A False Friends Exercise with Authentic Material Retrieved from a Corpus

Joachim Wagner, National Centre for Language Technology, School of Computing,
Dublin City University, Dublin 9, Ireland, jwagner@computing.dcu.ie

The Exercise

- Sub-field of vocabulary learning
  - Important for high proficiency in L2
  - Often disregarded
- Languages
  - L2: German
  - L1: related to German, at least sharing Latin character set
  - Instructions: English
- Application of NLP
- Use of corpora

Aim of the Exercise

- Raise awareness
  - High number of Cognates
  - Various levels of meaning correspondence
- Do not focus on notorious false friends

What are False Friends?

- Cognates with (important) differences in meaning
  - Pairs of words in two languages
  - Look or sound alike
  - Caveat: term "cognate" is only used etymologically in linguistics
- Misleading and causing errors

English	German
gift	Gift (poison)
brave	brav (honest)
prospect	Prospekt (brochure)
sensible	sensibel (sensitive)

Table 1: Some false friends

Task Description

Step 1 – Identification of Cognates

- Authentic texts
  - 197 book excerpts downloaded from a review website
  - Length restrictions
- Workload reduced by restriction to relevant part-of-speech (POS)

Blue words: available for selection
Pink words: selected
Right side: learner's annotation of English word associations

Step 2 – Classification

- Three classes of meaning correspondence
  - Similar meaning
  - Different meaning
  - Partly different meaning

Step 3 – View Results

Introduction Page

- Each step explained
  - Summary
  - Detailed breakdown
- Screenshots

Technology

- POS Tagger (IMS, Schmid 1994)
- String similarity measure (McEnery and Oakes 1996)
- CSS, JavaScript & dynamic HTML (Client GUI)
- CGI for substitution of texts and evaluation
- Static web-server / CD-ROM sufficient for exercise

Preliminary Evaluation

- Six subjects
- 20-minute aural introduction
- Exercise + questionnaire

Results

- Classification task is difficult
- Subjects remember either nothing or everything of the content of the text (depending on the time spent)
- Browser problems
- Subjects missed feedback

Future Work

### Ranking of texts for (semi-)automatic selection of suitable texts

- Density of cognates
- Lexical difficulty level (Wible et al. 2002)
- Stylistic features

Adaptation to other Languages

- ISO Latin 1 character set
- L1: word list
- L2: corpus and POS tagger

Integration of an Artificial Co-Learner

- Exploit inherent limitations of NLP to our advantage
- The learner "teaches" the artificial co-learner when it makes errors
- Improve both human's and computer's L2 knowledge
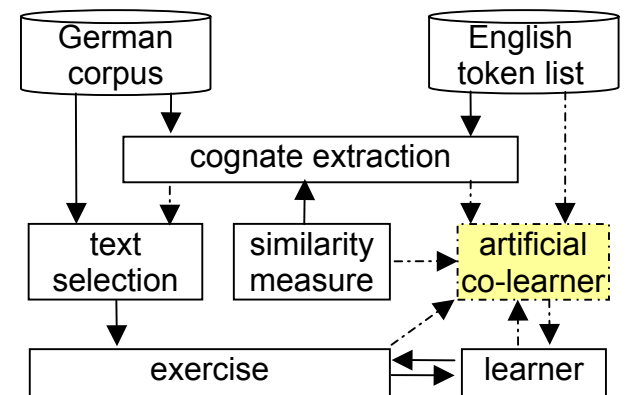
Improve Evaluation Methodology

- Assess increase of awareness
- What kind of feedback is expected?
- Impact of the text being used

References

B. Dretzke and M. I. Nester. 2001. *Student's Guide to false friends, old friends and new friends.* 3-464-037109-3 Cornelsen Verlag, Berlin

A. Erpenbeck, B. Koch, N. Kummer, P. Reuter, P. Tschorn and J. Wagner. 2002. KoKS – Korpusbasierte Kollokationssuche. Technical Report, Universität Osnabrück, Germany

C. Greene, K. Keogh, T. Koller, J. Wagner, M. Ward and J. Van Genabith. 2004. Using NLP Technology in CALL. *Proceedings of InSTIL / ICALL 2004,* Venice

R. Humphrey. 2003. *False Friends, Falser Friends, Flasest Friends.* 3-12-939613-6 Ernst Klett Verlag, Stuttgart, Germany

A. McEnery and M.P. Oakes. 1996. Sentence and Word Alignment in the CRATER Project: Methods and Assessment. In J. Thomas and M. Short (eds) Using Corpora for Language Research, Longman, pp 211-231

H. Schmid. 1994. *Probabilistic Part-of-Speech Tagging using Decision Trees.* Revised online version, http://www.ims.uni-stuttgart.de/ftp/pub/corpora/tree-tagger1.pdf

D. Wible, C. Kuo, F. Chien, C.C. Wang. 2002. A Lexical Difficulty Filter for Language Learners. In: B. Kettemann and G. Marko: Teaching and Learning by Doing Corpus Analysis. Editions Rodopi B.V.