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/

AACL 2014, Northern Arizona University, Flagstaff, US
September 25-28, 2014

Faculty of Science and Engineering, Waseda University

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

2

Background

3

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:**
  - Arabic English Parallel News, Babel English-Chinese Parallel Corpus, CEXI, COMPARA, CRATER, English-Chinese Parallel Corpus, EMILLE, English-Russian Parallel Corpus, ENPC, SPC, Europarl, Hansard French/English, Hong Kong Bilingual Corpus of Legal & Documentary Texts, ICE, IJS-ELAN corpus, INTERSECT, JOC, JRC-Acquis, KACENKA, Le corpus BAF, MLCC, OPUS, OMC, PKU Babel Chinese-English Parallel Corpus

4   (Bookmarks for Corpus-based Linguists, Routledge Handbook of Corpus Linguistics)

Background:

- **Using parallel corpora (in and beyond translation):**
  - Barlow, M. (2000); Botley, McEnery, and Wilson (2000); St. John, Elke (2001); Johanson, E. C., Kilimci, S., and Megyesi, B. (2010), Danielsson P. & Mahlberg, M. (2010), Chujo et. al. (2005-2013).
- **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
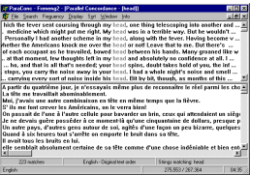 - leads to higher motivation

5   (Bernd Rüschoff, 2010)

Background:
Current offline parallel corpus tools

- **Multiconcord** (David Woolls)
  - released 2002 for Windows 3.x/95
  - pricing £40
- **ParaConc** (Athelstan)
  - released 2003 for Windows 3.x/95/98/NT/etc)
  - pricing $95
- **Limitations**
  - complex import for non-English corpora
  - designed for corpus linguists
  - single platform (Windows)
  - commercial
  - not updated

6

Slide 7

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
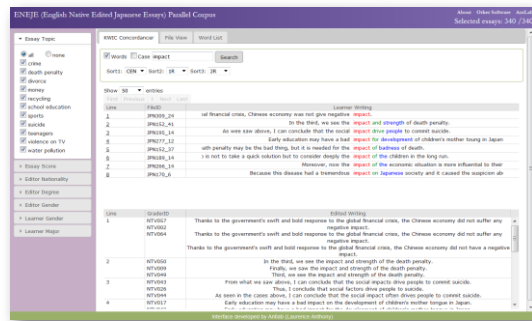 - Closed source (not portable)
  - Not always intuitive to use

7

Slide 8

Background:
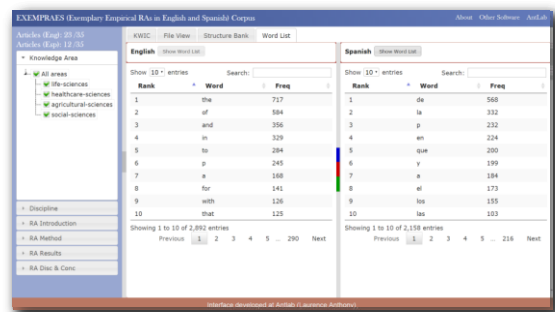Current web-based parallel corpus tools (*WebParaNews*)

8

Slide 9

Background:
Current web-based parallel corpus tools (*ENEJE Corpus*)

9

Slide 10

Background:
Current web-based parallel corpus tools (*EXEMPRAES Corpus*)

10

Slide 11

Overview of *AntPConc*
A new desktop parallel concordancer

11

Slide 12

Overview of *AntPConc*:
http://www.antlab.waseda.ac.jp/software/

- **Features**
  - parallel (multi) concordance
  - freeware
  - standalone (portable)
  - Unicode (UTF-8) compliant
  - simple/easy-to-use
- **Design**
  - Model-View-Controller (MVC)
  - N-gram database model (similar to BYU corpora)

12

Overview of *AntPConc*:
http://www.antlab.waseda.ac.jp/software/
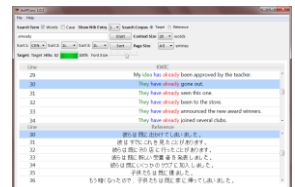
Model-View-Controller (MVC) Design

Python Programing Language

Controller
(KWIC engine)

A freeware parallel
Windows用フリー
ウェアの並列コン
コーダンスプログ
ラム。Windows用
フリーウェアの並
列コンコーダンス
プログラム。

View
(Browser Interface)

Model
(Corpus Database)

PyQt +
+ Webkit Internet Browser
+ JavaScript (+ JQuery)

SQLite

13

Case Study using *AntPConc*

14

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)

15

Nihon University, Japan (2012)

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

The girl looks.	その 少女 は 見ます 。
Two fish swim.	二匹 の 魚 が 泳ぐ 。
A bird sings.	鳥 が 歌う 。
The rabbits eat.	ウサギ は 食べます 。
The baby sleeps.	赤ちゃん は 眠る 。
Rusty stays home.	Rusty は 家 に いる 。
The boys walk.	少年 たち は 歩く 。
Mom waves.	母 は 手 を 振る 。
A cat looks out.	猫 が 外 を 見る 。
Birds eat food.	鳥 は 餌 を 食べます 。
The cow moos.	牛 は モー と 鳴く 。
Some ducks run.	何 羽 か の カモ が 走る 。
The dog barks.	犬 が ほえる 。
The kitten jumps.	子 ネコ が ジャンプ します 。

16

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

Target Grammar Items	Type	n	Pretest		Posttest	
			M (SD)	α	M (SD)	α
JHS (k = 62)	DDL	41	67.6 (12.7)	.96	79.1 (12.0)	.96
	Non-DDL	39	26.4 (17.6)		39.9 (19.8)	
SHS (k = 41)	DDL	88	65.9 (13.6)	.77	78.8 (10.6)	.72
	Non-DDL	44	45.1 (16.1)		62.7 (18.5)	

17

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

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.

18

Discussion and Summary

19

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, … ?

20

Summary

- *AntPConc* 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 *AntPConc* is ….
  - simple and easy
  - requires (almost) no knowledge of character encodings
- But…
  - *AntPConc* does not align corpus texts (relying on line break alignment)
  - *AntPConc* currently offers only KWIC concordancing

21