Academic English with the Electronic Theses Online Service (EThOS) at the British Library

Alannah Fitzgerald & Chris Mansfield
Workshop Overview

• FLAX Language Research & Development
  • Who we are

• Electronic Thesis Online Service (EThOS) at the British Library
  • Reuse of digital collections

• Abstracts
  • Tools for Search, Collocations, Word Lists, Lexical Bundles
  • Wikification and Linking to Open Resources
FLAX Language Project [flax.nzdl.org]
Greenstone Digital Library Lab
Waikato University NZ

Professor Ian Witten
FLAX Project Lead

Dr Shaoqun Wu
FLAX Project Lead Researcher & Developer
FLAX Open Language Research

Alannah Fitzgerald
FLAX Open Education Research
Concordia University

Chris Mansfield
Queen Mary Language Centre
University of London
FLAX DATA DRIVEN LANGUAGE LEARNING: MINING OPEN ACCESS PHD THESES FROM THE BRITISH LIBRARY
Indeed, by far the biggest impact of openness in the higher education sector has been with open access, showing the importance of government agencies in promoting accessible research (Finch Group, 2012) to ensure “enhanced availability of discoverable, reusable and repurposable academic open content.” (JISC, 2011)
EThOS at the British Library

http://ethos.bl.uk/Home.do;jsessionid=C0BD2D50495813E0DD83D5BD7E1341B6
Mahendra Mahey is the Project Manager of British Library Labs, an Andrew Mellon Foundation funded initiative which supports and inspires the use of the British Library's digital collections and data in exciting and innovative ways. British Library Labs, an extension of the Digital Research Team, encourages and supports scholars and innovators to work with us and our collections through competitions, awards and other engagement activities.
Alannah Fitzgerald: Here we are at the British Library on the 29th of October, 2016 in London. So, we’re going to talk about the EThOS collections and what your views are on reuse. I think you just started talking there about the complexity around the rights of reuse...complexity was the word that you used.

Mahendra Mahey: Okay, so there was a change in the law [2014], which allowed text and data mining. So, text and data mining for non-commercial purposes. [...] 

So, you know, obviously in terms of the projects I’ve worked on it’s all about trying to get, to open up digital collections for research and various other practices. So, when, obviously, when we saw that there was a change in the legislation that was being enforced, we were really excited about that.

And, one of the prerequisites we were kind of told we would have to do is to maybe do an internal project. As I was telling you earlier like dogfooding, to do an internal project which would allow us to try and do this.
**Mahendra:** What was agreed was that we would run internal experiments. However, we were told that we would need to know what the results of our questions would be before we did the experiments.

**Alannah:** Why?

**Mahendra:** Um, because they wanted to be able to understand...So, I think they would’ve been a bit more flexible because it was internal. But, especially, if a researcher from the outside wanted to do this they would need to know what they wanted to do with the data, and you know...especially around the idea of copying big chunks of it, that’s the biggest worry. So, that was probably the resistance about, you know, are you reproducing this stuff? Are you republishing this stuff? Because then you sort of get into a dodgy area.

PHD ABSTRACT COLLECTIONS: FLAX
Abstracts

• “gatekeepers” (Swales, 1990) of academic fields
• “sub-genre” (Swales and Feak, 2009)
• “self-promotional tools” (Hyland, 2000)
• Act as metadata (along with titles and keywords) for the improved searchability and ranking of a paper, thesis etc. via search engines
• Are often the only part of a paper read via abstracts databases
• Are often the only part of a paper that is accessible within subscription-based publications (Bordet, 2015)
flax.nzdl.org
Powerful yet simple interfaces for Data-Driven Learning
FLAX Academic English Collections

**PhD Abstracts Collections**

These collections come from the E-theses Online Service (EThOS) Open Access Initiative managed by the British Library.

- Arts and Humanities
- Life Sciences
- Physical Sciences
- Social Sciences
- Useful words for academic writing

**British Academic Written English Collections**

These collections come from the British Academic Written English (BAWE) corpus, which was developed at the Universities of Warwick, Reading and Oxford Brookes.

- Social Sciences
- Arts and Humanities
- Physical Sciences
- Life Sciences
- Useful words for academic writing

**LAW Collections**

The resources used in these law collections come from open podcasts, Massive Open Online Courses (MOOCs) and Open Access publications. They have been developed to support learners with Legal English and to demonstrate the types of domain-specific collections that can be built using the FLAX software.

- Age of Globalization MOOC (University of Texas at Austin with edX)
- English Common Law MOOC (University of London with Coursera)
- Environmental Politics and Law (Yale University with OpenYale)
- CopyrightX (Harvard University)
- Law Articles (Open Access Law Journal Publications)
- Legal Terms List
- British Law Report Corpus (BLaRC)
- ContractsX (Harvard University)
- Law PhD Theses Abstracts (EThOS at the British Library)

**Collections Created by Registered Users**

If you would like to build your own collections and put them here, please email us.

Note: we have moved some teacher created collections, which are currently under construction, to collections.flax.nzdl.org. However, if you would like your finalised collections to appear here, please email us.

Financial Crisis Lectures

**EThOS at the British Library**

- Social Sciences PhD Thesis Abstracts
- Arts and Humanities PhD Thesis Abstracts
- Law PhD Thesis Abstracts
- PhD Thesis Abstracts on Water Politics and Tourism Studies
- STEM PhD Thesis Abstracts

PhD Abstracts in FLAX

http://flax.nzdl.org/greenstone3/flax?a=b&rt=r&s=ClassifierBrowse&cl=CL1&c=PASS&if=flax
Search for Key Terms

PhD Abstracts (Life Sciences)

Search Words in Collection

Search Result: 40 sentences found

- The subcellular localization of AtPO81 and NbPO81 was analyzed by transiently overexpressing GFP-AtPO81 and GFP-NbPO81 fusion proteins in tobacco leaf tissue.
- The gene encoding the E. col flavodoxin NAD+ oxidoreductase (FLDR) and flavodoxin (FLD) have been overexpressed in E. coli and the gene products purified to homogeneity.
- Acquired resistance of the EGFR overexpressing colorectal cancer cells to treatment with afatinib, ICR62 or gefitinib was accompanied by the up-regulation of EGFR.
- Previous work to this thesis showed that CLL cells overexpress small pathway activation in normal B cells.
- Chronic lymphoma
- PKC-beta1 and c-Abl, kinases which have been implicated in mediating NF-kappaB pathway activation in normal B cells.

Search/“overexpressed”

http://flax.nzl.org/greenstone3/flax?a=b&rt=r&s=ClassifierBrowse&cl=CL1&c=PASS&if=flax
### PhD Abstracts (Life Sciences)

<table>
<thead>
<tr>
<th>academic Words</th>
<th>Wordlist/AWL</th>
</tr>
</thead>
<tbody>
<tr>
<td>data</td>
<td>3071 investig</td>
</tr>
<tr>
<td>thesis</td>
<td>2758 identify</td>
</tr>
<tr>
<td>significant</td>
<td>2077 potential</td>
</tr>
<tr>
<td>indicate</td>
<td>1602 evidence</td>
</tr>
<tr>
<td>individual</td>
<td>1478 assess</td>
</tr>
<tr>
<td>range</td>
<td>1372 affect</td>
</tr>
<tr>
<td>technique</td>
<td>1258 structure</td>
</tr>
<tr>
<td>concentration</td>
<td>1075 target</td>
</tr>
<tr>
<td>occur</td>
<td>1011 complex</td>
</tr>
<tr>
<td>isolate</td>
<td>916 outcome</td>
</tr>
<tr>
<td>primary</td>
<td>867 assessment</td>
</tr>
<tr>
<td>functional</td>
<td>843 participant</td>
</tr>
<tr>
<td>stress</td>
<td>791 evaluate</td>
</tr>
<tr>
<td>conduct</td>
<td>754 phase</td>
</tr>
<tr>
<td>normal</td>
<td>719 component</td>
</tr>
<tr>
<td>period</td>
<td>709 confirm</td>
</tr>
<tr>
<td>link</td>
<td>613 regulation</td>
</tr>
<tr>
<td>generate</td>
<td>602 aspect</td>
</tr>
<tr>
<td>theory</td>
<td>579 regulate</td>
</tr>
<tr>
<td>source</td>
<td>553 compound</td>
</tr>
<tr>
<td>project</td>
<td>543 select</td>
</tr>
<tr>
<td>cycle</td>
<td>528 survey</td>
</tr>
<tr>
<td>qualitative</td>
<td>520 inhibit</td>
</tr>
</tbody>
</table>

[http://flax.nzdl.org/greenstone3/flax?a=g&rt=r&s=FlaxWordList&sa=FlaxWordQuery&c=PALS&s1.listType=3000&if=]
Top 100 Collocations

http://flax.nzdl.org/greenstone3/flax?c=PALS&a=g&rt=r&sa=FlaxCollocationRetrieval&s=FlaxCollocationListRetrieve&s1.maxCount=100
These powerful collections draw on large reference corpora like the British National Corpus (BNC) and even larger datasets from Google and Wikipedia. More powerful than a dictionary, these collections show numerous examples of language in context for some of the most challenging areas of English language learning - collocations and phrases - where there are literally hundreds of thousands of possibilities for combining words.

**Learning Collocations**

**Book Phrases**

**Web Phrases**

**Web Collocations**

---

**About this collection**

There are many definitions of collocation. We think of collocations in the same way as expressed by Benson et al.: "In any language, certain words combine with certain other words or grammatical constructions. These recurrent, semi-fixed combinations, or collocations, can be divided into two groups: grammatical collocations and lexical collocations."
Link to the Collocation Learning System with the Wikipedia Corpus in FLAX (Wu, Li, Witten & Yu, 2016)

http://flax.nzdl.org/greenstone3/flax?
a=g&rt=r&sa=CollocationQuery&s=CollocationQuery&s1.title=&c=collocations&s1.threshold=0.5&s1.startNum=0&s1.perPage=20&s1.sampleNum=10&s1.type=&s1.wordType=&s1.colloType=&s1.query=role&s1.dbName=Wikipedia
Lexical Bundles
Biber et al. (2004, 2007)

http://flax.nzl.org/greenstone3/flax?a=g&rt=r&s=FlaxLexicalBundle&sa=FlaxLexicalBundle&c=PALS
Romantic posthumous life writing: inter-stitching genres and forms of mourning and commemoration

Contemporary scholarship has seen increasing interest in the study of *elegy*. The present work attempts to elevate and expand discussions of death and survival beyond the ambit of *elegy* to a more genre-inclusive approach. *Ethics* sensitive survey of Romanticism posthumous life writings. Constructing an *Ethics* of remembrance founded on mutual fulfillment and reciprocal care with the Romanticism tendency to *hybrid* (biology) different *genre* of mourning and commemoration, the study re-conceives 'posthumous life' as the 'inexhaustible' product of endless collaboration between the dead, the dying and the living. This *thesis* looks to the *philosophy* meditations of Francis Bacon, John Locke and Emmanuel Levinas for an *Ethics* framework of human protection, fulfillment and preservation. In an effort to locate the origin of posthumous life writing, the first chapter examines the *philosophy* context in which different *genre* and media of commemoration emerged in the eighteenth century. Accordingly, it will commence with a survey of *Age of Enlightenment* attitudes toward posthumous sympathy and the threat of death. The second part of the chapter turns to the tangled histories of *eulogy*, *biography*, portraiture, *epitaph* and *elegy* in the writings of Samuel Johnson, Henry Kett, Vicesimus Knox, William Godwin and William Wordsworth. The Romanticism culture of mourning and commemoration inherits the intellectual and religious legacies of *Age of Enlightenment*. Hence, Chapter Two will try to uncover *William Wordsworth* (1770 – 23 April 1850) was a major English Romantic poet. His *poetry* helped to launch the Romanticism/Romantic Age in English literature with the 1798 publication of *Lyrical Ballads.*

*Related topics in Wikipedia*
- Samuel Taylor Coleridge
- Robert Southey
- Edmund Spenser
- John Milton
- The Rime of the Ancient Mariner
- Romanticism

Wikify “William Wordsworth”
William Wordsworth

From Wikipedia, the free encyclopedia

"Wordsworth" redirects here. For other uses, see Wordsworth (disambiguation).
For the Scottish composer, see William Wordsworth (composer).

William Wordsworth (7 April 1770 – 23 April 1850) was a major English Romantic poet who, with Samuel Taylor Coleridge, helped to launch the Romantic Age in English literature with their joint publication Lyrical Ballads (1798).

Wordsworth's magnum opus is generally considered to be The Prelude, a semiautobiographical poem of his early years that he revised and expanded a number of times. It was posthumously titled and published, before which it was generally known as "the poem to Coleridge".[1] Wordsworth was Britain's Poet Laureate from 1843 until his death from pleurisy on 23 April 1850.[2]

Contents [hide]

1 Early life
2 Relationship with Annette Vallon
3 First publication and Lyrical Ballads
4 The Borderers
5 Germany and move to the Lake District
6 Marriage and children
7 Autobiographical work and Poems in Two Volumes
8 The Prospectus
9 Laureateship and other honours
10 Death

References


Thank You

Special Thanks:
Mahendra Mahey of British Library Labs
Sara Gould and Rosie Heather of EThOS at the British Library
The International Research Foundation (TIRF) for English Language Education

FLAX Language Project & Software Downloads: http://flax.nzdl.org/
FLAX Language Project Research:
https://www.researchgate.net/project/FLAX-Flexible-Language-Acquisition-flaxnzdlorg
The How-to eBook of FLAX:
FLAX Game-based Apps for Android via Google Play Store (free):
https://play.google.com/store/apps/developer?id=FLAX%20TEAM&hl=en

Ian Witten (FLAX Project Lead): ihw@cs.waikato.ac.nz
Shaoqun Wu (FLAX Research and Development): shaoqun@waikato.ac.nz
Alannah Fitzgerald (FLAX Open Language Research): a_fitzg@education.concordia.ca
Chris Mansfield (Queen Mary Language Centre): c.mansfield@qmul.ac.uk

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