Corpus Consultation for Academic Writing In English

Jamila K. Barton, Georgia State University
English Language Specialist Program
US State Department
AUCANI Webinar Series - July 17, 2020
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)

<table>
<thead>
<tr>
<th>CONCORDANCE LINES (more)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  SPOK: 2017: CBS: Face The Nation</td>
</tr>
<tr>
<td>2  BLOG: 2012: io9.com</td>
</tr>
<tr>
<td>3  SPOK: 1990: ABC Business</td>
</tr>
<tr>
<td>4  NEWS: 2004: CSMonitor</td>
</tr>
<tr>
<td>5  MAG: 2018: Daily Beast</td>
</tr>
<tr>
<td>6  BLOG: 2012: cs.bloodhorse.com</td>
</tr>
<tr>
<td>7  ACAD: 2012: GeneralMusicToday</td>
</tr>
<tr>
<td>8  SPOK: 1995: CNN, Burden</td>
</tr>
</tbody>
</table>
# Types of Corpora

<table>
<thead>
<tr>
<th>Type of Corpus</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Corpora</td>
<td>A large collection of texts from different registers, genres, and domains</td>
</tr>
<tr>
<td>Specialized Corpora</td>
<td>Collection of tests from a specific domain, register or genre (contextualized)</td>
</tr>
<tr>
<td>Written or Spoken Corpora</td>
<td>Collection of written texts or transcribed speech</td>
</tr>
<tr>
<td>Monitor Corpora</td>
<td>Collection of texts over time (language change)</td>
</tr>
<tr>
<td>Comparable or Parallel Corpora</td>
<td>Collection of similar or translated texts from 2 + languages</td>
</tr>
</tbody>
</table>
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](http://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 impact with the seawall occurred at 11:27 pm.*
  2. Search for ‘impact’ in [www.wordandphrase.info](http://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.

AHoliday 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 night." People set off the fireworks and firecracker and put the artificial couplet on their door that invoke 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 the Chinese calendar of the year, every people say to another "Happy New Year!"

**Link 1**
1. I did. Christmas Eve and New Year's Eve are two different days for us. I think everyday had one too many New Year's Eve. Because November holiday.
2. that's it. Fifty quid last New Year's Eve didn't we? Yeah but you
3. not only that I seen it's New Year's Eve, everybody else is bloody get
4. too. Crap, who'll be 75 on New Year's Eve. Yeah for New Year's Eve. I started at half past six a time in five years I had New Year's Eve off. And I was determined, an
5. 8. Yeah and had one on New Year's Eve or Christmas Eve or something
6. out of all the girls who wore New Year's Eve right. I told them on the mi
7. on it's a con right. Our New Year's Eve. She buys then presents.

**Link 2**
1. central bank would mean all the members of the EC having a say i
2. particular case is that all the members of the House of Lords ad
3. makes all the difference to the members of the public and the g

**Link 3**
1. a shaped like a lower lip is called a labellum. &8;1100; DONN
2. Am it is an existing it's called a Qualitative Data Review form
3. e "British sickness" as it is called abroad. The great strength of
4. i count, an organization called and we also count an organization
5. her who is the bride, she called Anne. You are called Joan --
6. This latter section is called 'Applied Sketching', but I fear
7. amongst this slip is called because its semi-transparent #
8. at 1) The common Dolly is called by at least three different n
9. e leaves. It is appropriately called 'Campanula' and you can b
10. 12 hats that's what it's called, is it called client contact form? Client
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
### 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


Useful Corpora for Teaching Writing

• Corpus for Contemporary American English (COCA)

• The Michigan Corpus of Upper-Level Student Papers (MICUSP)

• The Michigan Corpus of Academic Spoken English (MICASE)
Corpus Tools and Activities

• AntConc and Other Free Corpus Analysis Tools

• 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)

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