Investigating an open methodology for designing domain-specific language collections

http://opendesignnow.org/index.php/visual_index/events/
EuroCALL 2014

• Development of Tools and Language Corpora
  – Design-Based Research with the FLAX Project

• Openness in Corpus-Based Tools, Resources & Practices

• Research with Corpus-Based Approaches
  – Critical Reflection by Language Teachers with Open Do-It-Yourself ESAP Language Collections
WHO ARE WE IN THIS FLAX RESEARCH & DEVELOPMENT COLLABORATION?
FLAX Language at Waikato University
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 Corpora with FLAX

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/
Research with Queen Mary U. of London

http://language-centre.sllf.qmul.ac.uk/home
Be Free to Do Whatever You Want!

- Open Resources for ESAP Soup Dragons:
  - Building & Sharing Open ESAP Corpora to Promote DIY Corpus-Based Approaches
  - Developing Automated 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

http://en.wikipedia.org/wiki/The_Soup_Dragons
OPEN SOURCE LANGUAGE TOOLS DEVELOPMENT
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

1. **Word Guessing**
   - **Type:** Individual or Group collaboration
   - **Instructions:** Click on the gap and type in a word. Click on the light bulb icon (if any) for help.

2. **Completing Collocations**
   - **Type:** Individual
   - **Instructions:** Enter the missing words into the blanks.

3. **Scrambled Paragraphs**
   - **Type:** Individual
   - **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.

4. **Split Sentences**
   - **Type:** Individual
   - **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.
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 chocoholics).

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](http://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
    • Using All Rights Reserved content
• FLAX for MOODLE plug-in
• 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
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)
Domain-specific Collocations

We focus on lexical collocations with noun-based structures because they are the most salient and important patterns in topic-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 OpenYale Environmental Politics and Law:

• **noun phrase + of**
  
  *e.g.* So the idea of

• **prepositional phrase + of**
  
  *e.g.* on the part of

• **it + verb/adjective phrase**
  
  *e.g.* it’s going to be

• **be + noun/adjective phrase**
  
  *e.g.* is an example of

• **verb phrase + that**
  
  *e.g.* of the way that
## 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>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 - 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-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

- Achieving Ecological Objectives
  - Adaptation to Climate Change: Legal Challenges for Protected Areas
  - A Cayman Islands Perspective on Transborder Insolvencies and Bankruptcies: The Case for Judicial Co-Operation
  - A Cayman Islands Perspective on Transborder Insolvencies and Bankruptcies: The Case for Judicial Co-Operation Managing Ethical Risks and Crises: Beyond Legal Compliance
  - Climate Justice: A Constitutional Approach to Unify the Lex Specialis Principles of International Climate Law
  - Decomposing US Money Supply Changes since the Financial Crisis

- Achieving Ecological Objectives
  - Abstract
  - 1. Introduction
  - 2. Quality Standards
  - 3. References and Baselines - Essentiality Confused
  - 4. QualityEssential?
  - 5. Temporality
  - 6. Visualizing Consequences
  - 7. Conclusion
  - References
Collocations Within ESAP Collections

153 collocation(s) associated with the word environmental

Adjective (80)  Noun + of (38)  Verb (35)

- environmental law (44)
- environmental quality (37)
- environmental science (10)
- environmental impact statement (9)
- environmental tobacco smoke (8)
- environmental effects (6)

- Now, I'd like to close by having you think about damages and how we calculate damages. And as you think about every case that we've discussed in this term, when we think about environmental effects, you should be thinking about the magnitude of the effect and you should be thinking about the distribution of the effect. So think about this comparison between nuclear power on the one hand and wind energy on the other.

- And what we'll see today is that knowledge about the environmental effects of national security are produced predominately by the Defense Department. They control sites, they control the technology in weaponry and weapon delivery systems, making it extremely difficult for the public to understand really what the dangers might be.

- And the Environmental Protection Agency basically shuts down the public from understanding what that is. And it's all about how much of the chemical is produced, what the company knows about where it goes, and also what the health effects or environmental effects might be. So access to data, intellectual property rights, secrecy and confidentiality, these are all things that we really need to think about very carefully when we structure environmental law.

- So we could get a whole new level of understanding of energy and the environmental health of different products, including issues such as protein content, fat content, type of fat, amount of salt, et cetera. Right now, most of us walk through the marketplace really blind to these issues, blind to the environmental effects of food production, blind also to the energy consumption.
# Learning Collocations

## Linking to the FLAX Learning Collocations Collection (BNC, BAWE, Wikipedia)

### Family words:
- environment
- environmentalist
- environmentalists
- environmentally
- environments

### used as an adjective

<table>
<thead>
<tr>
<th>Environmental + noun</th>
<th>Count</th>
<th>Environmental concerns</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>environmental issues</td>
<td>2088</td>
<td></td>
<td></td>
</tr>
<tr>
<td>environmental protection</td>
<td>1146</td>
<td></td>
<td></td>
</tr>
<tr>
<td>environmental impact</td>
<td>882</td>
<td></td>
<td></td>
</tr>
<tr>
<td>environmental factors</td>
<td>830</td>
<td></td>
<td></td>
</tr>
<tr>
<td>environmental conditions</td>
<td>794</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Adjective + environmental</th>
<th>Count</th>
<th>Environmental concerns</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>social and environmental</td>
<td>79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>environmental and social</td>
<td>33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>various environmental</td>
<td>31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>local environmental</td>
<td>27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>significant environmental</td>
<td>26</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- One of the major environmental concerns facing the park is illegal fishing opera Chamo.
- The hills are mostly forested, although deforestation and the formation of fire-sava major environmental concerns in Trinidad.
- Though the route was basically set in stone in Connecticut, many issues remain biggest of which were major environmental concerns about how the freeway will Reservoir, which is the main drinking water supply for Providence.
- In order to bring global concerns to the forefront of public attention, Pyle has work collaborative relationships between major environmental concerns such as The
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 1962. 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 enmeshed 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
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 [45]. 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 [3,9,11].

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 [10,12,14]. The establishment of quality standards signifies 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 social and ecological systems, quality standards appear both as aquifer habitats, or mitigation of organic pollution, given the propensities [1,15,16]. This results in a legal instrument for a 'good ecological status'. It becomes important to focus on, and qualify dynamic, and complex ecosystems as though they were individual of biodiversity and ecosystem services, may be masked by the problemmatic aspects of quality standards emphasize the land, and ecological law, focused on the living components of the Directive, by scrutinizing the essential the

The precautionary principle or precautionary approach states that if an action or policy has a suspected risk of causing harm to the public or to the environment, in the absence of scientific consensus that the action or policy is harmful, the burden of proof that it is not harmful falls on those taking the action.

Related topics in Wikipedia
- Earth Summit
- Biosafety
- Biosecurity
- Rio Declaration on Environment and Development
- Genetically modified organism
Lexical Bundles

So I'm going to (22)
So that if you (16)
So the idea that (14)
And by the way, (13)
And I wanted to (13)

And I wanted to use this as a basis for reviewing a variety of different laws that apply to certain key provisions in the statutes that have met with more or less success.

And I wanted to just pause on this, because many people think that the idea of sustainability is fairly new.

And I wanted to just spend a moment to give you a sense of the kinds of problems that are experienced out in Wyoming in the northwestern corner known as the greater Yellowstone ecosystem.

And I wanted to spend just a couple minutes talking about course requirements.

And I wanted to come back to this, because I didn't explain it very carefully the other day.

And I wanted to refresh your memory about this idea that this industry evolved with the clear recognition that the product that they were selling was not aesthetically appealing on its own.

And I wanted to reemphasize a couple of points here.

And I wanted to remind you that next week, we're moving on to air quality.

And I wanted to make one thing clear about this.

And I wanted to bring your attention to the Wilderness Act of 1964.

And I wanted to start with this set of images from Thomas Cole, a well-known landscape artist from the nineteenth century, one of the Hudson River School painters.

And I wanted to mention a couple of kind of relatively recent technological designs.

And I wanted to pick up on some of the questions that were raised on Tuesday with respect to the property regime that was created by the State of New York in the
Hello. I'm Terry Fisher. This is the introduction to a series of 12 lectures on copyright. My hope in these lectures is to provide you with the following things. First, an understanding of the basic principles of copyright law.

Second, an appreciation of the ways in which that law affects, for better and worse, creativity and innovation in a wide variety of artistic and technical fields. To that end, I'll be providing along the way illustration of the ways that copyright works in literature, music, film, photography, graphic art, software, comedy, fashion, architecture, and so forth.

Third, I will try to provide you with a critical understanding of the main theories of copyright, by which I mean the arguments developed over several centuries by economists, philosophers, and political theorists concerning the purposes or functions of the copyright system. These are not near matters of academic curiosity and dispute. Copyright law is changing very fast, in part in an attempt to address new technologies. And the lawmakers who are making those changes are influenced by and in turn help to shape the competing theories of copyright. So to understand where the law is going, and certainly if you wish to understand its trajectory, you need to know a fair amount about theory. This brings me to the fourth.

Copyright law is changing very fast, in part in an attempt to address new technologies. And the lawmakers who are making those changes are influenced by and in turn help to shape the competing theories of copyright. So to understand where the law is going, and certainly if you wish to understand its trajectory, you need to know a fair amount about theory. This brings me to the fourth and last of the ambitions.
RESEARCHING RESOURCES AT THE INTERFACE OF OPENNESS FOR ACADEMIC ENGLISH
Key Data Sets Will Consist Of:

• Data for evaluation of collections and classroom teaching derivatives of the collections for ESAP
  – Survey and Think-Aloud Protocols to evaluate the FLAX Language System
  – Timed-writing exam preparation (Queen Mary University of London).

• Interview and focus-group data (f2f and online via Skype)
  – With stakeholders (language teachers, academics, MOOC providers) involved in the development of the academic language collections used in this research.
References


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