

# F-Lingo: Integrating lexical feature identification into MOOC platforms for learning professional and academic English

1<sup>st</sup> Alannah Fitzgerald  
Department of Computer Science  
University of Waikato  
Hamilton, Aotearoa / New Zealand  
alannah.fitzgerald@waikato.ac.nz

2<sup>nd</sup> Jemma König  
Department of Computer Science  
University of Waikato  
Hamilton, Aotearoa / New Zealand  
jkonig@waikato.ac.nz

3<sup>rd</sup> Ian H. Witten  
Department of Computer Science  
University of Waikato  
Hamilton, Aotearoa / New Zealand  
ihw@waikato.ac.nz

**Abstract**— F-Lingo is a chrome extension that works on top of the *FutureLearn* MOOC platform to support content-based language learning of domain-specific terminology for professional and academic purposes.

**Keywords**— MOOC delivery, learning support, domain-specific terminology, English for specific purposes, open content

## I. INTRODUCTION

We have developed an application called F-Lingo. Implemented as a Chrome extension, F-Lingo works on top of *FutureLearn* – an online MOOC platform – to help you learn about selected words, phrases, and concepts in the text you are reading. This presentation will introduce the F-Lingo Chrome extension, then describe the results of a study that has been conducted with learners enrolled in *FutureLearn's* *Practical Data Mining* courses. This study aims to address whether integrating language tools into an online MOOC platform can supplement learning domain-specific English for professional and academic purposes.

MOOCs have trended in various directions, most notably in their dual role to serve both workplace recruitment and workplace training [1, 2]. As EdTech and workplace recruitment continue to converge with increased stakes in the MOOC space (see *FutureLearn* and *Coursera* partnerings with the online job-matching service, *SEEK*), the need for scalable learning support that has been tailored for enhancing workforce diversity and inclusion has also increased. Reconsideration, therefore, of the design of learning environments for MOOC delivery that support an increasingly diverse learner population is one of the central premises of this research. In particular, we place emphasis on designing automated systems that support self-regulated learning [3, 4] for the understanding, uptake and utilization of domain-specific terminology that occurs in professional and academic contexts.

For the purposes of this research, we have developed domain-specific English language learning support for academic and professional purposes that we contend is useful for both native and non-native speakers of English [5]. Identifiable gaps with provision in Internet-based

learning schemes stack the chances of success against learning with MOOCs, especially when learners are unfamiliar with the use of domain-specific terminology in professional and higher education contexts. Access differentiation in higher education has been well documented in the research carried out into learner perceptions of self-efficacy when failure has been experienced with educational systems, making future attempts to engage in educational offerings less likely [6, 7].

Where informal and non-formal learning is concerned the success of reaching learners more than ever before with innovative Internet-based learning solutions – the mainstream MOOC being the latest in a long line of online distance education innovations – is at the same time isolating learners “in a world of text in an unfamiliar or semi-familiar language (usually English)” [8]. We contend, however, that many MOOC learners who are new to the subject areas they are studying, irrespective of their first language, are further isolated by unfamiliar or semi-familiar terms and concepts in the texts of MOOCs (video lecture transcripts and course readings) encountered through reading and listening that reflect domain-specific language features from target academic and professional communities [9, 10].

## II. F-LINGO

### A. Introducing F-Lingo

F-Lingo provides learners with the opportunity to further enrich their understanding of online MOOC content – both as a way to learn the features of domain-specific terminology, and as a way to gain a deeper understanding of the content and concepts being provided.

F-Lingo is a Chrome extension that was developed as an aid for second language acquisition. Although there is a plethora of existing computer assisted language software, F-Lingo is the first to explicitly take advantage of online courses. It has been developed to work with *FutureLearn*, an online MOOC platform that provides learners with free online courses from leading universities, specialist organizations, and institutions around the world.

*FutureLearn* has hundreds of active courses, three of which are run by the University of Waikato: *Data Mining with Weka*, *More Data Mining with Weka*, and *Advanced Data Mining with Weka*. The first – *Data Mining with Weka* – is made up of 95 pages of textual content that total 25,000 running words, plus over 30,000 running words from video transcripts. Using F-Lingo on top of *FutureLearn*, learners can view text augmented with additional lexical features, taking advantage of online courses for content-based language learning.

To trial, download F-Lingo from the Chrome store (<https://chrome.google.com/webstore/search/flingo>) and install it. Restart your browser and visit any page of the Data Mining with Weka MOOC (<https://www.futurelearn.com/programs/data-mining>) from the University of Waikato; the rest happens automatically. If you want to see what F-Lingo does without installing it, this 3-minute video (<https://www.youtube.com/watch?v=FRGwuexvkus&feature=youtu.be>) illustrates its facilities.

### B. Domain-specific terminology learning support

English for specific purposes (ESP) usually refers to learning features of domain-specific terminology for academic or professional purposes. In the field of linguistics, the study of terminology includes: “words, compound words or multi-word expressions that in specific contexts are given specific meanings, including conceptual meanings—these may deviate from the meanings the same words have in other contexts and in everyday language.” [11]. F-Lingo includes three main features: identifying vocabulary, highlighting vocabulary in text, and providing language resources.

F-Lingo uses a selection of natural language processing techniques to identify lexical features within the text of a *FutureLearn* MOOC course. It uses established frequency word lists to identify keywords within the text – classifying words as keywords only if they are absent from the General Service List [12] which contains 2000 headwords taken from a written corpus, representing words of the greatest ‘general’ use in English. Next, it uses syntactic patterns to identify phrases within the text, for example noun + noun (data mining), verb + noun (visualize data), and so on.

In particular, F-Lingo identifies two types of phrases: collocations and lexical bundles. Collocations are sequences of two or more words that occur together more frequently than by chance, and that hold semantic meaning [13, 14]. Lexical bundles are recurrent multi-word items that carry meta-discourse functions; the understanding of which improves learners’ accuracy and fluency [15]. It also uses the Wikipedia Miner toolkit of machine-learned approaches to detect and disambiguate Wikipedia concepts within a document. Wikipedia Miner is an information retrieval toolkit that can be used to examine and obtain usable data from Wikipedia. It includes a database of summarized Wikipedia content, and an API to access it [16].

Once they have been identified, F-Lingo highlights words, phrases, and concepts within the page. It traverses the HTML elements on a page and extracts textual content, then it tags the text with words, phrases, and concepts and re-

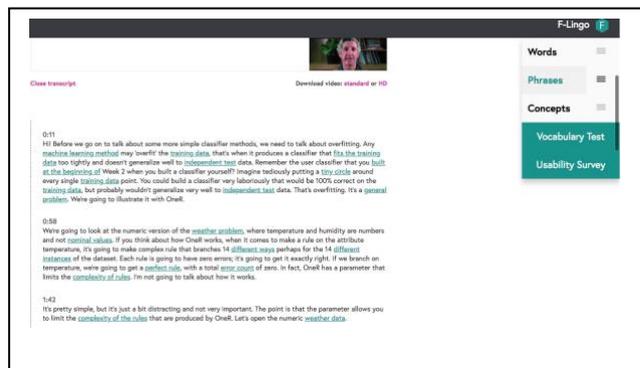


Fig. 1. F-Lingo highlighted phrases in MOOC video transcript

displays the tagged content on the page as can be seen in Fig. 1. Learners can select whether they want to view the content tagged with words, phrases, or concepts, each of which results in a selection of lexical items being highlighted and made into clickable links that provide additional lexical information.

Once words, phrases, and concepts have been highlighted, they can be clicked to view additional lexical information, specifically dictionary definitions, example sentences, expanded collocations, and disambiguated descriptions.

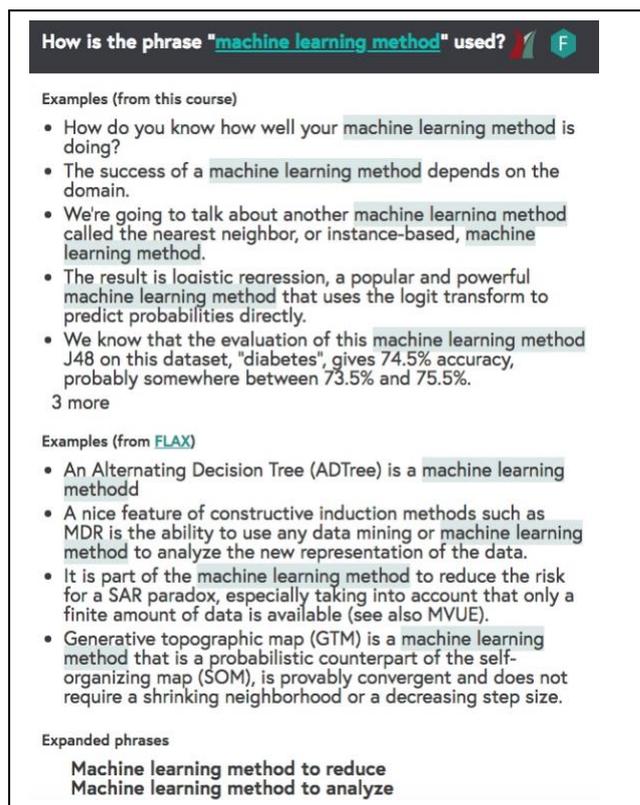


Fig. 2. F-Lingo phrase examples for “machine learning method” derived from MOOC content and the FLAX Wikipedia Learning Collocations collection

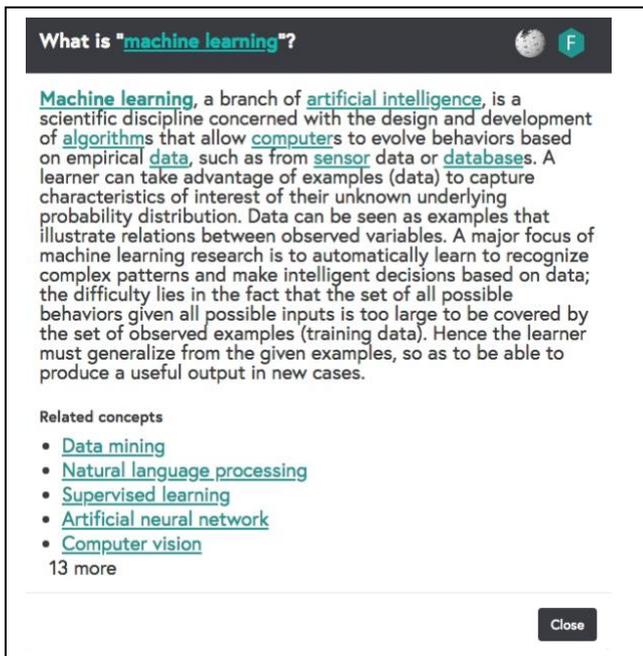


Fig. 3. F-Lingo MOOC concept examples for “machine learning” mined with the Wikipedia Miner Toolkit

First, F-Lingo retrieves dictionary definitions from Wiktionary. Once a definition has been retrieved and processed, it is re-formatted and returned to F-Lingo, where it is displayed in a modal dialog. If the word was derived from another, F-Lingo retrieves both definitions and display them in the dialog. Next, F-Lingo provides learners with example sentences, demonstrating how words and phrases are used within context. It provides learners with sentences from within the course, and from an external source called FLAX (Flexible Language Acquisition flax.nzdl.org).

FLAX is an online language learning application. It contains collections of digitized text from a variety of sources, including PhD abstracts from the British Library’s E-Theses Online Service [17], online course material from edX and Coursera, and a selection of texts created by their users. F-Lingo also uses the FLAX Learning Collocations database derived from a Wikipedia corpus to retrieve expanded collocations [18]. For example, *machine learning method to reduce* and *machine learning method to analyze* are collocations that expand the target collocation *machine learning method* as can be seen in Fig. 2. Finally, when learners click on concepts, F-Lingo uses Wikipedia Miner to retrieve the first paragraph and related articles. It then formats them, providing learners with the disambiguated description for the concept that was clicked as shown in Fig. 3.

### III. INTEGRATING F-LINGO INTO MOOCS

The F-Lingo Chrome extension is currently available for three *FutureLearn* courses: *Data Mining with Weka*, *More Data Mining with Weka*, and *Advanced Data Mining with Weka*. However, it can be applied to any *FutureLearn* course. This involves four steps:

1. Downloading the course content
2. Pre-processing the course content

3. Caching the data in the F-Lingo database
4. Adding the course title to the database

F-Lingo highlights words, phrases, and concepts each time a page is loaded and provides additional lexical information each time an item is clicked. However, providing these facilities first requires some pre-processing. For example, providing learners with example sentences from within the course, particularly from pages in the course that they have not yet visited, cannot be achieved without first processing the course content to extract sentences. This means that, before F-Lingo can be used on a course, it needs to have access to the content in its entirety.

Once the content is available, F-Lingo’s pre-processing can begin, splitting the course into sentences and indexing them, identifying words, phrases (using FLAX), and concepts within the text, and retrieving concept leads and related articles using Wikipedia Miner. This results in a set of text files that contain SQL statements that can be loaded into the F-Lingo database. Finally, F-Lingo checks the database for active courses before it integrates itself into a *FutureLearn* course. Adding a courses title to the database allows it to be recognized by F-Lingo.

The F-Lingo Chrome extension has been subjected to a data-based evaluation where learners enrolled in three *FutureLearn Data Mining with Weka* courses were given access to the application. 109 learners downloaded and used F-Lingo, and their interactions with words, phrases and concepts on the page were tracked. Any words, phrases, or concepts that were clicked, and the duration that learners looked at definitions, example sentences, and so on, was logged by the system. Five behavioural observations were noted as a result of the data gathered: (1) there were more L2 English speakers than L1 English speakers that downloaded and used the F-Lingo Chrome extension, (2) learners interacted with words more often than phrases or concepts, (3) learners interacted with items within course content more often than summary lists, (4) learners spent longer periods of time looking at additional lexical information for concepts than words or phrases, and (5) learners were more likely to click the same word than they were the same phrase or concept [19].

The results of this study have been used to interpret learner behaviour in relation to their use of F-Lingo as a language resource. The next intended study is an extensive longitudinal study, focusing on whether F-Lingo can be used to assist with language learning, and whether long-term use of an enriched course improves learners’ vocabulary.

### IV. THE IMPORTANCE OF LANGUAGE LEARNING SUPPORT IN THE MOOC SPACE

It is important to question how MOOC providers and MOOC course designers at leading universities and professional organizations around the world, many of whom invariably offer MOOCs in English, can support large and diverse learner groups with automated open data-driven language learning systems. Non-formal online learning is the activity of understanding, gaining knowledge or acquiring skills outside the remit of being a registered student at a formal educational institution. As with most

MOOCs, this kind of non-formal learning typically occurs without direct teacher or tutor support. The research we will present investigates whether or not stand-alone automated data-driven learning systems can assist with the learning of domain-specific academic and professional terminology.

In further iterations of the research, we plan to make F-Lingo widely operational through performance improvement campaigns with educational technologists working with the delivery of MOOCs who can be trained in data scraping methods to enable their course content to be enriched and parsed by the F-Lingo system software for automated language learning support. Our end goal is to make F-Lingo, which draws on linguistic databases from the open-source FLAX language project, interoperable with any online learning platform.

We also view F-Lingo as a work-around solution to carrying out text and data-mining methods on All Rights Reserved course content present in online learning. Although this content is currently deemed as Open Access in the sense of being read-only from an accessible outward-facing online learning platform (with MOOCs being a clear example), it has not, and most likely will not, be licensed openly with Creative Commons for reuse and remix by the wider education and learning community due to the current paucity of open education policy for content reuse and remix in higher education.

#### REFERENCES

- [1] A. W. Radford, B. Coningham, and L. Horn, "MOOCs: Not just for college students - How organizations can use MOOCs for professional development", *Employment Relations Today*, 41(4), 1-15, 2015.
- [2] P. Sreeleakha, and N. Manikandan, "The way forward with MOOCs – Professional development perspectives", *American International Journal of Research in Humanities, Arts and Social Sciences*, 15-514, 2015.
- [3] M. Egloffstein, and D. Ifenthaler, "Employee perspectives on MOOCs for workplace learning", *TechTrends*, 61(1), 65-70, 2017.
- [4] A. Littlejohn, N. Hood, C. Milligan, and P. Mustain, "Learning in MOOCs: Motivations and self-regulated learning in MOOCs", *Internet and Higher Education*, 29, 40-48, 2016.
- [5] A. Fitzgerald, S. Wu, J. König, I. H. Witten, and S. Shaw, "Designing and evaluating an automated open data-driven language learning support system for MOOCs," unpublished.
- [6] M.M. Chemers, L. Hu, and B.F. Garcia, "Academic self-efficacy and first year college student performance and adjustment," *Journal of Educational Psychology*, vol. 93, pp. 55-64, 2001.
- [7] A. Zajacova, S. Lynch and T. Espenshade, "Self-efficacy, stress, and academic success in college," *Research in Higher Education*, vol. 46, pp. 677-706, 2005.
- [8] T. Cobb, "Internet and literacy in the developing world: Delivering the teacher with the text," *Educational Technology Research & Development*, vol. 54(6), pp. 627-645, 2006.
- [9] P. Strevens, "ESP after twenty years: a reappraisal," in *ESP: State of the art*, M. Tickoo, Ed. Singapore: SEAMEO Regional Language Centre, 1988, pp. 1-13.
- [10] K. Hyland, "Specificity revisited: How far should we go now?," *English for Specific Purposes*, vol. 21, pp. 385-395, 2002.
- [11] Wikipedia, "Terminology," Retrieved from <https://en.wikipedia.org/wiki/Terminology>, 2019.
- [12] M. West, *A General Service List of English words*, London: Longman, Green and Co, 1953.
- [13] I. S. P. Nation, *Learning vocabulary in another language*, Cambridge: Cambridge University Press, 2001.
- [14] J. R. Nattinger and J.S. DeCarrico, *Lexical phrases and language teaching*, Oxford: Oxford University Press, 1992.
- [15] L. Li, "Sentence initial bundles in L2 thesis writing: A comparative study of Chinese L2 and New Zealand L1 postgraduates' writing, unpublished doctoral dissertation, The University of Waikato, New Zealand, 2016.
- [16] D. Milne, and I. H. Witten, "An open-source toolkit for mining Wikipedia," *Artificial Intelligence*, vol. 194, pp. 222-239, 2013.
- [17] S. Wu, A. Fitzgerald, I. H. Witten, and A. Yu, "Automatically augmenting academic text for language learning: PhD abstract corpora with the British Library". In B. Zou, M. Thomas (Eds.), *Integrating Technology into Contemporary Language Learning and Teaching*, pp. 512-537, Hershey, PA: IGI Global, 2018.
- [18] S. Wu, L. Li, I. H. Witten and A. Yu, "Constructing a collocation learning system from the Wikipedia corpus". *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT)*, 6(3), 18-35, 2016.
- [19] J. König, "Automating vocabulary tests and enriching online courses for language learners", unpublished doctoral dissertation. The University of Waikato, New Zealand, 2019.