Towards automated literacy diagnostics: Insights from spelling error analysis

Wieke Harmsen
Roeland van Hout
Catia Cucchiarini
Helmer Strik

15th AWLL International Workshops on Writing Systems and Literacy
29-31 October 2025
Pisa, Italy

Radboud University

Introduction

- Learning to spell is a complex task
  - Contribution and interaction of many different linguistic and cognitive components (Caravolas et al., 2012).
- Early detection of spelling problems is important
- Current standardized spelling tests yield limited diagnostic information
- Aim: Extract more detailed diagnostic information from children's spelling tests

Introduction

Recent developments open-up new opportunities to reach this aim:

1. Growing collection of digital spelling data
2. Development of tools to annotate these data

Introduction

Recent developments open-up new opportunities to reach this aim:

1. Growing collection of digital spelling data
2. Development of tools to annotate these data

Research questions

1. To what extent can we automatically measure children's ability to apply spelling principles?
2. What are problematic spelling principles for below-average Dutch spellers?
3. Can we divide a group of Dutch below-average spellers into clusters based on their performance on applying specific spelling principles?

Method: Data set

- Participants: Dutch 3rd and 4th graders (8-10 years old) with below-average spelling skills
  - Grade 3: 85 participants
  - Grade 4: 63 participants
- Word dictations from *Lexima's Letterster*[1] software
  - Grade 3: 53 words
  - Grade 4: 66 words
- Same dictation was administered max. 3 times over the course of one school year

[1] https://lexima.nl/lezen/programmas/letterster

Method: Automatic detection of PCUs

- What is a **Phoneme-Corresponding Unit (PCU)** (Laarmann-Quante, 2016)
  → A sequence of graphemes that corresponds to one phoneme (e.g., school = s, ch, oo, l)

- Employ the algorithm described in Harmsen et al. (2021)

Method: Automatic detection of PCUs

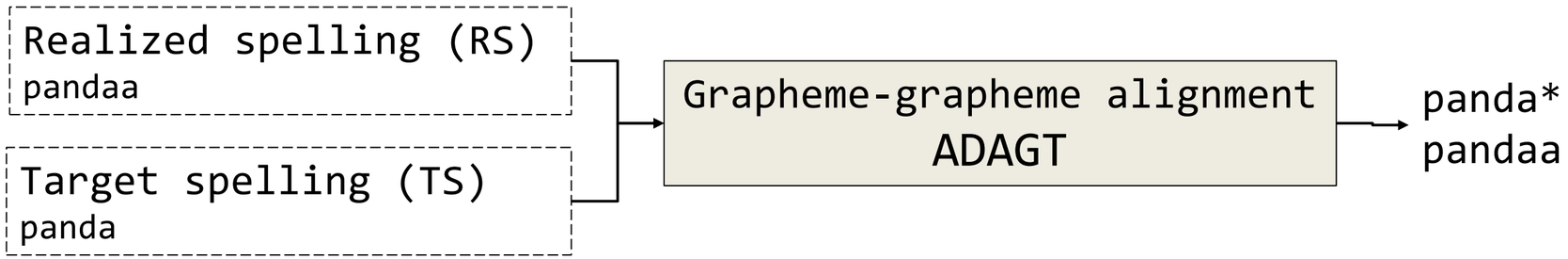
- What is a **Phoneme-Corresponding Unit (PCU)** (Laarmann-Quante, 2016)
  → A sequence of graphemes that corresponds to one phoneme (e.g., school = s, ch, oo, l)
- Employ the algorithm described in Harmsen et al. (2021)

Realized spelling (RS)
pandaa

Target spelling (TS)
panda

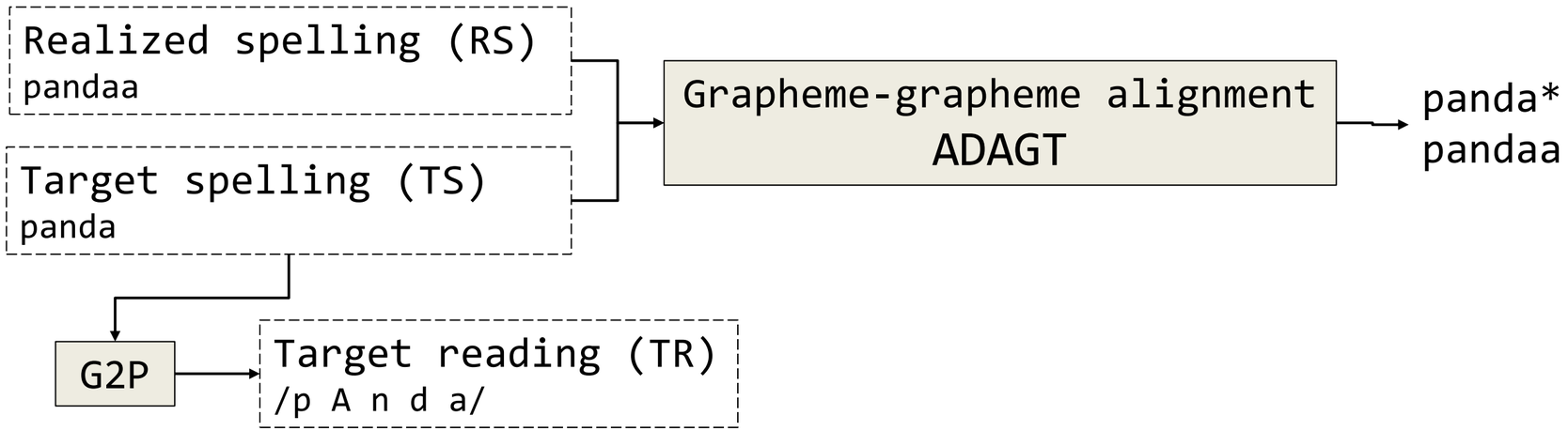
Method: Automatic detection of PCUs

- What is a **Phoneme-Corresponding Unit (PCU)** (Laarmann-Quante, 2016)
  → A sequence of graphemes that corresponds to one phoneme (e.g., school = s, ch, oo, l)
- Employ the algorithm described in Harmsen et al. (2021)



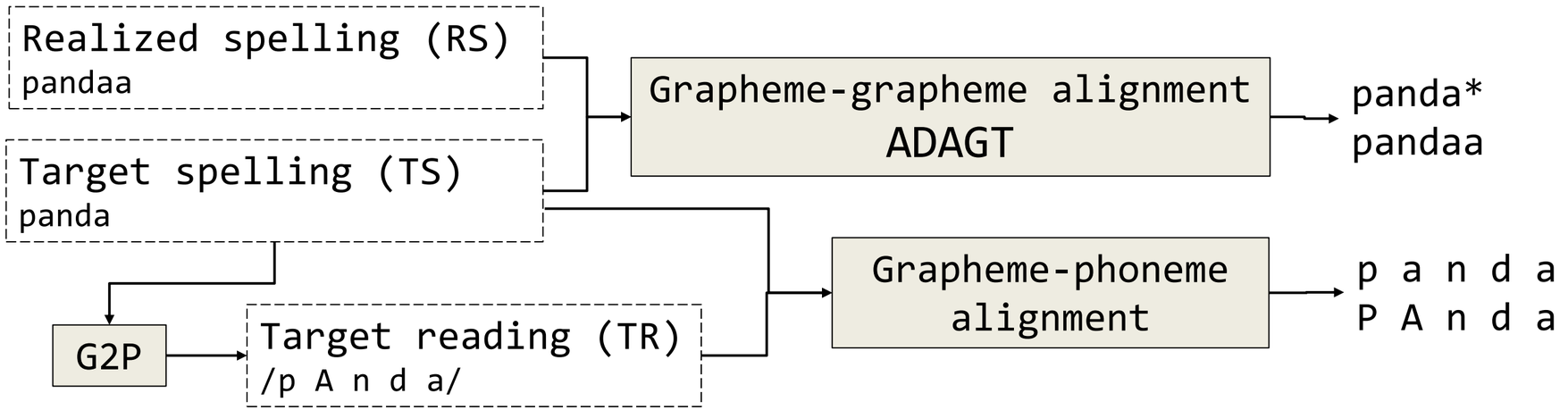
Method: Automatic detection of PCUs

- What is a **Phoneme-Corresponding Unit (PCU)** (Laarmann-Quante, 2016)
  → A sequence of graphemes that corresponds to one phoneme (e.g., school = s, ch, oo, l)
- Employ the algorithm described in Harmsen et al. (2021)



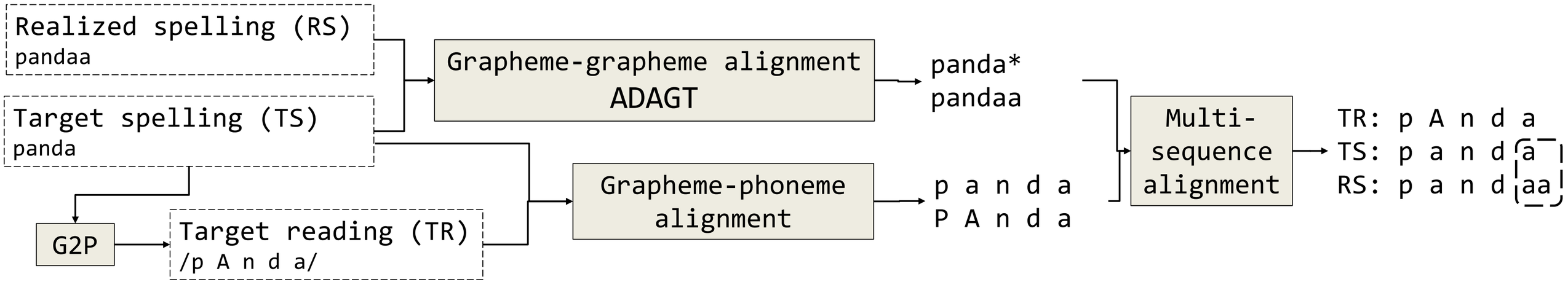
Method: Automatic detection of PCUs

- What is a **Phoneme-Corresponding Unit (PCU)** (Laarmann-Quante, 2016)
  → A sequence of graphemes that corresponds to one phoneme (e.g., school = s, ch, oo, l)
- Employ the algorithm described in Harmsen et al. (2021)



Method: Automatic detection of PCUs

- What is a **Phoneme-Corresponding Unit (PCU)** (Laarmann-Quante, 2016)
  → A sequence of graphemes that corresponds to one phoneme (e.g., school = s, ch, oo, l)
- Employ the algorithm described in Harmsen et al. (2021)



Method: Automatic annotation of PCUs with spelling principles

- Obtain lemma, POS-tags and morphemes of target word (FROG, van den Bosch et al., 2007)

Method: Automatic annotation of PCUs with spelling principles

- Obtain lemma, POS-tags and morphemes of target word (FROG, van den Bosch et al., 2007)
- Rule-based (Harmsen et al., 2021)

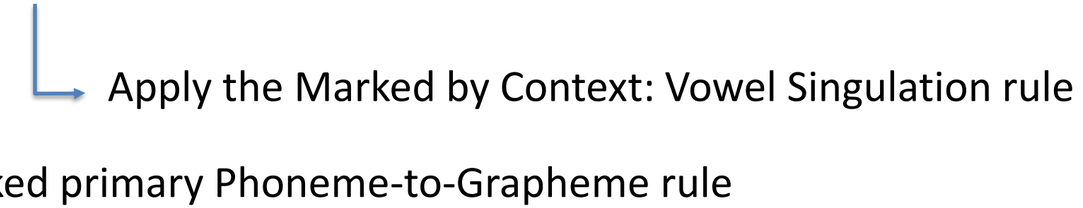
```
TR:/p    A    n    d    a /
TS: p    a    n    d    a
RS: p    a    n    d    aa
    Un1  Un1  Un1  Un1  CoVs
```

Apply the Marked by Context: Vowel Singulation rule

Apply an Unmarked primary Phoneme-to-Grapheme rule

Method: Automatic annotation of PCUs with spelling principles

```
TR:/p   A   n   d   a /
TS: p   a   n   d   a
RS: p   a   n   d   aa
    Un1 Un1 Un1 Un1 CoVs
```



- The annotation scheme has three main categories (Horbach-Kleijnen, 1992):
  - **Unmarked (Un)** Apply phoneme-to-grapheme conversion rule
    - Un1 PCUs Only one way in which this phoneme can be written (e.g., /m/ -> m)
    - Un2 PCUs Multiple ways in which this phoneme can be written (e.g., /EI/ -> ei or ij; /x/ -> ch or g)
  - **Marked by Context (Co)** The context of a PCU influences the spelling
  - **Marked by Morphology (Mo)** The morpheme containing the PCU influences the spelling

Method: Extraction of measures using the annotated PCUs

- For **Marked by Context** and **Marked by Morphology**:
  - Divide the number of times a rule is applied incorrectly by the number of times the rule had to be applied in the dictation.
- For **Unmarked Rules**
  - Label each error with an error type:
    - Sub_mir The incorrectly written PCU is a mirrored version of the target PCU (e.g., b-d-p; w-m).
    - Sub_rev The incorrectly written PCU is a reversed version of the target PCU (e.g., oe -> eo; ui -> iu).
    - Sub_pho The incorrectly written PCU maps to the same phoneme as the target PCU.
    - Sub_oth The incorrectly written PCU maps to another phoneme than the target PCU.
  - Divide the number of errors in each error type by either the number of Un1 PCUs or the number of Un2 PCUs.

Method: Data analysis

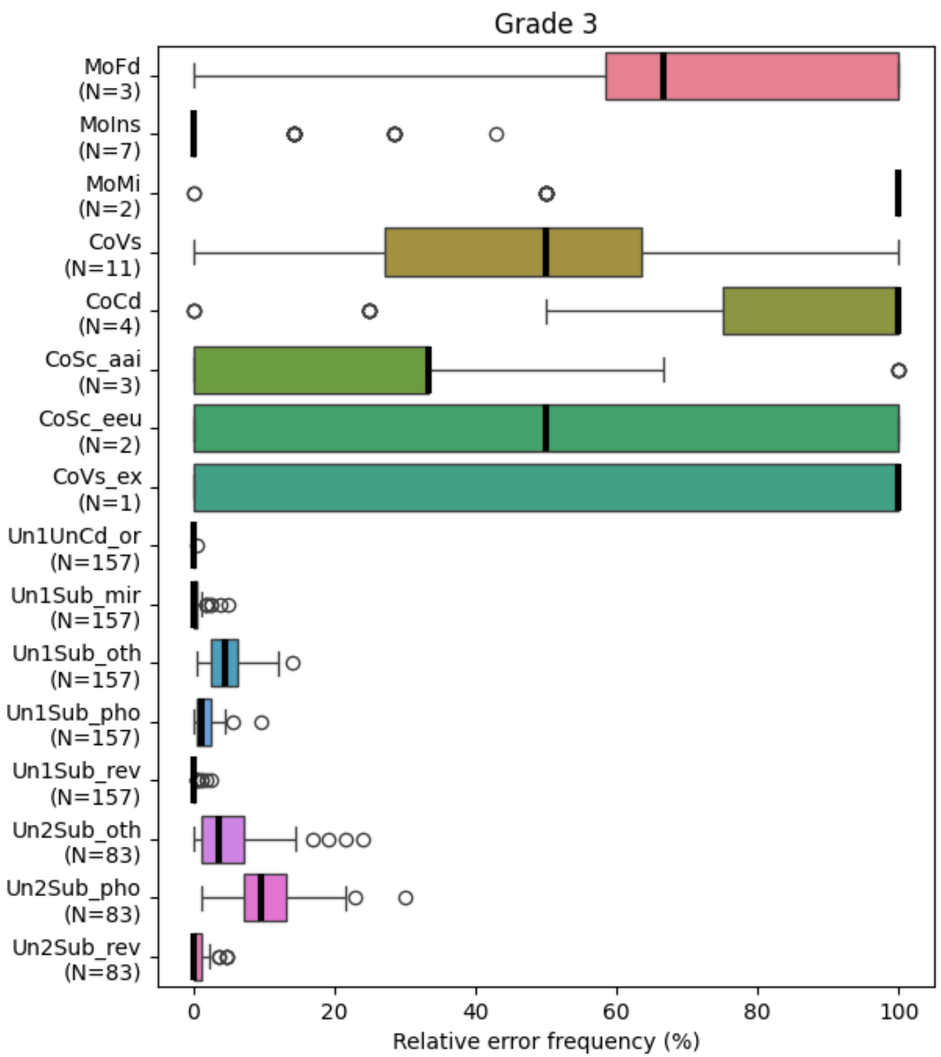
- Boxplots to visualize the performance on each spelling principle
- Hierarchical clustering
- Principle Component Analysis (PCA) to visualize clusters and identify differences between pupils.

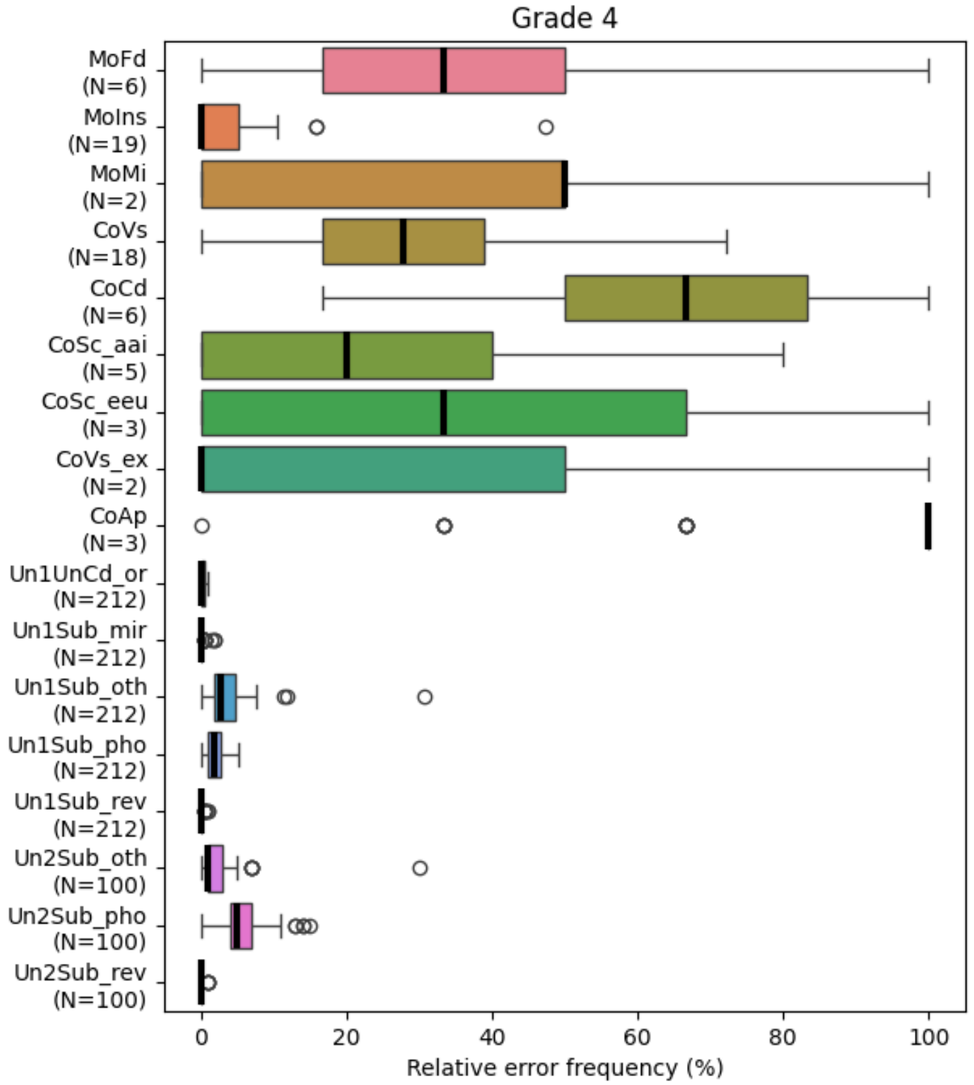
Results: Description of the data

grade	period	nr dictations	spelling score	
			mean	std
3	1	76	43.07	15.21
	2	60	56.40	16.23
	3	22	64.75	12.33
4	1	61	58.42	11.98
	2	52	68.46	13.73
	3	14	76.06	9.32

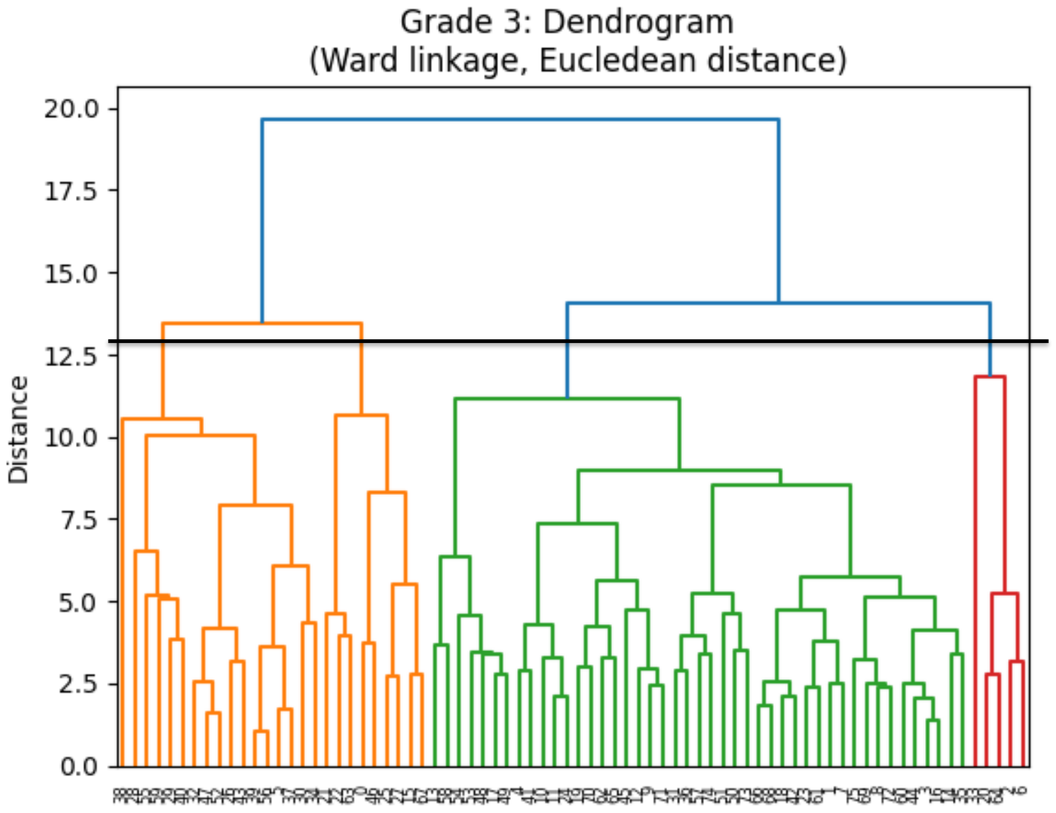
- Spelling score = percentage of correctly spelled words in dictation
- Not all children participated in all periods.
- On average, children become better spellers

Results: Proficiency in spelling error application

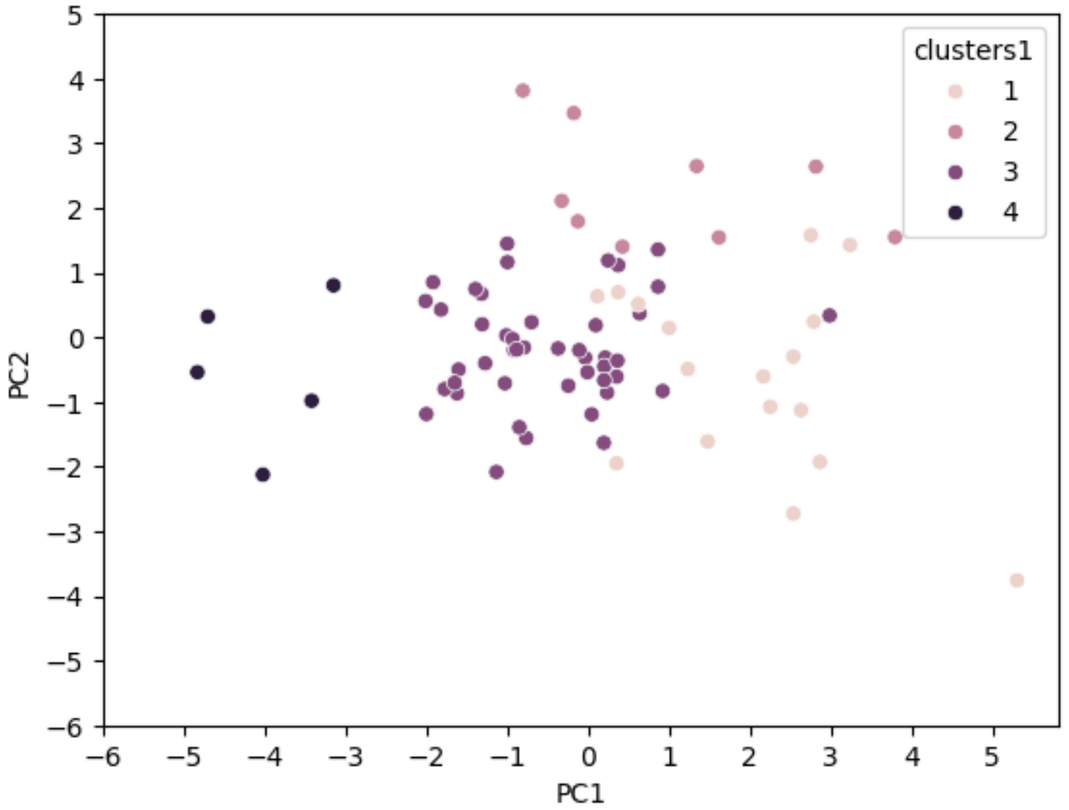




Clustering Grade 3



CoVs



Un2Sub_pho

Clustering Grade 4

- No clear clusters identified using the current method.

Discussion

RQ1: To what extent can we automatically measure children's ability to apply spelling principles?

- We presented a set of relative measures that quantify how often a spelling principle was applied incorrectly relative to the number of times a spelling principle had to be applied.
- Limitation:
  - Some spelling principles had to be applied only 1 or 2 times in the dictation. This is too little to measure a child's proficiency in applying this principle.

RQ2: What are problematic spelling principles for below-average Dutch spellers?

  - Grade 3: Consonant doubling (CoCd) rule -> ki**pp**en (*chickens*)
  - Grade 4: Apostrophe (CoAp) rule -> menu**'s** (*menus*)

Discussion

RQ3: Can we divide a group of below-average spellers into clusters based on their performance on specific spelling principles?

- Grade 3: Four clusters that differed on:
  - **Un2Sub_pho**: The percentage of phonetically correct errors made in PCUs that are phonetically ambiguous (e.g., writing *s**ch**ool* as *s**g**ool* (school))
  - **CoVs**: The percentage of incorrectly written PCUs where a vowel singulation rule had to be applied (e.g., writing *maken* as *maaken* (to make))
- Grade 4:
  - No clear clusters identified using the current measures and method
  - Future direction: Group the current measures differently

Conclusion

- First step towards obtaining more detailed information from children’s dictations
- Automatic evaluation of children’s ability to apply spelling principles
- This information can help teachers to personalize their instruction

Thank you!

wieke.harmsen@ru.nl

Radboud University Nijmegen, The Netherlands

Tools: https://github.com/WiekeHarmsen/multi_seq_align_nl

We would like to thank Martine Gijsel and Melissa in 't Zandt from Expertisecentrum Nederlands for their help with collecting the data.

This work is part of ASTLA project with project no. 06.20.TW.009, which is (partly) financed by the Dutch Research Council (NWO).

References

Caravolas, M., Lervåg, A., Mousikou, P., Efrim, C., Litavský, M., Onochie-Quintanilla, E., Salas, N., Schöffelová, M., Defior, S., Mikulajová, M., Seidlová-Málková, G., & Hulme, C. (2012). Common Patterns of Prediction of Literacy Development in Different Alphabetic Orthographies. *Psychological Science*, *23*(6), 678-686. https://doi.org/10.1177/0956797611434536

Harmsen, W. N., Cucchiarini, C., & Strik, H. (2021). Automatic Detection and Annotation of Spelling Errors and Orthographic Properties in the Dutch BasiScript Corpus. *Computational Linguistics in the Netherlands Journal*, *11*, 281–306. https://www.clinjournal.org/clinj/article/view/140

Horbach-Kleijnen, R. (1992). *Hardnekkige spellingfouten: Een taalkundige analyse.* Lisse: Swets & Seitlinger.

Laarmann-Quante, R. (2016). Automating multi-level annotations of orthographic properties of German words and children's spelling errors. *Proceedings of the 2nd language teaching, learning and technology workshop (LTLT 2016),* 14-22.

Van den Bosch, A., Busser, B., Canisius, S. and Daelemans, W. (2007). An efficient memory-based morphosyntactic tagger and parser for Dutch. *Computational linguistics in the Netherlands: Selected papers from the seventeenth CLIN Meeting*, 99-114.