An introduction to using corpora with EFL learners

Mark Donnellan

Kwansei Gakuin University

Nishinomiya

Japan

Overview

1. Introduction
2. What is Corpus
3. Overview of the course
4. Student feedback
5. Conclusion
6. Q&A and Demo (time permitting)

Introduction

- This presentation reports on a corpus linguistics course with EFL students at a Japanese university
- “How to” questions - please hold them until the end
- Feel free to ask other questions

What is a corpus?

A corpus is a collection of naturally occurring language text, chosen to characterize a state or variety of a language. In modern computational linguistics, a corpus typically contains many millions of words: this is because it is recognized that the creativity of natural language leads to such immense variety of expression that it is difficult to isolate the recurring patterns that alter clues to the lexical structure of the language.

(Sinclair, 1991, p. 171)

Overview of Corpus - Merits

- More reliable than native speaker intuition
- Aides judgment about collocations
- Aides judgment about frequency
- Semantic prosody revealed
- Reveals details of phraseology

(Hunston, 2002, p.20-22)

Overview of Corpus - Limitations

- Technological issues (historical, more written)
- Only tells us if something is frequent
- Only shows its own contents
- Only shows evidence (needs to be interpreted)
- Out of context (eg. intonation and body language)

(Hunston, 2002, p.22-23)

The British National Corpus (BNC)

- British English
- Written Component - about 90% (87,903,571 words)
- Spoken Component – about 10% (10,409,858 words)
- Written and spoken texts up to 1993
- BNCweb (Hoffmann & Evert 2013) was used

Your query "[word="corpus" %c]" returned 773 hits in 201 different texts (98,313,429 words [4,048 texts]

Clear (1990) suggests that perhaps the size of a	corpus	is more significant than its composition although the two parameters are inter-dependent
lower courts. However, the Supreme Court subsequently annulled the habeas	corpus	on grounds of procedural irregularites . Dr. Zúñiga was warned that the
held in Belgium at Liege — the city where the Feast of	Corpus	Christi originated. However, because of political circumstances it was held
imagined as a wet blanket. In those days the Fellows of	Corpus	were rather proud of the briskness of their conversation. Instead Ramsey
one scholar of native American languages calls the manuscripts 'the largest	corpus	of texts' of them and 'a remarkable resource'.
of this approach is the work on the effects of cutting the	corpus	callosum in humans (Gazzaniga 1985). Some thirty years ago
pay for him to study at the latter's old college of	Corpus	Christi at Oxford. Here his most influential teacher was John Reynolds
the linguist studying children's language, children have access to a	corpus	or sample of language in the utterances they hear. This appears
were. Evidently, the general collocation dictionary derived from the LOB	corpus	can make a significant contribution to the recognition of domain-specific documents.
the wealth of other Gnostic, Thomasine or Nazarean documents in the	corpus	of Nag Hammadi scrolls. Nazarean thought left an ineradicable imprint on
Since not every word in the lexicon is present in the	corpus	and those present need not appear with their rarer forms, the
apparatus would not be valid for use in normal subjects as the	corpus	collosum would allow the transfer of information from one hemisphere to the
swollen, hard, and erect. At the end of the	corpus	spongiosum (through which runs the urethra) is the glans penis
of error. Shinghal and Toussaint estimated their transitional probabilities from a	corpus	of English text containing 531,445 words. Uni-gram probabilities were used,
4.5.5. The LOB Corpus The LOB corpus has been the principal	corpus	used throughout the current project. The LOB corpus is made up
is really used today, drawing on the evidence of the COBUILD	corpus	devised by Professor John Sinclair's English Language Research team in the

- 8 out of the 16 concordance lines contain the meaning of corpus that we are interested in.

Patterns

- Size of a corpus
- Have access to a corpus
- Be present in a corpus
- A corpus of English

Overview of the course

- Two classes: class A - 5 students, class B – 9 students
- 14 ninety minute classes
- Window and OS X used

Overview of the course

Five main components

1. Theory
2. Corpus tasks
3. Teaching materials project
4. Corpus building project
5. Individual research

Overview of the course Theory

- Mainly in the 1st and 2nd weeks, a small amount in subsequent weeks
- To familiarize the students with the key concepts of corpus linguistics

Overview of the course
Corpus tasks

- Mainly weeks 2-5 of the course
- A series of tasks designed to give the students hands on practice with the BNC

Overview of the course
Corpus Tasks

Students' first searches

1. Can we find an alternative meaning of the word **wicked**?
2. How flexible/inflexible are these phrases?

be looking forward to

rain cats and dogs

3. Do men or women use the word **fuck** more?

Overview of the course
Corpus Tasks

- Error correction passages with errors chosen from An A-Z of common English Errors for Japanese Learners (Barker, 2010)

Example

Last week I played with my old sister. It was her birthday, she was 23. I want to pass more time with my sister because we are four families and I think it is important to be close to her. It wasn't funny because my sister thinks I am boring. We went to a live, AKB48 were playing.

Overview of the course Teaching materials project

- Weeks 5-8 of the course
- Students design teaching materials based on corpus data
- Other members of the class act as students for a teaching demonstration.
- Samples shown from *English Unlimited* (Tilbury, 2010)

Keyword ON

Keyword *on*

1 Add the highlighted expressions to the table.

1 Plates and mugs and stuff are up here on the shelf. Unit 5
2 Please use the wardrobe on the left. Unit 5
3 Talking to my husband and watching something good on TV with him. Unit 3
4 We have two New Years, one on January 1st and Seollal, in January or February. Unit 3
5 And can he come on Friday? Unit 3
6 On Sofasurfing.com you can read people's profiles, email them and go and stay in their homes. Unit 2
7 You can stay on a sofa or spare bed for one or two nights. Unit 2
8 I'm often on planes or trains and in hotels and offices around the world. Unit 1

places	days, dates	transport	media and communication
on the second floor	on March 25th	on the bus	on the phone

2 a Add on to the questions.

1 What did you do *on* / Friday evening?
2 How often do you go the internet?
3 What's your favourite programme TV?
4 What do you listen to the radio?
5 When was your first trip a plane?
6 Do you always work Mondays?

b Write four more questions with on.

c Ask and answer all the questions.

Keyword IN

1 a Add these highlighted expressions with in to the table.

1 I'm from Canada but I live and work in Japan. Unit 2
2 I live in a small house with my brother, Erkan. Unit 2
3 My mother lives in the same street. Unit 2
4 We were neighbours in Melbourne. Unit 1
5 We were in the same office. Unit 1
6 When I was at university in 2007 ... Unit 1
7 Can you do these things in English? Intro unit

Places	Times	Languages
in Istanbul	in the morning	in Japanese

b Add these expressions to the table.

in winter in the afternoon in German in Germany in March in the evening

2 a Add in to these sentences.

1 My birthday is *in* ~~/~~ October.
2 I work a small shop.
3 I was Athens in the summer of 2009.
4 People often visit my country the winter.
5 I live a flat with my wife and children.
6 My friend and I were in Spain together 1989.
7 I can say 'I love you' Korean.
8 I work the morning.

b Change the sentences so they're true for you.

My birthday is in ~~October~~ *March*.

c Compare your sentences.

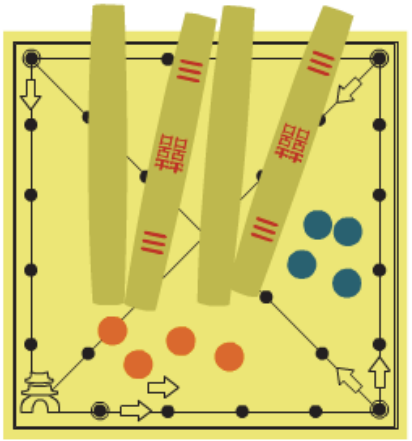
Keyword LIKE

1 a Look at sentences A–D from previous units. Then add *like* to sentences 1–8.

A I like Ed and I admire him in many ways. Unit 8
B I'd like to go to Cuba and Ireland. Unit 2 **(= want)**
C Our mum says we're like twins, just born ten years apart. Unit 8 **(= are similar to)**
D Soon fashionable men ... started wearing wigs, like the man in this picture. Unit 8 **(= for example)**

like

1 Would you / anything from home? Unit 2
2 I don't bad news. Unit 3
3 We all play yunnori. It's chess. Unit 3
4 There are more and more people José Luis all over the world. Unit 5
5 How many would you? Unit 6
6 But sometimes I do extra work, writing reviews. Unit 7
7 I going to bed late. Unit 7
8 Men in towns and cities in all regions of India usually wear western-style clothing, shirts and trousers. Unit 8



yunnori

b Are sentences 1–8 like sentences A, B, C or D?

2 a Complete the sentences with your own ideas.

1 I'm like my .... We're both ...
2 I like going ...
3 I usually wear ... , like ...
4 I'd like to buy ...

I usually wear smart clothes to work, like a jacket and tie.

b Listen to each other's sentences. Ask questions to find out more.

Keyword GO

1 We don't go out a lot but we're _____. Unit 3
2 Happiness is going fishing on a _____ with my friends. Unit 3
3 I guess I go to _____ at one or two. Unit 3
4 She usually goes out with friends from _____. Unit 3
5 We usually go for a _____. Unit 3
6 I'd like to go to Cuba and _____. Unit 2

Ireland bed meal
boat work happy

b Look at the highlighted expressions in the sentences. Then add these words to the table.

~~concerts~~ drink colleagues family parties shopping walk skiing Japan

go to	concerts
go for a	
go out with	
go + *-ing*	

c Cover the table. Test each other in pairs. How many expressions can you remember with:

1 go to ... 2 go for a ... 3 go out ... 4 go + *-ing* ?

2 a Match the questions and answers.

1 How often do you go for a walk?
2 Do you ever go to restaurants?
3 Would you like to go to Cuba?
4 Where do you go shopping?

a Yes. We always go for a drink on Friday evenings.
b Yes, I'd love to!
c Sometimes, but I like cooking at home.
d Not very often. I don't have time.

Overview of the course
Corpus building project

- Weeks 9-11 of the course
- Students collected texts related to their academic interests or hobbies.
- Examples included: Disney movies, American literature and Song lyrics
- Students formulated research questions
- Research questions were investigated using AntConc (Anthony, 2014)

Overview of the course
Individual research project

- Project introduced in week 9
- Research questions checked and refined weeks 10-11
- In-class work weeks 12-13
- Presentation and submission of written report week 14
- Topic chosen included the difference between *sweet* and *cute*, and translations of the Japanese verb *jistugen suru,* these included *achieve, realize, fulfill, materialize, come true and accomplish*.

Student feedback
Can-do self assessment

Students self assessments showed progress in relation to:

- Using the BNC
- Formulating research questions
- Using other tools to support corpus investigation

Student feedback Survey

- Student feedback collected about the 5 components of the course and the course as a whole
- Student indicated that they enjoyed the course overall
- Students indicated that the corpus tasks were the most beneficial part of the course
- The Corpus building project was the least well received

Student feedback
Survey comments

- Enjoyable course
- Will use corpus again
- The theory was difficult
- Reading concordance lines was difficult
- Learned why corpus was more useful than a dictionary

Conclusion

- Students actively participated in the course and felt they benefitted from it
- More corpus tasks involving reading concordance lines
- A lot of hands on practice for students
- Teachers hoping to undertake the course should become thoroughly familiar with the corpus tools to be used. Recommended text for BNCweb:

Hoffmann, S., Evert, S., Smith, N., Lee, D., & Belglund Prytz, Y. (2008). *Corpus Linguistics with BNCweb - a Practical Guide.* Frankfurt: Peter Lang.

References

Anthony, L. (2014). AntConc (Version 3.4.3) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/

Barker, D. (2010). *An A - Z of Common English Errors for Japanese Learners* (2nd ed.). Tokyo: BTB Press.

Hoffmann, S., & Evert, S. (2013). BNCweb (Version 4.3) [Computer Software]. Lancaster, UK: Lancaster University. Available from http://bncweb.lancs.ac.uk/

Hoffmann, S., Evert, S., Smith, N., Lee, D., & Belglund Prytz, Y. (2008). *Corpus Linguistics with BNCweb - a Practical Guide.* Frankfurt: Peter Lang.

Hunston, S. (2002). *Corpora in Applied Linguistics.* Cambridge: Cambridge University Press.

Sinclair, J. (1991). *Corpus, Concordance, Collocation.* Oxford: Oxford University Press.

Tilbury, A. (2010). *English Unlimited, CEF A2, Elementary: Coursebook with E-Portfolio DVD-ROM*. Cambridge University Press.