

Using Corpora in English Language Teaching – An Exploratory Learning Approach

Presenter: Dinh Thi Lam Tra

Email: dinhthilamtra@ukh.edu.vn

University of Khanh Hoa

Goals for the workshop:

- To provide some reasons for using corpora
- To show how to use corpora
- To provide some guidelines and suggestions

Dictionaries:

- <https://en.oxforddictionaries.com/definition/dictionary>
- <https://dictionary.cambridge.org/dictionary/english/kingdom?a=british>
- <https://www.macmillandictionary.com/dictionary/british/corpus>
- However, there was a tradition of using examples invented by lexicographers, rather than authentic materials.

Authenticity and Frequency information

- Core vocabulary is used to help learners of different modern languages.
- Frequency information is immensely useful in helping to prioritize what to teach - learn.
- COBUILD (HarperCollins, 1995) and Longman (1995) include such frequency information.

What is a corpus ?

- <https://www.macmillandictionary.com/dictionary/british/corpus>
- A large, principled collection of natural texts.

Corpus markup and annotation

- are encoded with:
 - textual (e.g. register, genre and domain)
 - sociolinguistic (e.g. user gender and age) metadata, which allows lexicographers to give a more accurate description of the usage of a lexical item

User-friendly online corpora

- British National Corpus (Brigham Young University; BYU-BNC), <https://www.english-corpora.org/bnc/>
<https://www.english-corpora.org/time/>
- Corpus of Contemporary American English (COCA),
<https://www.english-corpora.org/coca/>
- Corpus of Historical American English (COHA),
<http://corpus.byu.edu/coha/.>
- Michigan Corpora
- [MICASE](#) – a corpus of spoken academic English
- [MICUSP](#) - a corpus of written academic English

- The British National Corpus (BNC) contains 100 million words of text texts from a wide range of genres (e.g. spoken, fiction, magazines, newspapers, and academic).
- The Corpus of Contemporary American English (COCA) contains more than 560 million words of text (20 million words each year 1990-2017) and it is equally divided among spoken, fiction, popular magazines, newspapers, and academic texts.

TASKS

- Check if the mobile phones have internet connection.
- If not, use the the following wi-fi .
- **Name: lamtra, Password: 28091610**

Task 1:

- What are the plural forms of the words: ***dwarf, hoof, scarf, and wharf?***

Task 1:

Use the following corpora to carry out the task below:

- Corpus of Contemporary American English (COCA),
<https://www.english-corpora.org/coca/>
- British National Corpus (Brigham Young University; BYU-BNC),
<https://www.english-corpora.org/bnc/>

Check the total frequency of these alternatives in the BYU-BNC and COCA (*dwarfs, dwarves, hoofs, hooves scarfs, scarves, wharfs and wharves*)

Words	Plural forms	COCA frequency	BYU-BNC frequency
<i>dwarf</i>	dwarfs		
	dwarves		
<i>hoof</i>	hoofs		
	hooves		
<i>scarf</i>	scarfs		
	scarves		
<i>wharf</i>	wharfs		
	wharves		

Words	Plural forms	COCA frequency	BYU-BNC frequency
<i>dwarf</i>	dwarfs	1312	279
	dwarves	348	55
<i>hoof</i>	hoofs	168	48
	hooves	1451	206
<i>scarf</i>	scarfs	48	5
	scarves	1416	176
<i>wharf</i>	wharfs	31	10
	wharves	142	51

Uncertain usage of some plural forms.

- In several cases, the usage of /-vz/ in the plural form is uncertain, e.g. - ***dwarf, hoof, scarf, and wharf*** – both with /-fs/ and /-vz/.
- Non-corpus-based grammars can contain biases while corpora can help to improve grammatical descriptions.

LONGMAN

GRAMMAR

of SPOKEN

and WRITTEN

ENGLISH

Douglas Biber
Stig Johansson
Geoffrey Leech
Susan Conrad
Edward Finegan

LONGMAN

GRAMMAR

of SPOKEN and WRITTEN ENGLISH

'For the foreseeable future, anyone with a serious interest in English grammar will have to take into account the information this book contains.'

David Crystal

The *Longman Grammar of Spoken and Written English* is an entirely new grammar of American and British English – from the language of conversation to the language of academic textbooks. There are no made-up examples in this groundbreaking new grammar. The authors began, not with preconceived notions of the grammar of English, but with a huge bank of language data, the Longman Corpus Network.

A six-year research project brought together the linguistic expertise of an international author team – all acknowledged experts in the field of corpus linguistics and grammar.

The result of this research is the present volume. Many points of traditional grammar are confirmed, but now on the basis of much larger amounts of statistical data than ever before. Some aspects of traditional grammar are challenged by this book, and some new findings, not even suspected before now, will surprise and interest the reader.

What makes this book so special is that it turns English inside out. Professor Douglas Biber's research team logged and parsed the structures in the Corpus. This analysis revealed the degree to which different grammatical features of language vary according to the type of language. The way language is used in conversation is quite different from the way language is used in fiction, which in turn is often very different from the grammatical characteristics of newspapers or academic books.

But why is it different? The *Longman Grammar of Spoken and Written English* takes its reader into new and uncharted territory for a grammar by suggesting the reasons why we choose a particular structure in a particular context.

- Entirely corpus-based grammar of English
- Over 350 tables and graphs showing the frequency of constructions across different registers, from conversation to fiction to academic prose
- 6,000 authentic examples from the Longman Corpus Network
- British English and American English grammar compared
- New and challenging findings
- Reveals the differences between spoken and written English

C Consonant change

calf—calves

scarf—scarves

half—halves

sheaf—sheaves

knife—knives

shelf—shelves

leaf—leaves

thief—thieves

life—lives

wife—wives

loaf—loaves

wolf—wolves

Some words ending in *-f* take *-ves* in the plural, as shown above. The majority, including all nouns ending in *-ff*, take regular plural endings:

beliefs, chefs, chiefs, cliffs, proofs, puffs, reefs, roofs, serfs

There is variation with a few forms:

dwarf—dwarfs/dwarves, hoof—hoofs/hooves,

wharf—wharfs/wharves

(Biber, et al., 2010, p. 285).

Task 2:

- What are the patterns of RESULT with the prepositions “in” and “of” ?
- **Result (in): (v)**
- **Result (of): (n)**

Task 3:

- What does the word hand mean?
- Read task 3 and find out
 - patterns of uses of hand as a part of body.
 - if hand has other meanings than a part of body.
 - collocations with the word hand
 - if hand has figurative meanings

Task 4: Fill in all blanks with only ONE suitable word.

- <https://quod.lib.umich.edu/cgi/c/corpus/corpus?c=micas&cc=micas&e&type=simple&q1=hand>

-

HAND

Conclusion:
Learners can
interact with
corpora to:



Learn vocabulary



Explore extended
collocations



Discover patterns of use



(Rappen, 2012)



[https://www.youtube.com/watch?
v=Qf46lOnMCfs](https://www.youtube.com/watch?v=Qf46lOnMCfs)

Recommendations

- Corpus linguistics should be included in initial language teacher education to enhance teachers' research skills and language awareness (O'Keeffe & Farr, 2003).
- The traditional “**three Ps**” (Presentation, Practice and Production) approach to teaching may not be entirely suitable.
→ the more exploratory approach of “**three Is**” (Illustration, Interaction and Induction) (Carter and McCarthy, 1995) may be more appropriate:
 - “illustration” means looking at real data
 - “interaction” means discussing and sharing opinions and observations
 - “induction” means making one’s own rule for a particular feature

- The language learner should be encouraged to become “a research worker whose learning needs to be driven by access to linguistic data” (Johns, 1991, p. 2)
- This autonomous learning process “gives the student the realistic expectation of breaking new ground as a ‘researcher’, doing something which is a unique and individual contribution” (Leech, 1997, p. 10).



References:

- Aston, G. (2001). Learning with corpora: An overview. In G. Aston (ed.), *Learning with Corpora* (pp. 7–45). Houston: Athelstan.
- Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. (2010). *Longman student grammar of spoken and written English*. Essex: Pearson Education.
- Carter, R., & McCarthy, M. (1995). Grammar and the spoken language. *Applied Linguistics*, 16(2), 141–158.
- Leech, G. (1997). Teaching and language corpora: A convergence. In A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (Eds.), *Teaching and language corpora* (pp. 1–23). London: Longman.
- McEnery, T., & Xiao, R. (2011). What corpora can offer in language teaching and learning. In E. Hinkel (Eds), *Handbook of research in second language teaching and learning*, (pp. 364-380). New York: Routledge
- O'Keeffe, A., & Farr, F. (2003). Using language corpora in initial teacher education: pedagogic issues and practical applications. *TESOL Quarterly*, 37(3), 389–418.
- Reppen, R. (2012). Using Corpora in the Language Classroom. Retrieved from: <https://www.youtube.com/watch?v=Qf46lOnMCfs>