AntPConc: A Freeware Multi-Platform Parallel Concordancer

Laurence ANTHONY
Professor, Center for English Language Education, Waseda University, Japan
Honorary Research Fellow (2013-2015), Lancaster University
anthony@waseda.jp
http://www.antlab.sci.waseda.ac.jp/software/

Overview of presentation

- Background
- Definitions and uses of parallel corpora
- Current desktop and web-based parallel corpus tools
- The need for new desktop and web-based parallel concordance tools

Overview of AntPConc
- Design and features
- Case study using AntPConc
- Demonstration
- Discussion
- The future of parallel corpus tools in corpus linguistics

Background

- Parallel corpus (a definition):
  - "A set of texts in language A and their translations in language B" (Kenning, 2010)

- Related terms:
  - aligned corpus, bilingual corpus, comparative corpus, comparable corpus, multilingual corpus, translation corpus

- Existing parallel corpora:

- Using parallel corpora (in and beyond translation):

- Main advantage in translation studies
  - investigation (and avoidance) of translation errors related to collocations and semantic prosodies (Yepes, 2011)

- Main advantages in the classroom:
  - encourages independent learning via the Data-Driven Learning (DDL) approach
  - A focus on the exploitation of authentic materials
  - A focus on real, exploratory tasks and activities
  - A focus on learner-centered activities
  - A focus on the use and exploitation of tools
  - leads to higher motivation

- Limitations
  - complex import for non-English corpora
  - designed for corpus linguists
  - single platform (Windows)
  - commercial
  - not updated
Background: Current web-based parallel corpus tools

- **E-C Concord**: English-Chinese Parallel Concordancer
  - Developed by Wang Lixun
- **MOA**: Korean/English Parallel Concordancer
  - Developed by InJung Cho
- **Multi-Concordance**: Developed by T. Cobb
  - ...

Limitations
- Designed for specific corpora
- Closed source (not portable)
- Not always intuitive to use

Overview of AntPConc:
A new desktop parallel concordancer

Features
- parallel (multi) concordancer
- freeware
- standalone (portable)
- Unicode (UTF-8) compliant
- simple/easy-to-use

Design
- Model-View-Controller (MVC)
- N-gram database model (similar to BYU corpora)
Overview of AntPConc:
http://www.antlab.waseda.ac.jp/software/

Model-View-Controller (MVC) Design

- Python Programming Language
- Controller (KWIC engine)
- View (Browser Interface)
- Model (Corpus Database)
- PyQt+
+ Webkit Internet Browser
+ JavaScript (+ jQuery)
+ SQLite

Case Study using AntPConc

Teaching JHS/SHS vocabulary/grammar in a Japanese University context (Chujo et al. 2013)

- Students
  - Typical Japanese university students
  - 6 years of EFL study at JHS/SHS (remedial level)
  - Low motivation
- Ability level
  - TOEIC 350/TOEIC 300 or below
- Aim
  - Improve understanding of basic vocabulary and grammar items as taught in secondary school
- Approach (DDL)
  - AntPConc + Bilingual Corpus (e.g. American Grammar Texts)

Teaching JHS/SHS vocabulary/grammar in a Japanese University context (Chujo et al. 2013)

- CoBLE Parallel Corpus Overview
  - 10,352 English sentences (66,104 words)
  - Created from sentences found in American language textbooks (grades 1-6)
  - Included with a Japanese translation
  - Saved as UTF-8 text files

Target Grammar Items

<table>
<thead>
<tr>
<th>Target Grammar Items</th>
<th>Type</th>
<th>n</th>
<th>Pretest M (SD)</th>
<th>Posttest M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>JHS (k = 62)</td>
<td>DDL</td>
<td>41</td>
<td>67.6 (12.7)</td>
<td>79.1 (12.0)</td>
</tr>
<tr>
<td></td>
<td>Non-DDL</td>
<td>39</td>
<td>26.4 (17.6)</td>
<td>39.9 (19.8)</td>
</tr>
<tr>
<td>SHS (k = 41)</td>
<td>DDL</td>
<td>88</td>
<td>65.9 (13.6)</td>
<td>78.8 (10.6)</td>
</tr>
<tr>
<td></td>
<td>Non-DDL</td>
<td>44</td>
<td>45.1 (16.1)</td>
<td>62.7 (18.5)</td>
</tr>
</tbody>
</table>

Selected student comments about the DDL approach (translated)

I was able to study at my own pace.
I was able to approach grammar learning in many different ways.
Instead of looking at textbooks/reference books, I was able to enjoy studying with a computer and AntPConc.
I didn’t feel as though I was learning grammar.
Using AntPConc, I could study efficiently.
Selected student comments about improving AntPConc (translated)

It should be made into a commercial tool.
It was really easy to use.
Adding sound effects [voiceover] would be good.
Sometimes the results too a while to appear.
Discussion
The future of parallel corpus tools in corpus linguistics

- Have you used parallel corpus tools in your research?
  - If yes, what limitations/problems did you find with the software?
  - If no, why not?
- Why do you think so few parallel corpus tools are available today (for the desktop)?
  - alignment issues?
  - lack of tool developers?
  - no need for parallel corpus tools?
- What features would you want to see in the 'perfect' parallel corpus tool?
  - automatic alignment?
  - KWIC, plots, word frequencies, collocates, n-grams, ...?

Summary

- AntPCConc is ...
  - a freeware, parallel (multi) concordancer
  - a standalone (portable) application requiring no installation
  - a Unicode (UTF-8) and Right-To-Left compliant tool
- Loading parallel corpora into AntPCConc is ....
  - simple and easy
  - requires (almost) no knowledge of character encodings
- But...
  - AntPCConc does not align corpus texts (relying on line break alignment)
  - AntPCConc currently offers only KWIC concordancing