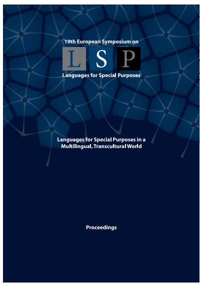

Languages for Special Purposes in a Multilingual, Transcultural World

Proceedings of the 19th European Symposium on Languages for Special Purposes, 8-10 July 2013, Vienna, Austria

http://lsp2013.univie.ac.at/proceedings

Language support for teachers and students in engineering via a lecture corpus interface

Nílson Kunioshi; Judy Noguchi; Kazuko Tojo; Hiroko Hayashi

Cite as: Kunioshi, N. et al. (2014). Language support for teachers and students in engineering via a lecture corpus interface. In G. Budin & V. Lušicky (eds.), *Languages for Special Purposes in a Multilingual, Transcultural World, Proceedings of the 19th European Symposium on Languages for Special Purposes, 8-10 July 2013, Vienna, Austria*. Vienna: University of Vienna, 380-385.

Publication date: July 2014

ISBN: **978-3-200-03674-1**

License: This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc/4.0/. This license permits any non-commercial use, distribution and reproduction, provided the original authors and source are credited.

Language support for teachers and students in engineering via a lecture corpus interface

Nílson Kunioshi
School of Creative Science and Engineering, Waseda University
Japan

Judy Noguchi
School of Pharmacy and Pharmaceutical Sciences, Mukogawa Women's University
Japan

Kazuko Tojo
Department of International and English Interdisciplinary Studies, Osaka Jogakuin University
Japan

Hiroko Hayashi
Center for International Education and Exchange, Osaka University
Japan

Correspondence to: nilson@waseda.jp

Abstract. In regions where English is not the native language, a shift to English-medium university instruction poses great challenges. For students and teachers who are nonnative English speakers (NNS), a good understanding of the academic lecture framework, as well as the ability to fluently receive/produce the academic content in real time is required to comprehend/deliver lectures in English. Student problems range from vocabulary comprehension to understanding of the information flow, while instructors need to mark clearly their discourse to aid student understanding. In order to support these efforts, we developed OnCAL (the Online Corpus of Academic Lectures, http://www.oncal.sci.waseda.ac.jp/), a web interface that helps NNS in science and engineering learn about words and expressions that are important in academic lectures. As of August 1, 2013, the corpus consists of 432 transcripts of lectures delivered at MIT and Stanford University. Words and expressions were identified as meaningful for academic lecture presentation and comprehension. Frequently used word clusters were identified and examined for their functions in the lecture as a whole. These words/expressions are found to be useful for guiding students through the information flow of the lectures and, therefore, can be used by instructors to prepare effective lectures.

Keywords. English-medium instruction, science classroom discourse, science higher education.

1. Introduction

The ability to attract international students is one of the criteria for ranking universities on a global scale (Salmi, 2009), and English-medium instruction seems to be particularly effective (Bologna Process, 2010). The number of education programs offered through English has grown significantly in Europe during the period from 2002 to 2007 (Wächter and Maiworm, 2008). Also in Japan, the Ministry of Education, Culture, Sports, Science and Technology (MEXT) recently launched the "Global 30" project to encourage universities to offer undergraduate and graduate programs taught completely in English to attract more international students (MEXT, 2009). However, a shift to English-medium instruction poses great challenges in a country like Japan where English is not the native language.

One of these challenges is the ability of students who are non-native speakers of English (NNS) to listen to and sufficiently understand lectures delivered in English, and another is the ability of instructors to deliver effective academic lectures in a second language. Both tasks require an understanding of the academic lecture framework, as well as the ability to fluently receive/produce the academic content in real time. On the part of the students, even 'simple' problems related to vocabulary can be stumbling blocks (Arden-Close, 1993). Other issues encountered by students learning in a second language have also been the subject of much research in different countries. For example, Evans and Morrison (2011) showed that first-year students in Hong Kong needed to have strong motivation, work hard, use effective learning strategies, and receive support to be successful in the adapting to a second language environment. For students studying physics concepts, an investigation of the effects of the medium of instruction on student performance led Airey and Linder (2008, 2009) to recommend that teachers encourage students to ask questions during or after class, give out lecture materials in advance, and use more visual illustrations in addition to oral explanations. Tan and Lan (2011) reported on the challenges faced by teachers and students in Malaysia around 2003, when English became the medium of instruction. In the classroom, instructors especially need to know how to mark their discourse so that students will know what to listen for (Deroey, 2012). Issues in teaching practice arising after implementation of English-medium instruction in higher education in the Netherlands were reported by Klaasen and de Graaff (2001); they noticed that NNS teachers needed to develop effective teaching skills using English and at the same time be aware of the language difficulties that NNS students may have when learning through English. In Denmark, Thøgersen and Airey (2011), analyzing differences between speaking rates and rhetorical styles in lectures delivered in Danish and English, found that a very experienced, English-fluent Danish professor teaching the same content in the two languages, spoke more slowly and used a more formal style with English.

Spoken language is just one of the modes which appear in a science classroom (Kress et al. 2001), and the quality of instruction does not depend only on the spoken mode (Neumann et al. 2012). However, the role of spoken language in the science classroom is still very important (Gee 2004, Llinares et al. 2012, Smit and Dafouz 2012). Especially in the case of English-medium instruction in a NNS environment, improvement in the quality of instruction cannot be achieved without a careful look at the classroom spoken discourse. We hypothesized that analyzing a corpus of lectures in science and engineering could contribute to improving the quality of instruction by offering linguistic options that NNS teachers may be less aware of, and also foster the lecture comprehension skills of NNS students. This led us to develop OnCAL, the Online Corpus of Academic Lectures (Kunioshi et al. 2012, http://www.oncal.sci.waseda.ac.jp/), a web interface that allows NNS in science and engineering to find words and expressions that are important in academic lectures. In this work, we show some words and expressions that are used frequently by university lecturers in the United States to signal specific linguistic functions during the logical flow of lectures. This should help NNS instructors prepare lectures in English and NNS students improve their listening comprehension by being aware of how these words and expressions are used in academic lectures.

2. Method

2.1. Corpus building

Transcriptions of lectures on courses related to science and engineering were downloaded from MIT Opencourseware (MIT OCW, http://ocw.mit.edu/index.htm) and Stanford Engineering Everywhere (SEE, http://see.stanford.edu/). The Creative Commons License allows full use of both MIT OCW and SEE contents as long as these are "shared alike" (http://creativecommons.org/licenses/by-nc-sa/3.0/us/legalcode).

Some relevant data related to the transcriptions that were downloaded from MIT OCW and SEE, and uploaded to OnCAL, are shown in Table 1. Some light editing, which did not affect the accuracy of transcriptions as, for example, substituting special characters used in the transcriptions (changing single quotation marks into double quotation marks, or standardizing all comments by the transcriber into a form like "[APPLAUSE]"), was done for consistency along different texts and sources. Lists with detailed data related to each single lecture in each course are available online (http://www.oncal.sci.waseda.ac.jp/lists.aspx). As of August 1, 2013, the total number of lecture transcripts uploaded to OnCAL is 432; the corpus comprises 3.5 million words, which correspond to a total lecture time (obtained from the length of the video recordings) of 397 hours.

Course No.	Field	Source	Student year	Lecture time (hh:mm:ss)	No. of words
1	Chemistry	MIT	1st	27:23:08	185,290
2	Physics: Mechanics	MIT	1st	28:47:53	229,006
3	Physics: Electricity/Magnetism	MIT	1st	30:11:40	248,620
4	Biology	MIT	1st	28:37:51	240,627
5	Math: Calculus	MIT	1st	28:20:14	201,194
6	Math: Differential Equations	MIT	1st	25:24:31	189,548
7	Computer Science: Programming I	SEE	Undergrad	22:17:30	292,165
8	Computer Science: Programming II	SEE	Undergrad	21:02:25	278,003
9	Computer Science: Programming III	SEE	Undergrad	22:27:22	214,539
10	Math: Fourier Transform	SEE	Graduate	25:38:05	222,721
11	Math: Linear Dynamical Systems	SEE	Graduate	24:26:52	238,649
12	Math: Convex Optimization I	SEE	Graduate	24:00:06	233,967
13	Math: Convex Optimization II	SEE	Graduate	21:58:28	209,853
14	Artificial Intelligence (AI): Robotics	SEE	Graduate	18:48:00	131,745
15	AI: Natural Language Processing	SEE	Graduate	22:01:12	193,299
16	AI: Machine Learning	SEE	Graduate	25:09:12	188,100
Total				3,497,326	3,497,326

Table 1: List of courses downloaded from MIT OCW and SEE and uploaded to OnCAL (as of August 1, 2013)

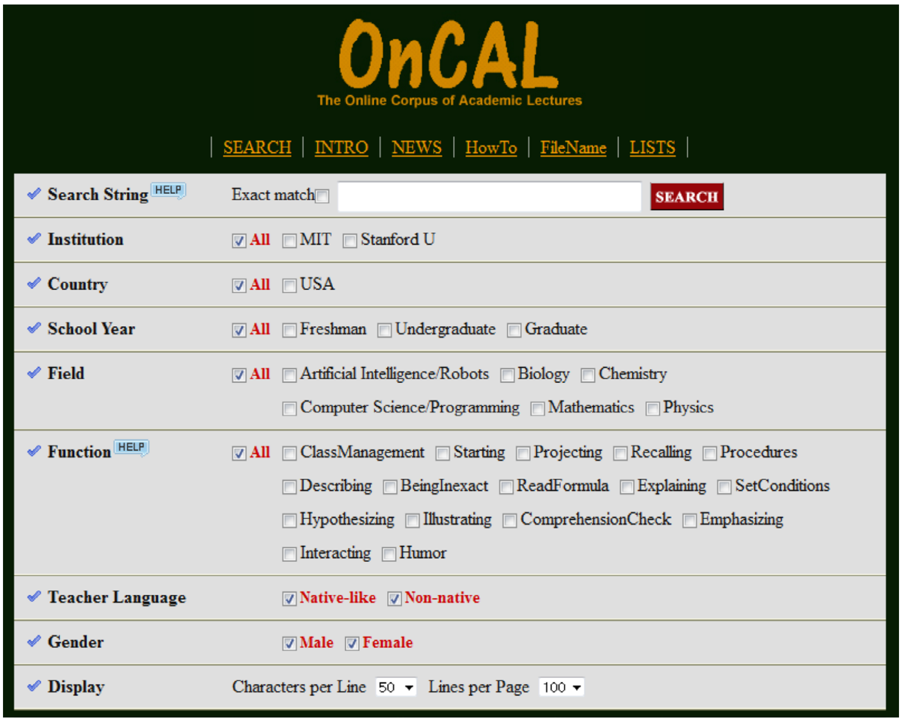

Figure 1: Top page of OnCAL, as of August 1, 2013

2.2. Interface design

The user interface was designed so that all functionalities can be "discovered" in an intuitive way, but at the same time we also assumed that users can learn the functions from trial and error. Figure 1 shows the homepage that is displayed when http://www.oncal.sci.waseda.ac.jp/ is accessed. The input box for the search string is clearly presented, the default conditions are displayed, and the search conditions can be changed easily. A search can be restricted to a particular field, or to a particular source, for example.

2.3. Linguistic functions

Linguistic functions were identified based on studies of classroom discourse (see, for example, Dalton-Puffer 2007) and an analysis of the corpus. Frequent 2-, 3-, and 4-grams related to each linguistic function were found using the "clusters" function of AntConc (Anthony, 2012) and counted carefully through the OnCAL interface. These expressions are registered in the OnCAL system, and displayed when the user chooses a specific function.

3. Results and discussion

When OnCAL was first released in 2010 and used in a workshop for content teachers, it had no linguistic functions. The users had to imagine words or expressions to search for, and ended up searching for technical terms such as "partial derivatives" or names of chemical compounds. Content NNS teachers did not consider the production of a logical flow in the classroom as a priority, and rather their main concern was on how to read a mathematical equation in English. We learned about their needs from their comments and the way they used OnCAL, and later decided to offer expressions for linking one technical concept to another, or for emphasizing the importance of a concept. We also wanted to offer the information that they explicitly asked for: how to read equations. Therefore, OnCAL was redesigned and can now display example sentences for a set of linguistic functions; the user can thus obtain example sentences for a specific function without having to imagine a word or expression in advance. For example, if the user chooses the linguistic function "Emphasizing" and press "Search", 153 sentences are displayed. Some of them are shown in Table 2.

	It makes sense	because if someone walks up to you
	It makes sense	both ways, but you can't drop the
	It makes sense	to do it once for the string, and
	It makes sense	to multiply t times the function
	It makes sense	to talk about the impulseresponse
in here basically because	it makes sense	that we group all the semaphores
hand side makes sense because	it makes sense	to multiply a distribution times a
Also	notice that the	peak value at resonance has not
I want you to also	notice that the	difference in energies between
You'll also	notice that the	gene-rich regions, here, are rich
That's something	you should remember	about planes.
If I use that label t2g	you should remember	the specific d-orbitals that contr
And, that's a formula that	you should remember	.
Not that	you should remember	this one by heart.
That is something that	you should remember	from your studies of chemical equi
A couple of things that	you should remember	.

Table 2: Some of the concordance lines displayed when function "Emphasizing" is chosen

"It makes sense", "notice that", and "you should remember" are expressions that were registered as frequent for the function "Emphasizing". All sentences that contain those expressions are displayed, but only some are shown in Table 2. From the sample sentences displayed, users should be able to gain clues for better searches.

OnCAL thus offers sample sentences without requiring the input of a word or expression. However, one limitation of the current version of OnCAL is that if the user wants to input a string and restrict the search to a specific function, very few sentences are found. This is because only sentences containing the registered expressions are considered in the search. Another limitation is that content teachers may not see linguistic functions such as "class management", "projecting", or "recalling" as important. Thus, changing the functions to "pedagogical functions" such as "making links to previous content", which may include a portion of the current "recalling", may be more meaningful for NNS content teachers and elicit more insights. One of the expressions registered for "recalling" is "last time"; some example sentences are shown in Table 3. In the sentences shown in Table 3, the teacher recalls what was mentioned in the previous session as a device for linking what was mentioned before to what is being taught or will be taught in the present session. The link is made for showing differences, or recalling previous knowledge that is needed for developing new content.

You recall that	last time we	were talking about the
You recall that	last time we	were talking about the process of
I think	last time we	already decided that this guy
And I think	last time we	started in the sense that we
Okay, well,	last time we	started looking at a little bit
have a quality constraints,	last time we	derived in terms of a little bit of
In other words,	last time we	were talking about the deviation

Table 3: Some of the concordance lines displayed when function "Recalling" is chosen

This is one example of pedagogical link-making, which "is concerned with the ways in which teachers and students make connections between ideas in the ongoing meaning-making interactions of classroom teaching and learning" (Scott et al., 2011). Because learning progressions and teaching sequences (see, for example, Duschl et al. 2011) are not possible without this type of link-making, such utterances should be used properly by teachers in the classroom to help students make connections in their meaning making process.

Searches through OnCAL can be restricted to a specific field, or even to a specific course to see, for example, how link-making utterances change from the first to subsequent lectures.

Other functionalities of OnCAL can be explored in many different ways, for many different purposes. Content teachers, content students, and applied linguists can hopefully obtain meaningful insights from this corpus.

4. Conclusions

Expressions frequently used by teachers (native speakers of English, mainly) in two universities in the United States were identified and linked to the linguistic functions they serve in the logical flow of lectures. Users of OnCAL, the Online Corpus of Academic Lectures, are now able to find these expressions by choosing a specific linguistic function with no need to think of words to search for. The expressions are expected to be useful for NNS teachers in their preparation of lectures in an English-medium program. These expressions can also help students comprehend lectures faster or in more depth. Applied linguists or language teachers may also find insights about the linguistic features of science and engineering lectures, or for supporting the language needs of NNS students learning through English.

Further work is in progress towards allowing users to combine string searches with specific linguistic functions and also to allow users to find expressions related to pedagogical functions.

5. Acknowledgements

This work has been supported by a Grant-in-Aid for Scientific Research (B, No. 24300273) from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) of Japan. The authors want also to express their gratitude to teachers and staff who delivered the lectures, recorded, transcribed, and uploaded the materials to MIT Opencourseware and Stanford Engineering Everywhere; their hard work and generosity in making such valuable data available to the world deserve recognition from all who benefit from the data.

6. References

Airey, J. & Linder, C. (2008). Bilingual scientific literacy? The use of English in Swedish university science courses, *Nordic Journal of English Studies*, 7, 145–161.

Airey, J. & Linder, C. (2009). A disciplinary discourse perspective on university science learning: Achieving fluency in a critical constellation of modes, *Journal of Research in Science Teaching*, 46(1), 27–49.

Anthony, L. (2012). AntConc3.2.4w [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.antlab.sci.waseda.ac.jp/.

Arden-Close, C. (1993). Language problems in science lectures to non-native speakers. *English for Specific Purposes*, 12(3), 251-261.

Bologna Process (2010). *The Bologna Process independent assessment. The first decade of working on the European Higher Education Area. Volume 2, Case studies and appendices*. Retrieved in June 22, 2011 from http://www.ond.vlaanderen.be/ hogeronderwijs/bologna/2010_conference/documents/.

Dalton-Puffer, C. (2007). *Discourse in Content and Language Integrated Learning (CLIL) classrooms*. Amsterdam: John Benjamins.

Deroey, K. L. B. (2012). "What They Highlight Is...": The Discourse Functions of Basic "Wh"-Clefts in Lectures, *Journal of English for Academic Purposes*, 11(2), 112-124.

Duschl, R., Maeng, S., & Sezen, A. (2011). Learning progressions and teaching sequences: a review and analysis. *Studies in Science Education*, 47(2), 123–182.

Evans, S. & Morrison, B. (2011), Meeting the challenges of English-medium higher education: The first-year experience in Hong Kong. *English for Specific Purposes*, 30(3), 198–208.

Gee, J. P. (2004). Language in the science classroom: academic social languages as the heart of school-based literacy. In E. W. Saul (ed.), *Crossing borders in literacy and science instruction*. Arlington: National Science Teachers Association Press, 13–32.

Klaassen, R. G. & de Graaff, E. (2001). Facing innovation: preparing lectures for English-medium instruction in a non-native environment. *European Journal of Engineering Education*, 26(3), 281–289.

Kress, G., Jewitt, C., Ogborn, J., & Tsatsarelis, C. (2001). *Multimodal teaching and learning: The rhetorics of the science classroom*. London: Continuum.

Kunioshi, N., Noguchi, J., Tojo, K., & Hayashi, H. (2012). OnCAL, the Online Corpus of Academic Lectures, an English language support system for university science and engineering instructors, in *International Session & AEESEAP Workshop 2012 Proceedings, Annual Conference of Japanese Society for Engineering Education*, 70–75.

MEXT, Ministry of Education, Culture, Sports, Science and Technology of Japan (2009). *Launching the project for establishing core universities for internationalization (Global 30)*. Retrieved in May 22, 2011 from http://www.mext.go.jp/ english/highered/1302274.htm.

Neumann, K., Kauertz, A., & Fischer, H. E. (2012). Quality of instruction in science education. In B. J. Fraser, K. G. Tobin, & C. J. McRobbie (eds.), *Second Handbook of Science Education*. Heidelberg: Springer, 247–258.

Salmi, J. (2009). *The Challenge of Establishing World-Class Universities*. Washington D. C.: World Bank.

Scott, P., Mortimer, E., & Ametller, J. (2011). Pedagogical link-making: a fundamental aspect of teaching and learning scientific conceptual knowledge. *Studies in Science Education*, 47(1), 3–36.

Smit, U. & Dafouz, E. (2012). Integrating content and language in higher education. An introduction to English-medium policies, conceptual issues and research practices across Europe. *AILA Review* 25, 1–12.

Tan, M. & Lan, O. S. (2011). Teaching mathematics and science in English in Malaysian classrooms: The impact of teacher beliefs on classroom practices and student learning. *Journal of English for Academic Purposes*, 10(1), 5–18.

Thøgersen, J. & Airey, J. (2011), Lecturing undergraduate science in Danish and in English: A comparison of speaking rate and rhetorical style. *English for Specific Purposes*, 30(3), 209–221.

Wächter, B. & Maiworm, F. (2008). *English-taught programmes in European higher education*. Bonn: Lemmens, 21–32.