# Use Case: searching for strong verbs in the historical CourantenCorpus

https://couranten.ivdnt.org

**Machteld de Vos | Radboud University | the Dutch Language Institute**

## INTRODUCTION

### Background: 17th-century Dutch

- Start of the standardisation process
- & language change in strong verbs from verb class III, e.g.:

**present** — **simple past** — **past participle**

*zingen* 'sing' – **sg**: *zang*, **pl**: *zongen* 'sang' – *gezongen* 'sung'

↓

*zong*

### Objectives

➢ Could standardisation (**norm**) have played a part in this change (**usage**)?

➢ **Use Case** for CourantenCorpus (CLARIAH tools)

## DATA & METHOD

### Data norm: NODE

- 10 normative grammars on Dutch
- Written between ca. 1550-1650

### Method norm:

- Identify normative comments on 41 class III strong verbs

https://couranten.ivdnt.org

### Data usage: CourantenCorpus (INT)

- Newspapers 1618-1700
- Ca. 19 million words
- Manually transcribed by volunteers
- **Automatically tagged & lemmatised (alpha version)**

### Method usage:

- **Corpus search (Blacklab)** for all possible verb forms of 41 class III strong verbs, incl. spelling variation & clitics

## USE CASE

### Tags & lemmas (alpha version)

- Automatic tagging & lemmatisation of insufficient quality, esp. for simple past singular (*song*: 2% correct lemma; 5% POS tag)

### Searching the corpus

- Query expansion via lexicon suggestions
- Use of wildcards:

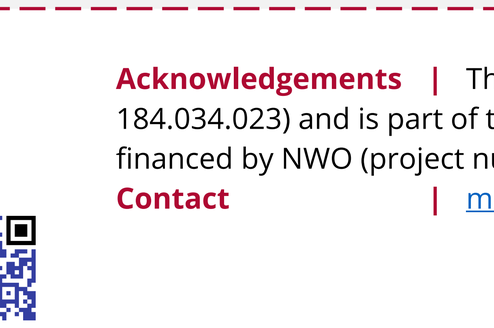
*sang*|*sank*|*sanc*|*zang*|*zanc*|*zank*
*song*|*sonk*|*sonc*|*zong*|*zonc*|*zonk*
*sing*|*sink*|*sinc*|*zing*|*zinc*|*zink*

- This led to large amounts of results to process manually:

| No. of results (total) | | Simple past sg. (total) |
|---|---|---|
| 563,014 | → | 9,398 |

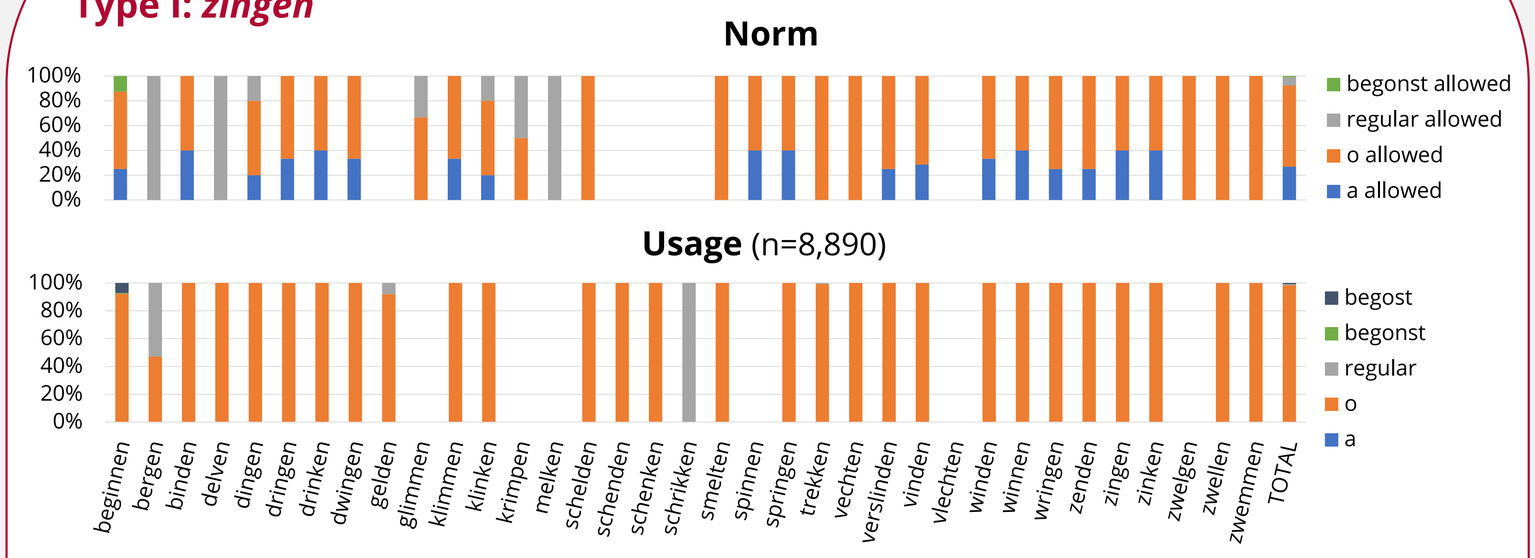
### Corpus specific

- The corpus is unbalanced



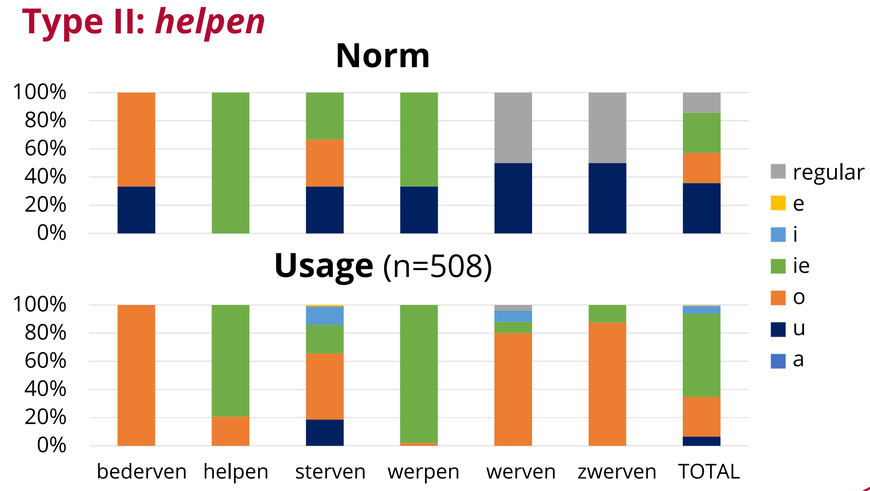

## RESULTS

### Type I: *zingen*

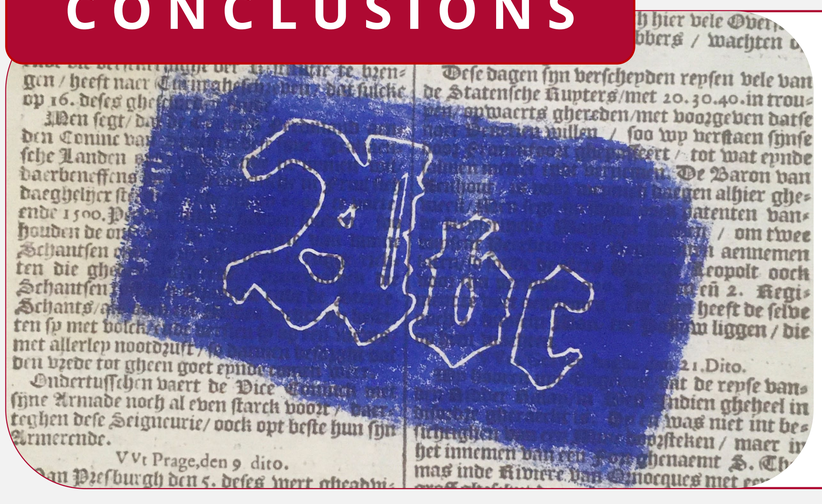



- Diffuse norm, making normative influence unlikely
- ***Zingen***: more variation in norm than in usage
- ***Helpen***: more variation in usage than in norm

### Type II: *helpen*




## CONCLUSIONS



➢ **Normative influence on strong verbs in 17th ct. newspapers unlikely**

➢ **List of requirements & wishes aimed at limiting the amounts of search results that need to be processed manually, e.g.:**

- Improve automatic tagging and lemmatisation (e.g. via crowdsourcing)
- Provide the possibility to create (balanced) subcorpora
- Create options to select and deselect within corpus search results

**Acknowledgements** | The research for this paper was made possible by the CLARIAH-PLUS project financed by NWO (Grant 184.034.023) and is part of the project *Spread the New(s). Understanding Standardization of Dutch through 17th-Century Newspapers*, also financed by NWO (project number 406.18.TW.005, research programme NWO Open Competition SSH).

**Contact** | machteld.devos@ru.nl | machteld.devos@ivdnt.org | @mtdevos

Radboud University /instituut voor de Nederlandse taal/ Clariah NWO