Using the ACORN (Aston Corpus Network) corpora for pedagogic purposes

Ramesh Krishnamurthy

Corpus-1: Why bother with corpora?

- "Language users cannot accurately report language usage, even their own" (Sinclair, 1987)

- "There are many facts about language that cannot be discovered by just thinking about it, or even reading and listening very intently" (Sinclair, 1995)

- As language teachers and professionals, we often have strong intuitions about language use… *Corpus-based research, however, shows us that our intuitions are often completely wrong.* (Biber 2005)

Corpus-2: Widening use of corpora

- From the 1960s to the 1990s, corpora were used mainly for research, in **language description and lexicography**.

- In the past decade, they are increasingly being used in other fields such as **translation, forensic linguistics, and computational linguistics**… and for **language learning and teaching**.

Corpus-3: Changes in language learning and teaching

- **goals**: from intellectual and literary to practical and communicative (business, media, tourism)
- **language content**: from made-up examples and classroom dialogues to authentic, real-world communication
- **learning and teaching methods**: from passive to active learning; from teacher-focus to learner-focus and learner autonomy; use of computer technology

Corpora are well-suited to these changes

Corpus-4: Corpora for language learning and teaching

- We <u>learn</u> our **mother tongue** by experiencing thousands of examples, in wide range of texts and situations, over a long period of time **(=inductive)**

- But we <u>teach</u> **other languages** through more abstract information (grammar rules, dictionary definitions), with fewer examples, limited in variety, over a shorter time period **(=deductive)**

- Corpora may **reduce this exposure deficit** (by concentrated exposure to variety of examples), and **offer new strategies / paradigms for language learning / teaching (=inductive)**

Corpus-5: How can corpus serve disciplines other than languages?

- **Improve general English academic writing skills**

- Differentiate concepts/terms of your discipline from general language use: **identification of subject terminology**

- Examine the **opinions and arguments** of **experts** in your discipline by looking at contexts around the key concepts/terms; how are opinions/arguments presented/discussed

- Compare the **discourses of different disciplines**

- Identification of suitable **quotes**

ACORN-1: Development History

- **2004-5: planning** meetings; **awareness-raising** LSS seminars; voluntary contributions from 2 PhD students
- **2005-6**: **LSS support**: computer, room; £1500 for Computer Science placement student; **pilot data and software**
- **2006-7**: HEFCE/**Aston**/FLDC **funding**: £26,380: my P/T secondment; 2 x P/T PhD students; 2 x undergraduate assistants; 1 x F/T PhD student; £5,200 for CS placement student; **data collection; web software**
- **2007-8: Live launch; dissemination; publications; external funding bids**; FLDC funding: £5,200 for CS placement student; enhancing interface; involving colleagues in corpus use with students and research
- **2008-9:** LSS funding: £5,200 for CS placement student; 3 x F/T PhD students; 2 x P/T PhD students

ACORN-2: Main Processes

- **Identify and select relevant texts** (staff questionnaires)

- **Obtain copyright permission**

- **Acquire texts** (download from Web; email attachments, etc)

- **Convert text formats** (PDF, DOC, HTML, etc to plain TXT)

- **Index data** (make texts available for corpus software)

- **Design web interface**

ACORN-3: Staff Questionnaires

- **Topics**: national identities and stereotypes, cultural differences, current affairs (social and political), institutions, international relations, history, economics, marketing, education, globalisation, cultural events, media, immigration
- **Language**: stylistics, terminology, dialects, grammar, discourse analysis, history, policy and planning
- **Text Types**: film/book reviews, commentaries, readers' letters, obituaries, abstracts, instruction manuals, résumés, academic writing, tourist brochures, recipes, fairy tales, short stories, novels, political speeches, medical texts
- **Academic Journals, Literature, Journalism, and a range of Websites**

ACORN-4: adding Student Data

- Obtaining material from students: consent forms; electronic submission
- Research will provide more information about the students:
- general academic development; mastery of topics, themes, subjects
- development in academic writing style
- language development; ab initio to PG
- successful learning and teaching strategies which can be shared by staff and students
- possible problem areas, and the need to use alternative learning and teaching strategies
- the strengths and weaknesses of the current syllabus, and the need to adjust the focus, alter the sequence, and add or omit elements

ACORN-5: 2007-2008 update

- Launched June 2007 to Aston staff and students
- over **4000 user-searches** since then

  (= c 400 per month)
- Added various **subcorpora**
- **Website** (Google Analytics): Apr20-May20: 171 visits from 14 countries (UK, Japan, USA, Germany, Italy, Austria, Spain, Indonesia, Canada, Ecuador); 79 visited only once; 92 visited 2+ times; Jun2-Jul2: 244 visits (22 countries; 158 returning, 86 new)

ACORN-6: 2008

- **ACORN** (Aston Corpus Network) currently contains c. **200 million words of texts**… …the equivalent of:
  - 7 years of non-stop conversation
  - 2000-3000 books
  - 15 years of a daily broadsheet newspaper
  - 12.5 million text messages
- **in English, French, German, Spanish**
- **and Parallel Texts for Translation Studies**
- **Pedagogic purposes:**
- '**expert**' (input) **texts** – selected in consultation with colleagues
- **student** (output) **texts** – to see how they progress, re-assess input
- **web access: in classroom, private study**
- **simple interface**

ACORN for pedagogic purposes:
Five brief Case Studies

1: English: pre-sessional Business students

2: English: assisted self-correction of Writing

3: English: UG Semantics: near-synonyms

4: Spanish: Grammar Clinics

5: Corpus Linguistics: UK political parties

Case Study-1: pre-sessional Business students

- Kaori, a Japanese-L1-speaker, Masters dissertation research

- Taught part of 5-week course; 2 x 30-minute sessions per week

- Session 1:

   - introduction to corpora / ACORN

   - raise awareness of **general** and **domain-specific vocabulary**

- Used Business English Corpus frequency list and concordance – grammar/general words and 'business vocabulary' [**see WBE corpus**]

Case Study-1: Evaluation

- **Positive Student Feedback**: *If I do not know how to use the word, corpus is helpful. I can check how the word is connected with other words such as verbs, adjectives and adverbs…It is useful for writing. When I write something, I check concordance… If similar examples are shown together in the concordance, it is helpful for language learner to study English effectively… Can I use the ACORN corpus after I graduate university?*

- **Positive Staff Observer evaluation**: *the students applauded at the end of the lesson, which you deserved.*

- **My suggestions**: *more **gradual introduction** (more sessions); train students to **use corpus themselves**; more careful **selection of examples** (for exercises); use **annotated frequency list** first; use **sorted concordances** (for multi-word units, patterns)*

Case Study-2: Assisted Self-Correction of Writing

- Steven, a Chinese-L1-speaker, Computer Science placement student working for ACORN
- He submits **weekly 1-page report** to me
- I spend c. **3 minutes** highlighting 'errors' in green
- He spends
  - c.**10>5 minutes** correcting 15%: 'silly mistakes'
  - c.**30>15 minutes** correcting 70%: **using the corpus**
- We spend c.**15 minutes** correcting the remaining 15%: determiners, countability, tense sequence, register, reported speech, conditionals, pragmatics, discursive logic (many contrastive Chinese/English features)…
- …and discussing software design and corpus search procedures (for help files, tutorials, etc)

Case Study-2: Examples and Evaluation

Examples:

- I will take ***a deep look into*** it next week. > ***close look at***
- I ***replied him***. > **replied to**
- He was not ***an expert with*** MySQL. > **expert in**
- The server might ***out put*** an error message.> **output**

[see ACADEMIC CORPUS]

Evaluation:

- Steven **enjoys** this method: **empowering, incidental learning** of other lexis and grammar; he makes **changes to software, from personal experience**: eg phrase search
- It **frees up my time** for more interesting / complex aspects
- The **procedure is speeding up**, as he remembers/checks previous weeks' problems; NB restricted, repetitive content
- Keep **more accurate records**: when is ACORN used; quickest, most useful searches; insights into learning process; when is external prompting needed; can use while drafting; reduces range/variety of errors

Case Study-3: UG Semantics: near-synonyms

- Pierre, a French colleague, 2ⁿᵈ year UG Semantics module

- Introduced ACORN; task: **disambiguate near-synonyms** using concordances and sorting: *town* and *city*

- **Student findings**: *large city, immense city, mighty city, small town:* is <u>size</u> the only difference?; *the town school* [sing] but *city schools* [pl]; often <u>name</u> + *city,* <u>not</u> often with *town;* <u>phrases</u> with *town: across town, around town, out of town; home town, market town, old town, town hall, town planning;* <u>phrases</u> with *city: capital city, inner city, city university, Holy city, City of god*

Case Study-3: Evaluation

- **Student feedback**: *useful way to ascertain more exact meanings and collocational nuances; dictionary often only gives a couple of examples; informative with respect to connotations; straight forward and easy to use*; *could be useful for language learning, translation studies, terminology*; *interesting; easy to use, very useful*
- **My observations**: need bigger general corpus, but system speed is already slowing; use fewer concordances; use sorting; may need 'para' function, more context for some synonyms

Case Study-4: Spanish Grammar Clinics

- Guadalupe, Spanish colleague, post-doc Fellow
- **2 groups: Beginners and Advanced**
- **Staff pre-selected topics**
- **Beginners: corpus-based**: teacher explains a grammar topic, students search corpus for examples matching each grammar statement
- **Advanced: more corpus-driven**: teacher specifies items (eg specific verbs, prepositions); students analyse concordances; and they find, list, and explain patterns themselves

Case Study-4: Examples

Beginners: *ser*: noun/pron + *ser* + noun/pron; noun/pron + *ser* + adj/phrase (=identity/nature); *ser* + *de* + noun (=origin, substance); in impersonal statements; *estar*: to describe state (not identity/nature); *estar* + *de* + adj/noun (= mood, temporary activity/situation); to indicate location

Advanced: preps: <u>students found:</u> *para*: purpose, object, destination; direction after verbs of motion; =*by* in time phrases; =*for* + period of future time; *por:* =*because/why;* =*by* in passives; =*in support of;* with *prices* and *amounts of money*

Case Study-4: Evaluation

Evaluation:

- Need better <u>corpora</u>, especially for beginners
- Software needs to <u>distinguish accented</u> and non-accented forms
- Advanced students <u>successfully discovered</u> most of the patterns in textbook
- <u>Positive reactions</u> of students to corpus
- Grammar <u>can</u> be taught in computer labs
- Software favours <u>lexical</u> analyses; <u>lemmatization</u>, <u>P-O-S tagging</u> may help for grammar; also need different <u>strategies</u>
- In future, can use <u>errors from student data to identify topics</u>
- Corpus can help to find <u>missing items in gapped exercises</u>

Case Study-5: Corpus Linguistics: UK political parties

- Iztok, PhD student, delivering Masters module
- Use corpora of **Manifestos** of Conservative, Labour, and Liberal(-Democrats) since 1900
- Compare **word frequencies and phrase (N-gram) frequencies:** *government, people, Britain, national, industry, education, tax, health, country, community; we will continue to, we will ensure that, in the interest of*

Case Study-5: Evaluation

- Useful to introduce **corpus functions**

My suggestions:

- Such corpora could be used as source of citations for Politics students, but would need accurate **bibliographic** information

- Data currently being collected for research projects (eg climate change) in Sociology could also be used in similar ways

Case Studies: Conclusions

- Corpora and software need improvement
- Students respond positively to corpus

  - if they are given initial training and tasks

- Corpus can be used successfully:

  - in classroom and for self-access

  - for lexis and grammar

  - for raising awareness of domain-specific vocabulary

Next steps

- Provide corpus **training sessions** for colleagues; more documentation, tutorials, help files
- Add corpus use in more **modules (UG and PG)**
- Make ACORN available within **VLE** (Blackboard)
- Process and analyse Aston **student data**
- Add simple **print and copy-paste** functions
- Create **exercise templates** (eg gap-fill exercises)
- Add **bibliographic data** for citational purposes
- Request **copyright permission** to allow **access beyond Aston**