 2003, 2004).

Bundles from British Law Report Corpus (BLaRC):

• noun phrase + of  
  e.g. In the course of his
• prepositional phrase + of  
  e.g. on the part of the
• it + verb/adjective phrase  
  e.g. it is common ground that
• be + noun/adjective phrase  
  e.g. be taken into account in
• verb phrase + that  
  e.g. There is no doubt that
Lexical Bundles

- It seems to me that (205)
- It is common ground that (141)
- I do not consider that (115)
- I do not think that (101)
- In the light of the (77)
- There is no doubt that (66)
- In the present case the (62)
- It is true that the (62)
- I do not accept that (61)
- It is clear that the (59)
- There is nothing in the (58)

There is nothing in the decision which deals in an adequate manner with this important topic.

There is nothing in the papers to indicate what actual steps would have been involved in preparing any information or data to be supplied by the Agency in response to the request by Mr Collie.

There is nothing in the papers indicating what those terms are.

There is nothing in the latter point as under the First-Tier Tribunal procedural rules representatives of the Commission are entitled to attend hearings as the decision maker is a respondent and so a party to the appeal and hence has a right to be heard (Tribunal Procedure (First-tier Tribunal) (Social Entitlement Chamber) Rules (2008/2685), rules 1(3) and 28).

There is nothing in the answers to suggest that he envisaged using either, or both, of the batons in such circumstances.

There is nothing in the reasons for decision to indicate how the tribunal reached the conclusion that such supervision was not required.
## Collocations Within Collections

**Top 100 collocations**

<table>
<thead>
<tr>
<th>Noun+Noun (100)</th>
<th>Adjective+Noun (100)</th>
<th>Noun+of+Noun (100)</th>
<th>Verb+Noun (100)</th>
<th>Verb+Preposition+Noun (100)</th>
<th>Adjective+to+Verb (100)</th>
<th>Adjective+Preposition+Noun (100)</th>
</tr>
</thead>
<tbody>
<tr>
<td>error of law (460)</td>
<td>findings of fact (357)</td>
<td>cause of action (346)</td>
<td>balance of probabilities (320)</td>
<td>grounds of appeal (316)</td>
<td>ground of appeal (292)</td>
<td>point of law (276)</td>
</tr>
<tr>
<td>circumstances of the case (258)</td>
<td>facts of this case (245)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- In my view, Ms. Lambert was correct to submit that on the **facts of this case** that is not an available option.
- I believe that a similar arrangement is presently operated on an informal basis, and an undertaking would provide some added value to the **facts of this case** which would in my view be of value.
- "(ii) Can Essex be said to have denied the Appellant's right to an education under A2P1 on the **facts of this case**?"
- The evidence adduced by NAS, and the **facts of this case**, suggest that there are insufficient trained staff in the education system; the **facts of this case** indicates the schools do not have the requisite quality and expertise to cater satisfactorily for the demands made by children with ASD.
- Had part of the delay in this case been caused by maladministration I would not, on the **facts of this case**, have held that this amounted to a denial of A's right to education under A2P1.
- The third question put to us was whether Essex can be said to have denied the Appellant's right to an education on the **facts of this case**.
- Other members of this Court have amply set out the **facts of this case**, all the relevant jurisprudence both domestic and from Strasbourg, and indeed the full terms of article 2. None of this need I repeat.
- The **facts of this case** are now before us in very considerable detail – much additional material having been adduced even since the Court of Appeal hearing.
- That does not seem to me to be a possible conclusion, on the **facts of this case**.
Linking to the FLAX Learning Collocations Collection (Wikipedia, BNC, BAWE)

### Learning Collocations

**Family words:** *cases*

#### used as a noun

<table>
<thead>
<tr>
<th>adjective + case</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>special case</td>
<td>2833</td>
</tr>
<tr>
<td>latter case</td>
<td>1075</td>
</tr>
<tr>
<td>particular case</td>
<td>772</td>
</tr>
<tr>
<td>general case</td>
<td>569</td>
</tr>
<tr>
<td>criminal case</td>
<td>511</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>noun + case</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>court case</td>
<td>2147</td>
</tr>
<tr>
<td>case studies</td>
<td>1223</td>
</tr>
<tr>
<td>murder case</td>
<td>1059</td>
</tr>
<tr>
<td>case law</td>
<td>972</td>
</tr>
<tr>
<td>landmark case</td>
<td>361</td>
</tr>
<tr>
<td>facts of the case</td>
<td>295</td>
</tr>
</tbody>
</table>

**Synonyms**

- legal case
- landmark legal case

- He was the central figure in a **landmark legal case** that took 13 years to settle.
- In 1986, FitzGerald was the defendant in a **landmark legal case** brought by Irish tobacco company Carroll Group.
- He was defendant in a **landmark legal case** in American jurisprudence that determined truth was a defense against charges of libel.
- Loizidou v. Turkey is a **landmark legal case** regarding the rights of refugees wishing to return to their former homes and properties.
- In 1998, Zednet was thrown into the media spotlight through what nearly became a **landmark legal case** testing UK copyright laws on the internet.
- The company was involved in a **landmark legal case**, Burmah Oil Co. v Lord Advocates, concerning the destruction of oil fields in Burma by British forces.
- United States v. Reynolds, was a **landmark legal case** in 1953 that saw the formal recognition of the State Secrets Privilege, a judicially recognized extension of presidential power.
- As a result of this demonstration and others led by the local Corps, The Salvation Army v. Zednet was thrown into the media spotlight through what nearly became a **landmark legal case** testing UK copyright laws on the internet.
Chapter 3. FIFRA Amendments, The Founding of the EPA and Dietary Diversity

The FIFRA amendments in 1964 came after Rachel Carson’s book, Silent Spring, raised the alarm and caused the population to be quite upset about pesticide residues, particularly their effect on wildlife, but also growing recognition that these chemicals could build in the human body. And also the Food and Drug Administration’s admission that they had found pesticides in human breast milk as early as 1952. The public wasn’t warned about this. And basically, if you find a chemical, regardless of what it is, you find it in another species of mammal’s breast milk, you can presume that it’s likely to get into human breast milk as well.

So Rachel Carson’s Silent Spring turned out to be a real watershed, not just legally for pesticides, because it really increased the sense of susceptibility to biocides or the economic poisons, but it really met with quite a bit of resistance in Congress. Again, this was the end of the nuclear weapons testing era in the atmosphere. And it was also a period of great unrest in the United States. The origin of the Civil Rights Movement may be traced to this period. The Civil Rights Act of 1964, recall that. Also, we were getting more deeply embroiled in the war in Vietnam at that point in time. And environmentalism was growing up, creeping up on the agenda. But it was really quite an interesting period. Congress, however, was preoccupied. Other than making some minor revisions to the statute that included adding these words: caution, warning, and hazard, depending upon the relative toxicity. This didn’t really help very much because of public confusion about what those phrases meant. And the Department of Agriculture’s secretary was given authority finally to remove pesticides from the market based upon a finding of imminent hazard to public.

Now, EPA was created in 1970, and it was given the responsibility to manage pesticides and it was consolidated from other agencies. Some fifteen or sixteen different subunits of different
RESEARCHING RESOURCES AT THE INTERFACE OF OPENNESS FOR ACADEMIC ENGLISH
Key Research Data Sets:

- Data for evaluation and impact of FLAX open language collections on learning and teaching
  - Survey and Think-Aloud Protocols to evaluate the FLAX Language System user experience
  - Analysis of written and spoken language
  - Interview and focus-group data (f2f and online)
    - stakeholders (language teachers, academics, MOOC providers) involved in the development of the academic language collections used in this research.
References


Thank You

FLAX Language Project http://flax.nzdl.org/
Shaoqun Wu: shaoqun@waikato.ac.nz
Alannah Fitzgerald: a_fitzg@education.concordia.ca
Ian Witten: ihw@cs.waikato.ac.nz

OER Research Hub http://oerresearchhub.org/
TOETOE Technology for Open English Blog: www.alannahfitzgerald.org
Slideshare: http://www.slideshare.net/AlannahOpenEd/
Twitter: @AlannahFitz