

Making use of and adapting MOOCs text resources for language learning

Shaoqun Wu

Computer Science Department, University of Waikato, New Zealand
shaoqun@waikato.ac.nz

Alannah Fitzgerald

Department of Education, Concordia University, Montreal, Quebec, Canada

Margaret Franken

Faculty of Education, University of Waikato, New Zealand

Abstract: *Massive Open Online Courses are becoming popular educational vehicles through which universities reach out to non-traditional audiences. Many enrollees hail from other countries and cultures, and struggle to cope with the English language in which these courses are invariably offered. Moreover, most such learners have a strong desire and motivation to extend their knowledge of academic English, particularly in the specific area addressed by the course.*

Online courses provide a compelling opportunity for domain-specific language learning, a growing trend in language teaching and learning. Typical MOOCs supply a large corpus of interesting linguistic material relevant to a particular area, including supplementary images (slides), audio and video. Such corpus provides an excellent context to study domain-specific lexico-grammatical features of any word or phrase, a challenging aspect of English productive use even for quite advanced learners.

We contend that this corpus can be automatically analysed, enriched, and transformed into a resource that learners can browse and query in order to extend their ability to understand the language used, and help them express themselves more fluently and eloquently in that domain. To illustrate this idea, an existing online corpus-based language learning tool (FLAX) is applied to a Coursera MOOC entitled English Common Law offered by University of London. We will illustrate how this resource has been augmented for language learning, and then review how learners can use it to explore language usage. This article uses a single running example, a Coursera MOOC course, but the approach is fully automated and can be applied to any collection of English writing.

Keywords: MOOCs, Corpus-based Language Learning, English for Academic Purposes, Data-Driven learning.

1. Introduction

A growing body of research aims to understand the linguistic features of academic text and their bearing on the problems faced by students learning to write. To support this work, corpora of academic writing have been built as a reference and research base. These include lists of academic words, syntactic patterns characteristic of academic writing, and distinctive linguistic characteristics of multi-word sequences that fulfill discourse functions. The research raises many pedagogical implications, and it should be possible to apply the findings to academic teaching practice. How we can bridge the gap between expert and student writing? Suggestions include helping students understand the importance of learning common collocates or recurrent lexical or grammar patterns in different contexts (Coxhead, 2007), making commonly used lexical bundles more accessible (Hafner and Candlin 2007), and providing students with more realistic models (Hyland, 2008). The use of domain-specific corpora is a growing trend in language teaching and learning (e.g. Gabrielatos, 2005). Most corpora are based on particular domains, genres, or collections of certain types of document from which recurrent phrases and grammatical patterns can easily be retrieved (Stubbs and Barth, 2003). Among other aspects of language, a corpus provides an excellent context in which to study collocations, a notoriously challenging aspect of English productive use even for quite advanced learners (Bishop, 2004; Nesselhauf, 2003).

Massive Open Online Courses (MOOCs) are becoming popular educational vehicles through which universities reach out to non-traditional audiences. They are generally offered by English-speaking universities in the US and UK, and proponents often express an explicit desire to reach out to other countries and cultures. For example, Coursera aspires to provide a “meaningful learning experience for the millions of students around the world who would otherwise never have access to education of this quality” (Ng and Koller, 2013). Clearly, many MOOC students will encounter a language barrier during their study. Moreover, they will be strongly motivated to improve their knowledge of English for Academic Purposes (Dudley-Evans and St. John 1988; Hyland, 2006) as it is used in the MOOC’s subject domain. Typical MOOCs constitute a vast corpus of multimedia information, consisting predominantly of text but supplemented by images in the form of slides, audio, audio transcripts, and video; all (usually) in the English language. This paper argues that the very same corpus, pre-processed appropriately and presented in a different way, provides a focused resource that allows second- language learners to improve their linguistic knowledge in the domain addressed by the MOOC. It is also helpful for native speakers of English.

We have designed and constructed a system called FLAX¹ that takes academic texts, automatically extracts linguistic features of text that have been identified in the research literature. Our design is principled and underpinned by three theories: noticing hypothesis (Robinson, 1995; Schmidt, 2001), inductive (discovery) learning (Bernardini, 2002), and data-driven learning (Johns, 1991). Noticing is facilitated through input enhancement and enrichment that has been proven to be effective in students' recognition and recall of language features like collocations (Sonbul & Schmitt, 2013; Szudarski & Carter, 2016). FLAX presents important components in academic texts—academic words, key concepts, collocations, and lexical bundles—in a way that draws them to the attention of students. External resources (Wikipedia) are linked to these components to give students opportunities to encounter them in various authentic contexts, and repeatedly. Simple interfaces are developed so that students can use information discovery techniques (e.g., searching and browsing) that they have familiarized with through search engines (e.g., Google, Bing) to discover and study the language features of their interests in the MOOC context. Johns (1991) used the term “data-driven learning” (DDL) to describe this approach that centers on fostering learners' skills in becoming a “language researcher”. Corpus-based language learning have been explored by many researchers and teachers to facilitate collocation learning with promising results as demonstrated in the literature (Daskalovska, 2015; Vyatkina, 2016; Boulton & Cobb, 2017; Wu, et al. 2018).

This paper uses as an example of Coursera offering entitled *English Common Law MOOC*, from which we have built a FLAX collection². We illustrate how this resource has been augmented for language learning, and review how learners can use it to explore language usage. Having established a specific context, we elaborate our position by showing how this approach might be used to facilitate language learning, and what organizational and teaching structures would be suitable to put such a proposal into practice. The aim of this paper is not to explain how the system works behind the scenes: that would be a technical discussion that is relatively uninteresting from the point of view of language education. Instead, we aim to illustrate what it does by describing the result of processing the text content of *English Common Law MOOC*. It works entirely automatically, without any human input, and can be applied to any collection of academic text—for example, samples of writing collected by an individual teacher; an entire textbook (provided it is available electronically); or essays written by students.

2. Building a Digital Collection from MOOC materials

Digital libraries created using MOOCs resources are the foundation of the FLAX system. They can be constituted from a variety of original sources: written text documents, visual images, and video and audio files. We adopt a broad and contemporary concept of text as “something that is continuous and, as such comprises a totality” (Aamotsbakken, 2008. p.24); and as inclusive of written, oral and visual modes, and so we consider all of these sources as texts.

Unlike other digital library collections, texts from MOOCs are already pedagogical in that they have been designed for teaching purposes, albeit for content learning, not language learning. Aamotsbakken (2008. p.30) defines a pedagogical text as one “linked to teaching situations where the primary intention is learning.” Mcleod, Haywood and Woodgate (2015, p.56) in describing the University of Edinburgh's MOOCs, say they are often seen as a place where pedagogical pioneers try out new ideas - “a sandpit for faculty to try out and debate new ideas in online learning.” This is a strength of MOOCs texts, and means that it is likely that there is:

- a substantial number of texts,
- a sequence of texts.
- a variety of types of texts, and
- a close relationship between the different textual modalities.

MOOCs texts typically include the *talking head* of the presenter speaking on a specific topic captured in a video, together with or supplemented by visual material such as maps or diagrams, the transcription of the lecture, and other written text sources also often with embedded visual material. The oral presentation and the transcript are alternative versions of input, and the visual images are related either to the lecture or to the surrounding written texts that they are embedded in.

From a comprehension point of view, these texts within MOOCs may avoid some of the problems associated with hypertext (Charney, 1994; Gervais, 2007; Shang, 2015), where “students may become lost in “a sea of information” (Shang, 2015, p. 296), or “an open-ended, perpetually unfinished textuality” (Landow, as cited in Shang, 2015, p.296). MOOCs texts tend not to be hyperlinked, and in fact operate largely like print based material.

What are the special affordances of MOOCs that make them such a potentially useful source for learning in general, and language learning specifically? Research has found that students in MOOCs often listen to lectures multiple times (Anderson, et al., 2014). The availability and easy access to such texts is of benefit to students. Because students can access the material and revisit the MOOCs FLAX site at any time, it supports self-paced learning, and repeated exposure to both language and content. It has been said that MOOCs are designed with an unknown learner in mind (Mcleod, et al., 2015), and thus are not tailored for a particular group of learners. However, if language learners seek MOOCs that align with their interests, then they are likely to be more engaged. The multimodal nature of these sources also allows learners to explore their preferences for visual, auditory or written text.

¹ <http://flax.nzdl.org>

² The collection is available at <http://flax.nzdl.org/greenstone3/flax?a=fp&sa=collAbout&c=englishcommonlaw>

The *English Common Law MOOC* was created by University of London and offered on Coursera. The following is the description of the course on the Coursera website.

“*This MOOC will give you an introduction to this influential legal system including its history, constitutional background, sources and institutions. You’ll learn about the different ways in which laws are made and interpreted, the English court system and the increasing importance of European Union and human rights law.*”

The course materials, published under a Creative Commons Attribution license, include lecture videos, quizzes, and extra recourses (video and transcripts, which are not part of lectures). All these resources were pre-processed before being built into FLAX collections. The lecture transcripts underwent simple editing, including division into subsections, and were reformatted into manageable chunks as HTML files to decrease cognitive load when listening and viewing. Scientific images and their labels from the lecturer’s PowerPoint slides were re-formatted for readability.

Figure 1 shows the main page of the *English Common Law MOOC* collection in FLAX where the *Lectures* button has been clicked, which brings up a list of lectures for each week. The following are what the buttons at the top do:

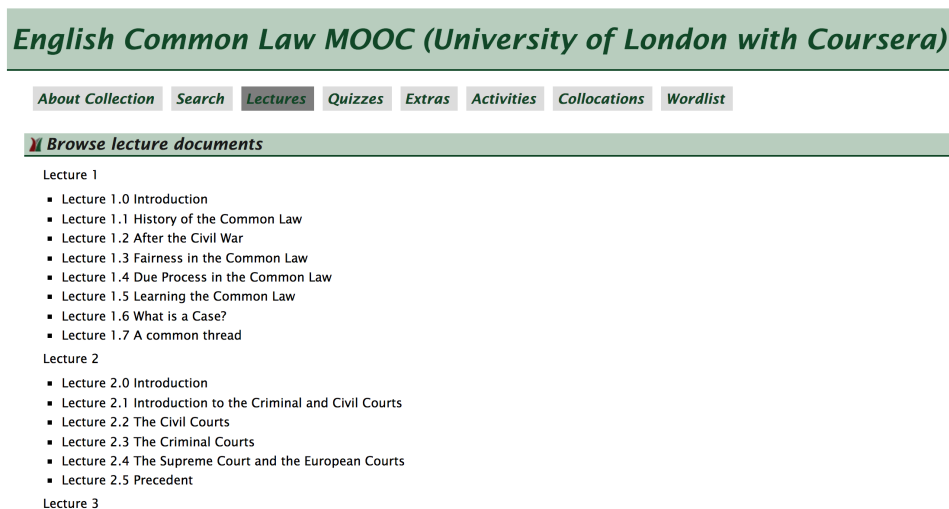


Figure 1. The main page of the *English Common Law MOOC* collection

- *About Collection* introduces the content of this collection.
- *Search* enables you to search the collection for particular documents, paragraphs, or sentences that contain particular words; or search for particular collocations
- *Lectures* lets you browse the lectures by title.
- *Quizzes* contains quizzes of this MOOC course.
- *Extra* contains supplementary materials of this MOOC course (videos and transcripts)
- *Activities* provides language activities that are created using the content of this course.
- *Collocations* lets you study the collocations in the documents.
- *Wordlist* presents the words in the collection, sorted by how often they occur.

3. Augmenting Text for Language Learning

FLAX takes text documents automatically extracts important language components—such as academic words and their usage patterns, key concepts, and collocations—and presents them in a way that draws the attention of students and gives them opportunities to encounter these components in various authentic contexts. This section shows how the lecture transcripts of the videos in *English Common Law MOOC* are augmented for supporting language learning.

3.1. Different Views of a Lecture transcript

FLAX provides students with simple interfaces to explore language features extracted from the course material. Learners can encounter and inspect words in their original context, or search for them by simply typing a word of interest, or browse them. Language activities built from course material reinforce what they have learnt.

Clicking on a lecture title (for example, “Lecture 2.3 The Criminal Courts” in Figure 1) brings up a page where students can watch or listen to the lecture video and read the transcript as shown in Figure 2. As well as viewing the

original article, to draw students' attention to the language features, other views can be obtained by clicking buttons along the right top of Figure 2:

- wordList View, which allows users to study the vocabulary used in the article;
- wikipedia view, which helps them grasp the meaning of concepts that are mentioned by the article;
- collocation view, which allows them to examine lexical compounds that occur in the article, divided into collocations that involve adjectives, nouns, prepositions, and verbs.

Lecture 2.3: The Criminal Courts



Original wordlist wikify adjective noun verb

The Criminal Courts

Turning my attention now to the criminal courts, I'm going to go through the same operation. I'm going to talk about the trial in the appellate jurisdiction of these courts and say a little bit more about the kind of work that they do. I've mentioned them before. Here, we're going to look at them in detail.

Magistrates Courts. Magistrates Courts are a key part of the criminal justice system, and the majority of criminal cases are completed in this court. In addition, Magistrates Courts can deal with some civil cases, for instance, anti-social behaviour, public health, and they're also responsible for the enforcements of fines, and community punishments. Whereas, cases requiring a penalty greater than the Magistrates' sentencing power will be sent to the Crown Court. So there's a relationship between the Magistrates Court, and the Crown Court.

Let me just pick up on that point I made a moment ago about the civil matters that the Magistrates can deal with. The Magistrates Courts have a very limited jurisdiction over certain kinds of civil matters. And this really reflects the historical powers that Magistrates used to have over certain areas of civic life. The Magistrates limited appellate jurisdiction over civil matters can be accounted for in similar terms. It's historical, in other words.

The Crown Court. The Crown Court, as I've said, deals with more serious criminal cases, such as murder, rape, robbery, some of which are on appeal or referred from the Magistrates Court. Trials are heard by a judge and a 12-person jury. Members of the public are selected for jury service or may have to go to court as witnesses. Crown Court cases originate from Magistrates Courts. The Crown Court, then, has an appellate jurisdiction over decisions of the Magistrates Court, in particular, relating to matters of sentence or conviction. Just to stress that is alongside the relationship between the Crown Court and the Magistrates Court that I was referring to. In other words, the Magistrates can refer cases to the Crown Court to be dealt with, for instance, if their sentencing powers are too restricted to deal with the kind of case that that is before them. The Crown Court also has a limited jurisdiction over civil matters. This includes a limited power to hear appeals from the Magistrates on civil matters.

Court of Appeal. The Court of Appeal Criminal Division. Remember there's a, the Court of Appeal has a civil division and a criminal division so that the criminal division hears appeals from the Crown Court.

Figure 2. A lecture video and transcript in the *English Common Law MOOC* collection

3.2. Wordlist view

The Wordlist view, which is shown in Figure 3, allows students to analyze the range of vocabulary used in the lecture. Language researchers have spent much efforts on identifying core words in academic text. Coxhead (2000) developed a list of 570 academic words from a 3.5M word corpus of academic writing, which has become a widely used resource for teachers and students. A number of competing vocabulary lists have been created—for example the University Word List (Xue & Nation, 1984), the Academic Words List (Coxhead, 2000), the Academic Keyword List (Paquot, 2012) and the Academic Vocabulary List (Gardner & Davies, 2014). Computer tools such as the Vocabprofiler available at the Compleat Lexical Tutor website help teachers and students analyze the vocabulary in a text with reference to this and other word lists. In Figure 3, the drop-down box above the text (currently showing “academic words”) provides five options: the most frequent 1000 and 2000 words, taken from wordlists used in language teaching (West, 1953); academic words included in the list by Coxhead (2000); and keywords. Figure 3 shows the academic words, and their proportion in the lecture is given beside the dropdown box (21%). Clicking a highlighted word leads to a page that shows all the sentences in the collection containing that word (Figure 10).

Figure 4 shows the keyword view, in which the words, Magistrates, Crown, anti-social, civic, jurisdiction, enforcement and so on, are highlighted in blue. The bar beside the dropdown box is a slider that the user can manipulate to reveal more words: moving it to the right makes the system less selective, highlighting more words; conversely, moving it to the left reduces the number of highlighted words. At the very left end only one keyword *Magistrates*, is given, while at the right end all content words are displayed.

Keywords are calculated by a heuristic method commonly deployed in information retrieval (known as TF-IDF, and described by, for example, Witten et al., 2010). First, documents are parsed, and the nouns, adjectives, verbs, and adverbs are designated as content words. For each such word, a score is calculated that reflects how important the word is to the document, based on the number of times it occurs in the document (which increases the score) and the number of times it occurs in the collection as a whole (which decreases it).

Original wordlist wikify adjective noun verb

academic words 4%

The Criminal Courts

Turning my attention now to the criminal courts, I'm going to go through the same operation. I'm going to talk about the trial in the appellate jurisdiction of these courts and say a little bit more about the kind of work that they do. I've mentioned them before. Here, we're going to look at them in detail.

Magistrates Courts. Magistrates Courts are a key part of the criminal justice system, and the majority of criminal cases are completed in this court. In addition, Magistrates Courts can deal with some civil cases, for instance, anti-social behaviour, public health, and they're also responsible for the enforcements of fines, and community punishments. Whereas, cases requiring a penalty greater than the Magistrates' sentencing power will be sent to the Crown Court. So there's a relationship between the Magistrates Court, and the Crown Court.

Let me just pick up on that point I made a moment ago about the civil matters that the Magistrates can deal with. The Magistrates Courts have a very limited jurisdiction over certain kinds of civil matters. And this really reflects the historical powers that Magistrates used to have over certain areas of civic life. The Magistrates limited appellate jurisdiction over civil matters can be accounted for in similar terms. It's historical, in other words.

The Crown Court. The Crown Court, as I've said, deals with more serious criminal cases, such as murder, rape, robbery, some of which are on appeal or referred from the Magistrates Court. Trials are heard by a judge and a 12-person jury. Members of the public are selected for jury service or may have to go to court as witnesses. Crown Court cases originate from Magistrates Courts. The Crown Court, then, has an appellate jurisdiction over decisions of the Magistrates Court, in particular, relating to matters of sentence or conviction. Just to stress that is alongside the relationship between the Crown Court and the Magistrates Court that I was referring to. In other words, the Magistrates can refer cases to the Crown Court to be dealt with, for instance, if their sentencing powers are too restricted to deal with the kind of case that that is before them. The Crown Court also has a limited jurisdiction over civil matters. This includes a limited power to hear appeals from the Magistrates on civil matters.

Court of Appeal. The Court of Appeal Criminal Division. Remember there's a, the Court of Appeal has a civil division and a criminal division so that the criminal division hears appeals from the Crown Court.

Figure 3. Academic words highlighted in the Wordlist view

Original wordlist wikify adjective noun verb

keywords

The Criminal Courts

Turning my attention now to the criminal courts, I'm going to go through the same operation. I'm going to talk about the trial in the appellate jurisdiction of these courts and say a little bit more about the kind of work that they do. I've mentioned them before. Here, we're going to look at them in detail.

Magistrates Courts. Magistrates Courts are a key part of the criminal justice system, and the majority of criminal cases are completed in this court. In addition, Magistrates Courts can deal with some civil cases, for instance, anti-social behaviour, public health, and they're also responsible for the enforcements of fines, and community punishments. Whereas, cases requiring a penalty greater than the Magistrates' sentencing power will be sent to the Crown Court. So there's a relationship between the Magistrates Court, and the Crown Court.

Let me just pick up on that point I made a moment ago about the civil matters that the Magistrates can deal with. The Magistrates Courts have a very limited jurisdiction over certain kinds of civil matters. And this really reflects the historical powers that Magistrates used to have over certain areas of civic life. The Magistrates limited appellate jurisdiction over civil matters can be accounted for in similar terms. It's historical, in other words.

The Crown Court. The Crown Court, as I've said, deals with more serious criminal cases, such as murder, rape, robbery, some of which are on appeal or referred from the Magistrates Court. Trials are heard by a judge and a 12-person jury. Members of the public are selected for jury service or may have to go to court as witnesses. Crown Court cases originate from Magistrates Courts. The Crown Court, then, has an appellate jurisdiction over decisions of the Magistrates Court, in particular, relating to matters of sentence or conviction. Just to stress that is alongside the relationship between the Crown Court and the Magistrates Court that I was referring to. In other words, the Magistrates can refer cases to the Crown Court to be dealt with, for instance, if their sentencing powers are too restricted to deal with the kind of case that that is before them. The Crown Court also has a limited jurisdiction over civil matters. This includes a limited power to hear appeals from the Magistrates on civil matters.

Court of Appeal. The Court of Appeal Criminal Division. Remember there's a, the Court of Appeal has a civil division and a criminal division so that the criminal division hears appeals from the Crown Court.

Figure 4. Keywords highlighted in the Keywords view

3.3. Wikipedia View

The Wikipedia view, illustrated in Figure 5, is obtained by clicking the button labeled *wikify*. It relates the terminology used in the article to the Wikipedia, highlighting concepts that are defined there. In Figure 5, the phrases *Criminal Courts*, *trial*, *appellate jurisdiction*, *public health*, *appeal*, *conviction*, etc., have been identified as pertinent concepts. These could be Wikipedia article titles, or “redirects” (i.e., synonyms) defined in the Wikipedia itself. Clicking any highlighted phrase in the document brings up information extracted from the Wikipedia in a popup window: the definition, hyperlinked to the Wikipedia article itself; followed by a list of related topics in Wikipedia that can also be clicked. In Figure 6, *appellate jurisdiction* has been selected, and in this case the list of related articles gives the names of prominent linguists in this area. Their full Wikipedia entries are only one click away.

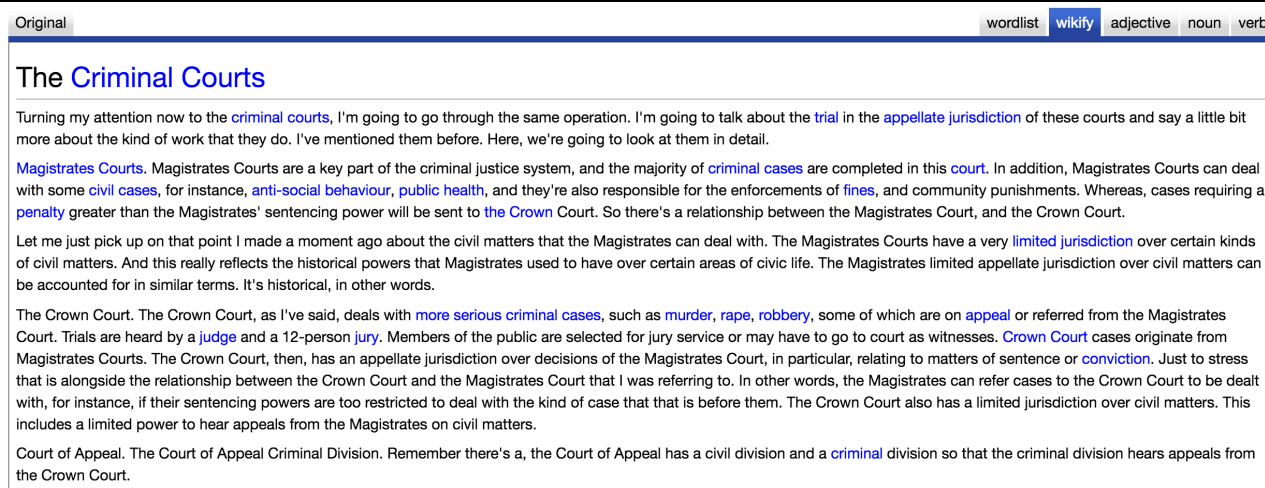


Figure 5. Wikipedia articles are highlighted in the Wikipedia view



Figure 6. A definition of *appellate jurisdiction* from Wikipedia in the Wikipedia view

The process of relating words and phrases running text to Wikipedia articles is called “wikification,” and Milne and Witten (2013) describe the method we use. It has three steps. First, sequences of words in the text that may correspond with Wikipedia articles are identified using the names of the articles, as well as their redirects and every referring anchor text used anywhere in Wikipedia. Second, situations where multiple articles correspond to a single word or phrase are disambiguated. For example, the word *kiwi* may refer to a bird, a fruit, a person from New Zealand, or the New Zealand national rugby league team, all of which have distinct Wikipedia entries. A machine learning classifier is used to make the appropriate choice, taking into account the prior probability of the mapping, semantic relatedness to other concepts in the same document, and some contextual information. The third step selects the most salient linked (and disambiguated) concepts to include in the output. Again a learning approach is used that combines prior probability, relatedness to context, disambiguation confidence, generality, location and spread. All this, including the machine learning methods and models, is described by Milne and Witten (2013).

3.4. Collocation view

The importance of collocation knowledge in academic writing has been widely recognized. Hill (1999) observes that students with good ideas often lose marks because they don’t know the four or five most important collocations of a key word that is central to what they are writing about. L2 student text tends to be cumbersome and error prone because of insufficient collocation knowledge. For example, in an essay on “Smoking” one student wrote *the smokers who rely on cigarettes and have to smoking everyday* instead of using phrases such as *heavy smoker* or *addicted to smoking*. Topic-specific corpora are valuable resources that help students build up collocation knowledge within the areas that concern them. When investigating semantic associations in a corpus of business English, Nelson (2006) discovered that the lexical environment surrounding business words is rich, diverse and semantically connected. He noted that collocates of certain common words are more formulaic than in general English. It is also true that the same word in different disciplines exhibits a different range of collocations. We used Benson et al’s (1986, p.ix) definition that “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.”

We focus on lexical collocations, which have structures *verb + noun*, *adjective + noun*, *noun + noun*, *noun + of + noun*, and *preposition + noun*.

Once an article has been reached, the collocation view is accessed by clicking one of the *adjective*, *noun*, *preposition* and *verb* tabs in Figures 2–6. Each tab shows collocations starting with that word type; for example, the *adjective* panel hosts collocations starting with an adjective. Collocations are highlighted in the text, to help students notice them and study their context. In the example shown in Figure 7, collocations related to the subject of the lecture, *The Criminal Courts*—*key part*, *criminal justice system*, *criminal cases*, *civil cases*, *historical power*, *civic life*—stand out from the surrounding text, attracting the student’s attention. The *criminal justice system* collocation has been clicked to reveal small icons. The underlined words, *criminal*, *justice*, *system*, are hyperlinked to entries that involve those words in an external collocation database,³ built from all the written text in the Wikipedia.

The Criminal Courts

Turning my attention now to the **criminal courts**, I'm going to go through the **same operation**. I'm going to talk about the trial in the **appellate jurisdiction** of these courts and say a **little bit** more about the kind of work that they do. I've mentioned them before. Here, we're going to look at them in detail.

Magistrates Courts. Magistrates Courts are a **key part** of the **criminal justice system**, and the majority of **criminal cases** are completed in this court. In addition, Magistrates Courts can deal with some **civil cases**, for in **criminal justice system** also responsible for the **enforcements** of fines, and community punishments. Whereas, cases requiring a penalty greater than the **Magistrates Court**. So there's a relationship between the Magistrates Court, and the Crown Court.

Let me just pick up on that point I made a moment ago about the **civil matters** that the Magistrates can deal with. The Magistrates Courts have a very **limited jurisdiction** over **certain kinds** of civil matters. And this really reflects the **historical powers** that Magistrates used to have over **certain areas** of **civic life**. The Magistrates limited appellate jurisdiction over civil matters can be accounted for in **similar terms**. It's historical, in other words.

The Crown Court. The Crown Court, as I've said, deals with more **serious criminal cases**, such as murder, rape, robbery, some of which are on appeal or referred from the Magistrates Court. Trials are heard by a judge and a 12-person jury. Members of the public are selected for jury service or may have to go to court as witnesses. Crown Court cases originate from Magistrates Courts. The Crown Court, then, has an appellate jurisdiction over decisions of the Magistrates Court, in particular, relating to matters of sentence or conviction. Just to stress that is alongside the relationship between the Crown Court and the Magistrates Court that I was referring to. In other words, the Magistrates can refer cases to the Crown Court to be dealt with, for instance, if their sentencing powers are too restricted to deal with the kind of case that that is before them. The Crown Court also has a limited jurisdiction over civil matters. This includes a **limited power** to hear appeals from the Magistrates on civil matters.

Court of Appeal. The Court of Appeal Criminal Division. Remember there's a, the Court of Appeal has a **civil division** and a **criminal division** so that the criminal division hears appeals from the Crown Court.

Figure 7. Collocations, highlighted in the Collocation view

criminal used as an adjective		criminal used as a noun		
criminal + noun	criminal law	1603	criminal charges	1289
	criminal cases	1159	criminal activity	826
	criminal justice	819	criminal investigation	647
	criminal record	567	criminal offence	532
	criminal organization	523	criminal prosecution	459
			>>> more	
adjective + criminal	civil and criminal	93	petty criminal	63
	American criminal	54	civil or criminal	40
	criminal and civil	39	international criminal	37
	professional criminal	36	serious criminal	30
	notorious criminal	25	former criminal	23
			>>> more	

Figure 8. Related collocations for the word *criminal*.

The system makes it easy for users to study further collocations related to these two words. For example, clicking *criminal* in Figure 7 generates a further popup, shown in Figure 8, that lists *criminal law*, *criminal record*, *criminal organization*, *criminal charges*, *criminal activity* etc., along with their frequency in that corpus; likewise, clicking *justice* in Figure 6 shows collocations such as, *social justice*, *economic justice*, *natural justice*, *civil justice system*, *chief justice*, *associate justice*, *environmental justice*, *fundamental justice*, *poetic justice* and so on. Figure 8 helps users learn about the word *criminal*. Furthermore, they can see samples of these collocations in context by clicking them in Figure 8, which shows relevant extracts from the Wikipedia articles—for example, clicking brings up ten sentences that use this phrase. Wu et al. (2016) have described the full functionalities of this collocation database.

4. More on learning words

Users can explore words in the *English Common Law MOOC* in the collection by clicking the *Wordlist* button in Figure 1, yielding the screen shown in Figure 9. The words as shown in Figure are academic words, and can be sorted alphabetically or by frequency; in either case, frequency in the collection is shown alongside the word. Clicking the word itself retrieves sentences containing it, on a page like that of Figure 10.

³ The database is available at <http://flax.nzdl.org/greenstone3/flax?a=fp&sa=collAbout&c=collocations>

academic Words		sort by frequency							
legal	153	civil	115	convention	95	principle	84	issue	82
fundamental	75	interpretation	73	obviously	71	area	57	previous	55
link	52	theme	52	process	51	community	51	precedent	51
legislation	49	interpret	48	institution	46	constitutional	41	hierarchy	39
role	38	involve	38	create	35	context	34	economic	33
individual	33	source	32	stress	32	lecture	32	domestic	32
authority	29	structure	29	tradition	29	constitution	29	require	28
define	27	period	27	contemporary	27	major	26	distinction	26
couple	25	method	25	concept	25	impact	24	section	24
paragraph	24	debate	23	integrity	23	inconsistent	23	consistent	22
similar	21	function	21	justify	20	focus	19	approach	19
primarily	19	evidence	18	circumstance	18	prior	17	presumption	16
quote	15	policy	15	chapter	15	instance	15	relevant	15
somewhat	15	legislative	15	grant	14	labour	14	ultimately	14
conventional	14	access	13	achieve	13	accurate	13	coherent	13
final	12	ensure	12	resolve	12	creation	12	ambiguity	12
quotation	12	so-called	12	definition	12	precedence	12	commentator	12
image	11	element	11	tension	11	prohibit	11	specific	11
commission	11	enable	10	complex	10	feature	10	category	10
restriction	10	affect	9	enforce	9	precise	9	abstract	9
analysis	9	document	9	establish	9	technical	9	commitment	9
regulation	9	significant	9	seek	8	prime	8	ratio	8
commit	8	factor	8	impose	8	theory	8	federal	8
primary	8	promote	8	vehicle	8	commentary	8	illustrate	8
appropriate	8	domesticate	8	prohibition	8	subordinate	8	secure	7

Figure 9. The academic words in the *Common English Law* collection

Search for that contain the word

You may be interested in [legally](#)

Search Result: 140 sentences found

Group by patterns

- Firstly, to insure consistency in the **legal** systems of the member states across the community.
- Moreover, the European Convention of Human Rights as incorporated into our law by the Human Rights Act of 1998, created a new **legal** order.
- Instead, it's as a **legal** order in which the United Kingdom assumes obligations to protect fundamental rights.
- Note that he says it creates a new **legal** order.
- What he says about the heart's concept of the rule of recognition is that this idea of a, of a **legal** rule of recognition gives way to what people have prepared to recognise as law.
- Can national identities exist within the European Union's **legal** system?
- It's full of quite difficult **legal** problems but also political issues that are very current, even current beyond Europe, especially if you're talking about mutual integration.
- And, to persuade you how much fun it can be is to ask you to read the new preamble to the new Treaty of the European Union from 1992 and the claims that the European Union makes in relation to what it will do as a **legal** system.
- So, the question is, do you think it's possible for a **legal** system to achieve that?
- And, it is primarily a **legal** system still and not a political system.
- I mean, that seems a lot unachievable, considering what we already know about the UK **legal** system.
- Hm, they all have important role in our **legal** system anyway, so, they all have to work together.
- In other words, in those areas which don't raise European **legal** issues.
- Now, Lord Bingham goes on to say that these interpretive tendencies have been encouraged by, quote: "The teleological and directed approach of European community jurisprudence and the influence of European **legal** culture generally".

Figure 10. Result of searching for sentences containing the word *legal*

Search queries can contain more than one word, in which case sentences are returned that contain all the query terms. For phrase searching, a query can be enclosed by quotation marks; for example, "*legal system*" returns sentences containing this phrase.

5. Conclusion

We have shown how academic text can be augmented to facilitate language learning. The process embodied in this system was guided by findings recorded in the research literature. It extracts useful language learning material from the input documents, including academic words, key words and concepts; collocations; typical word usage patterns; and lexical bundles. All these are made easily accessible through a unified searching and browsing interface.

Document text is presented in different views, each focusing on a particular linguistic aspect, the aim being to draw the student's attention to different aspects of language, and increase their awareness of how ideas are expressed. To further enrich and expand student knowledge, external resources—Wikipedia—are automatically linked in to the collection to provide additional context, both pragmatic and linguistic. Readers can get a better feeling for the facilities the system provides by exploring it online.

The MOOC language collections we have built demonstrate the affordances of the FLAX software. FLAX is open source and can be downloaded to build language support collections with any text-based content and supporting audio-visual material, for both online and classroom use. It is designed so that non-expert developers—whether language teachers, subject specialists, or instructional design and e-learning support teams—can build their own collections. Content varies in terms of licensing restrictions, depending on the publishing strategies adopted by institutions for their content. FLAX has been designed to offer a flexible suite of linguistic support options for enhancing such content across both open and closed platforms.

A recent review commissioned by the UK Department for Business Innovation and Skills (2013) tracks the progress of the MOOC phenomenon as it moves from experimentation into maturity. Current work focuses on meeting the accreditation needs of learners, and on devising and developing new pedagogical models to better support online learning.

FLAX's capabilities for building language collections with comprehensive facilities for search and retrieval, and customized interactive learning of key domain terms and concepts, addresses the needs of both native and non-native speakers who are interested in engaging deeply with specific academic resources in English while developing their receptive reading and listening skills.

We plan to investigate further MOOC collections to determine whether FLAX can assist not only with mastery approaches to learning and assessment like those employed in the *English Common Law MOOC* course, but also with constructivist approaches that support peer learning and assessment—where collections will be derived from student texts, seminar discussions, and peer-review texts, as well as from expert text and lecture transcripts. This will promote the aggregation of crowd-sourced content for collaborative peer learning.

References

- Aamotsbakken, B. (2008). The pedagogical text – an important element in the textual world? Reflections on the concepts of 'text', 'context' and literacy culture'. *Designs for Learning*, 1(1), 24-39.
- Anderson, A., Huttenlocher, D., Klienberg, J., & Leskovec, J. (2014, April). Engaging with massive on-line courses. Paper presented at the International World Wide Web Conference, Seoul, Korea. Retrieved from <http://cs.stanford.edu/people/jure/pubs/mooc-www14.pdf>.
- Benson, M. Benson, E., Ilsen, R.F. (1986). *The BBI combinatory dictionary of English: A guide to word combinations*. Amsterdam/Philadelphia: John Benjamins.
- Bernardini, S. (2002). Exploring new directions for discovery learning. In B. Kettemann & G. Marko (Eds.), *Teaching and learning by doing corpus analysis* (pp. 165–182). Amsterdam, Netherlands: Rodopi.
- Bishop, H. (2004) "The effect of typographic salience on the look up and comprehension of unknown formulaic sequences." In N. Schmidt (Ed.) *Formulaic sequences: Acquisition, processing, and use* (pp. 227-244). Philadelphia, PA, USA: John Benjamins.
- Boulton, A., & Cobb, T. (2017). Corpus Use in Language Learning: A Meta-Analysis *Language Learning*, 67(2), 348–393.
- Charney, D. (1994). The effect of hypertext on processes of reading and writing. In C. L. Selfe & S. Hilligoss (Eds.), *Literacy and computers: The complications of teaching and learning with technology* (pp. 238–263). New York, NY: Modern Language Association.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213–238.
- Coxhead, A., & Byrd, P. (2007). Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Second Language Writing*, 16, 129–147.
- Daskalovska, N. (2015). Corpus-based versus traditional learning of collocations. *Computer Assisted Language Learning*, 28(2), 130-144. doi: 10.1080/09588221.2013.803982
- Dudley-Evans, T., St John, M.J. (1988). *Developments in English for Specific Purposes: A multidisciplinary approach*. Cambridge: Cambridge University Press.
- Gabrielatos, C. (2005) "Corpora and language teaching: Just a fling or wedding bells?" *Teaching English as a second or foreign language*, 8(4). Retrieved Oct 21 2013 from <http://tesl-ej.org/ej32/a1.html>.
- Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, 35(3), 305-327. doi: 10.1093/applin/amt015.
- Gervais, B. (2007). Is there a text on this screen? Reading in an era of hypertextuality. In R. Siemens & R. Schreibman (Eds.), *A companion to digital literary studies* (pp. 183–202). Malden, MA: Blackwell
- Hafner, C. A., & Candlin, C. N. (2007). Corpus tools as an affordance to learning in professional legal education. *English for Academic Purposes*, 6(4), 303-18.
- Hill, J. (1999). "Collocational competence." *ETP*, 11.
- Hyland, K. (2002). Authority and invisibility: Authorial identity in academic writing. *Journal of Pragmatics*, 34(8), 1091-1112. doi: 10.1016/S0378-2166(02)00035-8.
- Hyland, K. (2008a). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, 18(1), 41–62.
- McLeod, H., Haywood, J., & Woodgate, A. (2015). Emerging patterns in MOOCs: Learners, course designs and directions. *TechTrends*, 59(1), 560-63.

- Milne, D., & Witten, I. H. (2013). An open-source toolkit for mining Wikipedia. *Artificial Intelligence*, 194, 222-239. doi:10.1016/j.artint.2012.06.007.
- Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge, United Kingdom: Cambridge University Press.
- Nelson, M. (2006). Semantic association in Business English: A corpus-based analysis. *English for Specific Purposes*, 25(2), 217-234.
- Nesselhauf, N. (2003) "The use of collocations by advanced learners of English and some implications for teaching." *Applied Linguistics*, 24(2), 223-242.
- Ng, A. and Koller, D. (2013) "The online revolution: education for everyone." *Proc ACM SIGKDD Int Conf on knowledge discovery and data mining*, p.2.
- Paquot, M. (2012). *Academic vocabulary in learner writing: From extraction to analysis*. London, United Kingdom: Continuum.
- Robinson, P. (1995). Attention, memory, and the "noticing" hypothesis. *Language Learning*, 45(2), 283- 331.
- Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3- 32). Cambridge, UK: Cambridge University Press.
- Shang, H. (2015). An investigation of scaffolded reading on EFL hypertext comprehension. *Australasian Journal of Educational Technology*, 31(3), 293-312.
- Stubbs, M., and Barth, I. (2003) "Using recurrent phrases as text-type discriminators." *Functions of Language*, 10(1), 61-104.
- Sonbul, S., & Schmitt, N. (2013). Explicit and implicit lexical knowledge: Acquisition of collocations under different input conditions. *Language Learning*, 63(1), 121-159.
- Szudarski, P., & Carter, R. (2016). The role of input flood and input enhancement in EFL learners' acquisition of collocations. *International Journal of Applied Linguistics*, 26(2), 245-265.
- UK Government Department of Business Innovation & Skills. (2013). *The maturing of the MOOC*. London: UK Government Publications.
- Vyatkina, N. (2016). Data-driven learning of collocations: Learner performance, proficiency, and perceptions. *Language Learning & Technology*, 20(3), 159-179.
- West, M. (1953). *A general service list of English words*. Longman, Green & Co., London.
- Wu, S., Li, L., Witten, I. H., & Yu, A. (2016). Constructing a collocation learning system from the Wikipedia corpus. *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT)*, 6(3), 18-35. doi:10.4018/IJCALLT.2016070102.
- Wu, S., Fitzgerald, A., Witten, I. H., & Yu, A. (2018). Automatically augmenting academic text for language learning: PhD Abstract corpora with the British Library. In *Handbook of Research on Integrating Technology Into Contemporary Language Learning and Teaching* (pp. 512-537). Hershey, PA: IGI Global. doi:10.4018/978-1-5225-5140-9.ch025.
- Witten, I.H., Bainbridge, D. and Nichols, D.M. (2010). *How to Build a Digital Library*. Morgan Kaufmann, Burlington, MA (second edition).
- Xue, G., & Nation, P. (1984). A university word list. *Language Learning and Communication*, 3(2), 215-229.