

University of Groningen

Representing Low-Resource Languages and Dialects: Improved Neural Methods for Spoken Language Processing

Bartelds, Martijn

DOI:

[10.33612/diss.812185119](https://doi.org/10.33612/diss.812185119)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2023

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Bartelds, M. (2023). *Representing Low-Resource Languages and Dialects: Improved Neural Methods for Spoken Language Processing*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen. <https://doi.org/10.33612/diss.812185119>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Stellingen behorend bij het proefschrift

Representing Low-Resource Languages and Dialects: Improved Neural Methods for Spoken Language Processing

1. Neural networks can model and help empower language varieties.
2. Intermediate layers of neural networks, especially the middle layers of Transformer models, capture information that is useful to automatically measure linguistic diversity.
3. Transformer-based speech representations capture more information from the speech signal than can be represented by phonetic transcriptions.
4. Closely-related language varieties can be distinguished using Transformer-based speech representations extracted from only six seconds of speech.
5. Speech representations that are useful for measuring linguistic diversity can also improve speech information retrieval performance.
6. Speech recognition performance for very low-resource language varieties improves when the training data is augmented through self-training or a text-to-speech system.
7. “Kiek ’es jongens, je huift niet per se uut Harvard te kommen of Cambridge ... ook gewoon in Grunningen ... kun je iets bereiken.” - *Ben Feringa*
8. “Het [PhD traject] heeft steeds weer een ander decor.” - *Annet Bartelds*

Martijn Bartelds, Groningen 2023