Changing English in a changing world

Unit 3
Corpora/English as a threat

Éva Illés and Tamás Eitler
Department of English Applied Linguistics

Supported by the Higher Education Restructuring Fund allocated to ELTE by the Hungarian Government
Task

In pairs or groups:

1. Read the characteristics of corpus-based analysis on Slide 4: explain what these characteristics entail.

2. Read the definition of Hymes’ communicative competence on Slide 5: which of the components can be related to corpus analysis?
Corpus

- Corpus: computerized;
  “empirical approach to linguistic analysis [...] based on naturally occurring spoken or written data as opposed to an approach that gives priority to introspection”

(Reppen and Simpson, 2002, p. 92)
Corpus

Characteristics of corpus-based analysis:
• Empirical
• Uses “a large and principled collection of natural texts”
• “Extensive use of computers for analysis”
• “Both qualitative and quantitative analytical techniques”

(Biber, Conrad and Reppen as cited in Reppen and Simpson, 2002, p. 93)
Corpus

• Hymes: Communicative competence

1. Whether (and to what degree) something is formally possible;
2. Whether (and to what degree) something is feasible in virtue of the means of implementation available;
3. Whether (and to what degree) something is appropriate (adequate, happy, successful) in relation to a context in which it is used and evaluated;
4. Whether (and to what degree) something is in fact done, actually performed, and what its doing entails.”

(Hymes, 1972, p. 281)
Task

Task based on home assignment

1. In groups of 3-4, collect the corpora you have found for each circle.

2. Compare your findings with the information on Slides 7-20.
Corpora

**Inner Circle:** British National Corpus (BNC), American National Corpus (ANC)

**Inner/Outer Circle:** International Corpus of English (ICE)

**ELF:** Vienna-Oxford International Corpus of English (VOICE); English as a Lingua Franca in Academic Setting (ELFA); Tübingen English as a Lingua Franca (TELF)
Corpora

• Specialised corpora

International Corpus of Learner English (ICLE)
Michigan Corpus of Academic Spoken English (MICASE)
Corpora of Inner Circle English

What is the BNC?

The British National Corpus (BNC) is a 100 million word collection of samples of written and spoken language from a wide range of sources, designed to represent a wide cross-section of British English from the later part of the 20th century, both spoken and written. The latest edition is the BNC XML Edition, released in 2007.

The written part of the BNC (90%) includes, for example, extracts from regional and national newspapers, specialist periodicals and journals for all ages and interests, academic books and popular fiction, published and unpublished letters and memoranda, school and university essays, among many other kinds of text. The spoken part (10%) consists of orthographic transcriptions of unscripted informal conversations (recorded by volunteers selected from different age, region and social classes in a demographically balanced way) and spoken language collected in different contexts, ranging from formal business or government meetings to radio shows and phone-ins.

The corpus is encoded according to the Guidelines of the Text Encoding Initiative (TEI) to represent both the output from CLAWS (automatic part-of-speech tagger) and a variety of other structural properties of texts (e.g. headings, paragraphs, lists etc.). Full classification, contextual and bibliographic information is also included with each text in the form of a TEI-conformant header.

Work on building the corpus began in 1991, and was completed in 1994. No new texts have been added after the completion of the project but the corpus was slightly revised prior to the release of the second edition BNC World (2001) and the third edition BNC XML Edition (2007). Since the completion of the project, two sub-corpora with material from the BNC have been released separately: the BNC Sampler (a general collection of one million written words, one million spoken) and the BNC Baby (four one-million word samples from four different genres).

http://www.natcorp.ox.ac.uk/corpus/index.xml
Corpora of Inner Circle English

https://www.anc.org/about/
Corpus of Outer Circle Englishes

http://ice-corpora.net/ICE/INDEX.HTM
Corpus of Outer Circle Englishes

Available ICE Corpora @ ICE-corpora.net

The following corpora are available free (under Licence) to download from this site:

- CANADA (ICE-CAN - 1m words, lexical)
- JAMAICA (ICE-JA - 1m words, lexical)
- HONG KONG (ICE-HK - 1m words, lexical)
- EAST AFRICA (ICE-EA - Kenya & Tanzania, v 2, 1m words, lexical, including a version for Wordsmith)
- INDIA (ICE-IND - 1m-words, lexical)
- SINGAPORE (ICE-SIN - 1m words, lexical)
- PHILIPPINES (ICE-Phl - 1m words, lexical)
- USA (ICE-USA, written component - c. 400,000 words, lexical)
- IRELAND (ICE-IRL - 1m words, lexical)
- SPICE-IRELAND (SPICE-IRL - c. 600,000 words with prosodic and pragmatic annotation)

The following corpora are also available, from the addresses shown:

- GREAT BRITAIN (ICE-GB - 1m words, POS-tagged and parsed, distributed with ICECUP retrieval software). Available from: Survey of English Usage, University College London, Gower St, London WC1E 6BT, UK. Order Form
- NEW ZEALAND (ICE-NZ - 1m words, lexical)
  Available from: School of Linguistics & Applied Language Studies, Victoria University of Wellington PO Box 600, Wellington, New Zealand. Order form
- SRI LANKA (ICE-SL - written component; lexical and POS-tagged with CLAWS C7 tagset)
  Available from the Department of English, the University of Giessen, Germany.
  To obtain a copy of the corpus, please email ice-sl@anglistik.uni-giessen.de

http://ice-corpora.net/ICE/avail.htm
Corpora of English as a lingua franca

Welcome to VOICE,
the Vienna-Oxford International Corpus of English!

In the early 21st century, English in the world finds itself in an “unstable equilibrium”: On the one hand, the majority of the world's English users are not native speakers of the language, but use it as an additional language, as a convenient means for communicative interactions that cannot be conducted in their mother tongues. On the other hand, linguistic descriptions have as yet predominantly been focusing on English as it is spoken and written by its native speakers.

VOICE seeks to redress the balance by providing a sizeable, computer-readable corpus of English as it is spoken by this non-native speaking majority of users in different contexts. These speakers use English successfully on a daily basis all over the world, in their personal, professional or academic lives. We therefore see them primarily not as language learners but as language users in their own right. It is therefore clearly worth finding out just how they use the language. This is exactly what VOICE seeks to make possible.

The VOICE project as such ran from 2005 to 2013, see News

https://www.univie.ac.at/voice/
Corpora of ELF

What is VOICE?

A computer corpus of English as a lingua franca

The most widespread contemporary use of English throughout the world is that of English as a lingua franca (ELF), i.e., English used as a common means of communication among speakers from different first-language backgrounds. A Hungarian educationalist coming to Copenhagen to discuss qualification equivalents in European higher education with her Danish, Finnish, and Portuguese colleagues; a Korean sales representative negotiating a contract with his German client in Luxembourg; a Spanish Erasmus student chatting with local colleagues in a student hall in Vienna; they all communicate in English as a lingua franca.

VOICE, the Vienna-Oxford International Corpus of English, is a structured collection of language data, the first computer-readable corpus capturing spoken ELF interactions of this kind.

VOICE, compiled at the Department of English at the University of Vienna, is funded by the Austrian Science Fund (FWF). These funds were further supplemented by a contribution from Oxford University Press in 2008. Supporting funds were also provided in the early pilot phase by Oxford University Press and by the Hochschuljubiläumsstiftung der Stadt Wien. The corpus currently comprises 1 million words of transcribed spoken ELF from professional, educational, and leisure domains.

It is the ultimate aim of the VOICE project to open the way for a large-scale and in-depth linguistic description of this most common contemporary use of English by providing a corpus of spoken ELF interactions which will be accessible to linguistic researchers all over the world.

https://www.univie.ac.at/voice/page/what_is_voice
Corpora of ELF

https://www.univie.ac.at/voice/page/corpus_description
Corpora of ELF

http://www.helsinki.fi/englanti/elfa/elfacorpus.html
Corpora of ELF

Description of the project "English as a Lingua Franca in Academic Settings"

The English language has established itself as a global lingua franca, a contact language spoken by people who do not share a native language. Most of its use today is by non-native speakers, who have far outnumbered its native speakers. English constitutes the main means of international communication in a variety of key domains in the world.

In view of this, there is surprisingly little empirical research into English as used internationally. The project "English as a Lingua Franca in Academic Settings" (ELFA) at the University of Helsinki offers a contribution towards an empirical basis for understanding this variety of English. The project falls into two main parts, the ELFA corpus project and the SELF project. The ELFA team has also compiled a database of written academic ELF (WrELFA).

Investigating English as lingua franca (ELF) serves three kinds of research interest; theoretical, descriptive, and applicational.

The theoretical interest arises from the nature of ELF as a contact language: the language contact situation is more complex with English than other languages, since virtually any other language in the world can be in contact with English. The theoretical interests around ELF centre on manifestations of features like simplification, evidence of universally unmarked features, hypothesised universals of communication, as well as evidence of self-regulative processes.

Descriptively, ELF research seeks to establish its characteristic features which deviate from Standard English, and look for possible "core" features of ELF. The description helps understand the ways in which English is currently changing and how its variability takes shape. Moreover, it contributes to an understanding of what second language use is like in authentic contexts, as opposed to learner performance in educational settings.

http://www.helsinki.fi/englanti/elfa/project.html
Corpora of ELF

Tübingen English as a Lingua Franca

TELF is an applied linguistics research project that systematically combines conversational ELF output data with introspective and retrospective accounts by speakers and interlocutors. This combination allows for differentiated insights into the nature of ELF communication; it provides rich examples of how speakers use their "own English" to cope with the communicative challenges of a lingua franca situation.

The TELF corpus is a collection of video-recorded and transcribed discussions by mixed groups of four to six native and non-native speakers of English from diverse linguistic and cultural backgrounds. Each discussion is followed up with introspective interviews addressing the participants’ learning history, English requirement profile, and performance in the discussion. Drawing on TELF-data, our research focuses on divergences between meaning and comprehension, co-construction and monitoring, and "pushing" one's limits of expression.

The TELF database is accessible via an online application supporting the detailed analysis of transcripts (including retrospective commentaries by participants in the form of interlinear glosses), questionnaire data, introspective interviews and retrospective accounts. A restricted demo access to the database application is available via the menu on the right.

http://projects.ael.uni-tuebingen.de/telf/
Specialised corpora

Specialised corpora

Michigan Corpus of Academic Spoken English

- Introduction
- Basic structure
- Annotation
- Background
- Bibliography

Michigan Corpus of Academic Spoken English (MICASE)

The Michigan Corpus of Academic Spoken English (MICASE) is a collection of nearly 1.8 million words of transcribed speech (almost 200 hours of recordings) from the University of Michigan (U-M) in Ann Arbor, created by researchers and students at the U-M English Language Institute (ELI). MICASE contains data from a wide range of speech events (including lectures, classroom discussions, lab sections, seminars, and advising sessions) and locations across the university.

Project leader: Dr. Ute Römer, University of Michigan
Time of compilation: 1997–2002
Size: 1.8 million words
Language: English
Number of texts/samples: 152 (over 190 hours)
Released: 2002; new interface 2007

http://www.helsinki.fi varieng/CoRD/corpora/MICASE/
Task

1. Identify a grammatical/lexical problem and consult available online corpora to solve the problem.

2. Class discussion:
How can corpora inform language teaching?
Limitations of corpus analysis (see also Widdowson, 2004, Chapter 7)
Task

• Is English a threat to other languages and multilingualism?

1. In pairs or groups of 3-4, collect your arguments for and against the above statement.

2. Read Slides 23-43. Identify arguments which are missing from your collection but which you find valid.
Is English a threat?

• The dominance of English in Europe

“So English impinges on the lives of all European citizens, in many different ways: academics, business executives and hip hoppers experience the language as a pervasive presence in their daily lives.”

(Seidlhofer, 2011, p. 135)
Is English a threat?

• The dominance of English in Europe

“English is a mandatory language in 14 countries or regions within countries (...). It is by far the most taught foreign language in nearly all countries at all educational levels. Trends since 2004/05 show an increase in the percentage of pupils learning English at all educational levels, and particularly at primary level (...).”

(Key data on teaching languages at school in Europe, 2012, p. 11)
Is English a threat?

• The dominance of English

“the competitive advantage which English has historically provided its acquirers (personally, organisationally, and nationally) will ebb away as English becomes a near-universal basic skill”

(Graddol, 2006, p. 15)
Is English a threat?

- The dominance of English in Europe
  “In the Netherlands there are many levels of society in which English is being used for communication.” (p. 115)

  “… people in the Netherlands are worried about the position of Dutch and the threat from English. They fear that English will take over the position of Dutch and that would be the end of our national identity.” (Nortier, 2011, p. 117)
Is English a threat?

• The dominance of English within the EU

“Without any doubt, English will be the working language” (van Els 2005: 277, italics in the original)

“In EU-meetings English is now being used more than ever before.” (Nortier, 2011, p. 127)
Is English a threat?

- English **is** a threat

“The more languages, the more English”

The global language system: 4 levels

1. Peripheral languages (98%)
2. Central languages (e.g., Yoruba in Nigeria)
3. Supercentral languages (e.g., Arabic, Chinese, Spanish, Russian)
4. Hypercentral language: English

(Based on de Swaan, 2001, in Nortier, 2011)
Is English a threat?

• “the language usually learned is higher up in the hierarchy”

• “In choosing the hypercentral language English, the level of supercentral languages French and German is skipped”

(Nortier, 2011, p. 123)
English is a threat?

Dangers of a global language:

• hastens the disappearance of minority languages; language death

• Linguistic power

• Linguistic complacency (monolingual bias)

(Crystal, 1997)
Is English a threat?

- English in the EU

“The preference for English is, naturally enough, seen as a threat to the maintenance of linguistic diversity. All languages are supposed to be equal but English is obviously ‘more equal than the others’. The protection of other languages is therefore assumed to involve resistance to English.”

(Seidlhofer, 2011, p. 139)
Is English a threat?

• One language, one state, one people

National pride (the French v. the Dutch)

“Exactly this attitude is at the heart of the fears of possible language – and identity – loss!”

(Nortier, 2011, p. 126)
Is English a threat?

- English – larger language communities (French, German, Spanish, Italian) – smaller language communities within the EU

Larger communities: upgrading English

\[ \downarrow \]

downgrading their own language, loss of prestige and international standing of the language

(based on Ammon, 2006, p. 332)
Is English a threat?

• Language laws

Toubon Law, France (1994): regulates the use of foreign languages in official documents, including commercial contexts. (Ruellot, 2011, p. 6)

• Civil movements

Netherlands: Stichting Taalverdediging (Foundation for Language Defence) (Nortier, 2011, p. 114)
Is English a threat?

• English is NOT a threat

“The spread of English is not the direct cause of language endangerment.” (Graddol, 2006, p. 60)
Is English a threat?

• Graddol (2006)
  “English is no longer the ‘only show in town’. Other languages now challenge the dominance of English in some regions. Mandarin and Spanish, especially, have become sufficiently important to be influencing national policy priorities in some countries.” (p. 62)
IS English a threat?

• House (2003)
  ELF = language for communication
  L1 = language for identification
Is English a threat?

House (2003)
Study: 60 English and German translation pairs

“The results of our project on the influence of English on discourse norms of other languages described above show that the massive borrowing from English lexis is not matched with (more insidious) shifts in discourse conventions.”

(p. 574)
Is English a threat?

House (2003)

“And finally, the project on the introduction of English as a medium of instruction in German universities has shown that in tertiary education – [...] – there are no signs (yet) of a threat to a native language (German) and to multilingualism.”
(p. 574)
Is English a threat?

• Seidlhofer (2011)

Who has to learn whose language?

issues of power: “... if a language is dominant, the nation that ‘owns’ it is bound to be dominant, too” (p. 138)
Is English a threat?

• Seidlhofer (2011)

“But English as a lingua franca is not like other European languages for it has been removed from native speaker ownership, appropriated and adapted for international use on a massive and unprecedented scale.” (pp. 139-140)

English in Europe:

ELF “rather than the native language of the British or the Irish” (p. 141)
Is English a threat?

• Seidlhofer (2011)

“What needs to be stressed is that this *natural* English is not the *national* English of native speakers.” (p. 141)

“It is only when English is conceived of as belonging to its native speakers and as a foreign language like any other that it constitutes a threat.” (p. 143)
Is English a threat?

- Nortier (2011)

Dutch not being overruled by English

“The influence of English is superficial (mainly lexical) + “Dutch is still the main language spoken in Dutch homes and in the broader society” (p. 131)
Task

Class discussion:
In two groups (one for, one against) have a debate about the threat English may pose. Use your own arguments as well as the arguments from the literature.
Task

Task for Unit 4

1. Research the Common European Framework of Reference (CEFR) and find out which variety/varieties of English it promotes for English language teaching/learning.

2. Carry out similar research with ELT coursebooks.

3. Read the article on linguistic landscape by Sayer (2009).


References


