MOOCs as environments for learning spoken academic vocabulary

Clinton Hendry[1] and June Ruivivar[2]

Abstract. Massive Open Online Courses (MOOCs) are easily accessible for anyone in the world to study any given subject, often for free. However, there is some question as to whether they are comparable to their real-world counterparts. The Academic Spoken Word List (ASWL) created by Dang, Coxhead, and Webb (2017) was designed to create a word list that is more representative of spoken academic English. To contrast the real-world academic context to MOOCs, we created a MOOC academic corpus and compared it with the Michigan Corpus of Academic Spoken English (MICASE). Last, we used both to test the effectiveness of the ASWL. Overall, we found that the ASWL had similar coverage in both the MOOC and MICASE corpora but interestingly saw slightly more coverage in the MICASE dialogic sections. We believe future research should address the slight discrepancy between dialogic and non-dialogic academic situations.

Keywords: academic English, spoken academic English, corpus linguistics, online courses.

1. Introduction

There are considerable lexical differences between written and spoken academic English. For example, Biber et al. (2002) illustrate that classroom teaching, previously thought to be highly informational and persuasive, is in fact conversational and dialogic. Conversely, materials intended to be accessible to students, such as university brochures, were found to be informationally dense. To fully prepare second-language users for English-medium instruction, English for

1. Concordia University, Montreal, Canada; clinton.hendry@concordia.ca
2. Concordia University, Montreal, Canada; june.ruivivar@concordia.ca

How to cite this article: Hendry, C., & Ruivivar, J. (2019). MOOCs as environments for learning spoken academic vocabulary. In F. Meunier, J. Van de Vyver, L. Bradley & S. Thouësny (Eds), *CALL and complexity – short papers from EUROCALL 2019* (pp. 180-185). Research-publishing.net. https://doi.org/10.14705/rpnet.2019.38.1006

Academic Purposes (EAP) and similar language training courses need to include the variety of registers they are expected to encounter.

To this end, Dang et al. (2017) created the ASWL to identify words that are most useful for L2 students in English-medium universities. The ASWL serves to complement the previous Academic Word List (AWL; Coxhead, 2000), which was based on written texts. Although the ASWL has been validated across several corpora, the authors recognize the need to validate it across a variety of academic contexts.

The present study responds to this call by examining the ASWL's coverage of MOOCs, an increasingly popular mode of instruction (Lederman, 2018). Although MOOCs are primarily offered in English, many online students speak English as an additional language (Haber, 2014). Therefore, to be on par with in-person courses, English-language MOOCs must be accessible to learners of different proficiencies and provide them with opportunities for further language development. In many ways, the spoken component of MOOCs resembles that of in-person courses, but with video lectures making up most of the content. One notable difference is the lack of interactive elements such as classroom management and service encounters, which Biber et al. (2002) have found to have distinct register characteristics. This raises two questions. First, how accessible are MOOCs to students who have limited knowledge of academic English? Second, do MOOCs offer the same opportunities for incidental vocabulary learning observed in classroom tasks (e.g. Newton, 2013)?

2. Method

2.1. Corpora

This study compares the ASWL coverage between two corpora: MICASE (Simpson, Briggs, Ovens, & Swales, 2002) and a new corpus of online courses from edX, a MOOC platform offering courses from universities around the world. MICASE consists of 1.8 million words across various speech events, classified on a scale of interactivity from mostly monologic (e.g. presentations and speeches) to mostly interactive (e.g. advising and tutorials). Our MOOC corpus consists of video lectures, interviews, and live streaming sessions from 18 courses, totaling 733,431 words.

The MOOC corpus was compiled from transcripts from a wide range of MOOCs available on the edX platform. Our goal was to create a corpus that was

representative of multiple academic fields from the natural sciences (247,247 words), social sciences and humanities (241,200), and vocational courses (126,000). Each transcript was individually cleaned of non-alphanumeric characters (e.g. @, #, $, etc.), html tags (i.e. < … >), and onomatopoeic language (e.g. [boom!] [loud crash] etc.).

The ASWL was created by Dang et al. (2017) and consists of 1,741 word families. It was compiled based on overall frequency and dispersion across academic domains found in a 13-million-word academic spoken corpus. To be included in the list, each item had to occur at least once in each academic subcorpus (hard applied sciences, soft applied sciences, etc.), 50% of all discipline-specific subcorpora, and at least 350 times overall. The list was then validated using a different but similarly sized academic spoken corpus.

2.2. Analysis

Analysis was conducted using AntConc v3.4 (Anthony, 2018) using the 'Stop List' function which allows us to remove all instances of ASWL items from a given text. We then calculated coverage by computing the difference between the original corpus and the corpus with all ASWL families removed.

3. Results

Table 1 shows the breakdown of ASWL coverage for the entire MOOC corpus and the speech events of the MICASE corpus. The ASWL saw comparable coverage of 86% for the MOOC corpus and 87% overall for MICASE, indicating that the ASWL is useful for learners in both in-person and online contexts. It also suggests that MOOCs' vocabulary requirements are comparable to that of real-world universities, and that online environments can offer opportunities for incidental learning of academic vocabulary at least at par with in-person courses.

Table 1. ASWL coverage by speech event type

	# words	ASWL coverage
MOOC corpus	746,231	107,211 (86%)
MICASE speeches	26,563	2,593 (89%)
MICASE thesis defenses	53,980	6,124 (89%)
MICASE campus tours	22,734	2,325 (90%)
MICASE seminars	99,685	9,560 (90%)
MICASE lectures	505,281	55,089 (89%)

MICASE meetings	68,062	8,353 (88%)
MICASE discussion panels	92,183	11,098 (88%)
MICASE interviews	12,097	1,062 (92%)
MICASE tutorials	26,670	2,303 (91%)
MICASE workshops	14,252	1,044 (93%)
MICASE study groups	61,300	6.458 (89%)
MICASE laboratory sessions	58,557	8,229 (86%)
MICASE service encounters	25,054	2,183 (91%)
MICASE advising sessions	43,828	4,912 (89%)
MICASE office hours	76,084	8,529 (89%)

4. Discussion

Our analysis shows that there is similar coverage across different speech event types in MICASE. The lack of interactivity therefore does not appear to substantially affect MOOCs' lexical content and accessibility. However, coverage exceeded 90% in interviews, tutorials, workshops, and service encounters, all of which are dialogic or interactive contexts that are lacking in the MOOC environment. Other interactive speech events, such as meetings and discussion panels, may have received less coverage because they occur in more formal academic environments, making them more informationally dense. These contexts also tend to have one or more dominant speakers at a time and are thus less interactive and more persuasive in nature. Classroom lectures, which are the closest in format to the MOOC courses, also received slightly higher coverage than MOOCs (86% vs. 89%). These findings suggest that the ASWL can adequately prepare English as a Second Language (ESL) students for a broad range of linguistic encounters at university, both within and outside of academic contexts. Despite the slightly lower coverage, it also covers a substantial part of the vocabulary requirements of MOOCs.

5. Conclusions and future work

This study revealed that the ASWL (Dang et al., 2017) provides adequate coverage for MOOCs, as well as a variety of academic contexts that ESL learners are expected to encounter in English-medium universities. Our results indicate that MOOCs' vocabulary requirements are slightly lower but comparable to those of real-world universities, and that online environments can offer opportunities for incidental learning of academic vocabulary at least at par with in-person courses. However, it also revealed some differences in coverage among interactive speech events.

Notably, the ASWL offered the greatest coverage in highly dialogic and interactive contexts and the least in more typically formal speech events, indicating possible differences along a dimension of formality or topic. However, these results may be indicative of the overlap (approx. 85%) between the ASWL and the 2,000 most frequent words in English (Dang et al., 2017). Consequently, the slightly higher coverage in the interactive contexts might indicate a slightly lower use of more discipline-specific technical vocabulary.

Further research is called for to investigate such differences in more detail. Such research can, for example, identify specific linguistic markers in formal versus informal academic contexts, or create specific word lists for spoken encounters in different academic disciplines. Indeed, Hyland and Tse (2007) have argued that the AWL's usefulness varies across academic disciplines, and similar research might reveal similar patterns in the ASWL.

In terms of pedagogy, EAP courses using the ASWL can further support vocabulary learning by offering practice in a variety of registers and non-classroom speech events with confidence, of which the ASWL provides substantial coverage. Universities can also emphasize online courses to support ESL courses to enhance international students' preparation for in-person academic studies.

References

Anthony, L. (2018). AntConc (Version 3.5.7) [Computer Software]. Waseda University. http://www.laurenceanthony.net/software

Biber, D., Conrad, S., Reppen, R., Byrd, P., & Helt., M. (2002). Speaking and writing in the university: a multidimensional comparison. *TESOL Quarterly*, *36*(1), 9-48. https://doi.org/10.2307/3588359

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213-238. https://doi.org/10.2307/3587951

Dang, T. N. Y., Coxhead, A., & Webb, S. (2017). The academic spoken word list. *Language Learning*, *67*(4), 959-997. https://doi.org/10.1111/lang.12253

Haber, J. (2014). *MOOCs (The MIT Press Essential Knowledge Series)*. MIT Press.

Hyland, K., & Tse, P. (2007). Is there an "academic vocabulary"? *TESOL Quarterly*, *41*(2), 235-253. https://doi.org/10.1002/j.1545-7249.2007.tb00058.x

Lederman, D. (2018, February 14). MOOCs: fewer new students, but more are paying. *Inside Higher Ed*. https://www.insidehighered.com/digital-learning/article/2018/02/14/moocs-are-enrolling-fewer-new-students-more-are-paying-courses

Newton, J. (2013). Incidental vocabulary learning in classroom communication tasks. *Language Teaching Research*, *17*(2), 164-187. https://doi.org/10.1177/1362168812460814

Simpson, R. C., Briggs, S. L., Ovens, J., & Swales, J. M. (2002). *The Michigan corpus of academic spoken English*. https://quod.lib.umich.edu/m/micase/

Published by Research-publishing.net, a not-for-profit association
Contact: info@research-publishing.net

CALL and complexity – short papers from EUROCALL 2019
Edited by Fanny Meunier, Julie Van de Vyver, Linda Bradley, and Sylvie Thouësny

Publication date: 2019/12/09

Typeset by Research-publishing.net
Cover theme by © 2019 Frédéric Verolleman
Cover layout by © 2019 Raphaël Savina (raphael@savina.net)

Fonts used are licensed under a SIL Open Font License

ISBN13: 978-2-490057-54-2 (Ebook, PDF, colour)
ISBN13: 978-2-490057-55-9 (Ebook, EPUB, colour)
ISBN13: 978-2-490057-53-5 (Paperback - Print on demand, black and white)
Print on demand technology is a high-quality, innovative and ecological printing method; with which the book is never 'out of stock' or 'out of print'.

British Library Cataloguing-in-Publication Data.
A cataloguing record for this book is available from the British Library.

Legal deposit, France: Bibliothèque Nationale de France - Dépôt légal: décembre 2019.