

EQTIS (<http://research.flw.ugent.be/en/eqtis>)

Lore Vandevoorde & Els Lefever

# PROFESSIONAL TRANSLATORS AS BILINGUALS: THE USE OF COGNATES AND NON-COGNATES

## INTRODUCTION

The field of Cognitive Translation Studies shares a number of research interests with bilingualism (Shreve, 2012). Cognates and non-cognates are a case in point. In Translation Studies, Shlesinger & Malkiel (2005) and Malkiel (2009) concluded that translators tend to choose a non-cognate translation over a cognate translation when both are (presumably) equally translationally equivalent (Kusmaul & Tirkkonen-Condit 1995). Evidence from psycholinguistic studies however suggests – based on the overwhelming evidence of the cognate facilitation effect (Schepens et al. 2012: 157-158) – that translators, when confronted with the possibility of translating an L2 source word by both an L1 non-cognate target form and an L1 cognate target form, would be more likely to choose the cognate form over the non-cognate equivalent (if both forms are equally frequent).

## RESEARCH QUESTIONS

- Do texts translated by professional translators contain more or fewer cognates than
  - non-translated texts,
  - texts produced by second language learners,
  - texts produced by novice translators?
- Can results from psycholinguistic research be generalized over *all* types of bilinguals?

How do professional translators deal with cognates compared to natural bilinguals/novice translators, taking into account:

- specific training of professional translators
- translation expertise (metacognition) ?

## METHODOLOGY (RESEARCH QUESTION 1)

**STEP 1: Establish an (as exhaustive as possible) list of cognate pairs for DUTCH-ENGLISH and DUTCH-FRENCH**

### SELECT TRANSLATION EQUIVALENT PAIRS FOR DUTCH-ENGLISH AND DUTCH-FRENCH

- Method:** Unsupervised statistical word alignment using GIZA++ (Och & Ney 2003)
- Corpus:** Dutch Parallel Corpus (Macken et al. 2011)
  - 10 million words
  - parallel and comparable
  - Dutch ↔ French and Dutch ↔ English
  - variety of textual registers

### Example:

SOURCE: anders nog waren de oem overeenkomsten met  
 verschillende partners in de grafische industrie  
 TARGET: another avenue was the oem agreements with other  
 partners in the graphics industry  
 0-0 anders another  
 0-1 anders avenue  
 1-1 nog avenue  
 2-2 waren was  
 3-3 de the  
 4-4 oem oem  
 5-5 overeenkomsten agreements  
 6-6 met with  
 7-7 verschillende other  
 8-8 partners partners  
 9-9 in in  
 10-10 de the  
 11-11 grafische graphics  
 12-12 industrie industry

### EXTRACT COGNATE PAIRS FOR DUTCH-ENGLISH AND DUTCH-FRENCH

- Apply measure(s) of orthographic (and phonetic) similarity to the list of candidate translation equivalents
- ORTHOGRAPHIC: Normalized Levenshtein Distance (NLD) calculated with R (Gries 2004)**

## RESULTS

- 14231 DU-EN candidate cognate pairs with NLD ≥ 0.5
- 11842 DU-FR candidate cognate pairs with NLD ≥ 0.5

- Example of raw output of DU-EN cognate pairs:

NL	EN	NLD (>=0.5)
speciaal	especially	0.625
speciaal	special	0.875
speciaal	specially	0.75
speciaalbieren	specialties	0.642857142857143
speciale	specialist	0.625
speciale	special	0.875
speciale	specific	0.625
specialiseerde	specialized	0.714285714285714
specialiseert	specialising	0.692307692307692
specialiseren	specialize	0.692307692307692
specialiseren	specialise	0.769230769230769
specialisering	specialisation	0.714285714285714
specialiste	specialist	0.909090909090909
specialisten	specialists	0.833333333333333
specialiteit	specialty	0.666666666666667
specialiteiten	specialities	0.785714285714286

- Lists can be used to calculate cognate ratios (→ STEP 2)
- Lists can possibly extend existing list (Schepens et al. 2013)
- BUT lists need manual validation to:**
  - exclude errors (alignment errors + homographs)
  - exclude proper names

### References

Brybaert, Marc, Michaël Stevens, Paweł Mandera, and Emmanuel Keuleers. "The Impact of Word Prevalence on Lexical Decision Times: Evidence from the Dutch Lexicon Project 2." *Journal of Experimental Psychology: Human Perception and Performance* 42, no. 3 (2016): 441–58.

Gries, Stefan Th. "Shouldn't It Be Breakfunch? A Quantitative Analysis of Blend Structure in English." *Linguistics*, 2004, 639–668.

Kusmaul, Paul, and Sonja Tirkkonen-Condit. "Think-Aloud Protocol Analysis in Translation Studies." *TTR* 8, no. 1 (1995): 177–99.

Malkiel, Brenda. "A Translation Process. Evidence from Cognates." *Babel* 55, no. 3 (2009): 228–43.

Och, F.J., and H. Ney. "A Systematic Comparison of Various Statistical Alignment Models." *Computational Linguistics* 29, no. 1 (2003): 19–51.

Schepens, Job, Ton Dijkstra, and Franc Grootjen. "Distributions of Cognates in Europe as Based on Levenshtein Distance." *Bilingualism: Language and Cognition* 15, no. 1 (January 2012): 157–66. doi:10.1017/S1366728910000623.

Schepens, Job, Ton Dijkstra, Franc Grootjen, and Walter J. B. van Heuven. "Cross-Language Distributions of High Frequency and Phonetically Similar Cognates." Edited by Kevin Paterson. *PLoS ONE* 8, no. 5 (May 10, 2013): e63006. doi:10.1371/journal.pone.0063006.

Shlesinger, Miriam, and Brenda Malkiel. "Comparing Modalities: Cognates as a Case in Point." *Across Languages and Cultures* 6, no. 2 (2005): 173–93.

The authors want to thank Stefan Th. Gries for providing us with his script to calculate Levenshtein Distances with R.

## STEP 2: Establish cognate ratios for corpora of translated texts and non-translated texts

	OrigDutch	TransDutchFR	TransDutchEN
corpus size (tokens in DPC)	4911944	2076443	2539248
DU-FR cognate types (Schepens et al 2013)	559	559	559
DU-EN cognate types (Schepens et al 2013)	1104	1104	1104
DU-FR cognates word tokens (based on types Schepens et al 2013)	78669	51677	37758
DU-FR cognates word tokens (based on types Schepens et al 2013)	226900	147840	105596
DU-FR cognate lemma tokens (based on types Schepens et al 2013)	104392	68459	50833
DU-EN cognate lemma tokens (based on types Schepens et al 2013)	275509	178458	127486
cognate ratio DU-FR (cognate-type/total token ratio) (based on types Schepens et al 2013)	0,0114	0,0269	0,0220
cognate ratio DU-EN (cognate-type/total token ratio) (based on types Schepens et al 2013)	0,0225	0,0532	0,0435
cognate wordtoken ratio DU-FR (cognate-tokens/total token ratio) (based on types Schepens et al 2013)	0,0160	0,0249	0,0149
cognate wordtoken ratio DU-EN (cognate-tokens/total token ratio) (based on types Schepens et al 2013)	0,0462	0,0712	0,0416
cognate lemmatoken ratio DU-FR (cognate-tokens/total token ratio) (based on types Schepens et al 2013)	0,0213	0,0330	0,0200
cognate lemmatoken ratio DU-EN (cognate-tokens/total token ratio) (based on types Schepens et al 2013)	0,0561	0,0859	0,0502

## FUTURE RESEARCH (RESEARCH QUESTION 2)

Experimental set-up with a translation task

### PARTICIPANTS

- student translators
- professional translators
- native bilinguals

### EXPERIMENTAL TASK

- translation of French/English text into Dutch (mother tongue)
- Text contains a selection of *n* cognates based on exhaustive list of corpus study (→ research question 1)
- Experimental items are French or English words which have both a cognate and a non-cognate translation equivalent in Dutch

### HYPOTHESES

- The higher the cognate ratio for two languages, the more cognate translations in the target text
- The more time pressure, the more cognate translations in the target text
- The less experienced a translator is, the more cognate translations he produces
- The more frequent and prevalent (Brybaert et al. 2016) a cognate translation is, the more frequently it is produced in the target text

### Contact

lore.vandevoorde@ugent.be  
<http://research.flw.ugent.be/en/lore.vandevoorde>

 Universiteit Gent

 @ugent

 Ghent University