

# Intra-language differences and translation quality assessment

Rudy Loock

Université de Lille 3 & UMR STL 8163 du CNRS

**STARTING POINT**

- Corpus-Based Translation Studies (CBTS) project (2011-):



- Researchers from the Universities of Lille, Limoges, Ghent, Helsinki

Bert Cappelle, University of Lille 3, STL CNRS

Gert De Sutter, University of Ghent

Cindy Lefebvre-Scodeller, University of Limoges

Rudy Loock, University of Lille 3, STL CNRS

Michaël Mariaule, University of Lille 3, CECILLE

Carmen Nuñez-Lagos, University of Lille 3, STL CNRS

Koen Plevoets, University of Ghent

Lea Huotari, University of Helsinki



- Quantitative approach of inter- and intra-language differences:
  - Automatic exploitation of comparable and parallel corpora to investigate **cross-linguistic differences** (e.g. between English and French) and **differences between original and translated language** (e.g. original vs. translated French)

Cappelle & Loock (2013), Loock (2013): existential constructions in original/translated English and French

Cappelle (2013): phrasal verbs in original vs. translated English

Huotari (2013): reference to parts of the human body in Finnish vs. English

Loock (2013): derived adverbs in original/translated English and French

De Sutter, Cappelle & Loock (in progress): impersonal passive in French and Dutch

...

- Quantitative approach of inter- and intra-language differences:

- Pedagogical aims:

1. Use of inter-language differences to define usage constraints

Higher frequency of existential constructions in English than in French

=> influence on translation behaviour (respect target-language norms for idiomaticity, naturalness)

2. Observation of intra-language differences in connection with TQA

Should there be linguistic homogenisation between translated and original language?

Does any under-/over-representation of a specific linguistic feature mean lesser quality?



# **HYPOTHESIS**

- Many studies have shown that the third code (Frawley 1984) is a reality:

Baker & Olohan (2000) on the ellipsis of *that* after reporting verbs in English

Olohan (2003) on contractions in English

Jiménez-Crespo (2010) on the presence/absence of the syntactic subject in Spanish

Laviosa (1996, 1997, 1998, 2002) on lexical variety and density in English

Xiao (2010) on lexical variety and density in Chinese

- But disagreement on interpretation:



- **Translation Universals (TUs)** (e.g. Baker 1993, 1995, 1996, Tirkkonen-Condit 2002, Huotari 2013)

- ⇒ Translation-inherent

- ⇒ Differences = natural phenomenon, should not be avoided

- **Source Language Interference (SLI)** (e.g. Rabadán et al. 2009, Loock et al. 2013)

- ⇒ Language-specific

- ⇒ ‘Translationese’, Possible link with translation quality



- If second interpretation:

- Can we say that original and translated language should show the same linguistic (lexical+grammatical) characteristics?
- Should there be linguistic homogeneisation between translated and original texts?

« *The smaller the disparity between native and translated usage in the use of particular grammatical structures associated with specific meanings, the higher the translation rates for quality.* » (Rabadán et al. 2009 : 323)



# **PILOT STUDY**

- Case study on *-ly* adverbs in English vs. *-ment* adverbs in French:
  - *properly, honestly, quickly, successfully* vs. *proprement, honêtement, rapidement, familialement*
  - generally presented as cases of **translational equivalence** (if we put aside morphological and semantic constraints)
  - except that *-ment* adverbs are discarded because of their length and poor style (e.g. Chuquet & Paillard 1987: 154-155)

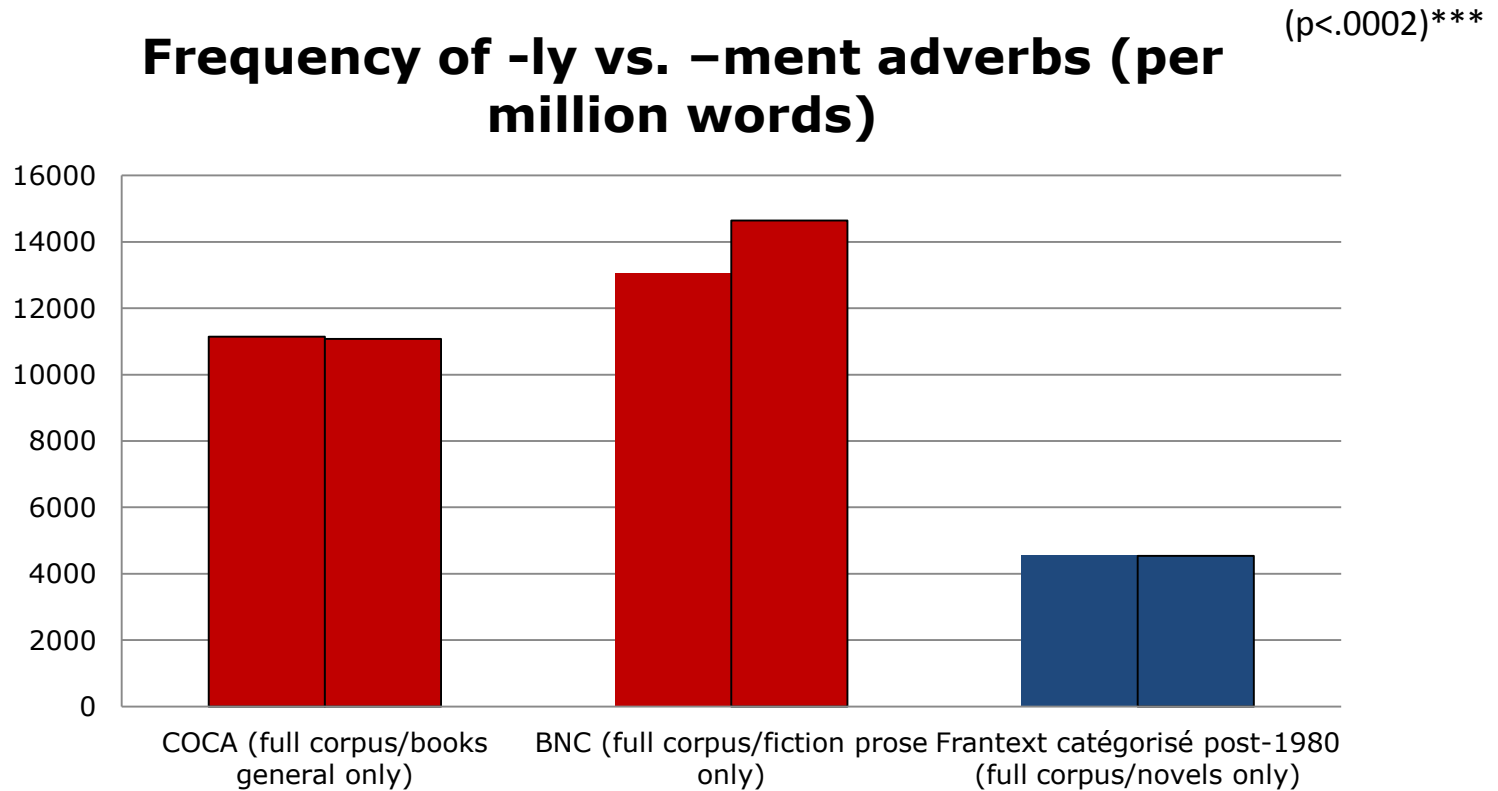
- **Hypothesis to be tested:**

any under-/over- representation of derived adverbs in translated language as opposed to original language = lesser quality

- **Methodology :**

- **Phase 1** : Observation of inter-language differences between English and French (comparable corpus of original language)
- **Phase 2** : Analysis of a learner corpus of students' translation tasks (EN>FR) that were previously and independently evaluated

- Phase 1: inter-language differences



- Phase 2 : Analysis of a learner corpus (i)
  - 1. Translations of short texts (1st- and 4th-year students)
    - English Department, University of Lille3
    - 2 in-class exams for 59 1st-year students, 32 4th-year students (2012)
    - Papers graded independently by a colleague (M. Mariaule, grade /20)
    - Presence of 3 –ly adverbs in each text
- ⇒ Specific analysis on the possible correlation between the way a student deals with a –/y adverb and the overall quality of their translation task

Text 1 : 1st-year students, 143 words

**Surely** the whole household couldn't have packed up and gone away, without Madeline even telling me about it.

I rang the bell again, long and **insistently**.

When you are in the country, and a dog barks in the middle of the night, it **merely** punctuates and emphasizes the silence.

(Jonathan Coe, *The Dwarves of Death*, 1990)

Text 2: Master's students, 282 words

The first few times I saw her after the conversation under the eaves, I was ready for at least a bit of huffiness, but no, she was **completely** civil, if a little flat.

The trouble was, because none of this had **actually** been talked about in the open, I couldn't find a way of bringing it all up with her.

As far as I remember, this suggestion wasn't taken up; maybe **hardly** anyone heard it, because a lot of us were talking all at once.

(Kazuo Ishiguro, *Never Let Me Go*, 2005)



SURELY			COMPLETELY		
Translation	Number of students	Average grade (/20)	Translation	Number of students	Average grade (/20)
-ment adverb (e.g. <i>sûrement, certainement</i> )	25	8.62	-ment adverb (e.g. <i>complètement, totalement, parfaitement</i> )	17	9.05
Impersonal structure (e.g. <i>il était certain que...</i> )	13	9.23	Adverbial expression (e.g. <i>tout à fait</i> )	9	10.33
Adverbial expression (e.g. <i>bien sûr</i> )	4	10.62	Adjective (e.g. <i>irréprochable, exemplaire</i> )	3	11.33
Personal structure (e.g. <i>j'étais certain que</i> )	3	13.16			

N.B.: semantically problematic translations omitted

INSISTENTLY			MERELY			ACTUALLY			HARDLY		
Translation	Number of students	Average grade (/20)	Translation	Number of students	Average grade (/20)	Translation	Number of students	Average grade (/20)	Translation	Number of students	Average grade (/20)
Prepositional Phrase (e.g. <i>avec insistance</i> )	37	9.675	Impersonal structures (e.g. <i>il/cela ne faisait que...</i> )	13	12.31	Adverbial expression (e.g. <i>en fait, en fin de compte</i> )	13	10	Adverb (e.g. <i>presque</i> )	11	11.36
Gerund (e.g. <i>en insistant</i> )	13	7.5	-ment adverb (e.g. <i>seulement</i> )	1	9.5	-ment adverb (e.g. <i>quasiment, pratiquement</i> )	9	11.55	-ment adverbs (e.g. <i>quasiment</i> )	6	10.5
Structure personnelle (e.g. <i>j'insistai</i> )	3	10.2							Adverbial expression (e.g. <i>à peine</i> )	3	8

- First conclusions :

- Translating a –ly adverb with a –ment adverb remains the most frequent translation (52% for 1st-year students, 55% for Master’s students)
- For *surely/completely*, there seems to be a correlation between non-literal translation and overall quality
- Mixed results for the other adverbs

- Phase 2 : Analysis of a learner corpus (ii)

- 2. Translations of longer texts

- English Department, University of Lille 3
    - 1st year of the MéLexTra Master's programme (selected students)
    - Translation of a short story/chapter from a novel from English to French
    - 17 translation tasks evaluated independently (2010; 2012) : 3-group classification by C. Oster
    - Comparison between the frequency of –ly adverbs in STs and frequency of –ment adverbs in TTs

⇒ Is there a correlation between the frequency of –ment adverbs and translation quality?

## 17 translation tasks:

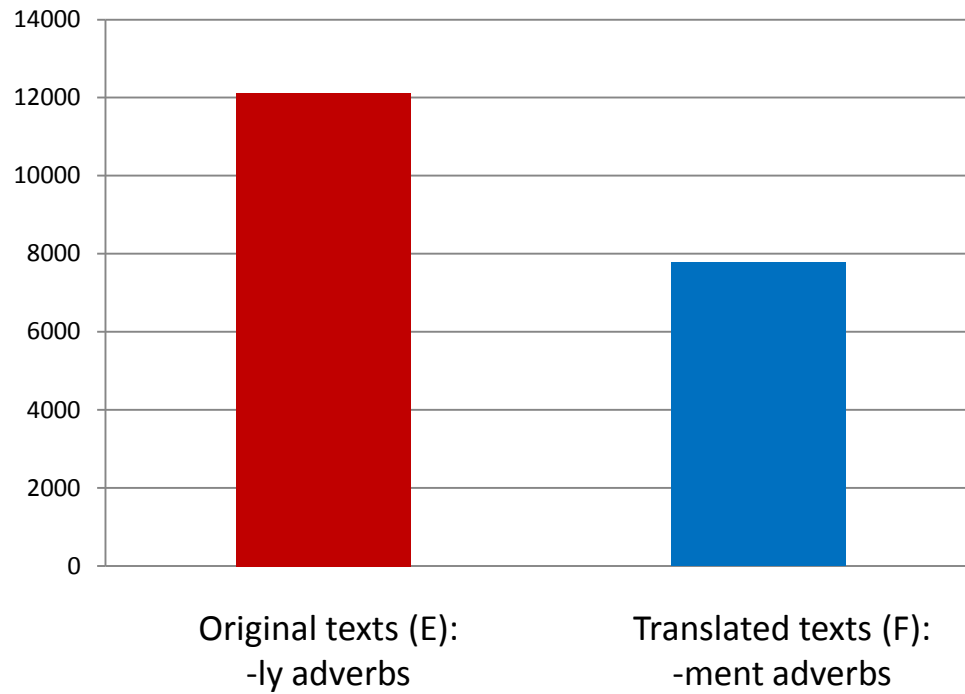
	Original English	Translated French
Minimum	7,448	8,458
Maximum	19,639	22,613
Average	11,155	12,801
Total number of words	<b>189,642</b>	<b>217,618</b>

Number of words in the corpus

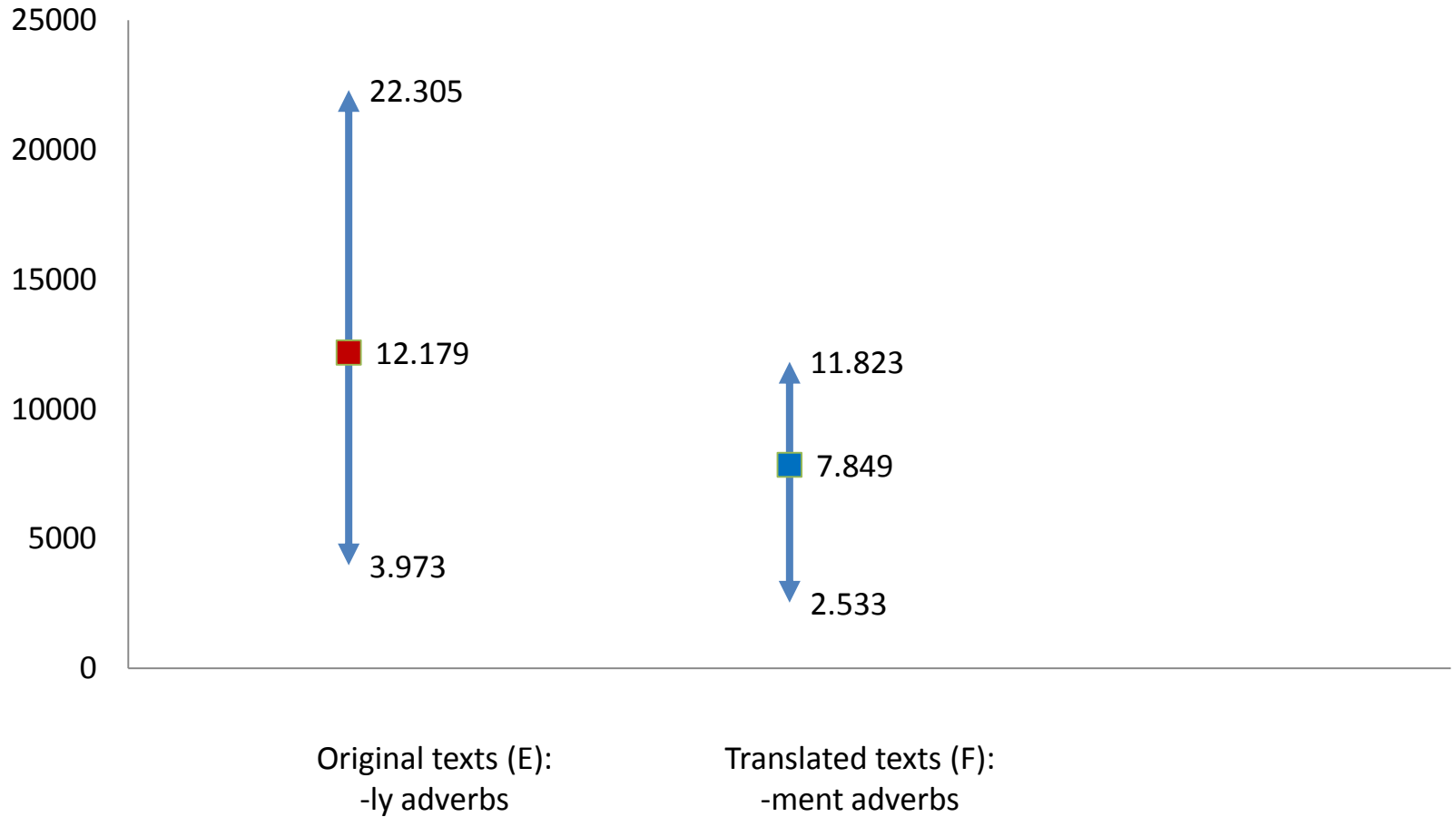
## 3 groups (independent classification by C. Oster):

A (good translation)	B (correct translation)	C (translation that can be improved)
4	9	4

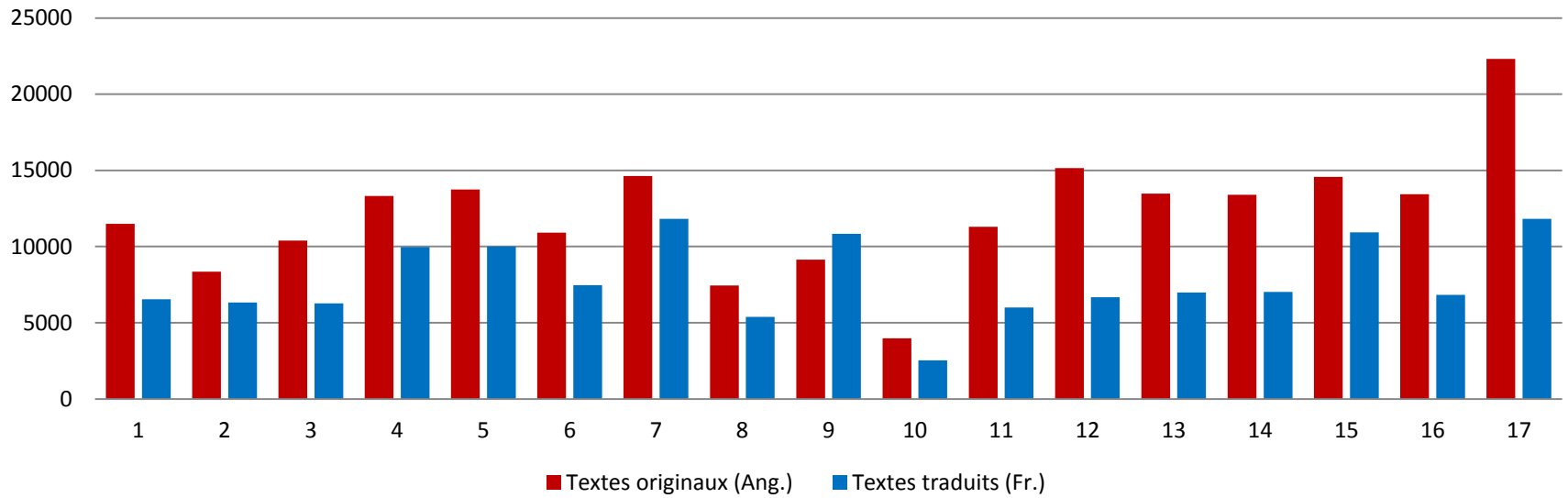
- Overall results



- Detailed results

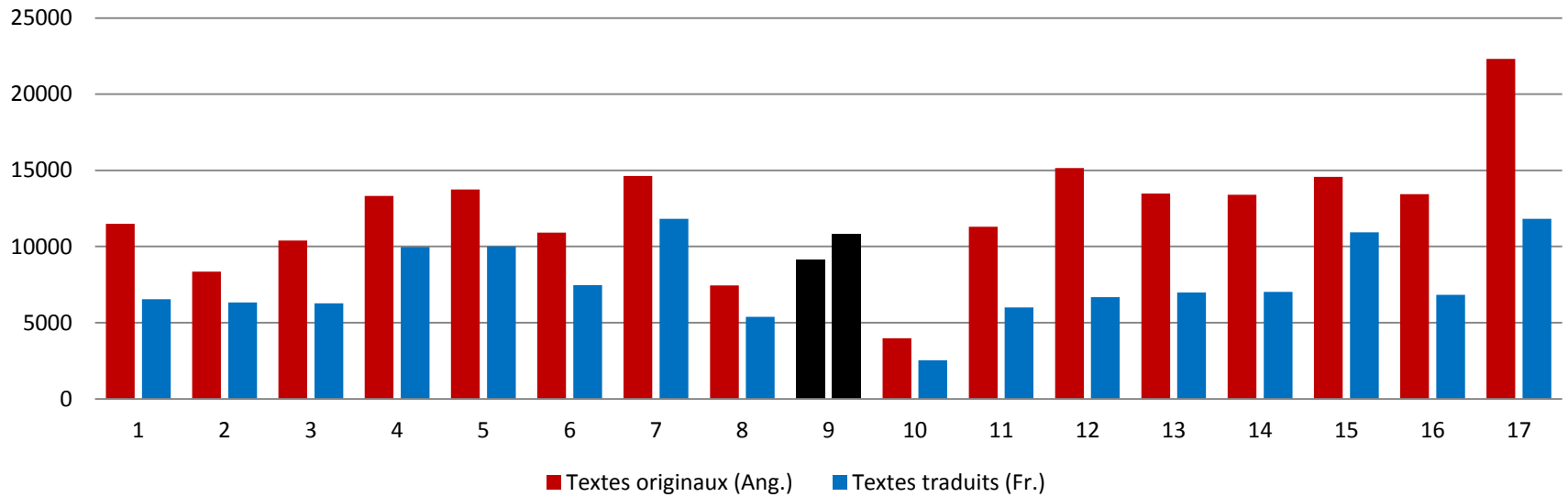


- Individual results

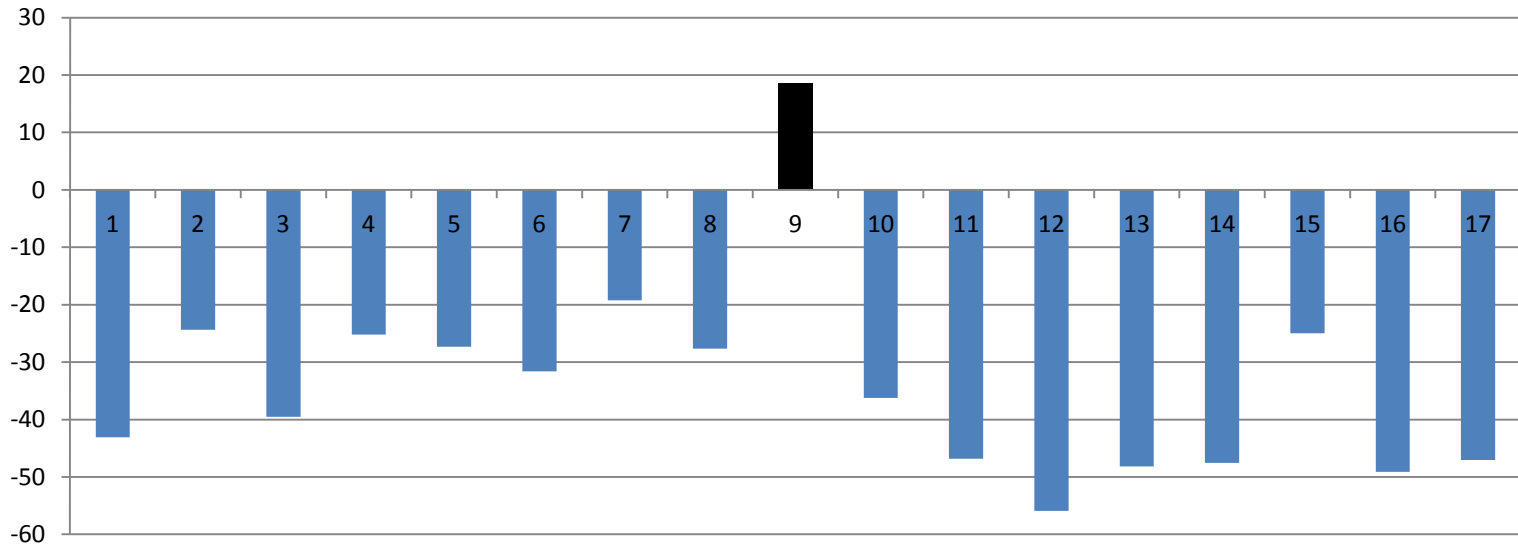




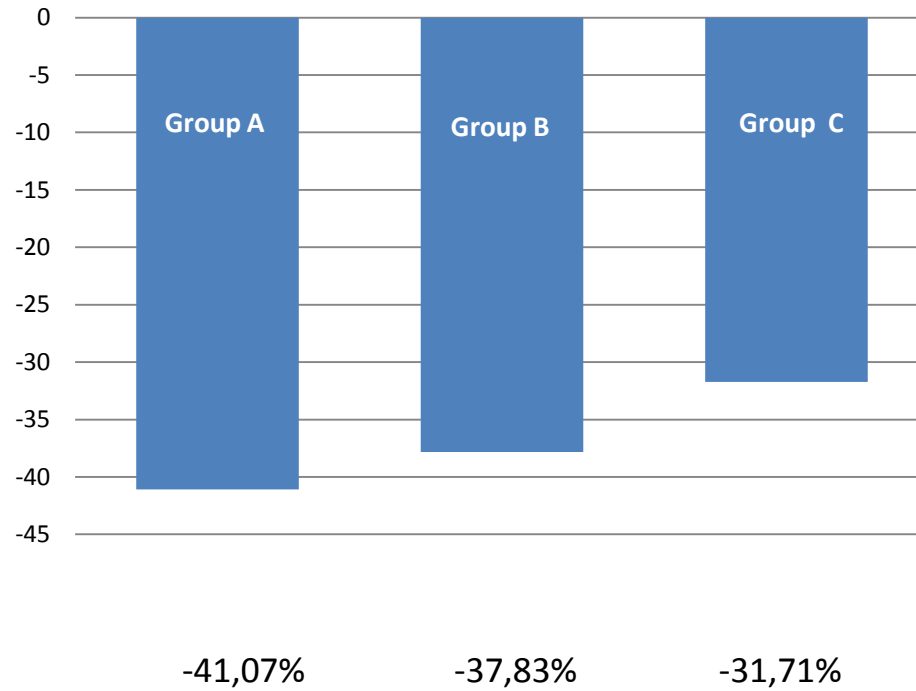
- Individual results



- Individual results (differential)



- Individual results (differential)



- So?...

The lower the use of *-ment* adverbs (compared to the number of *-ly* adverbs in the STs) in the student's translation tasks, the better the overall quality of the translation?

Can the frequency of *-ment* adverb in a translation as opposed to the frequency of *-ly* adverb in the original text be symptomatic of translation quality?

- Sounds nice...

But...

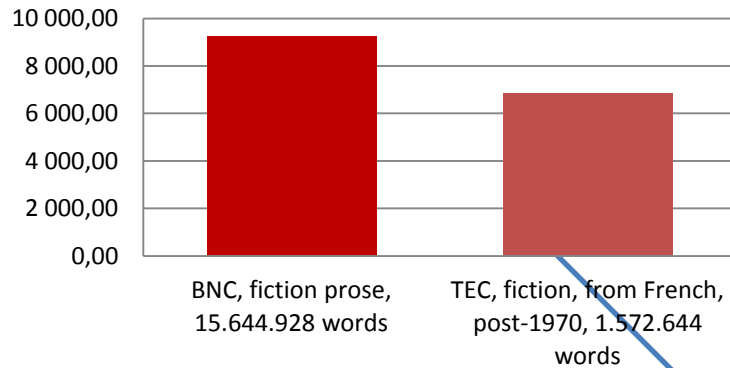
Group A	Group B	Group C
-31,581352	-43,0936766	-24,3457642
<b>-27,6685392</b>	-39,5467903	-19,2122544
-55,917241	-25,2110374	-36,2510612
-49,1199462	-27,2993623	<b>-47,0507374</b>
	-46,8014748	
	-48,1846434	
	-47,525997	
	-24,9691552	
<b>-41,0717696</b>	<b>-37,8290171</b>	<b>-31,7149543</b>



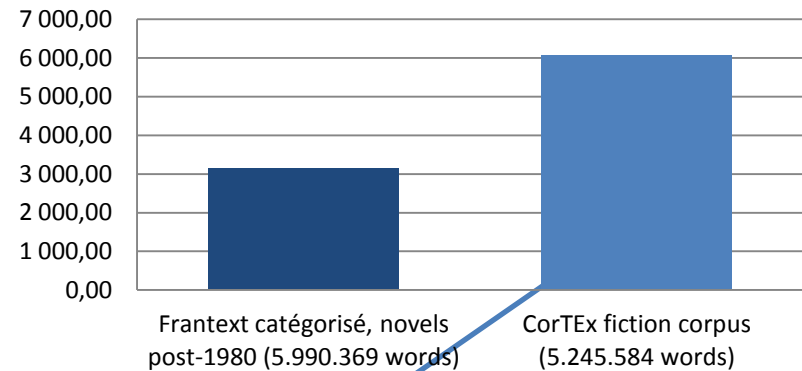
- Sounds nice...

And also...

Intra-language difference Top 100 -ly adverbs (pmw)



Intra-language difference Top 100 -ment adverbs (pmw)



Bad translations???



**WHAT NOW?**

- Study of other linguistic translationese-prone phenomena for EN>FR translations:

- Existential constructions (there is/are vs. il y a)
- -ing forms
- The passive voice
- Reporting verbs
- ...

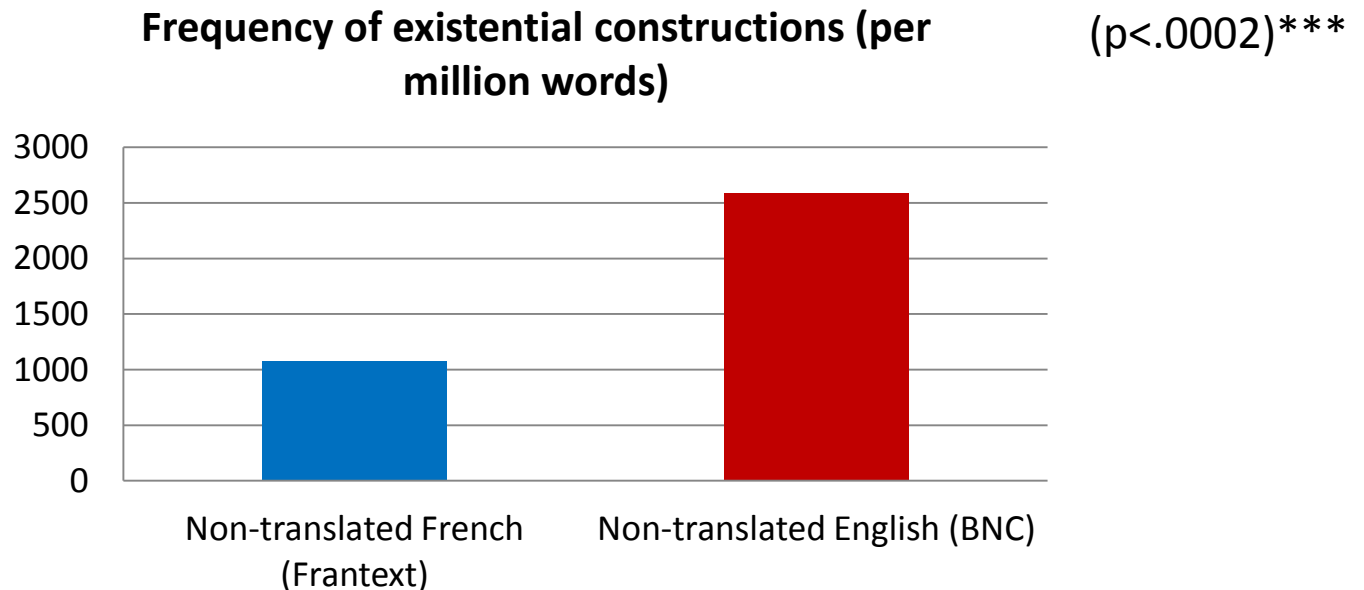
} = significant differences in language use

– Can we define a **check-list** of linguistic phenomena to observe to evaluate translation quality???

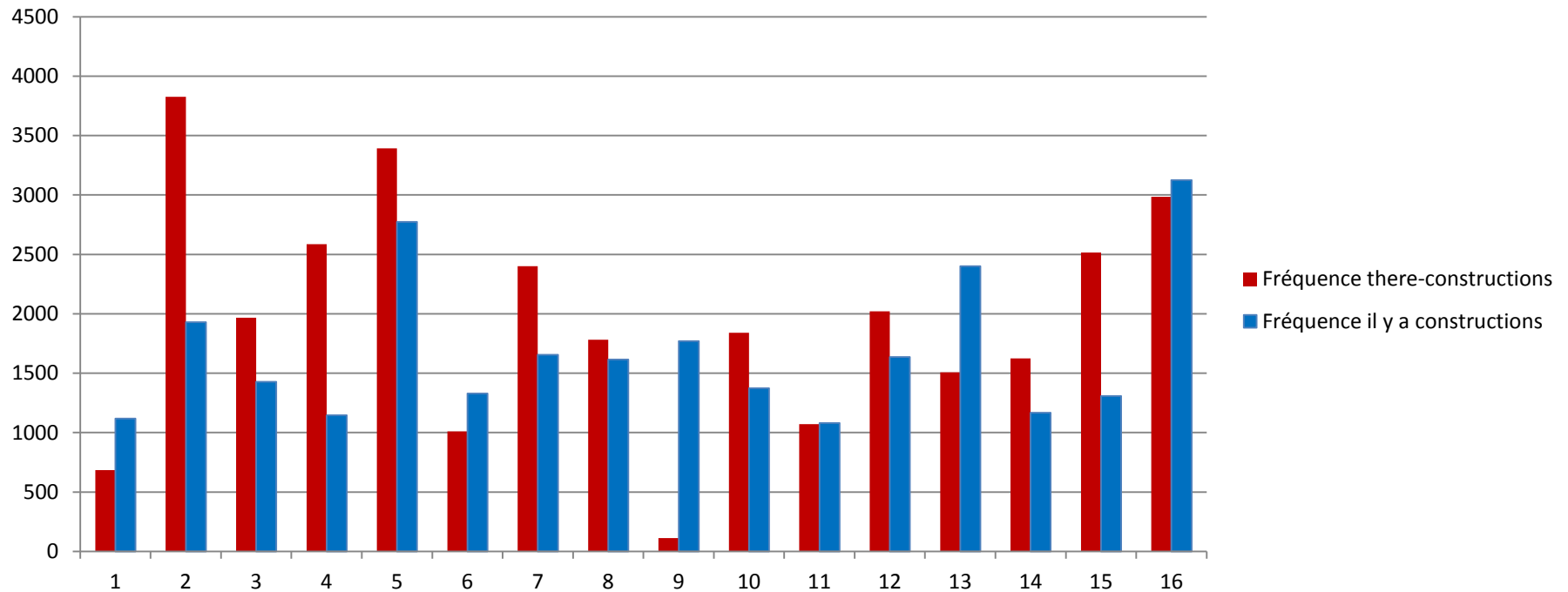




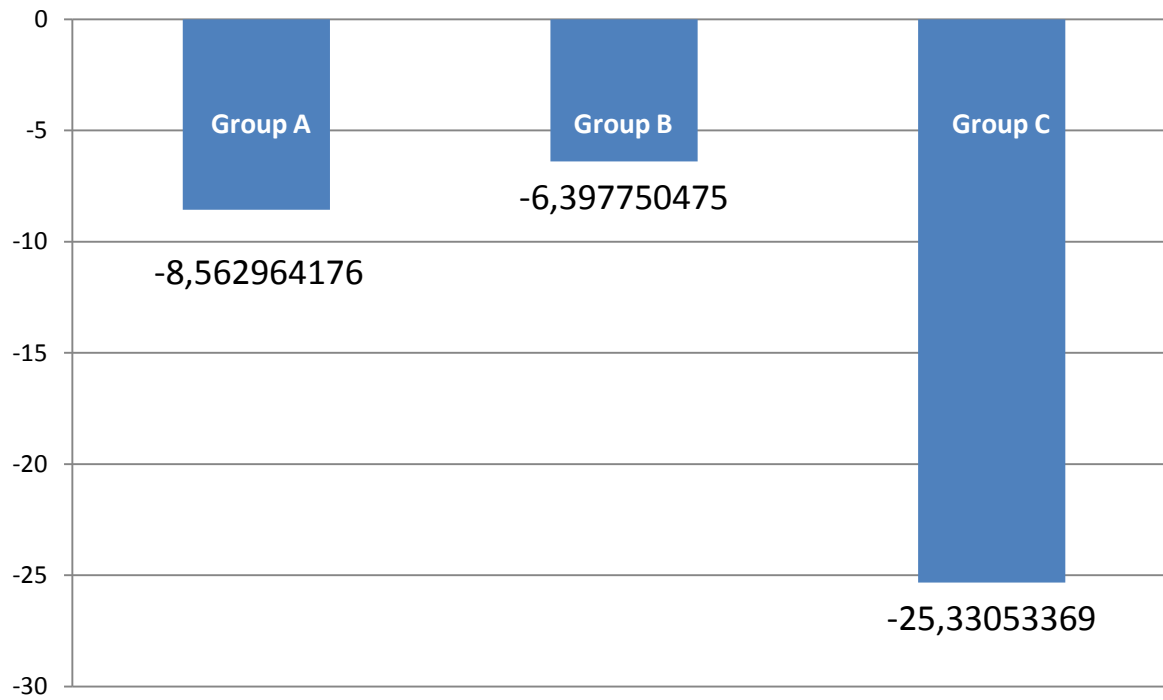
- Study of another linguistic phenomenon:
  - Existential constructions (there is/are vs. il y a) (cf. [Cappelle & Loock 2013](#))
  - Very similar to derived adverbs:



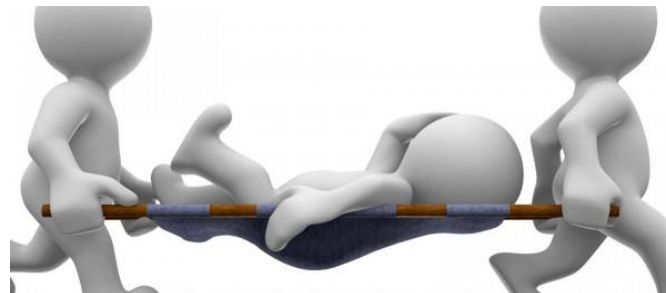
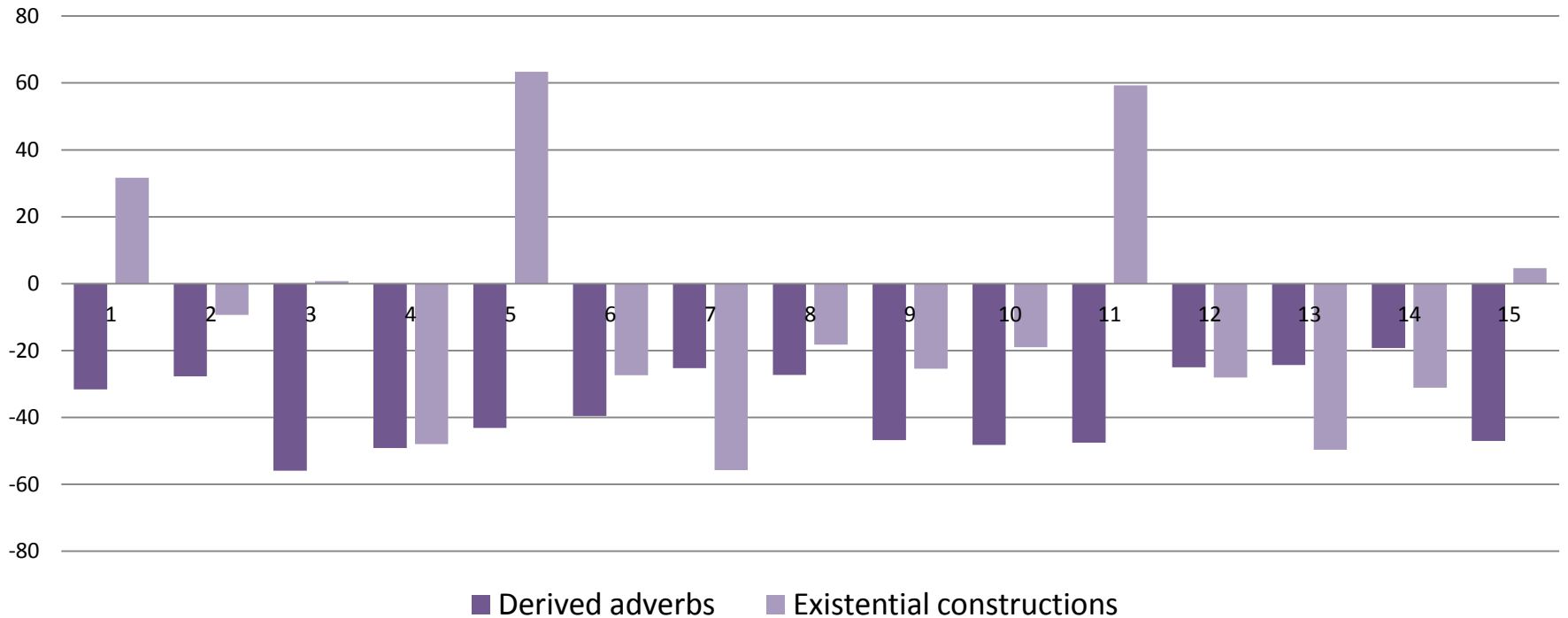
- If we check the learner corpus with longer texts (MéLexTra translation tasks):



NB: results for 9 can be explained by the repetition of "Y'a pas le choix" (= You got no choice)



- Comparison between adverbs and existential constructions:



# Conclusions

- There may be a correlation between linguistic homogeneisation and quality, but not for all linguistic phenomena
- Many questions remain unanswered
- Further research is required:
  - More linguistic phenomena
  - Increase corpus size

To be continued...



# Thank you for your attention!



Contact :

[rudy.loock@univ-lille3.fr](mailto:rudy.loock@univ-lille3.fr)

<http://stl.recherche.univ-lille3.fr/sitespersonnels/loock/>

[http://stl.recherche.univ-lille3.fr/CorTEx/CorTEx\\_index.html](http://stl.recherche.univ-lille3.fr/CorTEx/CorTEx_index.html)