Session 2. Corpora as an Authentic Resource of Language and Beyond

Time: 11 am – 12 pm Winnipeg Time (CST)
Presented by Yuliana Bagan, English Online Inc., Winnipeg, Manitoba

December 6th, 2014

Funded by English Online Inc. Corpora in the ESL Classroom by Yuliana Bagan is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.
Agenda

Corpus?
Why corpus: applications in the ESL classroom
Conclusions
More resources
Questions
Corpus?

What do you think a corpus is? Please type in the chat box.
Corpus?

- Corpus is a weird word.
- It has a weird plural form.
- It’s large?
- It shows how words are connected.
- It shows which words tend to go together.
- It is some kind of a tool that linguists use to research a language.
Corpus?

• Do you use corpus?
  – If not, then what do you expect to be able to do with corpus?
  – If yes, then how?
corpus
plural corpora [countable]

1 formal a collection of all the writing of a particular kind or by a particular person
the entire corpus of Shakespeare's works
2 technical a large collection of written or spoken language, that is used for studying the language:
a corpus of spoken English

(Online Longman Dictionary)
Examples of corpus

- Conventional
  - British National Corpus (BNC)
  - Corpus of Contemporary American English (COCA)

- Non-conventional
  - Google

- Corpus based tools
  - WebCorpLive
  - Word and Phrase
  - Just the Word
COCA: everything covered

THE CORPUS OF CONTEMPORARY AMERICAN ENGLISH (COCA)

450 MILLION WORDS, 1990-2012

BRIGHAM YOUNG UNIVERSITY
Types of corpus

1. **Specialised corpus** – e.g.
   - genre: the language of newspapers
   - time: 2005 to the present day
   - place: just texts published in China

2. **General corpus** – needs to be much larger. E.g. The British National Corpus (BNC) has about 100 million words of spoken and written British English

(Corpus Linguistics Course, Future Learn)
Types of corpora

3. **Multilingual corpus** – e.g. English and Spanish. Or American English and Indian English.

4. **Parallel corpus** – e.g. English and Spanish – exactly the same texts translated. E.g. the CRATER corpus.

5. **Learner corpus** – language use created by people learning a particular language. E.g. the International Corpus of Learner English.

   The *Longman Dictionary of Common Errors* is based entirely on a corpus of genuine students' writing - the *Longman Learner's Corpus*.

6. **Historical or Diachronic corpus** – e.g. Helsinki corpus – 1.5 million words of texts from 700AD to 1700AD.

7. **Monitor corpus** – continually being added to. e.g. the Bank of English.

(Corpus Linguistics Course, Future Learn)
Corpus: applications in the ESL classroom

Activities Examples
Corpus: from theory to practice

• What adjectives collocate with a word Canada?
Corpus: Let’s go to corpus and find out:

Click on the link: http://corpus.byu.edu/coca/

1. Type **Canada** in a **Word(S)** string
2. Click on **COLLOCATES**
3. Select 1 on the left side and 0 on the right side
4. In **POS LIST** Choose **Adjectives ALL** from the drop-down list
5. Click **Search**
## Corpus: results from COCA

<table>
<thead>
<tr>
<th>CONTEXT</th>
<th>FREQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>WESTERN</td>
<td>108</td>
</tr>
<tr>
<td>EASTERN</td>
<td>100</td>
</tr>
<tr>
<td>SCHOLASTIC</td>
<td>90</td>
</tr>
<tr>
<td>NORTHERN</td>
<td>65</td>
</tr>
<tr>
<td>SOUTHERN</td>
<td>50</td>
</tr>
<tr>
<td>ENGLISH</td>
<td>26</td>
</tr>
<tr>
<td>CENTRAL</td>
<td>25</td>
</tr>
<tr>
<td>UPPER</td>
<td>24</td>
</tr>
<tr>
<td>ENGLISH-SPEAKING</td>
<td>19</td>
</tr>
<tr>
<td>ONLY</td>
<td>2</td>
</tr>
<tr>
<td>SUBURBAN</td>
<td>2</td>
</tr>
<tr>
<td>TORY</td>
<td>2</td>
</tr>
<tr>
<td>UNLIMITED</td>
<td>2</td>
</tr>
<tr>
<td>NON-TErrORIST-PROMoting</td>
<td>1</td>
</tr>
<tr>
<td>NON-LOCAL</td>
<td>1</td>
</tr>
<tr>
<td>NOBLE</td>
<td>1</td>
</tr>
<tr>
<td>NINETEENTH-CENTURY</td>
<td>1</td>
</tr>
<tr>
<td>NEUTRAL</td>
<td>1</td>
</tr>
<tr>
<td>NESTING</td>
<td>1</td>
</tr>
</tbody>
</table>
#1 Collocations Matter: example

• Use corpora to find “just the right word”
  – Search word collocations and use the lists for enriching learners’ vocabulary
  – Brainstorm possible word combinations and check them against corpora
Collocations matter:

Example 1

• A learner looking for a word to describe the density of dough
  – tough, firm, hard?
Collocations Matter: dough

Aha!!!
And…

100 more adjectives to describe dough

- soft
- sweet
- real
- little
- heavy
- risen
- big
- wonderful
- frozen
- sugar-cookie
- leftover
- rough
- store-bought
- high-moisture
- prepared
- living
- paper-thin
- refrigerated
- pie-crust
- raw
- buttery
- brown
- flaky
- deep-fried
- wrapped
- deep
- fried
- premixed
- fried
- pink
- Danish
- brown
- overhanging
- crescent-roll
- 10-inch
- firm
- baked
- remaining
- fresh
- simple
- smooth
- moist
- 12-inch
- chilled
- moist
- cooked
- forgiving
- tender
- packed
- kneaded
- white
- cold
- filled
- unrolled
- sticky
- basic
- uncooked
- tart
- uncooked
- dried
- ready-made
- special
- shaggy
- unbaked
- unleavened
- round
- round
- premade
- pliable
- thick-rolled
Dough can actually be forgiving?!

Photo Courtesy of Adrian Scottow CC-BY-SA http://goo.gl/E7woXw
Collocations Matter

Example 2

• Present new vocabulary not as isolated words, but in chunks and collocations, or use it as a warm-up activity
  – Fork activities
Collocations Matter

Fork Activities

problems
  trouble
  damage
  difficulties
  cancer
  injury
  death
  confusion
  ...

? (verb)

? (verb)

Cause (verb)
Collocations Matter

Example 3

- Build on commonly used words learners know and expand their vocabulary by searching for synonyms
  - *a beautiful house*
Collocations matter

- They will be surprised to see the results:

<table>
<thead>
<tr>
<th></th>
<th>CONTEXT</th>
<th>FREQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>BEAUTIFUL HOUSE</td>
<td>43</td>
</tr>
<tr>
<td>2</td>
<td>LOVELY HOUSE</td>
<td>31</td>
</tr>
<tr>
<td>3</td>
<td>ATTRACTIVE HOUSE</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>MAGNIFICENT HOUSE</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>HANDSOME HOUSE</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>CHARMING HOUSE</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>PICTURESQUE HOUSE</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>WONDERFUL HOUSE</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>SUPERB HOUSE</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>DELIGHTFUL HOUSE</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>GORGEOUS HOUSE</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>EXQUISITE HOUSE</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>TOTAL</td>
<td>102</td>
</tr>
</tbody>
</table>

Data by British National Corpus
Collocations matter
Let’s see examples:

• In the 1780s, when Highgate Hill was so steep and deeply rutted that carriages regularly failed to make the grade, and the drive to town sufficiently dangerous that a wise man went with pistols, a merchant called Thomas Roxborough had constructed a **handsome house** on Hornsey Lane, designed for him by one Henry Holland. (Imajica. Barker, C. Glasgow: HarperCollins, 1992, pp. 7-131. 3030 s-units; BNC)

• The Bishop having departed, Treadwell had walked Theodora down the short drive of his **handsome house**. (Unholy ghosts. Greenwood, D M. London: Headline Book Pub. plc, 1991, pp. 1-142. 3531 s-units; BNC)

• Because although it's a **handsome house**, and the gardens are extensive, they in no way compare to those of the castle which is just up the road. (Love of my heart. Richmond, Emma. Richmond, Surrey: Mills & Boon, 1993, pp. ???. 4267 s-units; BNC)

• The new farm was majestic -- a **handsome house**, a huge acreage, a dairy herd as well as beef cattle, sheep as well as shire horses, and no tractor. (Country Living. London: The National Magazine Company Ltd, 1991, pp. 4-180. 2186 s-units; BNC)
Collocations matter

- But we can only use handsome when describing a male!?

<table>
<thead>
<tr>
<th>Word</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Man</td>
<td>81</td>
</tr>
<tr>
<td>2. Face</td>
<td>75</td>
</tr>
<tr>
<td>3. Features</td>
<td>16</td>
</tr>
<tr>
<td>4. Woman</td>
<td>16</td>
</tr>
</tbody>
</table>
Word and Phrase
Example 5

• **Word and Phrase** - run texts through a corpus based tool to see:
  • the range of vocabulary
  • word lists
  • definitions
  • all within ONE browser tab open
How do we deal with all this **data** without getting information **overload**? How do we use **data** to gain real **insight** into the world? Finding ways to pull interesting **information** out of **data** can be very **rewarding**, both **personally** and **professionally**. The managing editor of *Financial Times* observed on CNN’s *Your Money*: "The people who are able to in a **sophisticated** and **practical** way **analyze** that **data** are going to have **terrific** jobs." Those who learn how to present **data** in **effective** ways will be **valuable** in every field.
How do we deal with all this data without getting information overload? How do we use data to gain real insight into the world? Finding ways to pull interesting information out of data can be very rewarding, both personally and professionally. The managing editor of Financial Times observed on CNN's Your Money: "The people who are able to in a sophisticated and practical way analyze that data are going to have terrific jobs." Those who learn how to present data in effective ways will be valuable in every field.
Word and Phrase: as a dictionary

<table>
<thead>
<tr>
<th>OVERLOAD</th>
<th>SPOKEN</th>
<th>FICTION</th>
<th>MAGAZINE</th>
<th>NEWSPAPER</th>
<th>ACADEMIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLICK BAR TO LIMIT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STORED</td>
<td>18</td>
<td>23</td>
<td>52</td>
<td>38</td>
<td>71</td>
</tr>
<tr>
<td>MORE</td>
<td>76</td>
<td>106</td>
<td>223</td>
<td>154</td>
<td>255</td>
</tr>
</tbody>
</table>

**DEFINITIONS (WORDNET)**

1. an electrical load that exceeds the available electrical power
2. an excessive burden

**COLLOCATES**

*adj* sensory, visual, toxic, cognitive, emotional, thermal, psychosomatic, patient, nervous, suffering *noun* information, role, system, conflict, iron, stress, symptom, anxiety, protection, demand *verb* cause, suffer, avoid, result, prevent, experience, occur, refer, deal, link

**CONCORDANCE LINES**

<table>
<thead>
<tr>
<th>GENRE</th>
<th>SORT</th>
<th>CLICK WORD TO:</th>
<th>SEARCH AS COLLOCATE</th>
<th>QUERY THAT WORD</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACAD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FIC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPOK</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACAD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACAD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACAD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Word and Phrase: word lists**

- Use word lists for
  - pre-reading activities
  - adapting to different proficiency levels

<table>
<thead>
<tr>
<th>RANGE 3 (COCA LIST &gt; 3000) WORDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: analyze, insight, overload, personally, practical, professionally, rewarding, sophisticated, terrific, valuable</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RANGE 2 (COCA LIST 501-3000) WORDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>5: data</td>
</tr>
<tr>
<td>1: editor, effective, gain, managing, observed</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RANGE 1 (COCA LIST 0-500) WORDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>5: to</td>
</tr>
<tr>
<td>3: how, in, the</td>
</tr>
<tr>
<td>2: and, are, be, do, information, of, ways, we, who</td>
</tr>
<tr>
<td>1: a, able, all, both, can, deal, every, field, finding, getting, going, have, interesting, into, jobs, learn, on, out, people, present, pull, real, that, this, those, use, very, way, will, with, without, world</td>
</tr>
</tbody>
</table>

**ACADEMIC**

<table>
<thead>
<tr>
<th>DATA</th>
</tr>
</thead>
<tbody>
<tr>
<td>5: data</td>
</tr>
<tr>
<td>1: analyze, effective, insight, observed, practical, present</td>
</tr>
</tbody>
</table>
Corpus for learning Pragmatics

Example 6

- Use corpora to teach pragmatic expressions:
  - Thank you, I am good. (Politely saying “No” instead of “No, I don’t want it.”)
  - I am so over it. (spoken grammar)
  - I so want to get it. (Using so followed by a verb)
  - It’s way better. (“way better” for “much better”)
  - Are we there yet? (yet with Present Simple)
Other examples

• Spoken corpus for listening activities and spoken grammar
• Corpus of Academic English for EAP/ESP
• Building your own corpus
• And much more…
Corpus is not ...

- Replacing dictionaries, grammar books, course books or Google
- Substituting teacher in the classroom
- Denying everything we know about the language
Corpus is …

- Showing how language exists in reality
- Confirming or denying our intuitions about the language
- Revealing interesting facts about language
- Promoting teacher and learner autonomy
- Proving that language is rich, creative, allowing and forgiving
Use corpus to ...

- Promote exploratory learning and learner independency
- Trigger imagination, creativity and curiosity about language
Got interested?

More free resources here:

- Corpus Linguistics Community on Google +
- Corpus Linguistics Massive Open Online Course at Future Learn (coming in 2015)
- EFL Notes by Mura Nava
- Classroom Games From Corpora by Ken Lackman
- A lesson plan to introduce COCA activities in class
- Professor Geoffrey Leech at Lancaster University
Corpora list

1. British National Corpus 100 million words, 1980s-1993
   http://corpus.byu.edu/bnc/
2. Corpus of Canadian English (Strathy) 50 million words, 1920s-2000s
   http://corpus.byu.edu/can/
   http://corpus.byu.edu/coca/
4. Word and Phrase (on the basis of COCA) http://www.wordandphrase.info/
5. Word and Phrase (on the basis of academic texts from COCA)
   http://www.wordandphrase.info/academic/analyzeText.asp
6. Michigan Corpus of Academic Spoken English
   http://quod.lib.umich.edu/cgi/c/corpus/corpus?c=micase;page=simple
good enough. "And thank you for the present but what's wrong with it?": MELA: Mrs. Jervis, thank you for your kindness. Heaven reward you!

Pleasant Hall. PAMELA: thank you. Mr. MR. WILLIAMS: to MRS. JEWKES: formation so I can't thank you for it. JACKIE: Ods blood! PAMELA: I am GUIL: (Wry, gentle) Thank you; we'll let you know. (The PLAYER has much. Smiles.) Thank you. (The PLAYER turns and goes. ROS has bent for the no confidence in England. Thank you. (Thinks about this.) And even if it's the ability of life to thank you. He never gave commandment for their dirl Guides or church fest. Thank you. # THANK you to the great philosophy kettle means water. # all the other rhymesters, including behalf of them, we thank you. # KEEP IT COMING FOLKS # TWO million Services department at the council. Thank you. # THANK you, too -- for your department at the council? Thank you. # THANK you, too -- for giving me a youngsters cuddled up yesterday for a special thank you to the man who mungering kisses. The princess beam. Thank you. Psychiatric nurse Hub suggested she sit there. "No thank you," said Anne stiffly. 'We think you #328. # A song to say Thank You # MARJE PROOPS # MR K. Bailey, often she used a consultant's thank you letter to back her cheeky scam -- and gold crown. 'Father, thank you, thank you, thank you,' the newly-weds screamed. 'Father, thank you, thank you.' Thank you. The newly-weds screamed. At least 50 I text of the leaked n Thank you for taking time out of your busy sc
Have questions

- Email me at ybagan@myenglishonline.ca
- Skype me at yuliana_myenglishonline
- Phone me at (204) 946-5140 ext. 204
Stay tuned for upcoming events:

What: REALIZE National Realize Forum for EAL/ESL Professionals
When: January 23 & 24, 2015
Where: Online at English Online Inc.