

О.М. Демидова
Национальный исследовательский
Томский политехнический университет

MICASE – description and use in the language learning process

Abstract:

The article describes the on-line concordance MICASE – the Michigan Corpus of Academic Spoken English. Its features and some ideas about its use in studying English as a Foreign Language are represented.

Key words: MICASE; concordance; teaching English; internet – tools.

Introduction

MICASE – the Michigan Corpus of Academic Spoken English – is a collection of more than 1.7 billion words from across the University of Michigan. It comprises almost 200 hours of recorded authentic speech events of speakers from different status groups across the campus – faculty, staff members, undergraduates, graduates, speakers of different nations, proficiency levels – talking in different academic situations – lectures, small lectures, colloquies, presentation defense, and seminars and so on. Different disciplines are covered: Humanities, Social Sciences, Physical sciences. The starting interface is very simple, it shows two hyperlinks: 1) for browsing and 2) for researching.

See **figure 1**.



The University of Michigan
English Language Institute

MICASE
The Michigan Corpus of Academic Spoken English
Michigan Corpus of Academic Spoken English

Welcome to our NEW interface to the on-line, searchable part of our collection of transcripts of academic speech events recorded at the University of Michigan.
There are currently 152 transcripts (totaling 1,848,364 words) available at this site.

[Browse MICASE](#)
Browse the corpus according to specified speaker and speech attributes, returning quick file references.

[Search MICASE](#)
Search the corpus for words or phrases in specified contexts, returning concordance results with references to files, full utterances, and speakers.

For additional information, see one of the following:
[Explanation of Transcription Conventions](#) | [Speech Event and Speaker Categories](#)
[Search Tips](#) | [General Information about MICASE](#)

We want to hear from you!
We would like to know who is using this on-line corpus,
how you found or heard about this site, and what you think of it.
Please send questions, comments, or suggestions to us at micase@umich.edu.

Figure 1

Description of the tool

Browse option provides access to all the files and gives you the option of selecting specific groups of files. If you are interested in a particular event-type, such as a small lecture, discussion groups, large lectures you can retrieve only those files. Similarly, you can choose particular disciplines: Biology or Physical Education – their language may be drastically different at times. Also, you may select files containing the speech of the faculty as opposed to the speech of students, or natives as opposed to non-native speakers. See **figure 2**.

The screenshot shows the MICASE (Michigan Corpus of Academic Spoken English) interface. At the top, the logo 'MICASE' is on the left and 'Michigan Corpus of Academic Spoken English' is on the right. Below this is a navigation bar with 'Home', 'Search', 'Browse', and 'Help'. The main heading is 'Browse Transcripts'. On the left, there is a box with the instruction: 'First, choose the criteria using the menus on the right. Then click the button to see transcripts that fit the criteria.' Below this is a 'Submit Browse' button. To the right, there are two columns of filter menus:

- Speaker Attributes:**
 - Academic Position/Role: All, Junior Faculty, Junior Graduate Student
 - Native speaker status: All, Non-native speaker, Near-native speaker
 - First language: All, Arabic, Armenian
- Transcript Attributes:**
 - Speech Event Type: All, Advising Session, Colloquium
 - Academic Division: Not Applicable/Other, Physical Sciences and Engine, Social Sciences and Educator
 - Academic Discipline: Astronomy, Biomedical Engineering, Biology
 - Participant Level: (partially visible)

Figure 2

The right part of the picture shows all the criteria according to which every single file is classified. For non-native speakers you should also specify their native language background: German, Arabic or others. For the transcript attributes you can select the division, e.g. Biological Health Sciences that will give you more fine-grained distinction of the files you need. Here, you can select the interactivity level. If you are interested in large chunks of uninterrupted speech – pick up a file that is of low-interactivity. If you are interested in turn – taking, and how people hold the floor, discuss things – choose the high-interactive files. If you don't specify your choice – you will get a list of all the transcripts. It is shown in **figure 3** below.

MICASE		Michigan Corpus of Academic Spoken English	
Home	Search	Browse	Help
45 transcripts			
Transcript ID (click to view)	File Name	Recording Length	Transcript Word Count
COL285I0038	Education Colloquium	52 min.	9204
COL999I0040	Women in Science Conference Panel	105 min.	20099
COL575I0055	Golden Apple Award Statistics Lecture	45 min.	7253
COL999I0059	Problem Solving Colloquium	66 min.	9870
COL425I0075	Ecological Agriculture Colloquium	97 min.	17653
COL140I0114	Peking Opera Colloquium	74 min.	12152
DIS175I027	Biology of Birds Discussion Section	55 min.	8461
DIS175I081	Intro Biology Discussion Section	59 min.	7791
INT425I001	Graduate Student Research Interview 1	34 min.	5168
INT425I002	Graduate Student Research Interview 2	20 min.	2963

The hyperlinks on the left will transport you to a necessary file. The title informs you what it is about, the next column shows the duration of the file and number of words in it. By clicking the **left mouse click** you get the transcript on the screen. By clicking the **right mouse click** and choosing the option “save target as...” you can save the file in HTML-format on your computer. The picture of the transcribed file is shown in **figure 4**.

MICASE		Michigan Corpus of Academic Spoken English	
Home	Search	Browse	Help
<p>Title: Education Colloquium Transcript ID: COL285I0038 Academic Division: Social Sciences and Education Publisher: Michigan Corpus of Academic Spoken English, English Language Institute, University of Michigan Interactivity Rating: Mixed Number of Participants: Audience Members: 20 Speakers: 9 Recording Date: October 9, 1998 Recording Duration: 52 minutes Word Count: 7898 Download entire transcript in XML</p>			
<p>\$1: um, here at the U-of-M, the Center for the Study of Higher and Secondary Education.</p> <p>\$U-F: congratulations.</p> <p>\$S: yes <APPLAUSE></p> <p>\$1: i decided i'd rather have this recorded than the defense just to gimme a little bit of <SS LAUGH> because i talk to a bunch of linguists i'd s- be paranoid all over well, it was an interesting thing yesterday to start my defense with, asking them how many had read, the Chronicle of Higher Education article this week that talked about the fact that the United States is losing its market share of international students to other countries. [SU-F (obviously)] yeah, and uh it was a nice segue into that because it's it's a it's a timely topic to be looking at what we're doing with the international students, who we do currently enroll. my audience yesterday, was far less, knowledgeable about international students specifically so, i'm gonna use the same overheads, some of this is gonna be just a little bit i think it's gonna be, information that you already know but better to do that than to assume. so, title is Navigating within a Global Learning Community, Academic Departments' Response to Graduate International Students, as some of you may know, we know a lot more about international students and their experiences on college campuses, than we do about how the departments and institutions more generally respond to them. so that was where i took up my research, it needed to be an exploratory setting because there</p>			

Figure 3

Every file has a title, a brief description of the core characteristics and the transcription itself. From here you can save the file as XML. This type of file is useful for linguistic analysis, as it contains a lot of extra information.

If you use the search function at the very beginning you can type in any word or phrase that can be important for academic discourse, and look how it is used in a quick concordance shown in **figure 5**.



Figure 4

You have a search word or phrase in the middle and left and right context. In quick concordance the amount of context is limited as it displays as many examples as possible on the screen. If it is necessary you can look at the whole utterance for larger context.

Using MICAE in learning the English language

This concordance can be used in: 1) (Self) – teaching; 2) Integration of teaching materials; 3) Creating testing.

1) (Self) – teaching

- It may be helpful if you want to know how to use academic words in a particular context or academic discourse in general.

- The concordance may be used to find out about the frequency of the word. It is possible to compare the frequencies of the words or phrases, of synonyms, in order to see which one is favored in a particular speech event or academic speech in general.

- It allows looking at the word or phrase in the context and specifying the meaning of the unit. Here, you can use the frequency option and check which meaning is the most frequent one, or the most relevant one for learners to learn first, so they could handle academic disciplines.

- The students may study the context in order to understand the meaning of the whole lexical unit better. For example, the word “hand” – the meaning here is not in the word but in the unit. In academic writings the unit “hand forward” is frequently used. To understand this, you should study the context. It shows that the meaning of the word “hand” shifts. It is not the meaning of an individual word any more but of a unit.

- The context search may be useful in studying the prepositions that follow the key word.

- The students may focus on the context of a word or phrase in a particular speech event or discipline which may/may not be different from general English.

2) Integration of teaching materials

- The concordance helps to teach not only the isolated meanings of words, but their pragmatic function or a communicative function of the word/lexical unit.

- Teaching criticism strategies. For example, you may want to teach items that express negative evaluation. You may make a list of units that express negative evaluation and then search for them, and study the context in which they are used.

- “Snowball” – you may pick some of the items you are interested in and do a concordance for them to find more relevant examples. You may use “Statistics” to find the files most rich in the object you are interested in.

3) Creating testing materials

- With the click of the mouse you can retrieve lots of authentic speech samples, situations in different topics, real life examples of items you want to test.

- Cloze tests are useful for example for phrasal verbs. You could study the concordance at first to know what phrasal verbs occur in an academic discourse. For example, the verb “back” has only 4 prepositions in academic speech – “back, off, out, up”. After studying this information you may create relevant cloze tests.

Conclusion

The article describes the MICAЕ as an on-line tool of corpus linguistics. It gives instructions on how to search and browse with this instrument. The potential of the corpus linguistics is in the process of revealing itself in the methodology of teaching English as a foreign language that is why the examples and hint of teaching exercises may be of high practical importance.

References

1. Базарова Б.Б. Корпусная лингвистика и преподавание иностранных языков / Б.Б. Базарова // Вестник Бурятского государственного университета – 2015. – № 15. – С. 88–92.
2. Захаров В.П. Корпусная лингвистика: учебно-методическое пособие / В.П. Захаров. – СПб., 2005. – 48 с.
3. Кокорева А.А. Корпус параллельных текстов в обучении иностранному языку / А.А. Кокорева // Вестник Тамбовского госуниверситета.– 2013. – С. 57–62.

4. Michigan Corpus of Academic Spoken English Режим доступа: <https://quod.lib.umich.edu/m/micase/>, свободный.
5. Simpson R. Corpus Linguistics in North America / R. Simpson, J.M. Swales. – Ann Arbor: The University of Michigan Press, 2001. – 241 p.