A Case Study of the Characterizations of Spoken Academic English: The Adjective-Noun Combination

Fuying Lin (Free University of Berlin, Germany)

1. Introduction

"Academic English" constitutes a variant of English with properties unique enough to describe it as register (in theoretical linguistics) and to teach it with a specific case of English for Special Purposes (in applied linguistics) (cf. Brisk & Jeffries, 2008 and Bailey, 2012 for overviews and Benesch, 2001 for a critical appraisal). Academic English is largely treated as a written register with occasional exceptions, which often tends to be tiny portion of language materials; for example, spoken materials used in the New Academic Word list (NAWL 1.0, cf. Browne, C. et.al, 2013) is merely 1%, which is unlikely to have a noticeable effect on the word selection process. Although theoretical researchers are more inclusive of spoken language on average, researchers like Biber and Grey also equate Academic English with written academic prose in their investigation (Biber & Grey, 2016). Where spoken language is included, the definition of "Academic English" is often too broad, covering any type of classroom discourse, for instance, in the Corpus of English as Lingua Franca in Academic Settings (ELFA 2008) or in the Michigan Corpus of Spoken Academic English (MICASE) (Simpson, Briggs, Ovens, & Swales, 2002). And with the increasing popularity of Massive Open Online Courses (MOOCs), the visibility of the spoken contents in Academic English is worthy of more severe manifestation (see Baiely, 2012). This paper, seen as a pilot study of my dissertation, will narrowly aim at the investigation of the adjective-noun combination structure manipulated in Spoken Academic English.

2. Background

2.1. MOOCs and their significance

MOOC is an abbreviation of *Massive Open Online Courses*, a term that emerged in 2008 for a particular type of Internet-based long-distance learning format that has become a new extension of higher education in recent years (Fini, 2009; Kennedy, 2014). The year 2012 was even identified as the year of MOOC due to the enormous number of platforms, organizations and institutes joining the MOOC movement with offerings of their own (Pappano, 2012). There are at least 26 known providers of MOOCs, three of which can be considered major providers—Coursera, Edx and Udacity (Ha, 2014).

MOOCs clearly present a novel learning opportunity, making Higher Education available to many potential students all over the world. Unlike previous forms of self-study and/or long-distance learning, however, the participation in MOOCs requires a thorough command of academic English, not just in its written but also in its spoken form. This issue has not received enough attention in the research on academic English in general.

2.2 Definition and significance of Spoken Academic English

Several scholars have tried to distinguish academic English from general English (Coxhead and Nation, 2001; Schmitt, 2000; Xue and Nation, 1984). Academic English, different from everyday spoken English, has its own professional vocabulary (not limited to technical terminology), and is grammatically distinct with its complex, informationally-dense structures (Biber, Finegan, Johansson, Conrad, & Leech, 1999). There is an implicit equation of academic English with written English, even in recent research that challenge some of the stereotypical views of this register, such as Biber & Gray (2016). And although a few corpora that could be called Spoken Academic English have been constructed, such as 1.8-million-word Michigan Corpus of Spoken Academic English (MICASE) corpus (Simpson et al., 2002), the British Academic Spoken Corpus (BASE) (Nesi & Thompson, 2001), the Corpus of English as Lingua Franca in Academic Setting (ELFA 2008), and the 2-million-word Hong Kong Corpus of Spoken English (HKCSE)(Warren & Chen, 2004), they collected language data from various academic settings like classrooms, meetings, seminars, conferences and job interviews.

For the objective of my research, I will define Spoken Academic English more narrowly as *the spoken language used by academics in talking about their subject in academic settings*. This includes talks by and discussions amongst researchers at academic conferences as well as lectures by instructors in (physical or virtual) university classrooms, but excludes academics using language in non-academic university settings (such as faculty meetings), language dealing with organizational aspects of teaching, and classroom language by non-academics (i.e. students). The latter may be considered a learner variety of Spoken Academic English, worthy of study in its own right. The language of MOOCs, which is the focus of this research, clearly falls under this definition.

The Spoken Academic English found in the context of MOOCs is worthwhile not just in terms of characterizing it as an (emerging) register, but also with a view to improve ESP teaching and self-study resources. Some scholars have argued that an identifiable subgenre within Spoken Academic English from MOOCs justifies "EMP" (English for MOOC purposes, see Anthony, 2015).

3. Methodology

Because the current corpora of Spoken Academic English are either too small (such as BASE with 1.6 million tokens)(Thompson & Nesi, 2001), or proprietary and inaccessible (such as PICA, Ackermann et al., 2010), to achieve the aims of my research, I have specifically constructed a corpus, the MOOC corpus, consisting of 93 lecture transcripts from MOOCs, amounting to 8,716,104 tokens, in five broad subjects represented offered by the major MOOC providers (Coursera, EdX, Udacity and Futurelearn), as shown in the table 1 below.

The subjects are those widely agreed upon in studies of Academic English, represented in such large corpus as the British Academic Written English Corpus(BAWE)(Garnder & Nesi, 2012, 2013).

In order to identify the unique properties of Spoken Academic English, a corpus of general spoken English (the spoken part of COCA, the Corpus of Contemporary American English)(Davies, 2010) will be treated as the reference corpora.

Table 1. Overview of the MOOC spoken academic corpus

category	subcategory	course number	Total course number	Word count	Total word count
Arts and humanities	language & literature	5	25	252,219	1,855,055
	history	5		610,621	
	art, architecture & design	5		312,465	
	music and design	5		317,788	
	philosophy	5		361,962	
Social science	sociology	4	20	261,319	1,626,268
	political science	4		328,699	
	education	4		211,313	
	economics & finance	4		413,999	
	law	4		410,938	
Life science	biology	4	16	468,818	1,416,234
	Health science	4		294,882	
	food & nutrition	4		169,415	
	psychology	4		483,119	
Physical science	physics, astronomy	4	16	851,967	1,822,527
	chemistry	4		332,931	
	engineering	4		213,981	
	energy & earth science	4		428,648	
Information, computer science & mathematics	computer science	4	16	621,880	1,996,020
	robotics & electronics	4		347,827	
	mathematics	4		411,320	
	data science	4		615,533	
Total			93		8,716,104

4. Results & Conclusion

Academic English is assumed to be characterized both by domain-specific vocabulary and by vocabulary also found in colloquial language but used with specific technical meanings in academic settings. Terminological vocabulary may consist of single words, but more often it consists of multi-word expressions of sequences like Adjective-Noun or Noun-Preposition-Noun (Justeson & Katz, 1995). Comparing the frequency of, in this paper, Adjective-Noun combinations in the MOOC corpus against their frequency in the spoken part of the COCA and extracting combinations that occur significantly more frequently in the former than in the latter should thus allow us to identify academic terminology.

Table 2 shows the top 15 adjective-noun combinations that are significantly

more frequent in the MOOC data than in the spoken part of the COCA with their observed and expected frequencies, the log-likelihood value and the information whether they occur in both corpora or only in the MOOC data.

Table 2. Ad-N combinations significantly more frequent in the MOOC corpus compared to COCA

WORD	MOOC data Observed Expected		COCA spoken data Observed Expected		LL	Shared
1 random variable	1419	320.8	0	1098.2	4222.74	No
2 random variables	834	188.6	0	645.4	2481.05	No
3 special theory	722	163.9	3	561.1	2110.35	Yes
4 fair use	465	106.7	7	365.3	1313.77	Yes
5 next video	467	109.2	16	373.8	1256.68	Yes
6 next lecture	348	78.9	1	270.1	1021.78	Yes
7 differential equation	343	77.8	1	266.2	1006.94	Yes
8 gradient descent	338	76.4	0	261.6	1005.23	No
9 absolute value	315	72.4	5	247.6	887.87	Yes
10 initial state	298	67.8	2	232.2	863.24	Yes
11 solar system	529	175.9	249	602.1	725.50	Yes
12 white dwarf	240	54.5	1	186.5	701.28	Yes
13 kinetic energy	255	60.1	11	205.9	672.36	Yes
14 straight line	371	197.6	105	368.4	654.86	Yes
15 conditional expectation	219	49.0	0	169.5	651.27	No

The approach clearly yields mostly scientific terminology, either in a narrow (domain-specific) sense (like *random variabie(s)*, *differential equation*, *gradient descent, kinetic energy*) or in a general sense (*fair use, absolute value, initial state, solar system*). The former tend to be restricted to the MOOC data, while the latter tend to occur in the general language too, but have a special technical meaning in academic English.

References

Ackermann, K., De Jong, J.H.A.L., Kilgarriff, A. & Tugwell, D. (2010). The Pearson International Corpus of Academic English (PICAE).

Anthony, L. (2015). The Changing Role and Importance of ESP in Asia. *English as a Global Language Education (EaGLE) Journal*, 1 (1), 1-21.

Coxhead, A., & Nation, I.S.P. (2001). *The Specialized Vocabulary of English for Academic Purposes*. In. J. Flowerdew & M. Peacock (Eds.), Research perspectives on English for academic purpose. 252-267. Cambridge, U.K.: Cambridge University Press.

Bailey, L. (2012). Academic English. In James Banks (Eds.), *Encyclopedia of Diversity in Education*. 4-9. Thousand Oaks, CA: SAGE.

Benesch, S. (2001). *Critical English for academic purposes: theory, politics, and*

practice. Mahwah, N.J: L. Erlbaum Associates.

Biber, D., Finegan, E., Johnasson,S., Geoffrey N. L., & Conrad.S. (1999). *Longman grammar of spoken and written English*. Harlow: Longman.

Biber, D. & Gray, B. (2016). *Grammatical Complexity in Academic English: Linguistic Challenge in Writing*. Cambridge: CUP.

Brisk, M. E. & Julian J. (2008). Academic English. In Josué González (ed.), *Encyclopedia of Bilingual Education* .1-4. Thousand Oaks, CA: SAGE.

Browne.C., Culligan, B. & Phillips, J. (2013). *The New Academic Word List*. Retrieved from http://www.newgeneralservicelist.org

Cheng, W., Greaves, C. & Warren M. (2005). The creation of a prosodically transcribed intercultural corpus: The Hong Kong Corpus of Spoken English (prosodic). *International Computer Archive of Modern English* (ICAME) 29, 5-26.

Davies, M. (2010). The Corpus of Contemporary American English as the first reliable monitor corpus of English. *Literary and Linguistic Computing* 25(4), 447-464. doi:10.1093/llc/fqq018

ELFA. (2008). The Corpus of English as a Lingua Franca in Academic Settings. Director: Anna Mauranen. http://www.helsinki.fi/elfa/elfacorpus

Fini, A. (2009). The technological dimension of a massive open online course: The Case of the CCK08 Course Tools. *The International Review of Research in Open and Distance Learning* 10 (5). doi: http://dx.doi.org/10.19173/irrodl.v10i5.643

Gardner, S. & Nesi, H. (2013). A classification of genre families in university student writing. *Applied Linguistics*, 34 (1), 1-29.

Ha, T.H. 2014. MOOCs By the Numbers: Where Are We Now? Retrieved from http://ideas.ted.com/moocs-by-the-numbers-where-are-we-now/

Justeson, J. S. & S. M. Katz (1995) Technical terminology: some linguistic properties and an algorithm for identification in text. *Natural Language Engineering* 1, 9-27.

Kennedy, J. (2014). Characteristics of Massive Open Online Courses (MOOCs): A Research Review. *Journal of Interactive Online Learning*, 13.

Nation, P. & Xue, G. (1984). A university word list. *Language Learning and Communication*, 3 (2), 215-229.

Pappano, L. (2012). The year of the MOOC. *New York Times*. Retrieved from http://www.nytimes.com/2012/11/04/education/edlife/massive-open-online-courses-are-multiplying-at-a-rapid-pace.html?pagewanted=all&_r=1&

Schmid, H. (1994). Probabilistic part-of-speech tagging using decision trees. *Proceedings of International Conference on New Methods in Language Processing*. Manchester, U.K..

Schmid, H. (1995). *Improvements in part-of-speech tagging with an application to German*. Proceedings of the ACL SIGDAT-Workshop. Dublin, Ireland.

Schmitt, N. (2000). *Vocabulary in Language Teaching*. Cambridge, U.K.: Cambridge University Press.

Simpson R.C., Briggs, S.L., Ovens, J., and Swales, J.M.. (2002). *The Michigan Corpus of Academic Spoken English*, Ann Arbor, MI: The Regents of the University of Michigan.

Thompson, P. and Nesi, H. (2001). The British Academic Spoken English (BASE) Corpus Project. *Language Teaching Research* (3), 263-264.