

ACORN
(Aston Corpus Network)

PRACTICAL EXERCISES

Ramesh Krishnamurthy

1: A few corpus sites offer full-sentence examples

- <http://sara.natcorp.ox.ac.uk/lookup.html>
[British National Corpus]
- *Here is the first example. Can you guess which word I searched for?*
- **HH3 2316** They report to work at 8.30am on an empty stomach.
- *Here is another example for the same word. Can you work out which word it is now?*
- **G2Y 200** As English Heritage do not like us, it is with some glee that I report dissent in their own ranks.

Note

- A language learner may **come across/notice** a particular word (or a particular use of that word; or a phrase) for the first time one day, but may not **come across/notice** the same word again for another week, a month, a year...? By that time, they have often forgotten the first time...
- **A corpus can show you many examples of the same word, or use of the word, or phrase at the same time... = concentrated exposure; more impact**

Can you find the word in a full-sentence example?

- Can you find the same word in the next example?
- **ADB 39** While the younger members of the family are treated by the British press and public as stock characters in some soap opera, when one of them speaks, on architecture, AIDS or literacy, the news media none the less report their comments seriously.
- **And in the last example?**
- **B1E 308** Using data collected over a 2-year period from plots under a variety of management techniques, ranging from maize cultivation to a range of pasture grasses, they report that soil erosion and nutrient losses were greatest from maize (*Zea mays*) and guinea grass (*Panicum maximum* Jacq.) plots.

Note

- It is often difficult to **find** the search word in long, full-sentence examples.

2. Most corpus sites show examples in KWIC format: Key Word In Context

- <http://www.collins.co.uk/Corpus/CorpusSearch.aspx>
[Collins' Corpus Concordance Sampler]
- to make a sturdy coffee **table** with plenty of room for
- others who sit round her **table** include the business men
- MACPHERSON : Now on the **table** here we've got what looks
- a loveseat, a coffee **table** two large cardboard
- suitcases lay open on a **table**. The silver had already
- 1818? [p] Answer: From **Table** C, we find the year 1816
- I lowered my tray **table** from the back of the seat
- end of the dark shiny **table** and toasted each other in
- the size of a snooker **table**. Another room upstairs is

This is called a **CONCORDANCE**.

In KWIC format, the search word is obvious...but we now need to deal with bits of text, text fragments, incomplete sentences.

Note

- But this is quite **natural**. We actually **learn a lot of our L1** not from complete sentences, but **from such text fragments**: overheard bits of conversation when you enter a room, or on a bus; a torn poster on a wall, or a notice defaced by graffiti...

3. Retaining awareness of text

- *Take any KWIC example, and try to guess the word before the beginning of the example, and the word after the end of the example:*
- ...?... upon these cans for food. This did not **happen** in counties of North Georgia, where the rivers ...?...

Some corpus sites offer 'extended context' for each KWIC example: so you can check your guesses

- This is called '**para**' (paragraph view) in ACORN:
- vegetables from her garden was going to can soup, broth, hash, and stew against the winter. She had done it last year, and the year before, and the year before that, and she, and her people were **dependent** upon these cans for food. This did not **happen** in counties of North Georgia, where the rivers **run** and make rich the bottom land. Nor in South Georgia, where the summer sun shines warmly and gives early life to the things growing in the flat fields. This happened in Decatur, De*kalb County,

Note

- I often play a similar game with my 12-year-old daughter, when she reads to me at bedtime: **I stop her just before she turns the page of the book, and ask her to guess the word or phrase on the next page...**
- She often guesses right – sometimes not exactly, but this often leads to an interesting chat about her guess and the text, the author's choice of wording...

4. A simple corpus exercise: gapped concordances

- *Take a few examples from a concordance, and delete the search word:*

- to make a sturdy coffee
- others who sit round her
- MACPHERSON : Now on the
- a loveseat, a coffee
- suitcases lay open on a
- 1818? [p] Answer: From
- I lowered my tray
- end of the dark shiny
- the size of a snooker

with plenty of room for
include the business men
here we've got what looks
two large cardboard
The silver had already
C, we find the year 1816
from the back of the seat
and toasted each other in
Another room upstairs is

Can you remember the search word?

Note

- When you do this with an unknown word, it is more important to **focus on the CLUES** that help you to guess the word, rather than on getting the right answer...

A gapped concordance exercise

- - so it's generally best to feed a small
twice a day. Similarly, growing an
- fascinate me. You can tell how a is
feeling by his tail. If his tail is up,
- of days later I'm watching TV and this
food commercial comes on an it's the
- Year. For what remains of the Year of the
....., the Tiger should remain committed
- he was with a woman walking a black On
neither occasion did the man speak to
- I'm prepared if a policeman sees me with my
..... off a lead in a public place
- centuries. [A] Yes. [B] They were very-
eared lecture notes that he had in front

5: Log in to ACORN

- Open **Internet Browser**
- Preferably **Mozilla Firefox** (the “Edit – Find in This Page – Highlight all” function is very useful)
- [Internet Explorer is OK]
- Delete current URL and type:
acorn.aston.ac.uk/open
- In Username box, type: **rk_1**, **rk_2**, etc
- In Password box, type: **acorn1**, **acorn2**, etc

NB unfortunately, these logins are ***only valid for today***;
we initially obtained copyright permission for Aston only;
we are re-negotiating permission with data contributors

ACORN exercise 1: using frequency lists

- *You are going to teach English to medical students. You want them to be able to write abstracts for their research papers. First, you want to check their knowledge of vocabulary.*
- **Click on English** (Language selection)
- **Click on the box next to 2. Brown Corpus** (Corpus selection) and then **click Continue>>**
- **Click on Word Frequencies**
- **Click on Top 50**
- **Make a list of the content/vocabulary words (i.e. not the grammar/function words)**

ACORN exercise 1 (cont)

- The Brown Corpus is a corpus of **GENERAL ENGLISH**, so **grammar/function words are the most frequent words** (NB if you look at the Top 100 later, you will find *said, time, new, two, man, years* lower down)
- Click on **Corpus Selection** (top left, underlined)
- Click on the box next to **10. Medical Abstracts** (Corpus selection) and then **click Continue>>**
- Click on **Word Frequencies**
- Click on **Top 50**
- **Make a list of the content/vocabulary words** (i.e. not the grammar/function words)

Acorn Exercise 1: 'content words' in Medical Abstracts corpus word frequency list

patients	10-11 words in Top 50
treatment	patients + patient: human
clinical	treatment + therapy
<i>patient</i>	
therapy	
results	research
cases	not human
study	research
disease	cause
diagnosis	doctor's first action
surgery	2 nd action (cf treatment/therapy)

ACORN exercise 2: comparing domain-specific word frequency lists

- Click on Corpus Selection (top left, underlined)
- Click on the boxes next to 10, 11, and 16 (for Medical, Business, Labour Party=Politics) and then **click Continue>>**
- Click on **Word Frequencies**, then **Top 50**
- Make a list of the content words in the **Business and Politics corpora**
- Which words are unique to one domain?
- (Use the “Edit – Find in This Page – Highlight all” function in Mozilla Firefox)

ACORN exercise 3: comparing phrase frequencies in domain-specific corpora

- Click on **Phrase Frequencies** (top, underlined)
- Click on **4-grams** (= 4-word phrases)
- Click on **Top 50**
- Which phrases occur **only in one domain?**
- Which phrases occur in **more than one domain?**
- (Use the “Edit – Find in This Page – Highlight all” function in Mozilla Firefox)

ACORN exercise 4: looking at concordances, and sorting

- Click on **corpus selection**
- Click on **Brown Corpus** and then **click Continue>>**
- Click on **Concordance**
- **Type in *same* and click Search**
- Click on **Display Results**
- What words you notice to the left and right of *same*?
- Click on **Sort Results**
- **Click on ‘1st word before keyword’ in Main Sort line**
- **Click on the grey box Sort Results below**
- **What do you notice now?**

ACORN exercise 4 (cont)

- Which words in red are NOT *the*?
- *shrugged, That, these, those, Wednesday*
- Click Sort Results again and '1st word after keyword' in Main Sort line
- Click on the grey box Sort Results below
- What do you notice now?
- Click on corpus selection
- Click on the box next to 10. Medical Abstracts (Corpus selection) and then click Continue>>
- Click on Concordance
- Select one of the content words we extracted from the word frequency list, look at the concordances, sort them to left and right, and report back on what you found
- Which phrases/grammatical patterns would you teach your medical students? *I haven't looked at these before, so we will be genuinely discovering things together*