Agenda

Corpus Linguistics and Corpora

Language Insights for Corpus Driven Research

Data-Driven Learning and Corpus Methodology for Teaching L2 Writing

Teaching Vocabulary, Linguistic Features, and Genre for L2 Writing

Teaching L2 Writing with the Corpus of Contemporary American English (COCA)

What is Corpus Linguistics?

"an approach to research and teaching that makes use of computer assisted analyses of language" (Conrad, 2005, p. 393)

Belief in language description through on observations of naturally occurring written and spoken language use (NOT native speaker intuition)

Corpus research gives "prominence to those items that occur frequently in the language and that learners are hence most likely to encounter in real-life communicative situations" (Roemer, 2011, p. 208).

What is a corpus?

- A corpus - "a systematic compilation of naturally occurring language - serves as a primary dataset for linguists/ researchers interested in describing and analyzing linguistic forms, functions, and variation" (Friginal, 2018)

- Corpus analysis includes:
  - Word and phrase frequency
  - Range (in how many texts)
  - Distribution (where it occurs)
  - Collocations or Strength of Association
  - Concordances (the word or phrase in context)

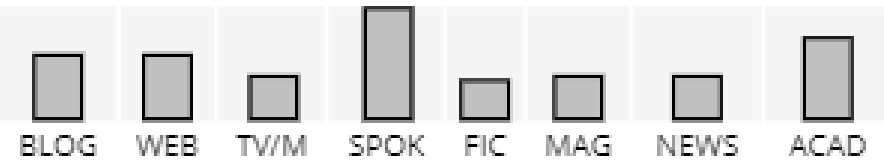# Sample Search for 'Question' (COCA)

Concordance for 'Question' (COCA)

CONCORDANCE LINES (more)

1	SPOK: 2017: CBS: Face The Nation	part of the debate . But let me ask you this	question	about how many people will lose coverage . What Speaker Ryan says
2	BLOG: 2012: io9.com	rock and start colonizing other solar systems -- there 's no	question	about it . Not only does our ongoing survival depend on it
3	SPOK: 1990: ABC_Business	as well as on the outside , and there 's no	question	about that . But there is , I think , at the
4	NEWS: 2004: CSMonitor	faculty and students . # Either way , it highlights some	questions	about the ethics of relationships , especially when they
5	MAG: 2018: Daily Beast	their trademark application . And after receiving a number of	questions	about their association with the Women 's March in light of the
6	BLOG: 2012: cs.bloodhorse.com	. # I had completely forgotten about Dashing Debbie . Good	question	about what happened to her . # Zookeeper , # I loved
7	ACAD: 2012: GeneralMusicToday	demonstrate perceptual skills by moving , by answering	questions	about , and by describing aural examples of music of various
8	SPOK: 1995: CNN_Burden	? JOSH COHEN : Yes . I wanted to ask a	question	about- I 'm sort of following up , I guess , in

Types of Corpora

General Corpora	• A large collection of texts from different registers, genres, and domains
Specialized Corpora	• Collection of tests from a specific domain, register or genre (contextualized)
Written or Spoken Corpora	• Collection of written texts or transcribed speech
Monitor Corpora	• collection of texts over time (language change)
Comparable or Parallel Corpora	• Collection of similar or translated texts from 2 + languages

Insights from Corpus-Driven Research

- Language choices are mediated by domain, register, and genre
- Sinclair (1991) *the idiom principle* "a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices" (p. 110).
- 58.6 percent of spoken and 52.3 percent of written discourse is formulaic (Erman and Warren, 2000)
- 'knowing' a word or phrase includes 'knowing' its collocations, semantic preferences, grammatical categories, and textual colligation.

Corpus Insights and the L2 Writer

- formulaic sequences serve special functions in genre and writing "because members of a speech community know the expressions" and use them (Schmitt and Carter, 2004)

- language learners struggle more with formulaicity and need a large amount of input in a variety of contexts for fluency in writing

Corpus Use and Teaching L2 Writing McEnery and Xiao (2011)

- **Indirect Use of Corpora**
  - Reference materials and language testing based on corpus research
- **Direct Use of Corpora**
  - Teachers use corpora to develop classroom materials
  - Students use corpora in classroom activities
- **Use of Specialized Corpora in specific contexts**
  - Collection and analysis of specialized corpora for EAP and ESP settings
  - Genre analysis classroom activities for EAP and ESP language learners

Data-Driven Learning (DDL)

- a data-driven learning approach "the computer acts as an informant and "gives learners access to the facts of linguistic performance" (Johns, 1991, p. 1).

- language learners search for and read concordances generated from corpus analysis software to identify patterns in actual written or spoken language

- an alternative to the traditional teacher-centered rules-based approach to language learning

DDL Pedagogy (Kennedy and Miceli, 2004)

teachers guide learners in using corpus tools to analyze features for patterns and draw conclusions

1. Formulate the question
2. Devise a search strategy
3. Observe the examples and select relevant ones
4. Draw conclusions

Criticism of DDL

- use of concordances 'decontextualize' the language from its original use
- Teachers are reluctant to adopt these methods because:
  - They don't learn corpus linguistics methodology in teacher education programs
  - They don't have access to freely available, user-friendly corpora and corpus tools
  - DDL can take up a lot of class time
  - L2 writers with lower proficiency find concordances overwhelming

Issues and Options in DDL Pedagogy

- Use of 'soft DDL', teacher-controlled deductive tasks – OR - 'hard DDL', student-led discovery tasks
- Use of paper-based corpus activities or online corpus activities
- Teacher mediation and training students to do corpus analysis
- With reading and writing, use of top-down or bottom-up approaches
- Simplifying concordance lines for L2 writers with lower proficiency

Learning Outcomes

- Studies mainly in undergraduate and postgraduate ESP and EAP settings
- Positive outcomes for vocabulary acquisition, phraseology, and academic writing
- Uses in EAP writing contexts
  - Use of concordances to teach highly frequent words, phrases, grammatical features, and moves in a genre or discipline
  - Raising learner awareness of language patterns not in textbooks
  - Inductive corpus tasks increase learner autonomy and critical thinking skills

Some Corpus Activities for Teaching L2 Writing

- Using www.wordandphrase.info to analyze text for frequent academic and technical vocabulary for reading and writing (Davies, 2015)

- Investigating frequently misused phrases in learner writing (Cresswell, 2007)

- Teaching students to consult corpora as they write (Gaskell and Cobb, 2004)

- Investigating writer positioning and stance with reporting verbs (Friginal, 2018)

- Teaching the use of lexical bundles in writing (Jones and Haywood, 2004)

- Raising genre awareness with concordances (Poole, 2016)

Using Word and Phrase for Academic Vocabulary (Davies, 2015)

- **Teacher Preparation:**
  - Analyze a sample text or model for writing for frequent words or phrases

- **Task:**
  - Use www.wordandphrase.info to analyze vocabulary in a model for writing

- **Discovery Activity:**
  1. Read the excerpt from sample text and analyze the underlined word: *The initial <u>impact</u> with the seawall occurred at 11:27 pm.*
  2. Search for 'impact' in www.wordandphrase.info
  3. What is the part of speech for impact in the sample text?
  4. Which definition most closely matches the context in the text?
  5. Look at some of the concordances. What adjectives frequently occur with 'impact'?
  6. Do you notice any other language patterns for this academic word?

Logical Connectors (Creswell, 2007)

- **Teacher Preparation:**
  - Analyze student writing for mistakes with logical connectors like *although, nevertheless, moreover, etc.*
- **Task:**
  - L2 writers use a sub-section of COCA (academic articles) to analyze a set of frequently misused logical connectors in concordances
- **Collaborative Activity:**
  - Groups receive a set of sentences with misused connectors underlined.
  - Groups use COCA to answer questions about meaning, collocation, and use.

Corpus Consultation While Writing (Gaskell and Cobbs, 2004)

- Teachers give links to **Compleat Lexical Tutor** website for common ESL writing errors in student writing as feedback

- Students go to the link and use the concordances to fix writing errors in their assignments.

A Holiday in My Country

Chinese New Year is a traditional holiday in my country. It is very important for Chinese. In Chinese New Year's Eve [Link_1], all the member of family [Link_2] will go to their parent's home and sit around to eat Chinese dumpling. That night called [Link_3] "reunion bight." People set off the firework and firecracker and put the antithetical couplet on their door that evoke a heavy holiday's atmosphere. In the reunion night there is a big evening party which was produced by CCTV show all the night. Almost Chinese watch it during the night. Some Chinese play the Mah-Jong or play card [Link_4]. All Chinese in the world do different thing [Link_5] to celebrate their holiday. In the first day of Chinese calendar of the year, every people say to another "Happy New Year!"

Link_1
```
1     e I did.      Christmas Eve and New Year's Eve are two double time cos he
2     think Gwynn had one too many New Year's Eve. Because      Never!     honest
3     that's it      fifty quid last New Year's Eve didn't we?      Yeah but you
4     ut not only that I mean  it's New Year's Eve, everybody else is bloody get
5     tion. Griggs, who'll be 75 on New Year's Eve, has been  helping Ruby Selle
6     r New Year's Eve.     Yeah for New Year's Eve. I started at half past six a
7     time in five years I've had    New Year's Eve off. And I was determined, an
8     es     Yeah     and had one on New Year's Eve or Christmas Eve or something
9     out of all the girls who work New Year's Eve   right, I told them on the ni
10    os it's a con right.    On er New Year's Eve.    She buys them presents.
```

Link_2
```
3     central  bank would mean all the members of the EC having a say i
5     particular case is that all  the members of the House of Lords ad
6     makes all the difference to the members of the public  and the g
```

Link_3
```
1     al shaped like a lower lip is called a labellum.    <48:1104>  DONN
2     Erm it is an existing    It's called a Qualitative Data Review form
3     <52:10>  What is this thing called a tear?    <52:11>  There are
4     e "British sickness" as it is called  abroad. The great strength of
5     ist we count, an organization called and we  also count an organiza
6     her who is the bride, the one called Anne. You are called  Joan --
7     il.    This latter section is called "Applied Sketching", but I fee
8     anatomical. This slip is so-called because its semi-ambiguous  RO
9     ne 1) The common dog-whelk is called by at  least three different n
10    s leaves. It is appropriately called   'Camelliifolia' and you can b
12    hat's what it's called. Is it called   client contact form?    Clien
```

Reporting Verbs (Bloch, 2010)

- **Teacher Preparation:**
- Teachers create a bank of sentences for concordances with reporting verbs used for academic positioning in research papers from COCA
- Use word frequency lists

- **Task:**
- Students read and classify concordance lines to develop understanding of how reporting verbs

- **Purpose:**
- Reporting verbs establish the credibility of both the writer and the claims so that there is a greater likelihood that the reader will accept the position the writer is taking.

Teaching Formulaic Phrases for Writing (Jones and Haywood, 2004)

- **Teacher Preparation:**
- Identify formulaic phrases to teach
- Select corpus extracts and remove phrases for fill-in-the-gap exercises.

- Task:
- Learners complete gap fills using COCA

Presenting supporting points

Read the passage below and fill the blank spaces with one of the expressions in the box

Stress is _____ one of the most serious modern diseases. _____ _____ the Institute of Management, approximately 270,000 UK workers take time off work every year because of work-related stress, at a cost to the nation in sick pay, lost production and medical bills of about £7 billion. _____ stress is less of a problem for bosses than for their subordinates, and _____ the survey, _____ only 9 per cent of junior managers looked forward to going to work. _____, only 7 per cent felt they were in control of their jobs.

which found that   arguably   this view is confirmed by
According to a survey carried out by   Furthermore
Experts have often suggested

From O'Connell. (2002). *Focus on IELTS*

Reconstruct the Context (Poole, 2016)

Stakeholder A
Read the sentences below carefully. Think about the keyword that is written in bold in each line.
Project
1 a final EIS for the Rosemont Copper **project** along with the Record of Decision.
2 Our leadership has always understood that our **project** would receive the highest scrutiny from
3 answered fully and successfully and that the **project** meets all standards addressed by the
4 and robust economics of the Rosemont **project**, propels Augusta to becoming a solid
5 The Rosemont Copper **project** is located in Pima County

*(10 more concordance lines with **project** were provided to students)*
Will

1 Rosemont Copper **will** set a high standard for sustainable mining
2 Phase 2 of the program **will** now focus on infill drilling
3 robust economics of this project, **will** propel Augusta to become a solid
4 the mine **will** produce over 400 direct jobs and about 1600

Discussion Questions:
- Who is the author of these sentences? Who is the audience?
- What is the purpose for writing?
- What rhetorical effect is achieved by the highlighted words?
- Are there other ways for the author to achieve the same effect?

Criteria for Evaluating DDL activities

- How well does the corpus materials fit the learner's needs and level of language?

- How well does the corpus activity teach linguistic features?

- Does the activity focus learner attention on particular forms?

- Is the activity collaborative?

- Does the activity provide Learners with feedback and provide teachers with assessment?

- Does the activity develop use of language learning strategies

Adapted from Friginal (2018)

Useful References

- Bloch, J. (2009). The Design of an Online Concordancing Program for Teaching about Reporting Verbs. *Language Learning and Technology, Vol 13, Iss 1, Pp 59-78 (2009)(1), 59.*

- Cobb, T., & Boulton, A. (2015). Classroom applications of corpus analysis. In D. Biber & R. Reppen (Eds.), *Cambridge Handbook of Corpus Linguistics (pp. 478-497): Cambridge University Press.*

- Conrad, S. (2005). Corpus linguistics and L2 teaching. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning (pp. 393-409). Mahwah: Routledge.*

- Davies, M., & Gardner, D. (2015). Corpus-based vocabulary support for university reading and writing. In N. W. Evans, N. J. Anderson, & W. G. Eggington (Eds.), *ESL readers and writers in higher education: understanding challenges, providing support (pp. 180-197). New York, NY: Taylor and Francis.*

- Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text-Interdisciplinary Journal for the Study of Discourse, 20(1), 29-62.*

- Flowerdew, L. (2008). Corpus linguistics for academic literacies mediated through discussion activities. In D. D. Belcher & A. Hirvela (Eds.*), The oral-literate connection : perspectives on L2 speaking, writing, and other media interactions* Ann Arbor : University of Michigan Press, c2008.

- Hunston, S. (2009). The usefulness of corpus-based descriptions of English for learners: The case of relative frequency. In K. Ajmer (Ed.), *Corpora and language teaching (pp. 141-154).*

- Johns, T. (1991). Should you be persuaded: Two examples of data-driven learning. In Tim Johns & Philip King, 1991 (eds) 6 Classroom Concordancing7. ELR Journal Vol. 4. 1-16.

- Jones, M., & Haywood, S. (2004). Facilitating the Acquisition of Formulaic Sequences: An Exploratory Study in an EAP Context. In N. Schmitt & R. Carter (Eds.), *Formulaic sequences: Acquisition, processing and use (pp. 269-300). Amsterdam, Netherlands: Benjamins.*

- Kennedy, C., & Miceli, T. (2001). An evaluation of intermediate students' approaches to corpus investigation*(3).*

- McEnery, T., & Xiao, R. (2011). What corpora can offer in language teaching and learning. *Handbook of research in second language teaching and learning, 2, 364-380.*

- Poole, R. (2016). A corpus-aided approach for the teaching and learning of rhetoric in an undergraduate composition course for L2 writers. *Journal of English for Academic Purposes, 21, 99-109. doi:10.1016/j.jeap.2015.12.003*

- Sinclair, J. M. (1991). *Corpus, concordance, collocation: Oxford University Press.*

Useful Corpora for Teaching Writing

- Corpus for Contemporary American English (**COCA**)
  - Davies, M. (2008 - 2020). The Corpus of Contemporary American English (COCA): One billion million words, 1990-2019. Retrieved from https://www.english-corpora.org/coca/. from Bringham Young University https://www.english-corpora.org/coca/

- The Michigan Corpus of Upper-Level Student Papers (**MICUSP**)
  - Michigan Corpus of Upper-level Student Papers. (2009). Ann Arbor, MI: The Regents of the University of Michigan.

- The Michigan Corpus of Academic Spoken English (**MICASE**)
  - Simpson, R. C., S. L. Briggs, J. Ovens, and J. M. Swales. (2002) The Michigan Corpus of Academic Spoken English. Ann Arbor, MI: The Regents of the University of Michigan.

Corpus Tools and Activities

- AntConc and Other Free Corpus Analysis Tools
  - Anthony, L. (2019). AntConc (Version 3.5.8) [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software

- Compleat Lexical Tutor
  - Cobb, T. *Compleat Lexical Tutor.* UQAM. Accessed 10 July 2020 at https://www.lextutor.ca/cgi-bin/range/texts/index.pl

- Tim Johns Kibbitzers (https://lexically.net/TimJohns/)

- MICUSP/MICASE Kibbitizers
  - (https://drive.google.com/drive/folders/1BNwXY_FSiYVR-9wVg55tZA043C5IsuvK)

- Davies, M. (2014) Analyze Academic Texts. Brigham Young University. Retrieved from www.wordandphrase.info

Upcoming Webinar

- Best Practices for Teaching an EAP course in Google classroom online.

- Increase student engagement and persistence by attending to universal design principles for course navigation, collaboration, assignments, assessments, and classroom management in an online environment.

- July 24, 2020 2:00 – 3:15 pm