

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 down-
 - loaded from a review website
 - Length restrictions
- Workload reduced by restriction to relevant part-of-speech (POS)

S Exercise - Konqueror	? _ 🗆 ×
Location Edit View Go Bookmarks Tools Settings Window Help	
]① ⊖ ∋ ⊉ © ⊗ ⊀ ʰ ℩ ≿ 웒 않 ♀ ♀ ≙ ⊙	÷
Lgcation: http://136.206.19.113/~jwagner/cgi-bin/ex/exercise.py	

Step 2 – Classification

- Three classes of meaning correspondence
 - Similar meaning
 - Different meaning 0
 - Partly different meaning 0

cation <u>E</u> dit <u>V</u> iew <u>Go</u> Boo	okmarks <u>T</u> ools <u>S</u> ettings <u>W</u> inc	ow <u>H</u> e	:lp						
Des <mark>u c</mark>		6	6	-	۲ ۰ ا	۲			
Location: http://136.20	6.19.113/~jwagner/cgi-bin/ex/ex/	ercise.p	у					•	
Navigation	Classification								
Introduction	Word	DK	s	D	Р	Class	Description		
Description	nun (nun)	0	0	۲	0	DK	Don't know		
Description	fast (fast)	0	\odot	۲	\odot	S	Same meaning		
1. Text	her (her)	0	0	۲	0	D	Different meaning		
2. Classification	kam (came)	0	۲	۲	\odot	P	Partly different meaning		
3. mesuits	vor (for)	0	0	۲	0	Outdalling			
Exercise	Tages (tags)	0	0	۲	\odot	Guide	Guideline		
1 Taut	drang (drank)	0	0	۲	0	• 0	decide for one word to co	mpare with	
2. Classification	sprang (sprang)	0	۲	۲	\odot	• f	 focus on important meanings ignore parts of speech ignore grammatical features (tense, number etc.) 		
3. Results	hastete (hasten)	\odot	۲	\odot	0	•			
	Treppe (trip)	0	0	۲	\odot	•			
What Next?	oben (oven)	0	0	۲	0		lambor oto.,		
Notes	Haus (house)	0	۲	۲	\odot				
	Kindergarten (kindergarten)	0	•	0	0				
	zu (zoo)	0	0	۲	0				
	< previous panel				_			next panel	
								÷	

Step 3 – View Results

2 Exercise - Konqueror									
Location Edit View Go Bookmarks Tools Settings Window Help									
Location:136.2C	Location: http://136.206.19.113/-jwagner/cgrbn/ex/exercise.py								
Navigation	Results								
Introduction	don't know	same meaning	different meaning	partly different meaning					
I. Text Classification Results Exercise I. Text Classification S. Classification S. (Results)		kam (came) sprang (sprang) hastete (hasten) Haus (house) Kindergarten (kindergarten)	 nun (nun) fast (fast) her (her) vor (for) Tages (tags) drang (drank) Treppe (trip) oben (oven) zu (zoo) 						
What Next?	Words you want to remember:								
Notes									
Open a result page in another window for printing:									

Introduction Page

- Each step explained
 - o Summary
 - Detailed breakdown
- Screenshots

Technology

- POS Tagger (IMS, Schmid 1994)
- String similarity measure ٠ (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

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 comput-• er'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,

	Navigation	Text	Annotation
		Es war <mark>nun</mark> schon fast zehn Jahre her und kam mir wie gestern vor,	nun
	Introduction	daß eines Tages Lauras leises Wimmern, das schnell in	nun
Der	Description	hemmungsloses Schluchzen überging, aus dem Kinderzimmer <mark>drang</mark> .	fast
	Description	Entsetzt sprang ich vom Schreibtisch auf und hastete über die Treppe	fast
	1. Text	nach oben. Stacy war nicht im Haus, Miriam im Kindergarten, und ich	her
	2. Classification	hatte hoch und heilig versprochen, mich um Laura zu kümmern. Sollte	her
	3. Results	ich etwa meine väterliche Fürsorge- und Aufsichtspflicht verletzt	kam
		haben? Wälzte sich die Vierjährige, nachdem sie gefallen war, sich	came
	Exercise	gestoßen, sich womöglich mit einer Schere oder dergleichen	vor
	1 Tout	verstümmelt hatte, bereits in ihrem unschuldigen Blut? Sie wälzte	for
	2 Classification	sich, in der Tat. Sie walzte sich auf dem Teppich vor dem	Tages
	3 Reculte	uberquellenden Kleiderschrank der Kinder und heulte Hotz und	tags
	o. mosuits	wasser. Eine verletzung oder gar Blut waren zum Gluck nicht	drang
	What Next?	auszumachen. Ich nahm sie in die Arme, versuchte sie zu berunigen,	drank
	AL	brüllenden Kummers. Die Antwort bervorgestoßen unter einem	sprang
	Notes	erneuten Heulschwall, brachte mir schlagartig eine unausweichliche	sprang
		Dimension meiner Zukunft als Vater zweier Töchter zu Bewußtsein.	hastete
		Sie lautete nämlich : Ich weiß nicht, was ich anziehen soll! Diese	hasten
		herzzerreißende Szene mit dem unvergeßlichen, seither S.	Treppe
		208-209	trip
			oben
		Excerpt from	oven
		http://home.t-online.de/home/booKinist/text/@movieru.htm	Haus

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

word associations

Preliminary Evaluation

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

Results

- Classification task is difficult
- Subjects remember either nothing ٠ or everything of the content of the text (depending on the time spent)
- Browser problems
- Subjects missed feedback

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.