Open Educational Principles for Designing & Developing Digital Language Learning & Teaching Collections

Created by Alannah Fitzgerald
Licensed for reuse under Creative Commons Share Alike
Workshop overview

- Open Corpora
  - FLAX, BNC, UKWAC

- Open Concordancers
  - FLAX, webBNC

- Open Collections
  - FLAX web collocations, phrases, pronoun phrases

- Open Language Teaching Resources & Communities
  - LORO, HumBox

- Open Licensing
  - Creative Commons, licensing scenarios
Support Centre for Open Resources in Education (SCORE)
Open corpora

UKWAC
FLAX
BNC
UKWAC
Focus on archiving

The UK Web Archiving Consortium (UKWAC)
Preservation of UK web activity using historical web dumps of content with the .uk domain (Bailey & Thompson, 2006)
Language learning resources?
About the BNC

The British National Corpus (BNC) is a 100 million word collection of samples of written and spoken language from a wide range of sources, designed to represent a wide cross-section of current British English, both spoken and written. [more]

Search the Corpus

Type a word or phrase in the search box and press the Go button to see up to 50 random hits from the corpus.

Look up: Enter text Go

You can search for a single word or a phrase, restrict searches by part of speech, search in parts of the corpus only, and much more.

The search result will show the total frequency in the corpus and up to 50 examples. [more information]

News from the BNC

- BNC sound files available
Focus on representation
The British National Corpus (BNC)
100 million-word static corpus 1978-1992
Spoken (10%); Written (90%); Domain representation
Flexible Language Acquisition Project (FLAX)

"... learning software as effective as a personal tutor ..." (from President Obama's address to the National Academy of Sciences, 27 April 2009)

Flexible Language Acquisition Project

Watch a video about FLAX! (38 mins 99 MB), or right click to download an AVI file (170 MB)

FLAX is now in Moodle. Try it out here

Objective: to automate the production and delivery of practice exercises for overseas students who are learning English. The exercises will involve students in a virtually endless supply of collaborative and competitive language activities that are interesting, compelling, and rewarding.

Strategy: to deploy digital library software to allow teachers and students to capitalise on top-quality prose and multimedia resources already present in the world's libraries. This yields an unprecedented supply of linguistic material for students to practise on.

Presentation: to provide a web-based social setting, matching in real time students in different locations who opt for a particular type of exercise, allowing them to discuss and negotiate its parameters using chat, and undertake activities that are competitive or collaborative.

License: All software produced by this project is open source issued under the GNU General Public License.

Focus on language learning
The Flexible Language Acquisition project (FLAX)
Greenstone

- Produced by the New Zealand Digital Library Project at the University of Waikato
- Developed and distributed in cooperation with UNESCO and the Human Info NGO
- Open-source, multilingual software
- Issued under the terms of the GNU General Public License.

http://www.greenstone.org/
The Flexible Language Acquisition Project (FLAX) by Shaoqun Wu & Ian Witten

Featuring FLAX collections for language learning and teaching
FLAX makers

How the FLAX open corpora were made:

Open concordancers

webBNC
FLAX
BNCweb concordancer – free download

<table>
<thead>
<tr>
<th>Query options</th>
<th>Standard query</th>
<th>Written restrictions</th>
<th>Spoken restrictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>User-specific functions</td>
<td>User settings</td>
<td>Query history</td>
<td>Saved queries</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Categorized queries</td>
<td>Make/edit subcorpora</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Upload external data file</td>
<td>Additional functions</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Browse a text (disabled)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Scan keywords/titles</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Explore genre labels</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Frequency lists</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Keywords</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>About BNCweb</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Standard Query</th>
<th>shall we</th>
</tr>
</thead>
<tbody>
<tr>
<td>Query mode:</td>
<td>Simple query (ignore case)</td>
</tr>
<tr>
<td>Number of hits per page:</td>
<td>50</td>
</tr>
<tr>
<td>Restriction:</td>
<td>Spoken Texts</td>
</tr>
</tbody>
</table>

Simple Query Syntax help

BNCweb (CQP-edition) © 1996-2008

http://bncweb.info/
### webBNC header information

<table>
<thead>
<tr>
<th>Title:</th>
<th>public context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spoken or Written:</td>
<td>Spoken</td>
</tr>
<tr>
<td>Number of Words (tagged items):</td>
<td>4,692</td>
</tr>
<tr>
<td>Average sentence length (&lt;w&gt;-tags per &lt;s&gt;-unit):</td>
<td>16.6975</td>
</tr>
<tr>
<td>Derived text type:</td>
<td>Other spoken material</td>
</tr>
<tr>
<td>Genre:</td>
<td>S: speech: unscripted</td>
</tr>
<tr>
<td>Text type:</td>
<td>Context-governed</td>
</tr>
<tr>
<td>Publication date:</td>
<td>1985-1993</td>
</tr>
<tr>
<td>Domain (spoken context governed texts):</td>
<td>Public/Institutional</td>
</tr>
<tr>
<td>Type of Interaction:</td>
<td>Monologue</td>
</tr>
<tr>
<td>Region where spoken text was captured:</td>
<td>Midlands</td>
</tr>
<tr>
<td><strong>Speakers</strong></td>
<td></td>
</tr>
</tbody>
</table>
FLAX vs BNC?

- Limitations with representativeness
  - Identifying register on the Web is difficult (Biber & Kurjian, 2007)
  - Successful corpora are based on domains, genres, collections of document types
  - The web is a “dirty corpus” (Kilgariff & Grefenstette, 2003)

- FLAX cleaned by 30% using BNC wordlist
  - Linked externally to BNC, Yahoo
    - Complementary sources, both with limitations
Open Collections

FLAX web collocations, phrases, pronoun phrases
FLAX web phrases collection

- Web n-gram corpora generated and supplied by 2006 Google web dump
- 500,000 words and 380 million five-grams
- Check function for student writing
  - Textbox type or cut n paste
  - GALL - Google Assisted Language Learning (Chinnery, 2008; Shei, 2008)
Enter a sentence in the box

1. please remove the brackets in the text.
2. please do not use hyphenated words (e.g. use "online" or "on line", not "on-line").

As the internet become all pervading, therefore human beings lives are changed.

Your sentence: As the internet become all pervading, therefore human beings lives are changed.

Parsed sentence: [ As the internet ][ become ][ all pervading ][ therefore human beings lives ][ are changed ]

As the internet become all occur 0 time(s) in the collection

As  ......................  200,000,000
As the  .....................  15,000,000
As the internet .................  22,000
As the internet become .................  0

search for phrases following As the internet

- Verb
  + As the internet improves (5,700)
  + As the internet evolves (3,000)
  + As the internet cultivates (2,900)
  + As the internet develops (1,600)
  + As the internet expands (1,400)
  + As the internet has (730)
Web Phrases OER

Library

Web Phrases

The web phrases collection was built with two- to five-words phrases gathered from the web, including:

1. 145,000 unique words
2. 14 million two-words phrases
3. 420 million three-words phrases
4. 500 million four-words phrases
5. 380 million five-words phrases

In this collection, you can

1. search for phrases containing particular word(s) in the natural or reversed order
2. check the popularity of your text

http://www.youtube.com/watch?v=n67FBqBFm6I
Google’s terms of services

“You agree not to access (or attempt to access) any of the Services by any means other than through the interface that is provided by Google, unless you have been specifically allowed to do so in a separate agreement with Google.”

http://www.google.com/accounts/TOS Clause 5.3
Web Pronoun Phrases

Search for **phrases containing** the word(s) **argued** in **phrases**

argued

- synonyms: noun, adjective, verb or adverb
- antonyms: noun, adjective, verb or adverb
- related words: noun, adjective, verb or adverb
- associated words
- collocations

102 documents matched the query.

**Modal (7)**

- It could be argued that ... (150,000)
- It can be argued that ... (120,000)
- It may be argued that ... (39,000)
- It might be argued that ... (34,000)
- It can also be argued ... (8,200)
So, on this basis, *it could be argued that* the *.. it could be argued that*. Efficient management of human resources, likewise, requires *..*

*.. basis, it could be argued that* the most efficient management of *.. it could be argued that*. Support comes from *.. references to expert opinion and ..*

*.. it could be argued that* diverting resources from the pink market in order to *.. While it could be argued that* eventually, those markets would grow to be able to *..*

*.. Allowing families to split incomes for tax purposes will cost the country *.. "It could be argued that* packages like Working for Families and other tax reforms *.."

*.. It could be argued that* Big Boi's first solo album was his half of Outkast's double, Speakerboxxx/The Love Below. *.. Film review: The Hopes & Dreams Of Gazza *..

*.. Paul Taggart It could be argued that* Westpac's criticism of the Labour Party *.. It could be argued that* Westpac's criticism of the Labour Party proposal to *..*

*.. It could be argued that* this home has been built on the best section on the *..*

*.. By reviewing the requirement for bylaws within the district, not only does *.. it could be argued that* environmental concerns could become evident should the *..*
FLAX web pronouns & web collocations collections

- Linking to individual web pages from the aggregations in the n-gram corpus
  - often unclean,
  - incomplete and,
  - repetitive

= “typographic salience” (Bishop, 2004)
Glenny's most amazing suggestion is that the recognition of Croatia and Slovenia deepened the problem. He does not explain that the fighting intensified because the army and Serbia wanted to discourage the recognition by escalating the conflict. It could be argued that the west lost the golden opportunity to contribute to the peaceful solution of the Balkan crisis because it did not de-recognise Yugoslavia (and, with it, outlaw the Yugoslav army), while recognising all the six former Yugoslav republics before the war started in earnest. This measure may well have to be considered now (as Washington is hinting), as it has become crystal-clear that the Bosnian war is being fomented by Milosevic and the extreme Serb leaders in Bosnia, and is primarily waged by the army.

The analysis of Lloyd LJ in Aswan was an exhaustive one of previous authorities and it is worth noting. His Lordship maintained that s14(6) did not alter the law and that "purpose or purposes" referred to the range of uses of the contract goods but not that they had to be fit for all these purposes. The question of suitability for a particular purpose was to be judged, as anticipated by s14(6), by reference to description and other relevant circumstances such as price. This approach is appropriate to a commercial buyer in that it forces the buyer to reveal his requirements if the purpose is over and above the most basic one for goods of that kind. Nevertheless, there is an element of incongruity here since the requirement of fitness for purpose is made the basis of s14(2) as well as s14(3). (e) Merchantability and quality The emphasis in s14(6) upon function and purpose can in one sense be regarded as undermining an important element in merchantability, namely, quality. Of course, it could be argued that purpose is not confined to use in a functional sense but also encompasses the enjoyment which the buyer can reasonably expect from his purchase. This is reflected in Mustill LJ's approach in Rogers v Parish (Scarborough) Ltd [1987] QB 233 where his
Technology for Open Education – Training with Open E-resources (TOE TOE)

Featuring OER training for language learning & teaching with open corpora & open tools
exercise one: collect useful phrases and structures using the Web Collocations and Phrases

topic

The threat of nuclear weapons maintains world peace. 
Nuclear power provides cheap and clean energy.

The benefits of nuclear technology far outweigh the disadvantages

1. Gather useful phrases related to the topic using the Web Collocations

   1. pick some keywords, for example threat, nuclear, weapons, energy, benefits, disadvantages
   2. search for collocations using those words.
      - go to Web Collocations
      - click search
      - type in nuclear
      - click show collocations
      - click nuclear weapons 
      - click nuclear weapons again to see Web samples
      - click the icon that follows nuclear weapons to see BNC samples
      - type in benefits
      - click show collocations
      - look up the collocations that of your interest

3. try out other words yourself
4. write down those you think might be useful
User guides & demos

http://flax.nzdl.org/resources/flax_video/flax_video.html
Teacher-created exercises

http://flax.nzdl.org/resources/flax_video/flax_video.html
Future of FLAX

FLAX is a fast-moving research and development project

Coming up …

... a web interface to cut and paste your documents into
... more activity types

FLAX: Creating practice exercises for language learners from digital libraries

Free open source software from http://flax.nzdl.org

http://flax.nzdl.org/resources/flax_video/flax_video.html
Web Pronouns Phrases OER

This collection contains five-words phrases that start with a pronoun word e.g. I, he, she, they, you, we and it. It is based on a massive collection of fragments of text gathered from the web, along with occurrence frequencies. You can study how a particular word is used in a pronoun context by looking at its frequently reoccurring language patterns. Through browsing, you can find out which words are most commonly associated with a particular pronoun.

Do you know which word is the "best friend" of the pronoun word I? Check it out, you will be surprised.

This is a part of Shaoqun's PhD project, please contact Shaoqun Wu for more information.

http://www.youtube.com/watch?v=Ns4nXsZQmUA
Open Language Teaching Resources & Communities
LORO
HumBox
OER as catalysts to professional development

Anna Comas-Quinn
Lecturer in Spanish and SCORE Fellow
Department of Languages, Faculty of Education and Language Studies, The Open University
Programme for Arts, Humanities and Languages, 12 July 2011

Created by Anna Comas-Quinn
Benefits of using LORO?

- For teachers?
- For students?
- For institutions?
- From a professional development angle?

Created by Anna Comas-Quinn
Perceived benefits

- Professional development (feedback from colleagues)
- Time saving
- Student support
- Authorship & showcasing your work

Survey of OU language teachers July 2009 (N = 129)

Created by Anna Comas-Quinn
Perceived challenges

- Quality and usefulness of resources
- System must work (search, browse, structure, file formats)
- Time consuming
- Lack of remuneration
- Reciprocity
- Recognition and authorship

Survey of OU language teachers July 2009 (N = 129)

Created by Anna Comas-Quinn
Benefits of using LORO

- Increased confidence in one’s own practice
  “Seeing other work enables you to judge your own, and reassures you that you are doing the right thing”

- Freedom to develop other aspects of one’s teaching practice
  “It gives us time and space to create some individual styles”
  “I can concentrate on how I will teach culture or how to teach through the asynchronous forum”

Created by Anna Comas-Quinn
Welcome to LORO
Welcome to LORO at The Open University -
LORO connects the OU to the wider languages community.
You can store your own resources, share resources with the OU community or the external community.

http://loro.open.ac.uk/
Welcome to HumBox

The HumBox is a new way of storing, managing and publishing your Humanities teaching resources on the web. Share handouts, exercises, podcasts, videos and anything else you can imagine!

Sign up for a free account, or take a look at what's already inside.

Recently uploaded resources

- Perspectives on International Politics: Discourse on Democracy and Diplomacy
  - Added On: 20 Feb 2011 11:03
  - Added By: Mr. Dave Savard
  - Tags: Democracy And Diplomacy, Ukooer
  - Languages: English

- Curriculum Vitae
  - Added On: 20 Feb 2011 10:57
  - Added By: Mr. Dave Savard
  - Tags: Dave Savard, Ukooer
  - Languages: French

Calling all HumBox users!

If you have used a HumBox resource in your teaching, we would love to know more about how you used it! What did your students think of it? How did you adapt it? Please tell us more about how you use HumBox by dropping us an email!

HumBox unaffected by cut of Subject Centres

HumBox came into being as the result of the effective collaboration of four Subject Centres and continues to be part of the activity of those centres. However, the Higher Education Academy, the body that manages the Subject Network, has decided to cease supporting the Network and so you may be worried about what this may mean for HumBox. Let me assure you that HumBox will continue! It will continue to be developed and managed by the University of Southampton, School of Humanities/School of Electronics and Computer Science, with no disruption of service at all.

http://www.humbox.ac.uk/
Web profiles

Mr Antonio Martínez-Arboleda
Senior Teaching Fellow
University of Leeds

I currently teach my specialist modules on Spanish Politics and Spanish in an Economic and Business Context. I am Programme Manager for our BAs in Spanish, Senior Language Co-ordinator, in charge of the whole Spanish Language Programme, and Instituto Cervantes Liaison Officer. I have been actively involved in the Humbox as an institutional partner representing the University of Leeds and I have been awarded a SCORE Fellowship (Support Centre for Open Resources in Education, based at the Open University) for the project "Review and Endorsement of Open Educational Resources by Graduate-Recruiting Employers".

Profile Details
- Name: Mr Antonio Martínez-Arboleda
- Institution: University of Leeds
- Occupation: Senior Teaching Fellow
- Interests: Autonomous Learning, OERs, Student Motivation, Employability, e-learning, Role-plays, Poetry, Translation and Performance, Spanish Politics, Spanish for Business and Spanish Language.
- Homepage: www.leeds.ac.uk/arts

Published Items
- Autonomous Learning Portfolio in Spanish Language
- Los miserables
- Professional Reported Speech in Spanish in a Political Context - C1-C2 Level
- Juan Ramón Barat - Selection of Poems Translated into English
- Enhancing Modern Languages Teaching: Student

Most Viewed Items
1. Attitudes towards Regional Autonomy in Spain: The Case of Lorca (Region of Murcia), Video 1 and Follow-up Video - Transcripts in English and Spanish - Summary of Methodology
   - Views: 4920
2. Attitudes towards Regional Autonomy in Spain: The Case of Lorca (Region of Murcia), Video 16 - Transcripts in English and Spanish - Summary of Methodology
   - Views: 2138
3. Los Cuatro Mulos - Spanish Popular Poetry in García Lorca’s work
   - Views: 3107
4. Attitudes towards Regional Autonomy in Spain: The Case of Lorca (Region of Murcia), Video 2 - Transcripts in English and Spanish - Summary of Methodology
   - Views: 2903
5. Attitudes towards Regional Autonomy in Spain: The Case of Lorca (Region of Murcia), Video 10 - Summary of Methodology - Transcripts in English and Spanish
   - Views: 2734

http://humbox.ac.uk/profile/12
Open Licensing with Creative Commons
Licensing scenarios
Copyright

http://www.flickr.com/photos/mikeblogs/3020966666/sizes/o/
Public Domain

http://www.flickr.com/photos/peterleth/4422431324/sizes/o/in/gallery-60810582@N04-72157626302755808/
Creative Commons

http://www.flickr.com/photos/peterleth/4422431300/in/gallery-60810582@N04-72157626302755808/
Attribution

http://www.flickr.com/photos/peterleth/4421664989/in/gallery-60810582@N04-72157626302755808/
Non Commercial

http://www.flickr.com/photos/peterleth/4421665015/sizes/o/in/gallery-60810582@N04-72157626302755808/
No Derivatives

http://www.flickr.com/photos/peterleth/442431384/sizes/o/in/gallery-60810582@N04-72157626302755808/
Lichôdmapwa v. Théâtre de Spa
Court of First Instance Nivelles
(Tribunal de Première Instance Nivelles)
26 October 2010

A Belgian band uploaded some songs on a freely accessible website under a non-commercial and no derivatives Creative Commons license. A Belgian theatre used one of the songs to create an advertisement for the next theatrical season, which was broadcasted on several national radios channels. The Court found that the theatre did not respect the license and consequently granted indemnities to the band.

Group work:

Read and discuss the following licensing scenarios as they would apply to language teaching and materials development practice.

(Adapted from copyright resources created by Bernie Atwell at the OU; adapted for language resource developers)
Use clearance

I’ve found an open access pre-publication article by Diane Nation on the web and this would be brilliant to exemplify some concordance lines from FLAX for teaching collocations. I intend to develop a language learning resource with these materials and then to upload it into LORO for open use. I’ve tried to contact Ms Nation twice and have been in touch with the web master of the site to see if s/he can help but have had no response so far. I’ve amended the article, as I didn’t agree with some of the points she was making. I think I’ve improved the work actually and I’ve obviously left her acknowledged as the author. As I’ve had no response I’m just going to use it anyway. Everyone’s always talking about risk so I’ll take one. Is this OK?
CC licensing worldwide

My institution has an online open learning resource and is based in the UK. We have selected an England and Wales UK licence for the use of our content. However, a user in China has asked us if the CC licence still applies? Does the CC licence refer to where the content is being used or where it is hosted?
Open software licenses

I have some software I would like to make available under a CC licence – would that be OK?
Logo protection

My institution is making some of its content available under a CC licence. How do we ensure that our trademarks/logos are protected?
Extended Licensing Scenario

The following scenario is intended to promote discussion around the areas of creative commons licensing, the collaborations involved, and any other issues the discussion may highlight.
Your educational institution is going to be working in collaboration with at least two other educational institutions in the UK.

You are going to create an innovative joint MA TESOL resource for Masters students studying and researching in the area of open corpora for teaching English for Academic Purposes (EAP). This facility will act as a provider of online resources. All institutions will provide some of their own existing materials that contain third party content (journal articles, images, extracts from books, and website content) which are made up of text and audio-visual content.

The collaboration would like to make the content openly available whilst ensuring that their intellectual property rights are not compromised.
Consider the following questions for discussion:

- How would you license this content to users?
- Would you consider using a Creative Commons licence, if so which one?
- Would you need to consider more than one type of licence?
- What would you need to take care of contractually in relation to the content?
- How would you ensure that the integrity of third party content is maintained?
References


Please contact us

- Alannah Fitzgerald (TOE TOE)
  a.t.d.fitzgerald@durham.ac.uk

- Shaoqun Wu (FLAX/GREENSTONE)
  shaoqun@cs.waikato.ac.nz

- Ian Witten (FLAX/GREENSTONE)
  ihw@cs.waikato.ac.nz

- Anna Comas-Quinn (LORO)
  a.comas-quinn@open.ac.uk