

# Corpus linguistics and language teaching

A teacher's guide to a working relationship

# What do we know about frequency?

- Which is more common in conversation, present simple (I live in Leeds) or present continuous (I'm staying in London)?

# What is a corpus?

- Principled and machine-readable collections of naturally occurring texts
- the British National Corpus (BNC) and the Corpus of Contemporary American (COCA) –<http://www.english-corpora.org>, a site
- <https://www.wordandphrase.info>
- ESP corpora

# Corpora and lexis

O'Keeffe, McCarthy and Carter (2007: 31):

- Most frequent words
- Vocabulary size (2,000 word threshold)
- Most frequent collocations e.g. Seriously consider
- Most frequent lexical chunks (formulaic sequences)  
e.g. To a certain extent

# Corpus Caveats

- Frequency v pedagogic convenience
- Corpus source v learner need
- Prodromou (1996): 'What is real in Nottingham is unreal in Athens and surreal in Delhi.'

# Collocation

‘This is a large theme in my thesis’

A combination of 2 lexical words frequently found together

- Dire ...
- Prior
- Severe
- Dulcet
- Carry out



Conzett 2000: 85):

- The single most important thing for teachers, more than worrying whether or not something is a collocation, is to shift their and their students' focus away from individual words to chunks of language.

# Lexical chunks

## **Lexical chunk:**

a frequent meaningful sequence of words which may include both lexical and grammatical words e.g. '*to a certain extent*' includes a preposition and an article.

- to a certain extent;
- by the way;
- last but not least;
- at the end of the day;
- play a role;
- not only ... but also;
- depend on



# The power of chunks

- Figment
- It is/was just/only a figment of his/your etc. (adjective)  
imagination

# The importance of lexical chunks

Lewis (1993): Language is grammaticalised lexis, not lexicalised grammar.

O'Keeffe, McCarthy and Carter (2007):

Integrated syllabus of words, collocations and chunks

# The ELT grammar syllabus: Tablets of Stone

Why is the ELT syllabus as it is?

What will this grammar structure enable my learners to do?

# Grammar: oversimplification

- Reported speech rules (Barbieri and Eckhardt 2007)
- 4 conditionals (Jones and Waller 2011)
- Clause-final and interactive use of 'though'

# Spoken v Written Use

- He's a bit of a nuisance. I quite like him though.

S1: Doing these reports is a pain

S2: They have to be done though

# The nature of spoken language

*Please put the following words in order of their frequency in **spoken British English**.*

- Think; House; Well; Money; Good; Just; Right; Weather.
- *Please put the following collocations in order of their frequency in spoken British English.*
- thank you; I think; you know; very nice; a bit; in fact; I suppose; this morning

# Spoken word frequency

- 1. Well
- 2. Think
- 3. Just
- 4. Right
- 5. Good
- 6. Money.
- 7. House
- 8. Weather
- 
-

# Spoken Collocation Frequency

- You know (1)
- I think (2)
- A bit (3)
- Thank you (8)
- In fact (11)
- I suppose (15)
- This morning (19)
- Very nice (84)



# The Lexis/Grammar Link

What tense are the following verbs attracted to?

- bleed
- chase
- shop
- starve
- joke
- kid
- moan

# The Grammar/Lexis Link

Most common verbs in Present Perfect aspect  
been; had; got; gone; done; made; seen; come; said;  
taken. (Biber et al 1999)

# The cumulative picture

- Lexis drives the meaning; grammar supports it
- A lot of language use is highly conventionalised
- Lexis and grammar are intertwined

# Learner Corpora

- Leech (1981, cited in Pravec 2002: 81) : ‘A learner corpus is a computerized textual database of the language produced by foreign language learners’.
- [Englishprofile.org](http://englishprofile.org)
- Syllabus design and level descriptors (e.g. CEF)

# Data-driven learning

- Learners access corpus data themselves to solve a problem
- Guided to autonomous continuum
- Viana, V. (2022) Teaching English with Corpora: A Resource Book. Routledge.

# The role of teacher education

- Teacher education needs to equip teachers with a critical awareness of what corpora have to offer.
- Teacher education needs to equip teachers with DDL skills and the critical awareness to judge when or whether it might be appropriate to use DDL.