

# Extracting present perfects from a multilingual corpus

Martijn van der Klis

Bert Le Bruyn

Henriette de Swart

# Why present perfects?

- Understudied in CompLing (Ritz 2012)
- Variation between languages
- Competition within languages
  - E.g. closely related to *simple past*
- Direct translation sometimes fails
  - Examples from children television
    - Mini-survey

# Mini-survey

## Set-up

- 16 participants
- 4 items
  - 2 ‘wrong’ uses of the *simple past* (OVT)
  - 2 ‘correct’ uses
- Participants were asked whether they would opt for a *present perfect* (VTT) alternative

# Mini-survey

## Example questions

[Context: Dora en Boots op het einde van één van hun avonturen delen hun vreugde over de goede afloop] Joepie! Het lukte!\*

Vergelijk met 'Joepie! Het is gelukt!'

- Het origineel is beter.
- Ik zou zelf het alternatief gebruiken.

**[Context: Dora and Boots share their joy over a successfully completed mission] Yippie! It succeeded! [original English: We did it!]**

Compare with 'Yippie! It has succeeded'

- The original is better.
- I would use the alternative.

[Context: Piet Piraat in de Griezelsgrot begint een griezelverhaal] Ik zat met de maatjes op het strand van een heel klein eiland...\*

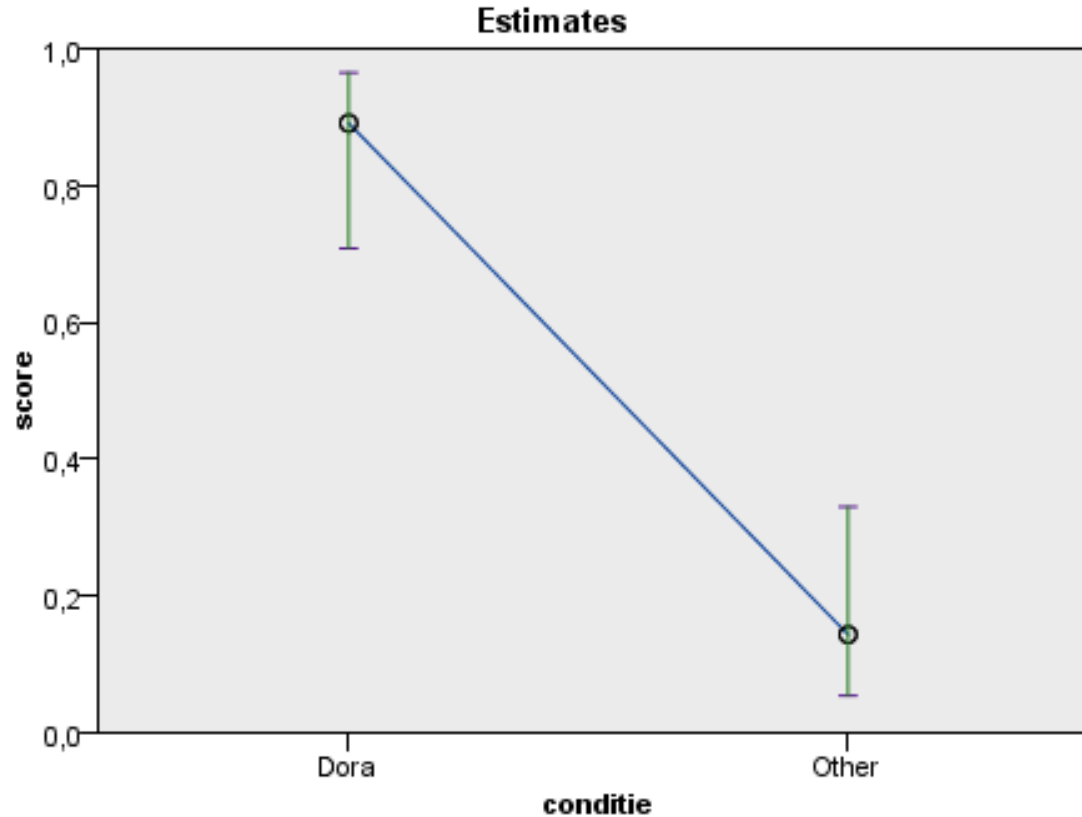
Vergelijk met 'Ik heb met de maatjes op het strand van een heel klein eiland gezeten.'

- Het origineel is beter.
- Ik zou zelf het alternatief gebruiken.

**[Context: Peter Pirate starts a story in the cave of horror!] I sat/was sitting with my buddies on a very small island...**

Compare with 'I have sat with my buddies on a very small island...'

# Mini-survey Results



Likelihood of going for the alternative *perfect* (VTT) was significantly different between the two conditions ( $p < .000$ ).

# Extracting present perfects Corpus

- Dutch Parallel Corpus (Paulussen et al. 2006)
  - 10 mln words, Dutch – English – French
    - Sentence-aligned
    - POS-tagged
    - Lemmatized
  - TEI XML-format

# Extracting present perfects

## Basic algorithm

- Look for an auxiliary verb
  - e.g. in Dutch: lemma *hebben* or *zijn* with a POS-tag stating present tense
- Find a neighboring past participle
  - e.g. in Dutch: look for the POS-tag *WW(vd,vrij,zonder)*
- Use a configuration file to discern between languages
- Output is a .csv-file which includes found sentence and translation.

# Extracting present perfects

## Words between aux. and participle

- Example:
  - Nobody has ever climbed that mountain.
- Affected languages: all (but esp. Dutch)
- Solution:
  - Allow for words in between
  - But stop looking at:
    - Verb forms that are not participles
    - Punctuation
    - Conjunctions



# Extracting present perfects

## Lexical restrictions

- Example:
  - Elle est arrivé.
    - lit. *She is arrived.*
- Affected languages: French, Dutch
- Solution:
  - Specify the verb(s) that have lexically restrictions
  - Compile a list of participles
    - For French: DR and MRS P. VAN DER TRAMP
    - For Dutch: list of ergative verbs from Wiktionary

# Extracting present perfects

## PP vs. PP continuous vs. PP passive

- Examples:
  - We have been here since 16 PM.
  - We have been presenting for 10 minutes.
  - We have been granted NWO funding.
- Affected languages: English
- Solution:
  - Check whether the participle “been” is followed by another participle (again possibly with words in between).

# Extracting present perfects

## Auxiliary after past participle

- Example:
  - Dat is de stad waar hij gewoond heeft.
    - lit. *This is the city where he lived has.*
- Affected languages: Dutch
- Solution:
  - Allow for a reversed search.

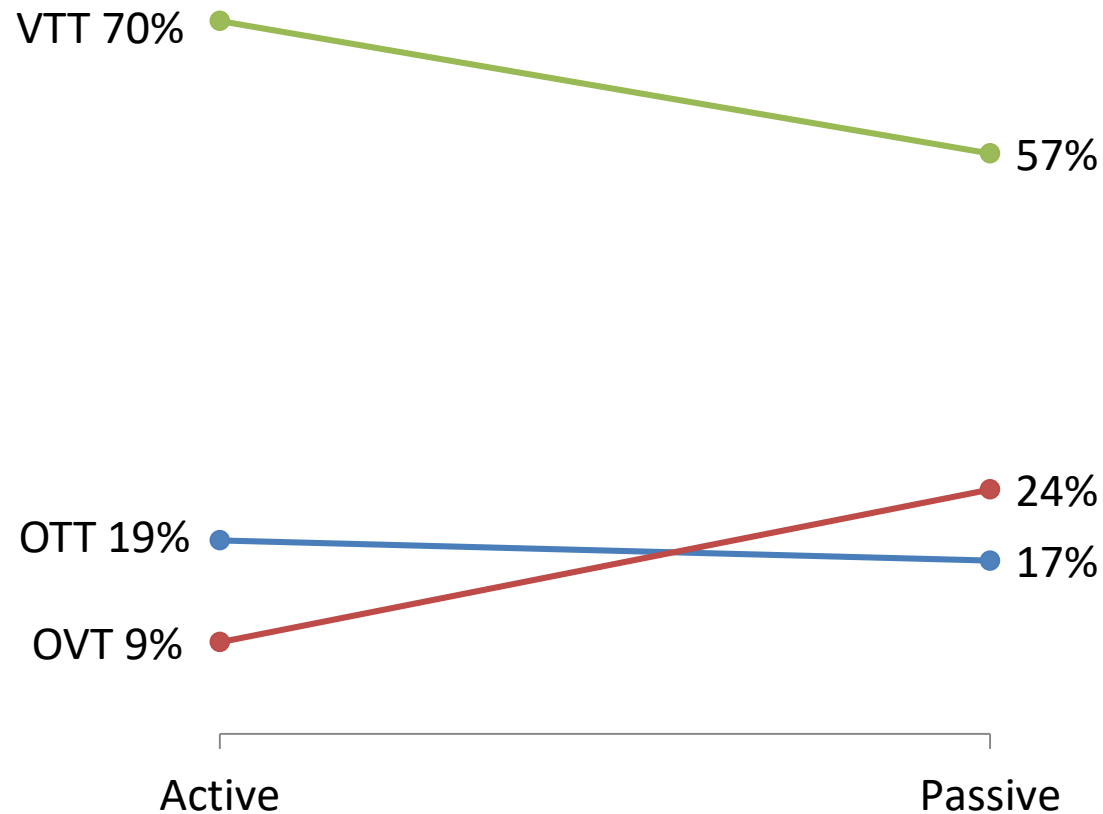
# Preliminary analysis

## Introduction

- Manual annotation of verb form in translation
  - is the form actually translated?
  - which tense?
  - active or passive voice?
    - Seems to be an important factor!

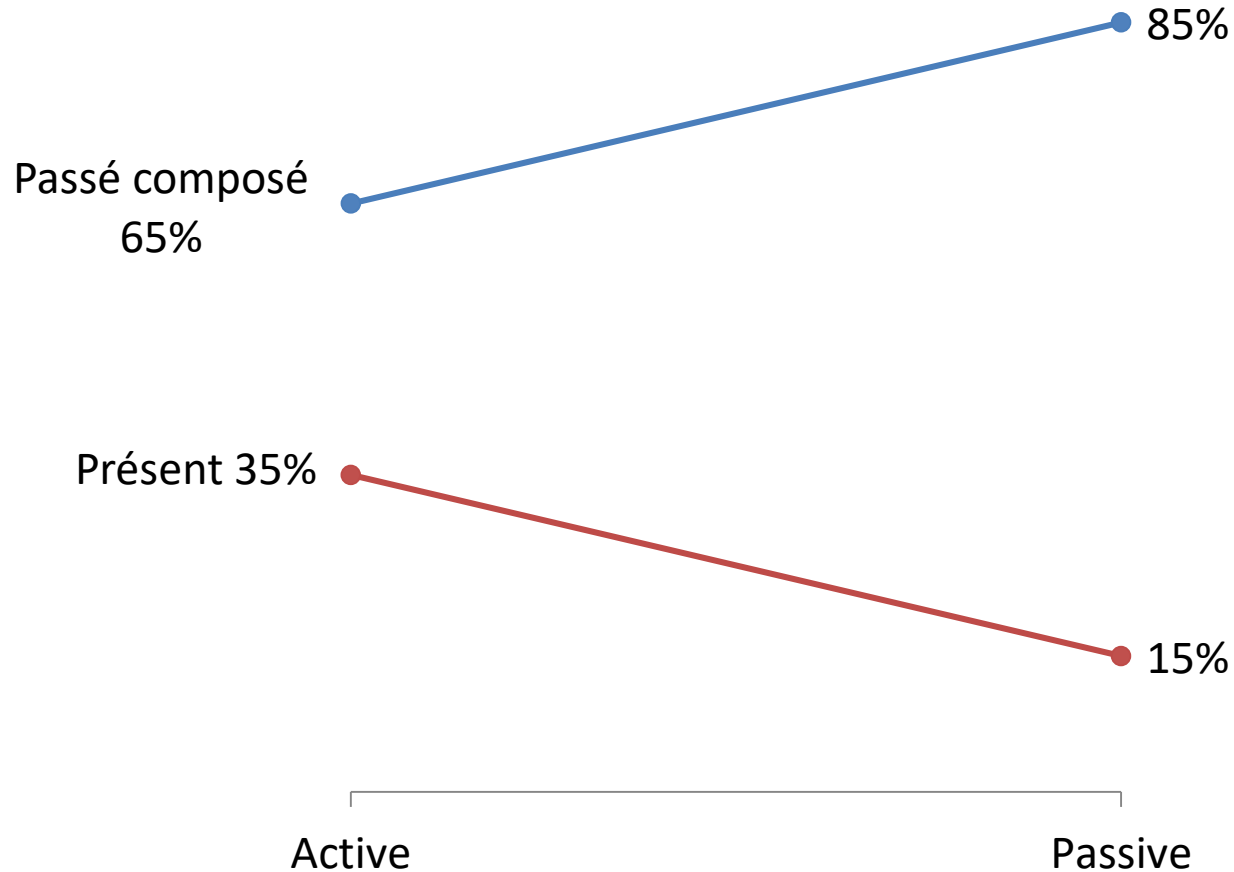
# Preliminary analysis

## English → Dutch



# Preliminary analysis

## English → French



# Preliminary analysis

## Tendencies

- Allergy of Dutch for passive present perfect (VTT)
  - Hypothesis: structural ambiguity
  - Example: *De deur is geopend* has two readings:
    - present perfect (the door has opened)
    - adjectival passive (the door has been opened)
- Continuative use of the perfect in English but not in Dutch/French
  - Example: *Mary has lived in London for five years.*
    - Translation: *Maria woont nu vijf jaar in Londen. (OTT)*
- Wider use of the Dutch and French present perfect (VTT/Passé Composé) not only in discourse (cf. de Swart 2007) but also for statement of fact
  - Example: *Vermoedelijk zijn deze dieren in visnetten verdronken.*
    - Translation: *These animals probably drowned.*

# Questions?

Martijn van der Klis

Digital Humanities lab, Utrecht University

M.H.vanderKlis@uu.nl

<https://github.com/UUDigitalHumanitieslab/time-in-translation>