Frisian and Low Saxon
in flux

RB

Frisian and Low Saxon in flux

university of groningen

CLCG
Center for Language and Cognition Groningen

bcn

The work in this thesis has been carried out under the auspices of the Center for Language and Cognition Groningen (CLCG) and the Research School of Behavioural and Cognitive Neurosciences (BCN). Both are affiliated with the University of Groningen.

Groningen Dissertations in Linguistics 253

ISBN: 978-94-6473-710-3

Document prepared with LaTeX and typeset with XeLaTeX. The main (slab) serif font, typeset with smaller letter spacing, is Charis SIL by SIL Global. The sans serif font used in graphs is Helvetica. The monospaced font is Roboto Mono by Google LLC. The Tengwar font is Alcarin Tengwar by Toshi Omagari (inspired by the Brill typeface by Brill Academic Publishers).

rijksuniversiteit
groningen

Frisian and Low Saxon in flux

Proefschrift

ter verkrijging van de graad van doctor aan de
Rijksuniversiteit Groningen
op gezag van de
rector magnificus, prof. dr. ir. J. M. A. Scherpen,
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

donderdag 13 februari 2025 om 14.30 uur

door

Raoul Sergio Samuel Jan Buurke

geboren op 3 april 1995
te Groningen, Suriname

Promotor(es)
Prof. dr. M. B. Wieling

Copromotor(es)
Dr. R. M. Knooihuizen
Dr. W. J. Heeringa

Beoordelingscommissie
Prof. dr. J. Grieve
Prof. dr. S. Marzo
Prof. dr. A. P. Versloot

PROPOSITIONS

1. The Low Saxon speaker population is in rapid decline while the Frisian speaker population appears stable. These population differences are primarily driven by a difference in intergenerational transmission rates.
2. Improving attitudes towards regional languages may boost their intergenerational transmission. This requires parents and governmental organizations to promote multilingual family policies that include regional languages.
3. In recent decades, Standard Dutch influence has been the main driver of pronunciation change in Frisian and Low Saxon varieties in the Netherlands.
4. Standard Dutch has a greater influence on Low Saxon than Frisian. However, convergence and divergence patterns vary considerably between lexical items and regional language users.
5. Regiolects appear to be forming in the Frisian and northern Low Saxon areas. Conversely, Low Saxon variants appear to become more inconsistent between speakers in, for example, the province of Overijssel.
6. Using a mobile laboratory enables reaching many participants, which is crucial for reliable language change analyses because language use between speakers can differ strongly.
7. *Roaden en gizzen binnen twij onwizzen.*
   'Guesswork and conjecture are two uncertainties'.
   — Aaldrik Sillius, *Grunneger Zegswiezen*.
8. *Meten is weten, maar perfecte meetmethodes bestaan niet.*
   'To measure is to know, but perfect measuring methods do not exist.'
   (Inspired by a common Dutch saying, which ranges back to Heike Kamerlingh Onnes' inaugural lecture in 1882.)

DANKWOORD

Mae govannen, ofwel: gegroet!

Het zal de meeste mensen niet ontgaan zijn dat ik een groot liefhebber ben van J. R. R. Tolkiens legendarium over Midden-aarde, dus hier en daar heb ik er wat inspiratie uit geput voor de vormgeving van dit proefschrift. Ik gebruik hiervoor graag als excuus dat mijn interesse in taalvariatie en kaarten al vroeg aangewakkerd is door Tolkiens werken en dat dit proefschrift me een geschikte plek lijkt om hier iets mee te doen.

Mensen willen graag weten welke mensen allemaal van belang geweest zijn voor zowel het onderzoek als de onderzoeker en daardoor is het dankwoord van een proefschrift doorgaans het meest gelezen stuk. Tijdens het schrijven van dit dankwoord merkte ik dat ik een behoorlijke lijst aan mensen moet bedanken, waarvoor mijn beste poging als volgt gaat.

Ten eerste wil ik graag de leden van de leescommissie bedanken: **Jack Grieve**, **Arjen Versloot** en **Stefania Marzo**. Hartelijk bedankt voor het lezen en beoordelen van dit proefschrift.

Ik had drie uitstekende begeleiders tijdens mijn promotieproject, die elkaar heel goed aanvulden. **Martijn**, mijn dank is groot. Niet alleen omdat je mijn interesse in de dialectometrie hebt aangemoedigd, maar ook voor je ijzersterke begeleidende rol door het project heen. Je weet altijd tien ballen tegelijkertijd in de lucht te houden, maar je verliest nooit de mensen in het team uit het oog. Je bent ook altijd bereikbaar voor advies, zelfs als ik meer onbenullige vragen had, maar je wist me ook aan te moedigen om me mijn eigen fouten te laten maken. Mede dankzij jou zijn onze gepubliceerde artikelen en dit proefschrift een stuk leesbaarder geworden. Je bent en blijft een aanwinst voor het wetenschappelijke veld, je collega's en voor de Nedersaksische taal en haar variëteiten.

Remco, ik denk dat ik je gepast mijn tweede dagelijkse begeleider kan noemen. We hebben regelmatig een half uur tot een uur na onze wekelijkse bijeenkomsten nog nagekletst in de gang en ik mis dit nu al. Uiteraard ging het

lang niet altijd over het project, maar dat hoeft ook niet. Ik ben voornamelijk geschoold in de meer beschrijvende dialectometrie en je hebt me herhaaldelijk aangemoedigd om de variatiepatronen in meer detail uit te pluizen, zodat we ook meer de verklarende kant op konden en meer diepgang konden geven aan het onderzoek. Daarnaast ben je een redactioneel tovenaar, waardoor onze teksten in alle details veel leesbaarder geworden zijn. Je bent in alle opzichten een uitstekende begeleider en het is ook te merken dat je er zelf veel voldoening uit haalt, dus ik hoop dat je veel van dit soort avonturen mag voltooien.

Wilbert, jouw rol in het project was ook vanaf het begin onverminderd waardevol. Je vond dit zelf niet altijd voldoende om medeauteur van de artikelen te zijn, maar niets is minder waar. Door jouw adviezen zijn flinke stukken van het onderzoek beter uit de verf gekomen, vooral doordat je vaak wist te wijzen op zaken die we eerder in het proces over het hoofd hadden gezien of doordat we te grote denkstappen maakten in de artikelen. Soms stond je inderdaad wat verder af van het onderzoek dan op andere momenten, maar juist de frisse blik die je op de zaken wierp als wij een artikel al bijna in steen gebeiteld hadden, was altijd erg nuttig. Je hielp ook met pas op de plaats houden met eventuele zijprojecten, zodat ik het gehele project goed af kon ronden. Tenslotte wil ik graag benadrukken dat ik je een bijzonder respectvol persoon vind en dat je enthousiasme voor taalvariatie aanstekelijk werkt. Het was een eer je erbij te hebben.

Ik heb ook het genoegen gehad om een congres te organiseren in Gent samen met **Cesko**, **Jos**, **Kristel**, **Lisa** en **Melissa**. Dit was een bijzonder prettige samenwerking en we hebben uiteindelijk, na een vrij lange tocht van twee jaar, ook nog een speciale uitgave van het blad Taal en Tongval weten uit te brengen. **Gunther**, ik bedank je ook ten zeerste voor je geduld en vlotte samenwerking met betrekking tot de speciale uitgave.

Zonder het uitstekende werk van de enthousiaste mensen die werken bij de Fryske Akademy was de kwaliteit van dit proefschrift ook minder geweest. Daarom bedank ik bij deze onder andere **Anne (Merkuur)**, **Anne-France**, **Eric (Hoekstra)**, **Hans**, **Henk (Bloemhoff)**, **Jelske**, **Nika** en **Willem (Visser)**. Ik kan natuurlijk ook niet **Lysbeth** en **Sannah** van de *Provinsje Fryslân* onbenoemd laten, die zich inzetten voor duurzaam Fries taalbeleid. De Friese taal en identiteit zijn buitengewoon waardevol en jullie doen die een eer aan met jullie werk, passie en onderzoek.

Ik zet me graag in voor het Gronings, de lokale varianten van het Nedersaksisch in Groningen, waarvan dit proefschrift een reflectie is. Gelukkig sta ik hier niet alleen in en werken de medewerkers van het Centrum Groninger Taal en

Cultuur en ook van Erfgoedpartners mee aan het behoud en promoten van ons regionale erfgoed. Hiervoor dank ik onder andere **Coby**, **Fieke**, **Goffe**, **Henk**, **Kim**, **Maarten**, **Merel**, **Nina**, **Olaf**, **Patricia**, **Roeli**, **Rosa** en **Sisi** en natuurlijk de vrijwilligers **Alie**, **Geesje**, **Leo** en **Nella**. *t Is nait altied makkelk om t Grunnegs te bescharmen, mor deur joen haarde waark kommen der nog altied stolde Grunnegers bie en doar doun wie t veur*.

Ik heb ook veel collega's in het veld ontmoet, die inzichten verschaften over taalvariatie in het noorden en oosten van Nederland en ook in zuidelijk Nederland en Vlaanderen. Ik bedank bij deze onder andere de volgende mensen voor hun harde werk met betrekking tot streektalen: **Amber**, **Anne (Breitbarth)**, **Anne-Sophie**, **Arja**, **Eline**, **Frans**, **Geert**, **Gerard**, **Henk (Wolff)**, **Jeffrey**, **Johan**, **John (Huisman)**, **John (Nerbonne)**, **Joske**, **Leonie**, **Martin**, **Melissa**, **Nicoline**, **Piet**, **Reinhild**, **Renate**, **Stef**, **Yoïn** en **Veronique**. Bedankt en wees altijd welkom in Groningen!

Ik heb tijdens het project ook regelmatig samengewerkt met het *Bilingualism and Aging Lab*, waarvan ik de leden hartelijk wil bedanken voor hun geduldige en prettige samenwerking. Een aantal mensen van de onderzoeksgroep wil ik in het bijzonder noemen.

Hanneke, we hebben elkaar relatief weinig gesproken tijdens mijn promotietraject vergeleken met daarvoor, maar ik vond het erg prettig dat je als vertrouwenspersoon voor mijn promotieproject wilde functioneren. Het is een goed teken dat we elkaar niet vaak hoefden te spreken in die context, maar je gaf me een sterk gevoel van verzekerdheid.

Jelle, het was een waar genoegen om samen te werken aan het Lifelinesonderzoek en ik hoop dat anderen de relatie tussen streektalen en sociale inbedding vaker zullen onderzoeken in de toekomst, want dit moet verder uitgeplozen worden. Buiten het werk kunnen we altijd genieten van elkaars kookkunsten, speciaalbierexpertise en wederzijdse flauwe humor en ik hoop dat we dit voort kunnen blijven zetten. Ik kan het ook specifiek waarderen dat als ik een flauwe grap je deze eigenlijk altijd aanvult met een andere flauwe grap, in plaats van dat je met je ogen rolt.

Floor, de analyses voor ons Lifelinesonderzoek waren erg pittig en veel zaken bleven relatief onzeker na de analyses. Desalniettemin heb je gepast juist de kracht van die analyses weten te presenteren en daar heb ik bewondering voor. Ik ben je verder in het bijzonder dankbaar voor het mij vergezellen bij de opnames in *Laorne* (Laren, Gelderland). Je bent een weerbare doorzetter en een warm persoon en ik wens je veel geluk toe in de toekomst.

Binnen *Speech Lab Groningen* heb ik veel geleerd, plezier gehad, samengewerkt aan projecten en bijgedragen aan wetenschapsevenementen. Hiervoor gaat mijn dank onder andere uit naar **Alexander**, **Caihong**, **Dan**, **Defne**, **Eva**, **Floris**, **Frank**, **Hedwig**, **Inge**, **Jidde**, **Katharina**, **Lisanne**, **Lourens**, **Martijn**, **Nikki**, **Reihaneh**, **Steven**, **Teja**, **Thomas**, **Valentina**, **Valentine**, **Valerie**, **Wietse** en **Yun**. Een paar mensen adresseer ik bij deze in meer detail, maar ik vind dat we als groep veel moois neergezet hebben de laatste paar jaar.

Defne en **Frank**, jullie hebben de groep aanzienlijker rijker gemaakt met jullie onderzoekslijnen. Het lab is hierdoor academisch diverser geworden en we hebben er ook nog eens twee stralende persoonlijkheden bij gekregen.

Floris, het was me een genoegen om je te begeleiden met je beginnende stappen in de dialectometrie. Het is duidelijk dat je dit onderwerp heel interessant vindt en daar kan ik uiteraard niets tegenin brengen. Veel succes en vooral plezier met je onderzoek in dit prachtige veld, want er is nog veel om uit te zoeken.

Hedwig, ik twijfelde van tevoren of ik je voor een aparte laatste paragraaf moest bewaren of je in de context van het lab al wilde noemen. Ik ben toch voor het laatste gegaan, omdat dit uiteindelijk toch een academisch proefschrift is en we ook in die capaciteit veel samen meegemaakt hebben. We hebben samen een artikel gepubliceerd, wat nu een hoofdstuk in dit proefschrift geworden is. We hebben ook samen veel gereisd naar conferenties, naar opnamelocaties met het mobiele laboratorium en naar veel wetenschapsevenementen waar we met het lab waren. Ik ben van mezelf geen reiziger, maar het reizen was altijd een stuk aangenamer met jou aan mijn zijde. Beter dan de meeste anderen in dit dankwoord weet en ervaarde je dat ik soms mijn kiezen flink op elkaar moest zetten tijdens dit promotieproject en ik ben erg dankbaar voor je geduld en medeleven, vooral tijdens deze momenten. Gelukkig kan ik er tegenover zetten dat je ook de vele positieve momenten meegemaakt hebt en er is niemand met wie ik mijn geluk liever deel.

Katharina, ik wil je graag bedanken voor je hulp bij de opnames in IJsselmuiden en voor de gezelligheid op en buiten het werk, bijvoorbeeld toen we in Wenen naar een *Heuriger* gingen. Ik zal vermoedelijk nooit een grote wijnliefhebber worden, maar de Veltliners van jullie boerderij waren de beste die ik ooit geproefd heb. Ik weet dat je naar eigen woorden iemand van weinig woorden bent, maar ik waardeer onze (best lange) gesprekken altijd enorm. Ik wens jou en je dierbaren, vooral Bruno natuurlijk, veel geluk toe.

Lourens, ik had aan het begin van het project nooit kunnen voorspellen hoeveel we zouden gaan samenwerken in de daaropvolgende jaren. *Streektaalstrijd*

begon achteraf naïevelijk als een zijproject en is geleidelijk aan uit de kluiten gegroeid tot een spel dat duizenden keren gedrukt is. Het is misschien maar goed dat we niet hebben bijgehouden hoeveel uren we hieraan gewerkt hebben in totaal, maar het resultaat mag er zeker wezen. Ik ben je ook zeer dankbaar voor het meerijden naar de opnames in Sint Annaparochie en ik heb veel respect voor hoe je jezelf ontwikkelt in het onderzoeksveld en als persoon. Ik ben benieuwd naar waar je je gaat vestigen uiteindelijk, maar je brengt er in ieder geval reusachtig veel kennis heen.

Reihaneh, het was volgens mij een aangename verrassing voor je om iemand in het lab te hebben die bekend was met Iraanse muziek. Die verrassing was insgelijks, want ik ken niet veel mensen die regelmatig muziek luisteren uit deze regio. Verschillende Iraanse muziekmeesters hebben we met veel plezier opgezet onderweg naar de opnamelocaties in Fryslân, waar we beiden opnames moesten maken met Friese sprekers. Ik wil je bij deze graag bedanken voor je hulp bij de opnames en dat je me hebt uitgenodigd voor (een onderdeel van) *Noroez*, wat een unieke en speciale ervaring was.

Teja, ik ben je uiteraard zeer dankbaar voor je hulp met het testen van mijn experimenten en voor het meerijden naar opnamelocaties. Je hebt zeker in de eerste paar jaar van mijn project veel van het praktische geregel binnen het lab op je genomen en daar zijn we je allemaal dankbaar voor. Je bent ook een heel vrolijk persoon en ik kan altijd goed met je lachen. Ik heb misschien nog wel het meest genoten van onze gedeelde interesse in zowel koken als eten, waarvan het hoogtepunt was toen je bij Fatoush simpelweg zuchtte en hardop zei *I love food*. Je hebt een warme en zorgzame persoonlijkheid waar mensen, waar je ook heengaat, zich gelukkig mee mogen prijzen.

Thomas, ik kan altijd goed met je lachen en je poëtische e-mails voorafgaand aan bijeenkomsten zal ik sterk missen. Je hebt ook bij verre de grootste rol gespeeld tijdens mijn opnameavonturen en ik ben je hier zeer dankbaar voor. Je hebt ook enkele sterke wijsheden gedeeld, zoals “remmen is angst” en “soms moet je voorrang nemen”, die ik tegenwoordig uiteraard volg. Je hebt het zelfs op je genomen om als een van mijn paranimfen te functioneren, dus het mag duidelijk zijn dat mijn promotieproject van begin tot eind minder leuk en succesvol was geweest zonder jou.

In mijn studietijd bij de Rijksuniversiteit Groningen hebben veel mensen deze tijd leuk en memorabel voor me gemaakt. Tijdens mijn bachelor waren dit onder anderen **David**, **Iris**, **Karin** en **Thomas (Seelemann)**. Ik bedank ook de hechte groep medestudenten van tijdens de onderzoeksmaster: **Alessandro**,

Anne-Marie, **Annika**, **Iris**, **Louisa**, **Matthias**, **Mieke**, **Nora**, **Penny** en **Tom**. De meeste van jullie zijn later zelfs collega's geworden op andere universiteiten of dezelfde universiteit.

Matthias, bedankt voor alle steun en gezelligheid in de afgelopen jaren en dat je als een van mijn paranimfen wilde functioneren. Doordat ik getuige was bij je bruiloft heb ik tevens, als waarnemer van je echtelijke plichten, een extra goede reden om de gezelligheid ver in de toekomst voort te zetten. Je bent een pittige gesprekspartner, maar ik verveel me nooit als we afspreken.

Penny, bedankt voor je eindeloze optimisme, aanstekelijke lach en onze hilarische gesprekken. Ik heb er verder bewondering voor hoe je van tijd tot tijd simpelweg besluit dat je iets gaat doen wat je spannend vindt, zoals zangles nemen, en dat gewoon doorzet en ervan groeit als persoon. De bekende Groningse uitdrukking *kop der veur* lijkt zeker op dat soort momenten goed bij je te passen.

Daan, het was erg leuk om onze gezamenlijke roots te herontdekken. Het lijkt erop dat onze paden toch eigenlijk altijd wel weer kruisen en misschien moeten we dat maar als een geschenk uit de hemel zien. Zo zie ik het in ieder geval zeker. Ik bedank je ook voor het meegaan richting Didam voor de opnames, wat de verste locatie was vanuit Groningen waar ik heen ben geweest. Verder vind ik het erg leuk dat je soms vraagt om gewoon eens ergens thee te gaan drinken of te lunchen, al is dat vooral om de thee te delen in plaats van te consumeren. Dat is niet iets wat ik regelmatig deed met vrienden voor jou, maar laten we het vooral blijven voortzetten in de toekomst.

Matthew, ik ben blij dat je een van de weinige mensen was die mijn poster bezocht bij een online congres tijdens de pandemie. Dat was geen toeval, want je wilde me blijkbaar al langere tijd eens spreken over mijn werk en dat kwam goed uit zo. Je passie voor het veld is indrukwekkend en je bent een uitzonderlijk enthousiaste en goede luisteraar, waar ik van heb genoten op congressen en toen we elkaar bezochten. Laten we maar gauw weer afspreken, zodat je me bij kunt praten over je meest recente interessante taalvariatiebevindingen.

Janine, ik hoop dat je ermee kunt leven dat ik je in het Nederlands adresseer en niet in het Gronings, want dat gaat me beter af. Ik schrijf wel met een grotere regelmaat korte berichten in het Gronings en soms komt een woord of gezegde zelfs eerder tot me in het Gronings dan in een andere taal, dus er komt langzaamaan wel wat schot in mijn Gronings. Je taalactivisme is inspirerend en ik twijfel er niet aan dat je dit voort blijft zetten in de toekomst. Ik wil je in het bijzonder bedanken voor het meegaan naar opnamelocaties en je interessante onderzoek naar Nedersaksische taalvariatie, waar ik ook een beetje aan bij heb

kunnen dragen. Ik wens je veel geluk toe in de toekomst en hoop dat je je maar lang mag bezigen met je vele interesses, van vogelspotten tot het Nedersaksisch in al zijn vormen.

Men zegt wel dat gedeelde smart halve smart is, dus vanzelfsprekend ben ik mijn lotgenoten die in hetzelfde jaar met hun promotieproject begonnen zijn ook dankbaar: **Lourens**, **Marjolein**, **Nelleke** en **Emma**. Dankzij onze onregelmatige lunches en patatavonden is de maximale smart toch viermaal gehalveerd naar 6.25% en ik hoop dat ook voor jullie gold. **Marjolein**, ik bedank jou nog even apart, omdat de combinatie van bijzondere gedachtesprongen en je gevoel voor humor altijd leiden tot heel grappige gesprekken. Je bent veel competitiever ingesteld dan ik, wat uiteraard heel nuttig was toen we *Streektaalstrijd* aan het testen waren en waarvoor ik je erg dankbaar ben. Zoals je weet is je iconische overwinning in het promofilmpje in Wenen bewonderd en via het internet misschien zelfs nog op veel meer plekken in de wereld. Je staat ook altijd klaar om anderen te helpen en ik vind dat bewonderenswaardig. Tenslotte wil ik je graag meegeven dat je weliswaar de verkeerde terminologie aanhangt voor gefrituurde staafjes aardappel aan deze kant van de rivieren, maar dat de koppigheid waarmee je *friet* aanhangt in plaats van *patat* je ook wel weer siert.

Mijn collega's van de GroNLP-onderzoeksgroep (*Groningen Natural Language Processing*) hebben me veel geleerd over de computationele tak van de taalwetenschap. De lijst van namen is flink gegroeid sinds het begin van mijn project en er zijn ook vaak mensen op bezoek vanuit andere universiteiten. Desalniettemin doe ik mijn best om iedereen bij naam te noemen: **Ahmet**, **Ana**, **Andreas**, **Antonio**, **Arianna**, **Beatrice**, **Chunliu**, **Federico**, **Francesca**, **Franziska**, **Gabriele**, **Gertjan**, **Gosse (Bouma)**, **Gosse (Minnema)**, **Greta**, **Huiyuan**, **Iris**, **Jirui**, **Johan**, **Khalid**, **Kun**, **Kyo**, **Leonidas**, **Lisa**, **Lukas**, **Malvina**, **Maria**, **Masha**, **Nastja**, **Noa**, **Pritha**, **Rik**, **Shaozhen**, **Silvia**, **Susan**, **Tommaso**, **Xiao**, **Xiaoyan**, **Yanan**, **Yevgen**, **Yongjian**, **Yuqing** en **Ze**. Ik wil jullie bedanken voor jullie interesse in en interessante vragen over mijn onderzoek. Jullie zijn ook als geheel een maatschappelijk betrokken groep, wat ik bijvoorbeeld heel duidelijk vond toen we *An Evening with ChatGPT* organiseerden destijds. Ik zal het t-shirt met de befaamde koe met trots blijven dragen. Ik bedank in het bijzonder **Rik** voor zijn actieve inzet om mij bij de groep te betrekken in het begin van mijn project. Ik bedank ook **Gosse B.** voor zijn geduldige bijdrage aan de pilot van mijn onderzoek, ondanks de irritante technische problemen die zich voordeden en ik niet op tijd genoeg opgelost had. Verder bedank ik ook nog specifiek **Ana**, **Andreas**, **Gosse M.** en **Khalid** voor de gezellige vrijdagmiddagborrels.

Naast de ondersteuning vanuit academische collega's heb ik ook vaak hulp nodig gehad van niet-wetenschappelijke medewerkers. Ik bedank bij deze onder andere de portiers, schoonmakers, collega's van de audiovisuele afdeling, de IT-afdeling, zalenbeheer en de universiteitsbibliotheek. Zonder jullie zou de wetenschap niet goed kunnen functioneren.

Ik wil ook graag enkele mensen in het kader van *Streektaalstrijd* bedanken, zoals **Thomas (van Halteren)** voor zijn waardevolle advies over de mogelijke spelmechanieken en **Marjolein (te Winkel)** voor haar zeer effectieve hulp met het bereiken van de pers. Mijn dank gaat ook uit naar de mensen van het Spellenmaakgilde en de Wirwar voor hun waardevolle adviezen over hoe het spel interessant te houden was voor een breed publiek. Verder bedank ik alle medewerkers in de journalistiek, in het bijzonder van regionale omroepen, bij wie ik vaak iets heb kunnen vertalen over het bordspel en streektalen in het algemeen. Jullie hebben geholpen om streektalen te belichten en hier ben ik erg dankbaar voor.

Ik wil ook de honderden deelnemers bedanken die aan mijn taalvariatieonderzoeken hebben bijgedragen de laatste paar jaar. Dit gaat zijn niet alleen de deelnemers die ik thuis bezocht heb met het mobiele laboratorium of naar ons toe kwamen, maar ook de vele festivalgangers van bijvoorbeeld de Zwarte Cross of het GRN Festival die ons hebben verblijd met hun regionale taalvarianten. Het was een eer om jullie regionale culturen en identiteiten gedeeltelijk vast te mogen leggen. Het was ook een genoegen om de persoonlijke verhalen te horen over hoe de streektalen een rol spelen in jullie levens. Ik heb mijn best gedaan om jullie een eer aan te doen in dit proefschrift.

Er waren veel inspirerende docenten die me op weg hebben geholpen richting mijn promotieproject. Ten eerste bedank ik **Mark (Huisman)** en zijn collega's van de onderzoeksgroep bij de afdeling sociologie, die mijn interesse in de statistiek hebben gevoed. Ik bedank ook **Jennifer (Spenader)**, dankzij wiens enthousiasme over dit enorme veld binnen de bachelor kunstmatige intelligentie ik uiteindelijk taalwetenschap ben gaan studeren. Verder wil ik ook onder andere **Bart**, **Dicky**, **Jack**, **Jan-Wouter** en **Simone** bedanken voor hun inspirerende colleges in de bachelor taalwetenschap. Ik bedank ook alle studenten die ik les heb gegeven, bijvoorbeeld over statistiek of programmeren in Python. Het verveelt nooit om te zien hoe jullie door een vak heen steeds meer leren en enthousiaster worden over jullie nieuwe kennis en vaardigheden.

Als laatste wil ik nog een aantal vrienden en familieleden benoemen die ik nog niet eerder genoemd heb. **Anton**, **Bart-Peter**, **Dian**, **Iris**, **Janko**, **Jeroen**,

Jikke, **Niels**, **Nilab**, **Renout**, **Teggo** en **Yde**: jullie geven het leven kleur en laten het sprankelen. Ik bedank ook mijn zussen **Angela** en **Warsha** en mijn vader en moeder **Siep** en **Elsbeth**, zonder wiens stevige fundering ik niet de juiste persoon geweest was om dit avontuur te beginnen en af te ronden. Ik bedank ook mijn bredere familie voor hun altijd bemoedigende instelling en in het bijzonder nog oma **Sara** en (mijn Groningse) oma **Henny**, die me altijd hebben aangemoedigd om mijn best te doen op school. Ik verbaas mezelf wel eens met de bijzondere dingen die ik meemaak en voor elkaar krijg, maar zij zouden hier nooit aan hebben getwijfeld.

𐑯𐑥𐑩𐑗𐑾𐑲

Namárië, ofwel: het ga jullie goed!

Raoul Buurke

in zijn geliefde Groningen
28 november 2024

CONTENTS

I Introduction **1**

1 Introduction **5**
1.1 Overview of regional languages in the Netherlands 6
1.2 Changing speaker populations 12
1.3 Language change 15
1.4 Research questions and chapter overview 22

II Speaker population changes **25**

2 Estimating speaker population sizes **29**
2.1 Introduction 29
2.2 Lifelines questionnaire 32
2.3 Applying post-stratification 34
2.4 Using reference works and intergenerational transmission 37
2.5 Discussion 46
2.6 Conclusion 52

3 Factors associated with regional language transmission **55**
3.1 Introduction 55
3.2 Methods 58
3.3 Results 61
3.4 Discussion 66
3.5 Conclusion 70

III Pronunciation changes **71**

4 Reducing transcriber variability in phonetic corpora **75**
4.1 Introduction 75
4.2 Methods 78
4.3 Results 86
4.4 Discussion 91
4.5 Conclusion 93

5 Estimating pronunciation change **97**
5.1 Introduction 97
5.2 Data 103

5.3 Methods 110
5.4 Results 115
5.5 Discussion 117
5.6 Conclusion 122

6 Re-estimating pronunciation change 125
6.1 Introduction 125
6.2 Data 128
6.3 Method 135
6.4 Results 138
6.5 Discussion 144
6.6 Conclusion 148

IV Conclusions 151

7 An ambispective view on Frisian and Low Saxon 155
7.1 Discussion 155
7.2 Suggestions for future studies 161
7.3 Conclusion 164

V Appendices 167

Appendix A Speaker count addendum 169
A.1 Speaker count components 169
A.2 Geographical distributions of language transmission 172
A.3 Estimates for the municipalities of Oost- and Weststellingwerf 174

Appendix B Reducing transcriber variability addendum 175
B.1 Phonetic symbols of the GTRP and RND corpora 175
B.2 Using more reduced symbol inventories 178

Appendix C Pronunciation change estimates addendum 181
C.1 Analysis using data from younger generation 181
C.2 Simulation of different sample sizes 183

Summaries in English and Dutch 185

About the author 197

Bibliography 201

Groningen Dissertations in Linguistics (GRODIL) 225

FIGURES

1.1 Map showing the major language varieties of Low Franconian in the Netherlands 8
1.2 Map showing the major language varieties of Frisian and Low Saxon in the Netherlands 9

2.1 Recruitment message sent to potential Lifelines respondents. 35

3.1 Marginal interaction effects of regional language use with the other parent, educational attainment, and age of acquisition. 63
3.2 Marginal interaction effects of environmental address density and positive language attitude. 64
3.3 Marginal effects of regional language use frequency in different environments and the number of languages a respondent speaks. 64
3.4 Marginal geographical effect smooth of the final model predicting regional language transmission. 65

4.1 Overlapping recording locations between the RND and GTRP corpora. 80
4.2 Predicted proportions of pronunciation change in the Netherlands and Flanders. 89

5.1 Traditional dialects in the province of Fryslân. 99
5.2 Dialect map of the Low Saxon area. 100
5.3 Recording locations of the GTRP and DIAREG corpora. 107
5.4 Estimated marginal interaction effect between change direction and a binary distinction between Westphalian Low Saxon and the other regional varieties. 116

6.1 Photo of the exterior of the mobile laboratory. 130
6.2 Map of the 32 overlapping locations between the GTRP and SPRAAK-LAB corpora. 131
6.3 Dialectometric maps of the pronunciation variation. 139
6.4 Estimated marginal effects of the final model. 142
6.5 Marginal geographical effects of the final model. 143

A.1 Geographical distribution of the speakers who transmit Frisian to their children. 172

A.2 Geographical distribution of the speakers who transmit Low Saxon to their children. 173

B.1 Proportions of pronunciation change predicted in the Netherlands and Flanders using different frequency constraints. 180

C.1 Estimated marginal interaction effect between change direction and a binary distinction between Westphalian Low Saxon and the other regional varieties (based on data from GTRP speakers and younger women in the DIAREG). 182

C.2 Simulation of different target word sample sizes in dialectometric studies. 183

TABLES

2.1 Multiplication factors for the age variable to match the sample ratios to population ratios in 2021. 36
2.2 Estimated percentages for speaking and using Low Saxon in Groningen and Drenthe. 43
2.3 Reference percentages for speaking and using Low Saxon. 43
2.4 Estimated percentages for speaking and using Frisian in Fryslân (excluding Oost- and Weststellingwerf). 45
2.5 Reference percentages for speaking and using Frisian. 45

3.1 Coefficients for a logistic model predicting regional language transmission. 61
3.2 Smooth terms for a logistic model predicting regional language transmission. 62

4.1 Overlapping target words between RND and GTRP corpora. 79
4.2 Levenshtein alignment between variants of Dutch *straat* 'street'. 81
4.3 Example of a PMI-based segment distance matrix. 83
4.4 IPA symbols that occur only in a subset of the GTRP (Dutch part: -NL, Flemish part: -BE) and RND. 86
4.5 Combined phonetic symbol inventory 88

5.1 Overlapping target words between the GTRP and DIAREG corpora. 106
5.2 Number of available words per comparison. 108
5.3 Characteristics of speaker age and recording year per corpus. 109
5.4 3D Levenshtein alignment example for variants of *straat*. 112
5.5 Recording locations associated with each dialect group. 115
5.6 Coefficients for a generalized additive model predicting change based on a binary distinction between Westphalian Low Saxon and the other regional varieties. 116

6.1 Metadata per reference location. 132
6.2 Standard Dutch target words for the word list translation task. 133
6.3 Phonetic symbols used in the combined corpus. 136
6.4 Parametric coefficients of a model predicting convergent (conv.), divergent (div.), and neutral (neutr.) segments. 141
6.5 Smooth coefficients and random effects of a model predicting convergent (conv.), divergent (div.), and neutral (neutr.) segments. 141

A.1 Percentages per level of Low Saxon speaking proficiency across different age groups. 169
A.2 Percentages of Low Saxon use at home across different age groups. 169
A.3 Adjusted percentages of Low Saxon speakers in Groningen and Drenthe. 170
A.4 Adjusted home use percentages of Low Saxon (LS) speakers in Groningen and Drenthe. 170
A.5 Percentages of speakers living in Fryslân able to speak Frisian and using Frisian at home in Fryslân across different age groups. 170
A.6 Adjustment components for Low Saxon speakers in Groningen and Drenthe. 171
A.7 Adjustment components for Frisian speakers in Fryslân. 171
A.8 Adjustment components for Low Saxon speakers in the municipalities of Oost- and Weststellingwerf. 174
A.9 Estimated percentages for speaking and using Low Saxon in the municipalities of Oost- and Weststellingwerf. 174

B.1 Occurrence of all IPA symbols in the GTRP and RND corpora. 175
B.2 Combined phonetic symbol inventory with a 2.5% frequency constraint. 178
B.3 Combined phonetic symbol inventory with a 5% frequency constraint. 179

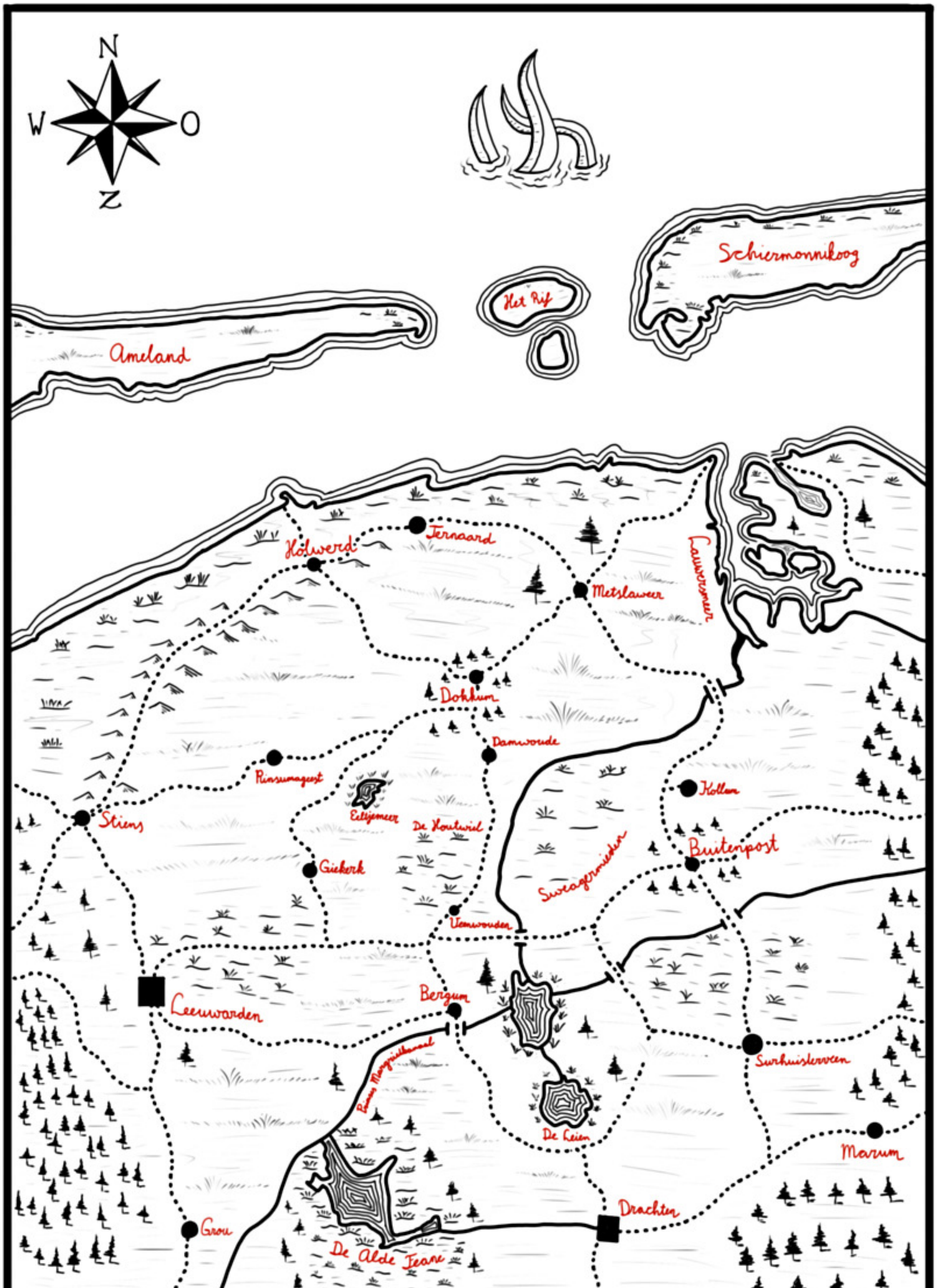
C.1 Coefficients for a generalized additive model predicting change based on a binary distinction in Low Saxon (based on data from GTRP speakers and younger women in the DIAREG). 181

Part I

Introduction

Dokkum, Fryslân

CHAPTER 1


N
W
O
Z
Schiermonnikoog
Het Rif
Ameland
Ternaard
Holwerd
Metslaweer
Lauwersmeer
Dokkum
Damwoude
Rinsumageest
Kollum
Stiens
Eeltjemeer
De Houtwiel
Buitenpost
Giekerk
Sweagermieden
Veenwouden
Bergum
Leeuwarden
Surhuisterveen
Prinses Margrietkanaal
De Leien
Marum
Drachten
Grou
De Alde Feanen

Voor Thomas

INTRODUCTION

Recent estimates indicate that approximately 1,500 languages may no longer be spoken by the end of the 21st century (Bromham et al., 2021), which is a significant reduction of the approximately 7,000 languages in the world at the beginning of the century (Austin & Sallabank, 2011, p. 1). Most of this linguistic loss has resulted from colonization and ongoing globalization, which has caused some of the world's languages to be viewed as more valuable than others. Not everyone may view this large-scale language loss as a problem, because it appears more economical if everyone speaks the same language. Economically speaking, however, language loss leads to reduced cultural diversity and cultural diversity is an important driver of innovation (e.g., in firms, due to a more varied knowledge base; Østergaard et al., 2011). Additionally, local languages promote a community's cohesion and self-confidence (Crystal, 2014, p. 40) and are often perceived as an important part of regional identity (Van der Star & Hochstenbach, 2022), so their loss can be felt both in societies and smaller regions.

The Council of Europe acknowledged in 1992 that the regional and minority languages of its member states contribute to Europe's cultural wealth and proposed the European Charter for Regional or Minority Languages (ECRML), which consists of three parts describing different degrees of protection. Regional or minority languages are, according to Article 1 of the ECRML, languages that are traditionally used in a specific territory by nationals of that member state, and languages that are not an official language (everywhere) in the member state. The governments of EU member states were invited to sign and ratify the Charter and could choose under which parts to include specific languages (Part I is always applicable, so either Part II or Parts II and III could be chosen). As of 2024, seven languages are included under Parts II and III of the Charter, and 75 languages are included only under Part II.[1]

The Netherlands is one of the 25 EU member states that signed and ratified the ECRML (in 1996). An overview of the regional languages in the Netherlands follows below, but we focus mainly on two regional languages in this dissertation: Frisian and Low Saxon. These languages neighbor each other in the Netherlands, but Frisian is included under Parts II and III of the ECRML, while Low Saxon is only included under Part II. Consequently, Low Saxon has less support from

[1] See https://rm.coe.int/november-2022-revised-table-languages-covered-english-/1680a8fef4 for a recent overview.

the government than Frisian, although support for Frisian is arguably still limited compared to Dutch (the standard language of the Netherlands).

Given this context, the main aim of this dissertation is to compare how well Frisian and Low Saxon fare nowadays based on estimated speaker counts of these languages and how likely their speakers are to pass on the language to the next generation. Additionally, we aim to assess how the pronunciation variation of these languages is changing. We focus on how varieties of these languages are changing in the context of both the ubiquitous Standard Dutch and other local varieties, providing a more complete picture of the current state of these languages.

Investigating these population and pronunciation change processes in conjunction is beneficial, because tracking speaker population developments facilitates the interpretation of pronunciation change patterns. For example, the local varieties in a particular area may show relatively much pronunciation change. If the corresponding population has recently declined rapidly, the speakers may shift to a more prestigious variety and actively borrow from that variety. However, when intergenerational transmission is stable, it becomes less likely that the observed pronunciation changes derive primarily from the influence of a more prestigious variety. In that case, additional explanations should be explored, such as the presence and relevance of a regional standard. In other words, population and pronunciation metrics provide insights into the state of regional languages, but they are most informative when considered together.

This introduction is structured as follows. We provide an overview of the regional language variation in the Netherlands in Section 1.1. Afterward, we discuss prior speaker counts of Frisian and Low Saxon and which factors are likely associated with language maintenance in Section 1.2. The dynamics between Standard Dutch and regional languages in the Netherlands are delineated in Section 1.3, and we also shortly describe the corpora we use. Finally, we summarize the research questions and provide an overview of the remainder of the dissertation in Section 1.4.

1.1 Overview of regional languages in the Netherlands

The languages traditionally spoken in the Netherlands are varieties of the Low Franconian, Low Saxon, and Frisian branches of the West Germanic language family. The maps in Figures 1.1 and 1.2 indicate where these languages are spo-

ken in the Netherlands, based on several established classifications (e.g., Hinskens & Taeldeman, 2013; Van de Velde et al., 2019; Bloemhoff et al., 2020). None of these languages are exclusively spoken in the Netherlands, because varieties of Low Franconian are also spoken in Belgium and Germany, and varieties of Frisian and Low Saxon are spoken in in Germany. The standard language spoken in the Netherlands, Standard Dutch (see Section 1.3.1 for details), is part of the Low Franconian branch and is usually also spoken by regional language users. We will describe the main groups within these languages in the Netherlands below, but admit that the definitions of such groups always depend on the metrics used for the grouping. For each language, we offer references for further reading.

1.1.1 Varieties of Low Franconian — The major dialect groups within Low Franconian are presented in Figure 1.1. Borders between dialects are not indicated on the map, because they are usually gradual in the dialect landscape of the Netherlands (Wieling & Nerbonne, 2015). The Low Franconian dialects are mostly not discussed in this thesis (except in Chapter 4), and we only provide a short overview of them below. We refer the interested reader to Hinskens and Taeldeman (2013) for an extensive overview of the linguistic and sociolinguistic features of these varieties.

Wieling et al. (2011) showed that the Low Franconian varieties in Figure 1.1 can be ordered from least to most similar to Standard Dutch as follows. The dialects of *Hollands* 'Hollandic' comprise the largest part of this area, and the pronunciation patterns of these varieties are the most similar to Standard Dutch (Wieling et al., 2011; Breder Birkenes & Pheiff, 2022), because Standard Dutch is derived substantially from the historical Hollandic dialects. Another contributor to the emerging Dutch koine (a common dialect arising from dialect contact and mixing) of the 16th and 17th centuries were the dialects of *Brabants* 'Brabantish', when there was large-scale migration to the Holland region (Howell, 2006; Goss & Howell, 2006). Nowadays, the dialects of Brabantish are still highly similar to Standard Dutch, also due to the extensive language contact in recent centuries (De Schutter, 2013). The dialects of *Zeeuws* 'Zeelandic' are similar to West Flemish varieties, which are spoken in Flanders, but the pronunciation of Zeelandic is sufficiently different to be detected as a separate dialect group (Heeringa, 2004, p. 228).

The pronunciation of the dialects of *Limburgs* 'Limburgish' is the most distinct from Standard Dutch in the Low Franconian group, as Limburgish includes

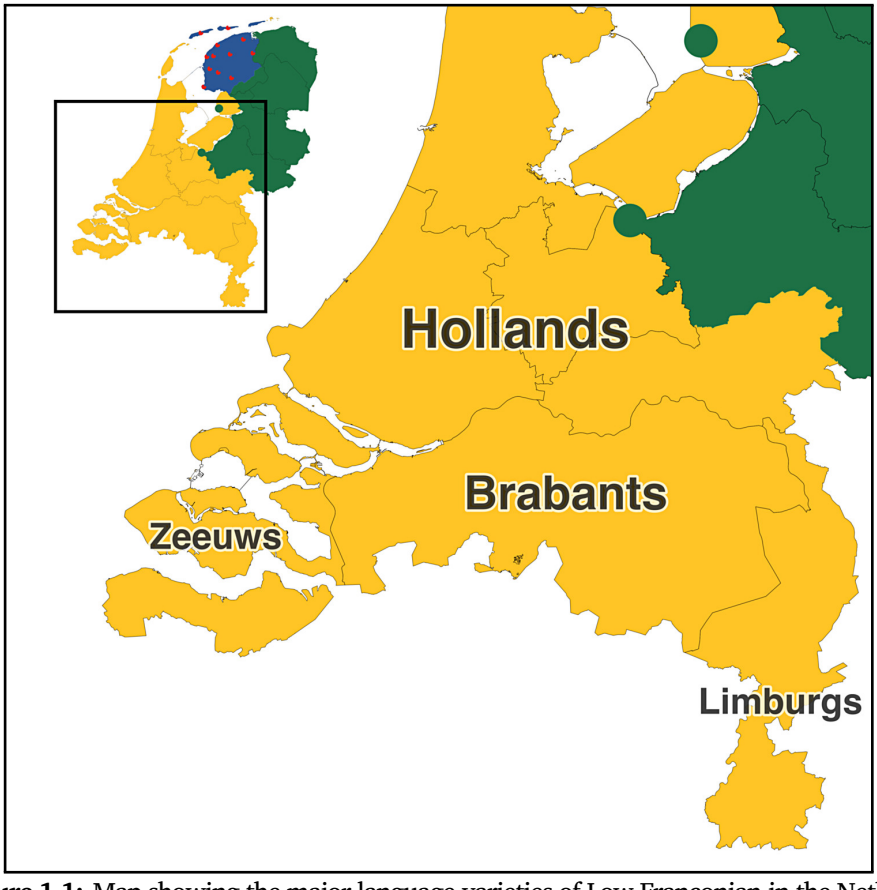


Figure 1.1: Map showing the major language varieties of Low Franconian in the Netherlands. The Low Franconian and Low Saxon varieties are marked in yellow and green, respectively.

phenomena that are rare in the Netherlands (e.g., tonal accents and voicing across word boundaries; Hermans, 2013). Pronunciation patterns differ substantially between Limburgish dialects, however, especially between the northern and southern dialects (Bakker & Van Hout, 2017). Limburgish is the only Low Franconian regional language that is included under the ECRML, specifically under Part II, which means it has substantially more governmental funds for promotion and protection than Zeelandic or Brabantish (Swanenberg, 2013).

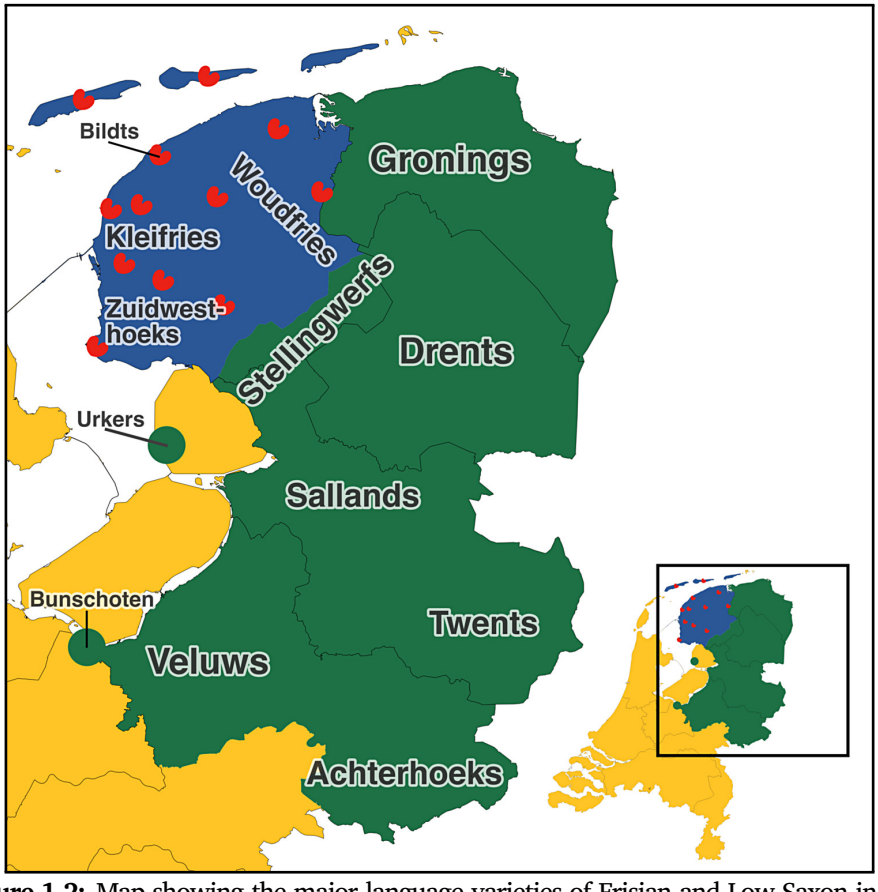


Figure 1.2: Map showing the major language varieties of Frisian and Low Saxon in the Netherlands. The Low Franconian, Low Saxon, and Frisian varieties are marked in yellow, green, and blue, respectively. The red icons indicate Frisian-Hollandic contact varieties.

1.1.2 Varieties of Frisian — The major varieties of Frisian in the Netherlands are presented in Figure 1.2 and are mainly spoken in the province of Fryslân. Frisian is historically and linguistically distinct from Standard Dutch. Old Frisian, for example, already developed separately from the languages that would become Standard Dutch and Low Saxon in the Early Middle Ages and was spoken in a large area along the coast of the North Sea (Bremmer, 2008). The West Frisian varieties currently spoken in the province of Fryslân (referred to as 'Frisian' throughout this dissertation) are usually clustered into three groups (Hoekstra, 2003): *Wâldfrysk* (Dutch: *Woudfries* 'Wood Frisian'), *Klaaifrysk* (Dutch: *Kleifries*

'Clay Frisian'), and *Súdwesthoeksk* (Dutch: *Zuidwesthoeks* 'Southwestern Frisian').[2] These varieties are mutually intelligible (Stefan, 2022, p. 53) and typically cluster as a single language variety when regional language varieties in the Netherlands are aggregated (Nerbonne et al., 1996). Frisian is perceived as the regional language that sounds most distinct from Standard Dutch, and mutual intelligibility between Dutch and Frisian speakers is lower than between Dutch and Low Saxon (Van Bezooijen & Van den Berg, 1999). Frisian is also documented and studied relatively well as a regional language in the Netherlands (see, e.g., Munske et al., 2001), while research on regional varieties of Low Saxon and Low Franconian is much rarer. Frisian is the only language that is included under Parts II and III of the ECRML, so its promotion and protection are financially supported most extensively by the Dutch government.

There are also numerous Frisian-Hollandic contact varieties in Fryslân (see Figure 1.2). These varieties are collectively called Town Frisian in this dissertation, because they often cluster together (Heeringa, 2005). The varieties of *Bildts* in the northwest of mainland Fryslân can be seen as relatively distinct within this group (Duijff, 2002). The origin of the Frisian-Hollandic contact varieties is typically dated to the early 16th century (Fokkema, 1937; Van Bree, 2001), and they are spoken in traditionally Frisian areas. Many Hollandic settlers arrived in these regions in the Middle Ages, and the extensive language contact resulted in periods of change toward Frisian at some times and toward Dutch at other times (Versloot, 2021b). Town Frisian varieties were relatively prestigious until the 20th century, but they lost prestige afterward (Bloemhoff et al., 2010, p. 724). These varieties are generally more similar to Standard Dutch than Frisian and Low Saxon varieties are similar to Standard Dutch.

1.1.3 Varieties of Low Saxon — The Low Saxon language is the largest regional language in the Netherlands, being spoken in multiple provinces in the Netherlands (see Figure 1.2). Varieties of Low Saxon used to be spoken in a much larger area in Europe, when it still was the lingua franca of the Hanseatic League and widely used as a written language (usually referred to as Low German in this context; Langer, 2003). Nowadays, Low Saxon varieties are mainly spoken in the northern and eastern Netherlands and northern Germany. The Low Saxon varieties around the border between the Netherlands and Germany are becoming more similar to their respective standard languages (Smits, 2011),

[2] Some authors also distinguish *Noordkleifries* 'North Clay Frisian' (Hof, 1933), but these varieties cluster together with Clay Frisian (Heeringa, 2005).

which causes the local Low Saxon dialects to diverge from each other. All Low Saxon varieties in the Netherlands are relatively distinct from Standard Dutch, though they differ in their degree of similarity to Standard Dutch (Wieling et al., 2011). We refer to Bloemhoff et al. (2008b) and Bloemhoff et al. (2020) for detailed descriptions of Low Saxon pronunciation patterns.

The main varieties of Low Saxon in the Netherlands are *Gronings* 'Groningen dialect', *Stellingwerfs* 'Stellingwerven dialect', *Drents* 'Drenthe dialect', *Sallands* 'Salland dialect', *Twents* 'Twente dialect', *Veluws* 'Veluwe dialect', and *Achterhoeks* 'Achterhoek dialect'. *Urkers* 'Urk dialect' is spoken in the province of Flevoland and was included in Low Saxon protection efforts under the *Convenant Nedersaksisch* 'Covenant Low Saxon' in 2024.[3] Urk dialect is still actively used across generations and using the language is viewed as an important part of local identity (Kruize, 2012; Bloemhoff et al., 2019, p. 106). Finally, the Low Saxon dialect in the municipality of Bunschoten in the province of Utrecht has also recently been included under the Covenant Low Saxon.

Substantial pronunciation differences appear within the Low Saxon group when Low Saxon varieties in the Netherlands are clustered. A northern and a southern cluster can be distinguished, which we dub the Northern Low Saxon and Westphalian Low Saxon clusters (Nerbonne & Heeringa, 2001; Heeringa, 2004). The Northern Low Saxon cluster includes the varieties of Groningen dialect and the northern varieties of Drenthe dialect. These varieties originated in the Late Middle Ages in an area where Frisian was spoken (Van Bree, 2017) and were relatively similar to Low Saxon varieties in Northwestern Germany (Bloemhoff et al., 2008b, p. 160) until the varieties on different sides of the national borders started to converge to the respective national languages (Smits, 2011). The other Low Saxon varieties are part of the Westphalian Low Saxon cluster, as these varieties generally have more linguistic innovations in common with language varieties from Westphalia (in Germany; Heeroma, 1953; Bloemhoff et al., 2013a).

Varieties of Drenthe dialect are often perceived as relatively similar to Standard Dutch, partially because there is no umlaut in diminutives and plural forms in these varieties, which is common in other Low Saxon varieties (Van den Berg & Van Oostendorp, 2012, pp. 61–62). The northern varieties are more similar to *Gronings*, while the southern varieties are more similar to those in the provinces of Overijssel and Gelderland.

The varieties of Stellingwerven dialect are situated around the Frisian-Low

[3] See https://www.rijksoverheid.nl/onderwerpen/erkende-talen/de-nedersaksische-taal.

Saxon border and likely originated as varieties of Drenthe dialect that became influenced by Frisian over time (Winkler, 1874; Sassen, 1953, p. 101). The pronunciation patterns of these varieties are commonly found to be influenced by Frisian, including the occurrence of word-initial [g] instead of [x], although this sound is rarer among younger generations (Bloemhoff, 2008a, p. 180).

The varieties of Salland dialect and Twente dialect in the province of Overijssel and varieties of Achterhoek dialect in the province of Gelderland are included in the Westphalian Low Saxon group. The varieties of Veluwe dialect in the province of Gelderland (and also in Bunschoten) may also be included in this group, although their similarity to Standard Dutch is stronger (Wieling et al., 2011) and the Westphalian innovations are not as prevalent for the westernmost Veluwe dialect varieties (Bloemhoff et al., 2019, p. 49).

1.2 Changing speaker populations

The European Union assesses the linguistic diversity of its member states, stressing that its protection is incorporated in both the European Charter of Fundamental Rights and the Treaty of the European Union (European Commission, 2018, p. 5). Despite the efforts of the European Union, it cannot be denied that regional languages in Europe are widely in decline (Auer et al., 2005). It is difficult to reliably assess how many regional or minority language speakers there are currently, because asking everyone in a country about their language use is very costly and labor-intensive.[4] Reliable estimates can be obtained through a sample survey instead, but such surveys require the population to be properly represented in a sample. It is generally rare for surveys to be fully representative, so samples often need to be corrected (e.g., using post-stratification according to known population parameters; Holt & Smith, 1979; Bethlehem, 2009, p. 250), mainly due to increasingly low response rates (especially in the digital age; Manzo & Burke, 2012). Assessing the size of regional and minority speaker populations is important for language policies concerning these languages, because it can determine their visibility and recognition in their respective societies (Duchêne & Humbert, 2018, p. 3).

[4] But note that it happens at regular intervals, for example, in the United States, countries formerly part of the Soviet Union, and the United Kingdom (Stevens, 1999; May, 2000; Silver, 2002; Gledhill, 2020; Duchêne & Humbert, 2018).

1.2.1 Speaker counts — The size of speaker populations of regional languages in the Netherlands declined in the 20th century (Goeman & Jongenburger, 2009; Versloot, 2021a). However, the rate of this decline strongly differs between the regional languages. Frisian and Limburgish have been declining, but at a considerably slower pace than Low Saxon or other regional Low Franconian varieties (Versloot, 2021a, p. 11). Different authors draw different conclusions regarding the fate of dialects in the Netherlands. Goeman and Jongenburger (2009, p. 62) remain agnostic about the future of dialects in the Netherlands, stating that their findings cannot be used for inference. However, Versloot (2021a, p. 12) explicitly posits that traditional dialects in the Netherlands and Flanders are now in irreversible decline, with possible exceptions of Frisian and Limburgish.

The size of the Frisian speaker population is regularly assessed by the *Fryske Akademy* 'Frisian Academy' and the provincial government of Fryslân (e.g., Pietersen, 1969; Jonkman & Gorter, 1995; Klinkenberg et al., 2018; Provincie Fryslân, 2020). However, estimates of the speaker population size can vary substantially depending on whether people were interviewed directly or filled in a survey (see Klinkenberg et al., 2018, p. 116). For example, 77% of the interviewed participants indicated using Frisian at home, while only 55% indicated doing so in the survey (arguably due to Frisian speakers being overrepresented in the interviews; Klinkenberg et al., 2018, p. 86).

The size of the Low Saxon speaker population has rarely been estimated. The most recent large-scale survey dedicated specifically to the use of Low Saxon is from nearly two decades ago (Bloemhoff, 2005). The reported overall percentage of people indicating they could speak Low Saxon was approximately 71%, while around 53% indicated using the language at home (covering the regions in Figure 1.2, except Urk and Bunschoten; Bloemhoff, 2005, p. 88). Versloot (2021a, p. 9) argues that these numbers seem high, and this appears to be the case when these estimates are compared to other estimates partially covering Low Saxon (e.g., Driessen, 2012; Schmeets & Cornips, 2022). Especially given the strong decline in the 20th century, it remains unclear how many Low Saxon speakers are left in the respective provinces today (see Figure 1.2). We therefore aim to determine this number in this dissertation. To determine the validity of our estimation approach, and to compare this information across language communities, we assess this question simultaneously for Low Saxon and Frisian.

1.2.2 Language transmission factors — Speaker counts are useful snapshots of ongoing language maintenance or decline, but exploring what moves people

to maintain their language is equally relevant. Language maintenance is affected by processes at the societal level (e.g., government policies; Edwards, 1992), but personal developments and preferences also play a role. For example, language maintenance may be negatively impacted by someone emigrating to another country (Schmid & Keijzer, 2009), negative language attitudes toward the regional variety (e.g., superimposed by society; Dragojevic et al., 2021), or the perception that using and maintaining the language is not economically useful (Harbert, 1999; Gao, 2009). Not all of these factors are equally important, but exploring the relative effects of these factors is relevant for organizations trying to protect and promote the languages (e.g., the Council of Europe and the Dutch government, as indicated by their inclusion of Low Saxon and Frisian under the ECRML).

One of the main mechanisms of language maintenance is language transmission from parents to their children, which is potentially the strongest indicator of language maintenance for all languages (Wölck, 2004, p. 7). A substantial drop in intergenerational transmission in a single generation can significantly contribute to language death in the long term (Fishman, 1997). Fewer people in that 'missing' generation can then teach the language to new speakers, which may lead to fewer potential teachers in the subsequent generation (unless there is a perfect transmission rate across generations). Family language policy studies show how parental inclination to transmit a minority language can depend on beliefs (e.g., incorrectly believing that children cannot successfully acquire multiple languages; Bialystok, 2011; Costa & Sebastián-Gallés, 2014), but also on the receptiveness of the child, who is also subject to external social pressures (King et al., 2008; Fogle & King, 2013).

Given the decline of Frisian and (especially) Low Saxon, we explore which social and personal factors are associated with the choice of parents to transmit their regional language to their child(ren) in recent generations. These neighboring languages present an interesting case, because they are similar in several ways (e.g., officially recognized as regional languages and being societally subordinated to Standard Dutch), but their speaker populations also seem to differ in certain respects. For example, positive language attitudes toward Frisian are relatively common among its speakers (Klinkenberg et al., 2018), while the situation is less straightforward for Low Saxon. Its speakers are also reported to have positive attitudes toward the language (Ter Denge, 2012; Braat, 2020), but the language is at the same time more commonly ridiculed (by both the speakers themselves and others in society; Bloemhoff et al., 2019, p. 103). Frisian speak-

ers also overwhelmingly view their language as a language separate from Dutch (Klinkenberg et al., 2018, p. 68), while people often either view or name Low Saxon dialects as dialects of Dutch (Bloemhoff, 2008b, p. 298). The transmission of Frisian furthermore benefits from the fact that it is taught in schools (in Fryslân), and its strong embedding in daily life in the province of Fryslân (where most of its speakers grow up).

1.3 Language change

From the regional language overview in Section 1.1, it is clear that the pronunciation similarities to Standard Dutch differ substantially between the regional varieties (see also Wieling et al., 2011; Breder Birkenes & Pheiff, 2022). The similarity to Standard Dutch also appears to have increased for many regional varieties in the 20th century (Heeringa & Nerbonne, 2000; Heeringa & Hinskens, 2015). Dutch influence on regional varieties is not new and was already ongoing in the 19th century, likely as a result of increased migration patterns due to industrialization (e.g., for Low Saxon; Bloemhoff et al., 2019, p. 101). This process may have accelerated in the past century due to the increased mobility of the Dutch population (Bek, 2022, p. 11) and the rise of mass media (Willemyns, 2003, p. 110), introducing Standard Dutch and the accompanying standard language ideology more widely in the Netherlands.

Language change usually occurs for individuals, but also for larger communities. The different processes driving these changes are complex and can be intricately linked, so it is unlikely that we can derive a simple model of which factors cause pronunciation change in the northern and eastern Netherlands. However, we can try to disentangle the different processes and discuss their joint contributions to the overall ongoing pronunciation change. In this dissertation, we focus on estimating pronunciation change in regional varieties at the level of communities (instead of smaller groups or individuals). We discuss the dynamics between regional language communities and between these communities and Standard Dutch below, but we first expand on the relative prestige and power of Standard Dutch.

1.3.1 Standard Dutch — Before discussing the political status and power of Standard Dutch, it is worth noting that Standard Dutch itself is a difficult concept to define. People view Standard Dutch as ‘correct’ Dutch and ascribe it

aesthetic qualities (Smakman, 2006, pp. 108, 160), which is a common finding for standard languages (Milroy, 2001; Smakman, 2012). This shows that Standard Dutch exists in the mind of Dutch speakers, but its pronunciation eludes a consistent linguistic description (Smakman, 2006). This is likely a consequence of the fact that the codification and standardization of Dutch has historically focused on writing and not on pronunciation (Willemyns, 2003; Stegeman, 2021). Another associated characteristic of Standard Dutch pronunciation is that it is perceived as non-regional and widely used (Smakman, 2006, p. 277). However, (regional) variation in spoken Standard Dutch is also accepted to a certain degree in the Netherlands nowadays (Grondelaers & Van Hout, 2010; Grondelaers et al., 2016), and Standard Dutch pronunciation also changes (e.g., Van de Velde et al., 1997; Stroop, 1998). Taken together, these findings indicate that people in the Netherlands are oriented toward this perceived Standard Dutch, but it is unclear whether people have a consistent view of which pronunciations are part of it and which are not.

Politically, the status of Standard Dutch is clear and cemented. The regional languages discussed in this dissertation have been subordinated to Standard Dutch since the country was formed in 1813, because the Netherlands followed the trend of 'one nation, one language' observed in many European countries in their nation-building practices (Haugen, 1966; Milroy, 2001; Burke, 2004). Standard Dutch is the language that Dutch citizens are taught from a young age onward, which they are expected to be proficient in, and it is expected they use the language for daily communication. It is also the only official language that applies everywhere in the Netherlands, although co-official languages exist in smaller areas of the country.[5] The influence of Standard Dutch on Frisian and Low Saxon speakers was relatively limited until primary school education became obligatory in the early 20th century, because Standard Dutch was not widely visible for or used by people who were not educated in it. Standard Dutch remained the only language allowed in education until Frisian became allowed in primary school in 1955.[6] The situation has not changed drastically for the most part, as Standard Dutch is essentially the language of instruction throughout all edu-

[5] By law, Frisian is only an official language in the province of Fryslân (see https://wetten.overheid.nl/BWBR0034047/2024-01-01) and not elsewhere in the Netherlands.

[6] This was mostly achieved in the aftermath of *Kneppelfreed* 'Baton Friday', which was a collision between Frisian speakers and the local police following a case of language discrimination against a Frisian farmer. This increased the awareness of Frisian, also leading to Frisian becoming an accepted language in court proceedings (Van der Velden, 2004).

cational phases in the Netherlands (although English is also commonly used for university courses). There are exceptions for Frisian, but its success as a language of instruction is limited (see Varkevisser & Walsweer, 2018; Bayat et al., 2023). Teaching materials are available to some degree for other regional languages, but there is little financial support for such purposes in these regions and many teachers are unaware of their existence (Doreleijers, 2021).

The sociolinguistic circumstances of regional languages in the Netherlands are unlikely to change, because the *Nederlandse Taalunie* 'Dutch Language Union', the governmental body responsible for language policy for Dutch, has only recently started tracking regional languages[7] and still mistakenly denotes Low Saxon as a dialect of Dutch instead of as a recognized separate language in a recent research report (Rys et al., 2021, p. 17). While the Dutch Language Union representatives admitted in the report (the version aimed at the general public) that the regional languages deserve more attention (Taalunie, 2021, p. 21), the emancipatory language policies of the Dutch government appear unlikely to be buoyed to new heights. The Dutch government's negligence and lackluster support of its regional languages have been noted in the ECRML evaluation reports of the Council of Europe.[8] Regarding Low Saxon, the Council recommended the Dutch government to take immediate action to set up a body to represent Low Saxon interests at the national level and to facilitate the teaching of Low Saxon in primary school. Additionally, they recommended ensuring a full-time professorship for Frisian studies and facilitating the use of Frisian in court. Failure to comply with the ECRML commitments has no legal consequences for the Dutch government, however, so only limited pressure can be applied by the Council.

1.3.2 Dynamics between Standard Dutch and regional languages — At the group level, there are three principal directions in which language varieties can change. Regional language varieties can become more similar (i.e., converge), less similar (i.e., diverge), or stay equidistant. Clusters of local dialects in the Netherlands have become more similar to each other in recent decades (Heeringa & Hinskens, 2015), resulting in a loss of local linguistic diversity together

[7] The Dutch Language Union also explicitly stated they will not support the recognized regional languages, as they cannot be supported as varieties of Dutch. See the 2019 report on language variation policy: https://taalunie.org/publicaties/88/visie-op-taalvariatie-en-taalvariatiebeleid.

[8] See https://www.coe.int/en/web/european-charter-regional-or-minority-languages/reports-and-recommendations.

with the retention of more regional dialectal features. These convergence patterns between local dialect clusters may lead to a new regional variety, bounded by a larger geographical region than traditional dialects, such as Koine Greek in classical antiquity (e.g., Bubenik, 1993). Such regional varieties do not necessarily need to replace the traditional dialects and can be used in tandem.

Another relevant dimension in the context of regional dialects is whether changes occur with respect to a 'roof variety' or the standard language, or whether they occur between dialects. Regional dialects in Europe often become more similar to the dominant standard language (Auer, 2018, p. 163). Heeringa and Hinskens (2015) found that convergence to Standard Dutch is generally stronger than divergence from Standard Dutch, and they also found that convergence to Standard Dutch usually coincides with convergence between dialects (which is commonly observed in Europe; Auer, 2005, p. 26). So-called 'regiolects' (a term usually ascribed to Hoppenbrouwers, 1990) can form in situations where this process continues structurally. Regiolects can be seen as regional varieties located at the level between the traditional dialects and the standard language, with features from both dialects and standard language. Some examples of regiolect formation follow hereafter, focusing on varieties in the Netherlands and Flanders.

A widely known variety that can be classified as a regiolect in this part of Europe is *Tussentaal* 'intermediate language' in Flanders (Vandekerckhove, 2009), which may be a consequence of Belgium's tumultuous sociolinguistic history since the 19th century. French was the language of power in the country until the Coremans-De Vriendt Law in 1898 (even though Belgium was already formed in 1830), which granted Dutch and French equal legal status. In the 1930s, Dutch became the only official language in Flanders. There was no competition between the standard language and traditional dialects until then, because there was no widely accepted standard. Afterward, Standard Dutch in Belgium was strongly oriented toward Standard Dutch in the Netherlands until the 1980s, but since then the Belgian standard has partially developed away from the Dutch one and has been adopted by the national news broadcaster (Haeseryn, 2010, p. 712). Traditional dialect features have receded from the 1970s onward, and are increasingly replaced by either standard language or regional dialect features (particularly from Brabantish varieties; Vandekerckhove, 2009, p. 76). It is unclear whether *Tussentaal* should be seen as a single variety or as a collection of regiolects (De Caluwe & Van Renterghem, 2011; Ghyselen, 2015), but it is perceptually salient to speakers in Flanders. These varieties usually take over the

function of the disappearing traditional dialects, especially in urbanized environments.

In the Netherlands, some regional varieties have also shown patterns that suggest the formation of regiolects. Regiolect use was first reported by Hoppenbrouwers (1990) in the Brabantish area (see Figure 1.1), and it appears widely in use by young people in the city of Eindhoven (Wilting et al., 2014). Doreleijers and Swanenberg (2023a) showed that hyperdialectisms (i.e., the exaggerated use of dialectal features that deviates from traditional grammar) frequently occur among users of the regiolect, because they are not sufficiently proficient in the traditional dialects. For the Limburgish area, Van de Velde et al. (2008) report that a regiolect exists, even though Standard Dutch is usually preferred when the social context does not allow local dialect use. Furthermore, De Tier et al. (2008) report that people in the Zeelandic area prefer using the local dialect, but that regiolect use is also common. To our knowledge, there is no evidence for regiolectization in the Hollandic area, where overall dialect use is generally low (with the possible exception of West Frisian Dutch; Berns & Steusel, 2004; Van Bree, 2004). Overall, there seems to be some evidence for regiolects in the Netherlands, although Versloot (2021a) argues that *Tussentaal* fundamentally differs from the regiolect patterns found in the Netherlands.

The case of Frisian is peculiar in the context of regiolectization, because two possible roof varieties exist for most Frisian speakers: Standard Frisian and Standard Dutch. However, Standard Frisian is primarily a written standard instead of a spoken one and few people are proficient writers of this standard (i.e., around 20% of the speakers; Stefan, 2022, p. 3), so it is unclear to what degree it functions as a roof variety. Standard Dutch is the more prominent, or perhaps only, spoken roof variety for Frisian speakers (Stefan, 2022, p. 23). This also results in widespread lexical influence from Dutch, especially through loan words (Gorter, 2001). At the same time, Dutch influence on Frisian pronunciation specifically seems to be relatively limited (Feitsma, 1989; Van Bezooijen, 2009; Heeringa & Hinskens, 2014), leaving the Frisian varieties relatively distinct from Dutch in this regard. In this dissertation, we reassess whether Standard Dutch influence on Frisian pronunciation has remained limited. We also assess whether Frisian varieties have become more similar over time, which may indicate the formation of a regiolectal variety.

Low Saxon is understudied compared to Frisian, so we are unsure about regiolect formation in this area. It is unlikely that a single Low Saxon regiolect would arise in the Netherlands, because its speaker population spans across a

much larger geographical area than Frisian and most Low Franconian varieties (see Figures 1.1 and 1.2). If dialects become substantially more similar, this likely occurs more locally in regions with substantial contact. There is no reason to suspect multiple potential roof varieties for Low Saxon varieties, because Low Saxon is not standardized in either script or pronunciation. Written standards have been developed for smaller regions, such as in Groningen (Reker, 1984), or even the entire Low Saxon region across national borders (i.e., the *Nysassiske Skryvwyse* 'New Saxon Writing Method'),[9] but none of these writing methods are widely used, let alone taught at school in the Netherlands. If Low Saxon speakers write their language at all, they likely use a localized ad-hoc spelling instead (e.g., similar to Frisian and Limburgish speakers; Jongbloed-Faber et al., 2016, 2017). The roof variety for Low Saxon varieties is Standard Dutch, given its prominent status and the absence of alternatives. As a consequence, Low Saxon varieties likely become more similar to Standard Dutch over time (following the general patterns in the Netherlands; Heeringa & Hinskens, 2015). Similar to the Frisian varieties, we reassess to which degree convergence to Standard Dutch occurs for Low Saxon varieties, and evaluate whether the dialects have become more similar to each other. We assume that these effects may show different patterns for different Low Saxon areas, because of the large area in which Low Saxon is spoken in the Netherlands.

The spatial dimension should not be ignored when assessing language variation and change patterns within and between regional languages. Linguistic innovations usually diffuse gradually (Nerbonne, 2010), which we also expect for the changing pronunciations of Frisian and Low Saxon. For example, we can expect changes to occur more rapidly between areas with extensive networks of contact or progress first to areas that are densely populated (see, e.g., Trudgill, 1974). This can lead to cities serving as centers of innovation, such as Antwerp in the case of *Tussentaal* (Vandekerckhove, 2009, p. 82). Furthermore, there are also differences in language contact within the Low Saxon area (e.g., more contact with Frisian in the northern Low Saxon area), and there are differences in how similar varieties currently are to Standard Dutch (see Wieling et al., 2011). It is also well known that regional languages and their traditional dialects are substantially more prevalent in rural areas than in urbanized ones (Goeman & Jongenburger, 2009; Tagliamonte, 2016, p. 3). Given these many spatial determinants, we also assess how pronunciation variation and change are geographically distributed in this dissertation.

[9] See https://skryvwyse.eu/.

1.3.3 Dialect corpora — Researchers in dialectology and related fields often rely on phonetic corpora to investigate pronunciation variation. To construct these corpora, regional language users are usually presented with written prompts (such as single words or entire sentences) and asked to translate them into their local language variety. This tradition appears to have started in Germany (Wenker & Wrede, 1889), resulting in a substantial amount of data on language variation based on sentential prompts. The approach was adopted by linguists in the Netherlands and Flanders within a few decades, after which the *Reeks Nederlandse Dialectatlassen* 'Dutch Dialect Atlas Series' was constructed based on translations of 139 sentences and a few conjugations (abbreviated as RND; Blancquaert & Pée, 1930). Field workers collected data from an impressive 1,956 locations from 1923 onward, but the work was only concluded in 1982, resulting in large time differences between recordings from different regions.[10] The successor to the RND was the Goeman-Taeldeman-Van Reenen project (GTRP; Taeldeman & Goeman, 1996), which was organized more efficiently, and the data was therefore collected in a shorter time span (1979 to 2001, with most recordings made between 1985 and 1989).[11]

Dialect speakers were prompted with short sentences and individual target words in Standard Dutch for the RND and GTRP, although most of the GTRP targets were single words. Prompting single words is attractive, because it allows for a controlled experimental approach, keeping the time required for data collection relatively short. The disadvantage of prompting only single words is that they are not the most realistic reflection of someone's natural speech. For another dialect corpus, the From Dialect to Regiolect project (DIAREG; Heeringa & Hinskens, 2015), sentences were prompted based on a short silent movie. The participants discussed how to translate the story into the local dialect, after which the translated story was pronounced by one of the speakers. These translated sentences are more realistic representations of local dialect use, but the disadvantage of sentential prompts is that these also result in more phonetic reduction effects and are therefore more difficult to analyze systematically (see Chapter 5).

Phonetic transcriptions are available for the RND, GTRP, and DIAREG. However, inter-transcriber agreement of phonetic transcriptions is usually limited, especially when transcriptions are relatively narrow and detailed (Amorosa et al., 1985; Shriberg & Lof, 1991). For example, the transcriptions of the GTRP recordings lacked consistency due to the many transcribers involved in the project. Wie-

[10] See https://www.dialectzinnen.ugent.be.

[11] See https://projecten.meertens.knaw.nl/mand/GTRPdatata.html.

ling et al. (2007) showed that the number of phonetic symbols used for the GTRP transcriptions of dialects in the Netherlands was much higher than for the GTRP transcription of dialects in Belgium (73 versus 44), and transcriber effects were reported for the GTRP (Hinskens & Van Oostendorp, 2006). This is problematic, because such corpora are of key importance in dialectology, and phonetic transcriptions are the established way to encode the observed pronunciation variation. We therefore also assess whether it is possible to compare corpora that differ substantially in terms of their transcription practices.

1.4 Research questions and chapter overview

Given the overviews of the decline in speaker numbers for Frisian and Low Saxon and the ongoing language variation and change in both languages presented above, the research questions (RQs) for this dissertation can be summarized as follows:

RQ 1a. What percentage of the population in the traditional areas of Frisian and Low Saxon is able to speak the regional language?
RQ 1b. What percentage of the population in the traditional areas of Frisian and Low Saxon uses the regional language at home?
RQ 2. Which factors are associated most strongly with parental language transmission of Frisian and Low Saxon?
RQ 3. Can phonetic corpora that differ in transcription conventions be used to estimate pronunciation change accurately?
RQ 4a. To what degree is the pronunciation of Frisian, Town Frisian, and Low Saxon becoming more similar to Standard Dutch?
RQ 4b. To what degree is the pronunciation of Frisian, Town Frisian, and Low Saxon becoming more similar to neighboring language varieties?
RQ 4c. Is there evidence of regiolect formation for Frisian, Town Frisian, and Low Saxon varieties?

The chapters of this dissertation address these research questions in the same order. Chapters 2 and 3 are concerned with changes in the speaker population, and the three subsequent chapters (i.e., Chapters 4 to 6) address pronunciation change.

In Chapter 2, we describe how we construct speaker counts for Frisian and Low Saxon to answer RQ 1a and 1b. We rely on a sample from a pre-existing

participant pool of a large-scale longitudinal study in the northern three provinces of the Netherlands (Fryslân, Groningen, and Drenthe), to whom we sent out a regional language questionnaire in 2021. This limits our geographical scope for Low Saxon, but it enables us to obtain a much larger sample than if we would try to find Frisian and Low Saxon speakers directly. Unfortunately, the ratio of regional language users to non-regional language users among the respondents is unrealistically high in the obtained sample. As post-stratification is not useful in our situation, we estimate the metrics of interest for several generations by combining regional language transmission numbers extracted from our questionnaire data with speaker estimates from earlier reports.

In Chapter 3, we rely on the same participant sample as in Chapter 2 to explore which factors are associated with parental language transmission to answer RQ 2. We model whether someone uses their regional language (Frisian or Low Saxon) with at least one of their children based on a set of variables derived from the regional language questionnaire. These variables include whether someone has positive language attitudes toward their regional language, whether someone perceives their variety as a dialect of Dutch, how often someone uses their language and in which contexts, and whether someone uses their language with the other parent or caregiver of their children. We also include variables related to a speaker's wider environment, such as where someone grew up, and whether someone lived in a rural area at the time of receiving the questionnaire.

Turning to pronunciation change methodology and focusing on RQ 3, Chapter 4 is concerned with inter-transcriber inconsistencies in large dialect corpora, because different transcribers often arrive at different transcriptions of the same target words. More specifically, we focus on the RND and GTRP corpora as the main corpora of dialects in the Netherlands and Flanders in the 20th century. Even when we ignore diacritics, a much larger phonetic symbol inventory is used for the GTRP transcriptions than for the RND transcriptions. We propose that replacing infrequently occurring symbols with their closest alternative symbol in phonetic space can be used to make these corpora more comparable for analysis, while ideally minimizing the loss of phonetic detail and retaining the main pronunciation variation in the data. We subsequently estimate how much pronunciation change occurred in the Netherlands and Flanders using the adjusted corpora, but acknowledge that the aforementioned time span differences in the RND remain problematic for drawing conclusions about pronunciation change.

To answer RQs 4a through 4c, we compare the GTRP corpus to the more recent DIAREG corpus in Chapter 5, using a single Standard Dutch speaker (Astrid

Kersseboom) as a reference point. She was selected as the most representative Standard Dutch speaker from a set of news presenters by native Dutch speakers. We measure pronunciation change again, but zoom in on the geographical region of interest in this dissertation, namely specifically the Frisian, Town Frisian, and Low Saxon areas in the Netherlands. A single transcriber made all transcriptions to avoid the inter-transcriber issues observed in Chapter 4. The GTRP and DIAREG corpora were collected during a much smaller time span than the RND and are consequently better comparable than the GTRP and RND. However, the tasks used to elicit dialectal pronunciations differed between the GTRP and DIAREG, so the comparison is again suboptimal.

We therefore revisit and re-estimate pronunciation changes for the same geographical area in Chapter 6, again focusing on RQs 4a through 4c. We take the GTRP corpus as a reference point and use a new dialect corpus from the 2020s (collected using a mobile laboratory in the context of this dissertation). The recordings were made in a time span of less than 18 months and we ensured that the new speakers matched the speaker characteristics of the GTRP for 32 overlapping locations in our area of interest. We ensured that a sufficiently large number of target words overlapped between the GTRP and the new corpus, and a single transcriber again made all transcriptions. These parameters therefore ensure an optimal comparison to assess whether Frisian and Low Saxon varieties converged or diverged with respect to neighboring varieties and whether they converged to or diverged from Standard Dutch.

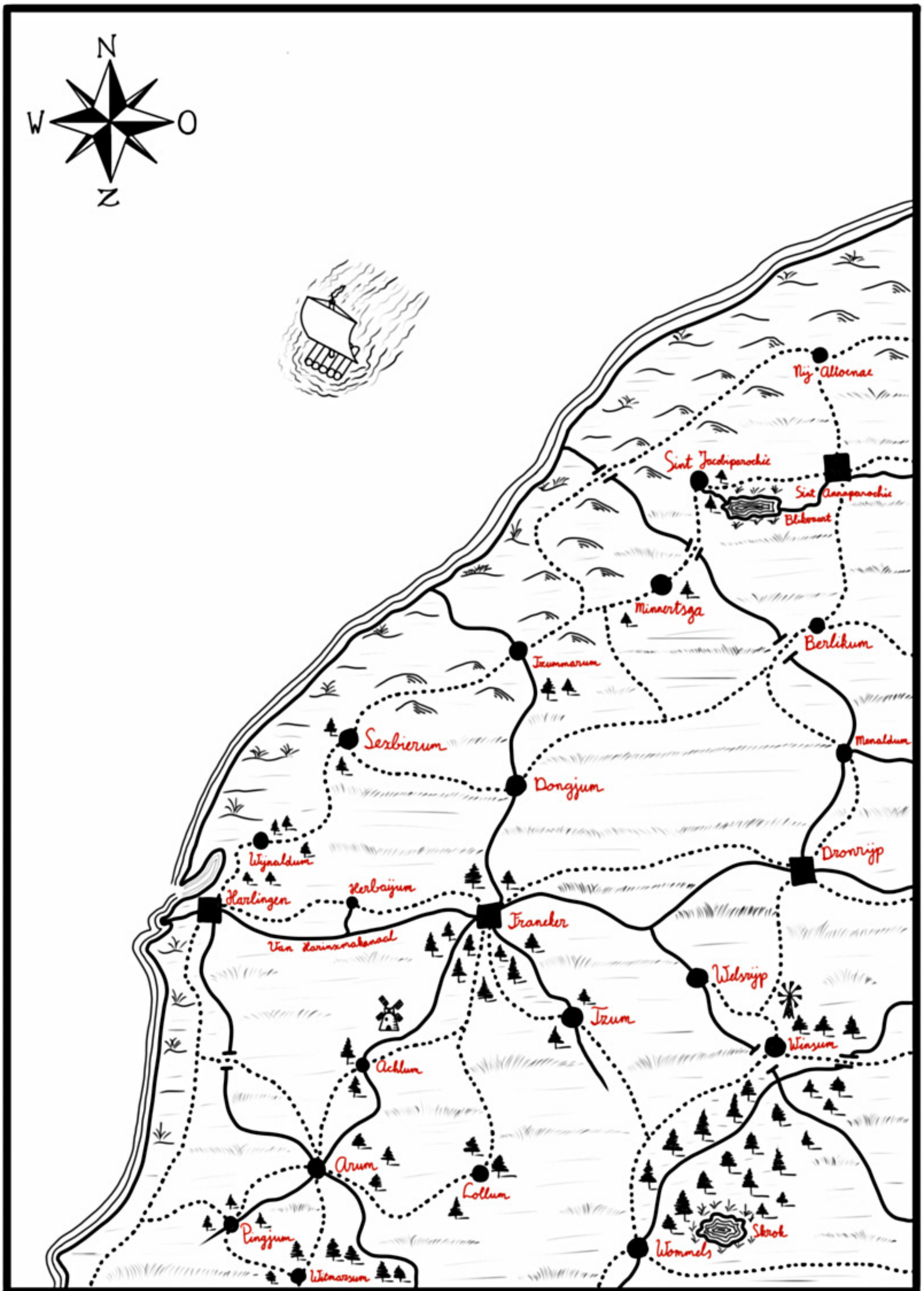
Finally, we discuss our findings and methodological shortcomings in Chapter 7. We end this dissertation by presenting possibilities for future studies, with a particular focus on Frisian and Low Saxon.

Part II

Speaker population changes

Sexbierum, Fryslân

CHAPTER 2



Voor Reihaneh

ESTIMATING SPEAKER POPULATION SIZES

Abstract

Language questionnaires are often used to approximate the size of linguistic communities, which we attempt for the regional languages Frisian and Low Saxon in the Netherlands in this chapter. Through an existing large-scale longitudinal study, we distributed a questionnaire about various topics (e.g., language use in different contexts, proficiency, intergenerational transmission, and the respondent's language learning context). This resulted in 38,500 respondents across the three northern provinces in the Netherlands (Fryslân, Groningen, and Drenthe) where the two regional languages are spoken. We specifically estimate how many people in these provinces are able to speak the regional languages and how many use the language at home.

The sample we obtained through the questionnaire likely suffered from sampling bias, because the prevalence of dialect speakers was unrealistically high. Initially, we applied post-stratification to account for differences between ratios in the sample and the northern population (e.g., adjusting for gender, age, domicile population density, or educational attainment). This only had a limited effect on our metrics, so we opted instead for an intergenerational transmission approach. Earlier regional language usage estimates were used as reference points, and we derived estimates for the generations that followed these reference generations to obtain estimates for more recent generations.

Our results showed that the Low Saxon speaker population size is declining across recent generations, with around 350,000 speakers in 2021 aged between 6 and 69 (around 41% of the population in that age range) and 140,000 people using it at home (around 17%). The Frisian population appears stable, with around 250,000 speakers aged between 5 and 60 (62% of the population in that age range) and 195,000 people using it at home (around 48%). As these estimates seem plausible compared to other speaker counts, our intergenerational estimation approach may be suitable for obtaining speaker estimates when transmission information is available and more common methods are ineffective.[1]

2.1 Introduction

SURVEYS AND QUESTIONNAIRES are essential tools in academic and governmental research, because they can offer valuable insights into societal patterns. However, ensuring accurate and representative results poses a

[1] This chapter is adapted from: Buurke, R., Bartelds, M., Knooihuizen, R., & Wieling, M. (2024b). Multiple estimates of the Frisian and Low Saxon speaker population size in the Netherlands. *Linguistic Minorities in Europe Online*. https://doi.org/10.1515/lme.28672125

significant challenge. In this chapter, we delve into some complexities of survey methodology, particularly addressing biases arising from question phrasing and procedures surrounding the distribution of regional language surveys. We highlight the limitations of a conventional solution such as post-stratification in more extreme cases of sampling issues, as it appears inadequate for the case study presented here. We propose a novel approach based on using prior surveys in combination with intergenerational transmission to overcome the issue of sampling bias.

Ó Riagáin (2018, p. 5) notes that there are "no universally accepted survey measures of key sociolinguistic concepts such as language proficiency, language use, or language attitudes". This absence of a standardized methodology can enable a wide range of biases (see Delgado-Rodriguez, 2004, for an overview of known bias types in research), both due to internal and external aspects of the questionnaire. Internally, bias can be introduced through the inability to minimize variability in the possible interpretations of the survey questions, because informants may perceive the intentions behind questions differently. Externally, a non-representative sample may be obtained due to a suboptimal approach to reaching the relevant social groups.

The internal bias of a questionnaire can be minimized by asking questions reliably, but formulating reliable questions for language questionnaires is notoriously difficult (Duchêne & Humbert, 2018). The relevant question for our case study seems trivial: "Do you speak the regional language?". However, the interpretation of this question by informants may be difficult to predict. Speakers find it difficult to determine what constitutes a (proficient) speaker of a specific language, and they tend to be reluctant to classify themselves as one (Moore et al., 2010). Informants can also be asked to rate themselves on how well they speak the language, as opposed to indicating whether they classify themselves as a speaker. In that case, it is still unclear what they perceive as speaking a language well and what they compare their proficiency to (e.g., their proficiency in the majority or standard language). Furthermore, speakers of regional languages with a standardized variety (e.g., Frisian) are known to report themselves as having low language proficiency, because they never learned the standardized variety (Stefan, 2022, p. 52). There is no obvious solution to these problems for speaker counts, because such ambiguities can probably only be avoided by making the questionnaire unattractively long and approaching the concept from many angles.

A less ambiguous question is whether informants use the language at home,

which is a relatively simple and closed question (i.e., yes or no). This may also be a better determinant of language maintenance and shift, as it reflects the structural embedding of a language in everyday life. However, this metric may be conservative, because people may also use the language exclusively in contexts other than the home (e.g., with friends or colleagues). This is particularly relevant if someone's partner does not use the regional language, but their friends and family do use the regional language. In addition, the exact wording still matters, as respondents may give different answers depending on whether they only consider the language used at home if they use it with everyone at home (e.g., if the language is used with someone's partner, but not their children; Driessen, 2012). Nonetheless, the fact that fewer definitions are necessary (e.g., what it means to be a speaker of a language) to answer this question makes it an attractive option combined with speaker proficiency questions.

Questions relating to self-reported language proficiency and language use at home are reported in this study, because both are informative about regional language use. Both questions relate to the concept of being a language user, although how they relate to this concept differs. If a language is used at home, this suggests a structural embedding of the language in someone's life, while this embeddedness is not a given when someone perceives themselves to be highly proficient in a particular language. The proficiency metric indicates something about someone's knowledge of a language, but someone may be able to speak a language and still refuse to do so. For example, the majority language may be strongly preferred when someone joins the conversation who cannot speak the minority language. In that case, the minority language may be used much less than expected based on speaker numbers. Similarly, someone who can speak the regional language (possibly even very well) may have moved to a different area, potentially resulting in less regional language use by that speaker. For these reasons, we consider it appropriate to include both metrics in this study.

In this chapter, we assess whether the speaker populations of Frisian and Low Saxon are declining or stable based on how many people can speak the languages and use them at home (see research questions 1a and 1b in Section 1.4). We explore how to derive a speaker count for these languages through a language questionnaire. However, Our sample does not accurately reflect some prior known facts about the distribution of regional language users in the northern Netherlands, as it contains an overrepresentation of regional language users. We argue that this is an issue of non-response (of those not speaking a regional language), which is a common survey problem (Groves & Peytcheva, 2008) and is typically

solved by applying corrections to the obtained numbers based on known population parameters. We attempt this in Section 2.3 using post-stratification (Holt & Smith, 1979; Bethlehem, 2009, p. 250), but the effects appear too limited. We therefore propose an alternative novel method combining previously established speaker counts with intergenerational transmission information from our survey that is unaffected by the sampling bias. We then evaluate our estimated speaker counts by comparing these to other sources.

2.2 Lifelines questionnaire

The northern Netherlands is considered an ideal laboratory for multilingual research, because there is considerable language variation in the three provinces of Fryslân, Groningen, and Drenthe, even though only a relatively small part (approximately 10%) of the total Dutch population resides here. People in this geographical region may speak a variant of the regional languages Frisian or Low Saxon in addition to Standard Dutch (varieties of Town Frisian, see Section 1.1.2, are not considered in this chapter). Multilingualism including multiple regional languages also occurs in these geographical areas, such as in the municipalities of Weststellingwerf and Ooststellingwerf in Fryslân on the Frisian-Low Saxon border.[2]

In many scenarios where speaker population sizes of minority languages are systematically investigated, the aim is often to inform language policy through a census (e.g., in Ireland, Belgium, and Canada; Duchêne & Humbert, 2018). However, no regular census in the Netherlands covers regional language use. A representative sample can be drawn instead, but this is only effective if a sufficiently large sample can be drawn. Fortunately, there was an opportunity to distribute a regional language questionnaire to a large existing participant pool in the northern Netherlands: the Lifelines Cohort Study of the Lifelines biobank (Scholtens et al., 2015; Sijtsma et al., 2022). More than 167,000 people have contributed their data to Lifelines, which corresponded to almost 10% of the northern population in 2022. The multi-generational cohort has been followed over time from 2006 onward, which means we may be able to use this data to estimate usage and intergenerational transmission rates of the regional languages.

[2]For our questionnaire, respondents were asked to choose the regional language they spoke best and fill in the questions for that language. Participants were prompted to fill in the questions again for their second regional language at the end of the questionnaire, but no one did so, presumably due to the length of the questionnaire.

We designed a questionnaire about regional language use covering various topics (e.g., where speakers use their regional language, a speaker's self-reported speaking proficiency, and statements about a speaker's regional language attitudes), which was added to Lifelines. The supplementary material for this chapter can be found at https://osf.io/pxtmb/, which includes the complete set of questions used in the questionnaire. The Lifelines team consequently implemented and distributed the questionnaire, so we did not have direct contact with the respondents. Once the Lifelines team completed the data collection, we gained access to the data. This is less ideal than having full control of the survey process, but supporting the entire survey process is often too expensive for individual researchers or groups, so working with secondary language survey data is increasingly common (Ó Riagáin, 2018, p. 112).

The Lifelines team conducted the data collection for our study in November 2021. The Lifelines team was given the questions, structure, and logic for the questionnaire, and they made sure it was presented accurately to respondents along with a recruitment message using their internal systems. At that time, approximately 132,000 people were active in the Lifelines participant pool to which our questionnaire was sent out. Around 38,500 people responded to the questionnaire, so the response rate was approximately 29%. Around 14,500 respondents lived in Fryslân, around 12,000 in Groningen, and around 10,500 in Drenthe. The questionnaire was also filled in by respondents who no longer lived in the North (around 1,500 people). These respondents were left out of our analyses, because we focus on Low Saxon and Frisian language use specifically in their traditional context and geographical area.

Adding a questionnaire to Lifelines is attractive, because its large pool of potential respondents presents a unique opportunity to reach many regional language users. However, the participant pool does not completely reflect the demographic composition of the northern population in the Netherlands. For example, Klijs et al. (2015) showed that Lifelines informants were more likely to be female, middle-aged, married, a native speaker of Dutch, and living in a semi-urban area than the population of the northern provinces at large. There is also an education bias in the participant pool, because more higher-educated people are represented in the data, which is a common occurrence for questionnaires (Christoffersen, 1987; Korkeila et al., 2001). This is unfortunate, because regional languages tend to be more widely used by people with lower educational attainment and people who live in rural areas (Driessen, 2005; Goeman & Jongenburger, 2009).

On top of these general issues of representativeness in Lifelines, it became clear upon inspection of the collected sample that the relative number of users of regional languages was disproportionately high. Within the group of respondents, around 63% in Groningen, 65% in Drenthe, and 70% in Fryslân indicated that they were able to speak the most prominent regional language in that region (i.e., Low Saxon in Groningen and Drenthe, and Frisian in Fryslân). These percentages are likely overestimates when we compare these values to recent speaker counts, especially for Low Saxon (Provincie Fryslân, 2020; Versloot, 2021a). The sample is therefore not representative of the northern Dutch population, and estimates based on the data need to be amended in some way.

We cannot be certain about the cause of the overrepresentation bias of regional language users, but we suspect that the balance of the information in the message to potential respondents discouraged self-reported monolinguals and non-dialect speakers. The core part of the recruitment message excluding salutation and practical information (see Figure 2.1) focuses on multilingualism, especially in combination with dialects. The message included an explicit statement at the end that responses from people who do not speak a second language or a dialect are still useful for the sake of comparison. It is possible that including this encouragement was insufficient to ensure participation from those who did not use a regional language. The Lifelines team sent a reminder if someone did not participate within 14 days. This probably ameliorated the issue, but insufficiently so, because the recruitment message and encouragement remained the same.

2.3 Applying post-stratification

2.3.1 Method — When sampling issues occur in the research process or due to problematic survey design, it is possible to correct for imbalances in the sample according to known population distributions by using post-stratification (Holt & Smith, 1979; Bethlehem, 2009, p. 250). For example, if the percentage of men in the population is about 50%, but 40% in our sample, the sample estimate of the relevant metric for men should be multiplied by 1.25 (i.e., 50% ÷ 40%). Likewise, the percentage for women should be multiplied by 0.83 (i.e., 50% ÷ 60%). Estimates can be further improved when marginal distributions are available (e.g., of the educational attainment in each age group).

For the post-stratification approach, it is necessary to obtain reliable popula-

Original core part of recruitment message in Dutch:

Het spreken van twee talen kan op latere leeftijd positief effect hebben op het geheugen en concentration. Onderzoeker [*naam van de onderzoeker*] ([*naam van de universiteit*]) wil inzicht in de cognitieve voordelen van tweetaligheid, in het bijzonder van het spreken van talen en dialecten die veel op elkaar lijken. Hiervoor willen we u uitnodigen om de vragenlijst Streektalen in te vullen. Ook wanneer u geen andere taal of dialect spreekt zijn uw antwoorden heel waardevol, zodat er een vergelijking mogelijk is.

Translation in English:

Speaking two languages can have a positive effect on memory and concentration later in life. Researcher [*name of the researcher*] ([*name of the university*]) wants to understand the cognitive benefits of bilingualism, especially of speaking languages and dialects that are very similar. To this end, we would like to invite you to complete the Regional Languages questionnaire. Even if you do not speak another language or dialect, your answers are very valuable, so that a comparison is possible.

Figure 2.1: Core part of the recruitment message sent to potential Lifelines respondents.

tion information for the provinces of Fryslân, Groningen, and Drenthe. Statistics Netherlands is a governmental organization that collects population statistics and provides them online. We apply post-stratification here for a few key factors that are suspected to possibly influence regional language use, and for which data are available from Statistics Netherlands. These include the respondent's age, gender, educational attainment, and the domicile population density surrounding someone's address when they filled in the questionnaire (i.e., a higher density of postal codes indicates a more urbanized area; Centraal Bureau voor de Statistiek, 2023c). We adjust our sample ratios accordingly to assess whether this yields more realistic estimates of regional language use.

We divided the data into three age categories: 0 to 30, 31 to 60, and 61 to

90.[3] The scale of domicile population density is divided into five levels, ranging from 'not urban' to 'strongly urbanized'.[4] Educational attainment is divided into low, middle, or high educational attainment according to Dutch education levels (see Centraal Bureau voor de Statistiek, 2023a, for details).

2.3.2 Results — When the ratios are compared between our sample and the population, the following groups seem to be underrepresented to different degrees: men, people aged 0 to 30, people in (strongly) urbanized areas, and people with low or middle educational attainment. The other categories of these variables are therefore overrepresented. Statistics Netherlands does not provide marginal distributions across all variables of interest, so we cannot apply post-stratification across all variables simultaneously. The strongest decrease in estimated regional language use percentage points is when we adjust for population age distribution, for which we report the necessary sample-to-population adjustments in Table 2.1. The other sample ratio multiplication factors can be found in the supplementary material.

Table 2.1: Multiplication factors for the age variable to match the sample ratios to (Statistics Netherlands-based) population ratios in 2021.

Province	Multiplication factor		
	0 – 30	31 – 60	61 – 90
Drenthe	12.72	0.73	0.80
Fryslân	14.29	0.72	0.72
Groningen	11.09	0.73	0.63

The age-adjusted estimates of regional language users (i.e., Low Saxon in Groningen and Drenthe and Frisian in Fryslân) are 55% (in Groningen), 62% (in Drenthe), and 70% (in Fryslân). While the age-based post-stratification has the strongest effect on reducing the percentage of regional language users, the resulting percentages are still too high when we compare them with published speaker counts. Versloot (2021a, p. 11) estimates that Low Saxon usage is approximately between 35% and 5% (for people born between 1957 and 2006),

[3] Women have their first child around the age of 30 in the Netherlands (Centraal Bureau voor de Statistiek, 2021), so these categories encompass three generations. Narrower age categories (e.g., ten-year ranges) would have been problematic, because there are relatively few Lifelines respondents who are either young or relatively old.

[4] See https://www.cbs.nl/nl-nl/onze-diensten/methoden/begrippen/stedelijkheid--van-een-gebied-- for details.

which is much lower than our age-adjusted estimates. The corresponding Frisian estimates approximately range between 55% and 40%, which differ to a smaller extent (but still considerably) from our age-adjusted estimates. Estimates that rely on the sample ratios between regional language users and non-users are therefore too biased and do not allow for accurate population inference. Even if a more advanced post-stratification approach had been possible (i.e., if the relevant marginal distributions had been available from Statistics Netherlands), the estimates would have only become higher and even less realistic.

2.4 Using reference works and intergenerational transmission

2.4.1 Method — The post-stratification approach was shown to be insufficient for obtaining realistic estimates of regional language use, at least for our specific case. We therefore investigate combining data from different sources to more extensively alleviate the bias present in the current sample. If a suitable and more reliable reference work is available, it can serve as a starting point for obtaining more reliable estimates. We can estimate how many children acquired the language in the next generations (i.e., the generations after the ones reported in the reference work), which provides information about the current generations. To do so, we require information about the intergenerational transmission rates of the generations reported in the reference work. The estimates can be evaluated afterward by comparing the obtained estimates to estimates of the same metrics in other works, if they are consistently operationalized and available for the correct age groups (i.e., those of the generations following the reference generation).

2.4.1.1 Reference works. In this study, we take Bloemhoff (2005) as a reference work for Low Saxon, and we take Klinkenberg et al. (2018) as a reference work for Frisian. The regional language use (i.e., self-reported speaking proficiency and language use at home) of subsequent generations can then be estimated based on the intergenerational transmission rates of the reported reference generations, which can be obtained from our survey. These estimates allow us to construct a speaker count from different sources and simultaneously explore a new approach to quantifying regional language use of speaker populations.

Naturally, this approach only works well when the reference works themselves are (more) representative of the speaker population. Versloot (2021a, p. 9) expresses doubts about the Low Saxon reference work, because the reported percentages seem high. We share these doubts, but acknowledge that no good alternative speaker count is available. Furthermore, the sampling procedure of the reference work appears trustworthy (i.e., an external company called random phone numbers) and we cannot explore in detail what could have caused relatively high estimates. For this reason, we think Bloemhoff (2005) is still the most reasonable point of departure for this method. For the reference work of Frisian, we assume that the estimates are accurate, because the sampling procedure also appears sound. The questionnaire was sent to 30,000 addresses in Fryslân provided by the provincial government without specific selection criteria, and the response rate was 15%. Klinkenberg et al. (2018) acknowledge that there are some representation issues (as younger people were underrepresented), but they adjusted their estimates using post-stratification.

As noted before, how respondents interpret survey questions depends on their formulation. Fortunately, the exact question formulations are available for the Low Saxon and Frisian reference works, so we can assess their comparability and reliability. We provide the questions and translate them to English from Dutch or Frisian. For the self-reported speaking proficiency metric, Bloemhoff (2005) used this formulation (translated from Dutch): "To which degree can you speak [regional language variant]?", with the possible proficiency answer options ranging from no proficiency to being able to speak the language very well. The question formulations differed for each Low Saxon area included in the study. For the province of Groningen, the regional language variant was "Gronings of Westerkwartier-Gronings" [Groningen dialect or Westerkwartier-Groningen dialect].[5] For Drenthe, it was "Drents of Stellingwerven-Drents" [Drenthe dialect or Stellingwerven-Drenthe dialect]. For Frisian, Klinkenberg et al. (2018) used the following formulation (translated from Frisian): "Can you speak Frisian?". The range of options (i.e., not very easily to completely) is similar to Bloemhoff (2005), although they did not provide the names of regional varieties of Frisian in different areas.

Bloemhoff (2005) used the following formulation to refer to the use of Low Saxon at home: "What do you mainly speak at home?", with the answer options

[5]The *Westerkwartier* (lit. 'western quarter') is located in the west of the province of Groningen. Around the border of the municipalities of Weststellingwerf and Ooststellingwerf and the province of Drenthe, the local Low Saxon dialects are Stellingwerven dialects.

of Low Saxon (again using localized names), Dutch, both Low Saxon and Dutch, or something else. For the use of Frisian at home, Klinkenberg et al. (2018) used the following prompt: "Which language do you speak with your spouse?", with the options Frisian, Dutch, and something else. This is different from Bloemhoff (2005), who provided a separate answer option for using both the regional languages and Dutch at home and did not specify with whom the regional language is used at home. The percentages reported in the Frisian reference work are slightly lower if Frisian speakers are asked whether they speak the regional language with their children. This is in line with Driessen (2012), who showed a similar pattern across all regional languages in the Netherlands. We use the reported percentages about speaking the language with someone's spouse for Frisian, because we assume that the question in the Low Saxon reference work was intended and likely also interpreted as language use with someone's partner. The reference data for both metrics for Low Saxon are presented in Tables A.1 and A.2 in Appendix A, and the Frisian reference data are presented in Table A.5.

Additional remarks on Low Saxon. A few caveats need to be addressed for the Low Saxon reference work. In Bloemhoff (2005), there are three age groups: 18 to 39, 40 to 60, and 61 and older. The oldest group has no upper limit, but it can be assumed that most of the data concern the category of 61 to approximately 81, as only approximately 2.5% of the Dutch population was aged above 81 in 2005 (Centraal Bureau voor de Statistiek, 2023b). In other words, each age group covers a range of approximately 20 years.

As indicated before, we are interested in the self-reported speaking proficiency of Low Saxon and the use of Low Saxon at home. The reference estimates from the reference work are described for the entire Low Saxon area across the three age groups in Tables A.1 and A.2 in Appendix A. As the data from Lifelines only cover the Low Saxon provinces of Groningen and Drenthe, we would need the age-based usage values for both provinces separately. These data are not available in Bloemhoff (2005), but usage data for these provinces irrespective of age are available.

It is clear from the provided distributions (see Bloemhoff, 2005, p. 88, and also the supplementary material) that a higher proportion of the population uses Low Saxon in Groningen and Drenthe than in the other provinces in the Low Saxon language area. We therefore adjust the Low Saxon age-group-based percentages upward. We do this by comparing the average percentages across the whole Low Saxon area (not separated by age group) for each level of the variables

to the corresponding percentage of each region. As a hypothetical example, let us say that the average percentage of 'speaking Low Saxon very well' is 20% for the entire Low Saxon area and 25% for the province of Groningen. In that case, the age-group percentages reported for the entire Low Saxon area for 'speaking Low Saxon very well' are multiplied by 1.25 (i.e., 25 ÷ 20) to obtain the adjusted age-group percentages for Groningen. We repeat these steps for all levels of self-reported speaking proficiency and language use at home. We do not provide all calculated percentages here, but they can be found in the supplementary material. Note that this calculation is a simplification, as it assumes that the higher number of regional language speakers in Groningen and Drenthe (compared to Low Saxon in general) is not more pronounced in one age group than another.

2.4.1.2 Intergenerational transmission. Three main components are necessary to obtain a reliable speaker count in the context of this study. The first is a (more) representative reference point of how many people use the regional language (either Frisian or Low Saxon) in different generations. From these reference works, we use the computed percentages (based on reported percentages) of the following metrics as starting points: (1) how many people indicate speaking the regional language (at least at a reasonable level), and (2) how many people indicate using the regional language at home (potentially together with Standard Dutch).

The second component in our method is the intergenerational transmission rate of the generations for which speaker count information is reported in the reference works. Due to the large time depth in the Lifelines data and the reference works, we match the information from both sources by the birth years of the generations of interest. The Lifelines respondents were prompted to indicate whether they transmit their regional language using the following formulation: "Do you speak (or did you speak) your regional language/dialect with your child(ren)?". If respondents indicated that they spoke the regional language with at least one child, we took this as evidence of intergenerational transmission.

As the third and final component, we need to account for regional language speakers who acquired the language not from their parents, but through other means and in different contexts (e.g., from grandparents, friends, or others in their close environment). If 80% of the speakers indicate that they learned the language from their parents, we need to increase the estimate by a factor of 1.25 (i.e., 100% ÷ 80%) to account for the group that learned the language outside the home. In our questionnaire, we inquired whether people acquired the regional

language mainly in or outside the home environment (i.e., "How did you learn your regional language/dialect?"), and use this to adjust the transmission rate upward.

These components are combined to construct estimates of regional language use for both the Frisian and Low Saxon speaker generations in 2021. All components are percentages, but the last component becomes an adjustment ratio after dividing the percentages. The formula below shows how to estimate the use of the regional language based on the three components. In this formula, three variables are used. The first two variables, x and y, represent the minimum and maximum ages (in years) of a particular generation, such as 30 and 50, whereas the third variable, z, equals the difference in years between the parental generation and that of the children. The value of z is always 30 in our case, because that has been the average age at the first childbirth of women in the Netherlands in recent decades (Centraal Bureau voor de Statistiek, 2021).

$$\begin{aligned} \text{metric estimate}_{(x,\,y)} = {} & \text{metric estimate in reference work}_{(x\,+\,z,\,y\,+\,z)} \\ & \times \text{parental transmission rate}_{(x\,+\,z,\,y\,+\,z)} \\ & \times \frac{100\%}{\text{rate of acquisition through parents}_{(x,\,y)}} \end{aligned} \tag{2.1}$$

As an example, consider a hypothetical scenario in which 70% of the speakers between the ages of 30 and 50 from Bloemhoff (2005) use the regional language at home. We first identify the data from that same generation in Lifelines. We need to focus on the transmission rate of respondents aged between 48 and 68 in our Lifelines dataset, because the Lifelines questionnaire data were collected 18 years after the data for our Low Saxon reference work were collected (namely in 2003; Bloemhoff et al., 2008b, p. 298). The children of this reference generation are approximately 30 years younger, so we also determine the language acquisition patterns of the Lifelines respondents aged between 18 and 38. If we identify a (hypothetical) parental transmission rate of 50% and also that 80% of the respondents learned the regional language at home, we need to adjust the estimate by 1.25 (i.e., 100%/80%). This results in an estimate of 44% (i.e., $0.70 \times 0.50 \times 1.25 \times 100\%$) for those aged between 18 and 38 in the Lifelines data (i.e., the children of the reference generation).

2.4.1.3 Evaluation of results. The calculated estimates also need to be evaluated. For this evaluation, we both use the reference works themselves (which contain usage data for multiple generations) and additional resources. These additional resources consist of results from the study of Driessen (who investigated the language in which parents and their children interacted; 2012) and an extra dataset obtained from Statistics Netherlands (SN; *Centraal Bureau voor de Statistiek* in Dutch, and abbreviated as CBS) based on a regional language usage study carried out in 2019 (Centraal Bureau voor de Statistiek, 2019; Schmeets & Cornips, 2022).

As mentioned before, we convert all age groups to birth year ranges. When the birth year ranges differ between those in our study and those in the dataset we compare our estimates with, we combine information from multiple birth year ranges and adjust the percentages according to the overlap between our estimates and those of other works. For example, consider evaluating a (hypothetical) estimate we obtained in our study for people born between 1973 and 1993. To evaluate our estimate, we could compare it to estimates based on data collected by Statistics Netherlands in 2019. However, the evaluation dataset only provides information for people born between 1980 and 2004 (for which the regional language use was 16.9%), and people born between 1960 and 1980 (with a regional language use of 29.7%). Out of the 21 years covered by our age group (i.e., from 1973 to 1993), the 1980–2004 group overlaps 14 years (1980–1993), whereas the 1960–1979 group overlaps seven years (1973–1979). We therefore adjust the estimate in the data of Statistics Netherlands for 1973–1993 by multiplying the percentage of the younger group with 2/3 (16.9% $\times$ 14/21), the older group with 1/3 (29.7% $\times$ 7/21), and then adding these two values together ($\approx$ 21%). When data from multiple groups cannot be combined (e.g., due to an age cut-off in the dataset), we compare our estimates with the estimates for the generation with the largest overlap.

2.4.2 Results — Estimates can now be constructed based on the Lifelines data and the reference works according to the outlined procedure. The derived estimates and the corresponding reference percentages from other speaker counts are presented in the following sections.

2.4.2.1 Groningen and Drenthe. The components necessary to calculate the estimates for Low Saxon (according to Equation 2.1) are reported in Table A.6 of Appendix A. The final estimates of the percentages of people who indicate

speaking Low Saxon and speaking both Low Saxon and Dutch at home in 2021 are shown in Table 2.2.

Table 2.2: Estimated percentages for speaking and using Low Saxon (LS) in Groningen and Drenthe.

Birth years	Able to speak LS		Using LS at home		Using LS and Dutch at home	
	Groningen	Drenthe	Groningen	Drenthe	Groningen	Drenthe
1952 – 1972	51	55	22	28	30	37
1973 – 1993	39	46	14	19	24	33
1994 – 2015	28	39	6	10	14	23

As indicated before, the estimates are evaluated by comparing them to estimates from other works. For the self-reported speaking metric, we use the percentages from Bloemhoff (2005), while we compare our findings to those of Driessen (2012) and Statistics Netherlands for the language use at home metric. The percentages reported in the works we use for evaluating the obtained estimates (adjusted in line with Section 2.4.1.3 when necessary) are presented in Table 2.3. Note that the birth years in the tables reflect those of the estimates and not (always) the data reported by the works used for evaluation.

Table 2.3: Reference percentages for speaking and using Low Saxon (LS). Italicized values are adjusted based on multiple generations in the Low Saxon reference work (see Section 2.4.1.3).

Birth years	Able to speak LS	Using LS at home		Using LS and Dutch at home	
	Bloemhoff 2005	Bloemhoff 2005	SN 2019	Bloemhoff 2005	Driessen 2012
1952 – 1972	*75*	*58*	*33*	*46*	*23*
1973 – 1993	71	53	*21*	39	15
1994 – 2015	-	-	17	-	-

For the evaluation, the reported percentages from the reference work and Statistics Netherlands are averaged across Groningen and Drenthe, but the percentages from Driessen (2012) represent the entire Low Saxon region. The data from Driessen's study are therefore likely an underestimate for Groningen and Drenthe, as the regional language usage in those provinces was higher compared to the others (Bloemhoff, 2005). There were no data available for the birth years from 1994 to 2015 in the reference work of Bloemhoff (2005) and Driessen (2012), but they were available in the data from Statistics Netherlands.

Our estimates for both metrics are substantially lower than those reported in the Low Saxon reference work (see Bloemhoff columns in Table 2.3). When our estimates are compared with the percentages for generations following the reference generation in Bloemhoff (2005), the differences for different generations range from 20% to 32% for the self-reported speaking proficiency metric. Similarly, our estimates differ between 30% to 39% for the language use at home metric and between 6% and 16% for the regional language use at home (together with Dutch) metric.

The differences (specifically for the language use at home metric) compared to the other works used for evaluation range between 2% and 11% (i.e., compared to Statistics Netherlands), and 7% and 15% (i.e., compared to Driessen, 2012). Except for the values obtained from Driessen (2012), the estimated percentages obtained through our method are lower than those reported in the works used for evaluation. Possible explanations for these patterns are explored in Section 2.5.

2.4.2.2 Fryslân. For Frisian, we rely on Klinkenberg et al. (2018) as a reference work (abbreviated in the tables as KJS 2018). There is no need to adjust these values for differences between language areas as we did for Low Saxon, because we ignore Frisian speakers outside of Fryslân for our estimates. The municipalities of Weststellingwerf and Ooststellingwerf in Fryslân are not fully included in the reference work and the Lifelines data, because these areas are traditionally Low Saxon areas instead of Frisian ones. Some Frisian villages in Ooststellingwerf are included in the reference work, but the other works used for evaluation provide no subdivisions in Fryslân. Consequently, we have left the municipalities of Weststellingwerf and Ooststellingwerf out of the main analyses reported in this chapter.[6]

The estimates for Frisian and the percentages used for evaluation are presented in Tables 2.4 and 2.5, respectively. As before, the reported birth years in Table 2.5 reflect those for the final estimates shown in Table 2.4, and they may not overlap fully with those used in the reference works. For example, the reference work for Frisian did not include respondents under 18.

It is immediately clear that Frisian language use is declining to a lesser degree than Low Saxon language use, and it seems that the use of Frisian at home may even be increasing slightly. The difference in percentages between the esti-

[6] There were relatively few data for these areas in Lifelines, but the estimates and necessary components that could reliably be determined for this region are reported in Tables A.9 and A.8 in Appendix A.

Table 2.4: Estimated percentages for speaking and using Frisian in Fryslân (excluding Oost- and Weststellingwerf).

Birth years	Able to speak Frisian	Using Frisian at home
1961 – 1981	65	46
1982 – 1996	60	47
1997 – 2016	59	51

Table 2.5: Reference percentages for speaking and using Frisian (at home) in Fryslân (excluding Oost- and Weststellingwerf). Italicized values are adjusted based on multiple generations in the reference work (see Section 2.4.1.3).

Birth years	Able to speak Frisian	Using Frisian at home		
	KJS 2018	KJS 2018	Driessen 2012	SN 2019
1961 – 1981	*71*	*60*	*44*	*35*
1982 – 1996	*68*	*62*	-	38
1997 – 2016	66	62	-	38

mates for both metrics and the works used for evaluation is also less substantial for the Frisian case than the Low Saxon one. However, the differences between the estimates and evaluation works are more pronounced for the language use at home metric (i.e., between 2% and 15%) than the self-reported speaking proficiency metric (i.e., 8% or less). This is partly because Klinkenberg et al. (2018) report noticeably higher percentages of Frisian language use at home than Driessen (2012) and Statistics Netherlands.

2.4.2.3 Speaker count. It is now possible to calculate a speaker count for the age groups for which we have estimates in the different geographical areas. Statistics Netherlands provides the population count for each age in 2021 (Centraal Bureau voor de Statistiek, 2023a). By summing the population count for each age range in each generation in Tables 2.2 and 2.4, we can obtain the number of language users for each regional language by multiplying the population counts with the estimated percentages. We do not provide these calculations here, but they are available in the supplementary material.

For Groningen, we estimate that there are about 183,000 people who are able to speak Low Saxon (out of a total of 467,624 people between the ages of 6 and 69; around 39%), and about 65,000 people (around 14%) who mainly use the regional language at home. These estimates for Drenthe are, respectively,

about 180,000 people (out of a total of 380,226 people between the ages of 6 and 69; around 47%) and 75,000 people (around 20%). In Fryslân (excluding the municipalities of West- and Ooststellingwerf), there are about 251,000 people who are able to speak Frisian (out of a total of 406,896 people aged between 5 and 60; around 62%) and about 195,000 people (around 48%) who use the language at home.

2.5 Discussion

Determining the relative number of speakers of regional or minority languages is a laborious process, because there is no established standardized methodology, and different types of biases may be introduced. We set out to estimate the number of speakers of Frisian and Low Saxon in the northern Netherlands to address research questions 1a and 1b (see Section 1.4), but there was an overrepresentation of regional language users in our sample that did not align with the established literature. As a solution, after post-stratification appeared insufficiently effective, we combined data from existing sources and intergenerational transmission data from a new large-scale questionnaire and evaluated these metrics. Our findings align with patterns in other studies (Driessen, 2012; Versloot, 2021a), namely that the decline in speaker numbers of the Low Saxon population is steeper than that of the Frisian one. The obtained percentages seem plausible when contrasted with other speaker counts, with most estimates deviating less than 15% from other speaker counts concerning the same populations. Where the differences are larger than 15%, it seems that the used Low Saxon and Frisian reference works are reporting a higher number than is expected compared to the other sources.

The metrics under scrutiny in this study were whether regional speakers consider themselves able to speak the regional language (from reasonably to very well) and whether they mainly used the regional language at home (or in addition to Standard Dutch). The former type of question has the potential to be confusing for informants (Duchêne & Humbert, 2018, p. 9), which may partially explain the differences between the estimated values differ substantially from what is reported in other works. At the same time, the alternative language use at home metric is also susceptible to the issue of question phrasing. Respondents were asked about language use at home with their spouse for the Frisian reference work, while this was left unspecified for the Low Saxon reference work. For

Driessen (2012), both language use with the spouse and children were asked, and we relied on the indicated language use with someone's spouse. However, this was unspecified in the case of the Statistics Netherlands data. In future studies, it would be good practice to ask for both types of interlocutors, because the indicated use is likely lower when language use at home includes the use of the regional language with children.

For Low Saxon, the differences between the estimated values and the reference work were up to 18 percentage points (for the language use at home metric) and 32 percentage points (for the self-reported speaking metric). For Frisian, the differences were overall smaller, with differences up to 15 percentage points for the self-reported speaking metric and at most 8 percentage points for the language use at home metric. While these differences are not excessive, the differences between our estimates and the reference works remain relatively high in some cases. We discuss several potential sources of these differences, but we cannot eliminate any possible sources, because the data were already collected.

A possible source of the larger differences between the estimates for Low Saxon and Frisian is the degree of representativeness of the reference works. Versloot (2021a, p. 9) expressed doubts regarding the reported percentages of the Low Saxon reference work (Bloemhoff, 2005), because they seemed high. While the sampling method for the Low Saxon reference work seems unbiased (i.e., a company called randomly selecting phone numbers), it is possible that it did not prevent bias sufficiently. For example, people who do not use Low Saxon may have stopped participating in the questionnaire after they heard it was about regional language use. Non-response bias in general is known to be an increasing problem for telephone questionnaires (Montaquila et al., 2008). Klinkenberg et al. (2018, p. 28) also suffered from representativeness issues to some degree, such as younger people being underrepresented in the sample. The authors applied post-stratification to correct for population differences, so these issues are partially accounted for. Regardless, it is clear that the language use at home metric seems more reliable than the self-reported speaking metric across different studies estimating the size of the regional language-speaking population.

The intergenerational transmission in the Lifelines dataset shows a downward trend for both the Frisian and Low Saxon speaker population, which is consistent with trends on intergenerational transmission reported by Driessen (2012). However, the overall rate remains higher for Frisian than for Low Saxon. Frisian also seems to be used at home slightly more often in younger genera-

tions than older ones. This is clear from both Klinkenberg et al. (2018) and the data from Statistics Netherlands. It contrasts with the number of self-reported speakers of Frisian in Fryslân, as that seems to be decreasing slightly. This is not a universally established pattern (cf. Driessen, 2012), but Frisian appears relatively stable and in the most optimal condition compared to other regional languages in the Netherlands. Frisian is protected in full under the European Charter for Regional and Minority Languages and teaching Frisian in schools is facilitated (and people who move to Fryslân also tend to learn Frisian; Kircher et al., 2023), although the Dutch government has also been criticized for its lackluster support of teaching Frisian (Bayat et al., 2023). If the Frisian language at home is increasing, it is a rare case where regional language decline has partially halted, and Frisian language use may become more widespread over time. This is in stark contrast with many other regional languages in the Netherlands and elsewhere in Europe (Auer, 2005; Versloot, 2021a), as there are only a few cases of successful revitalization at the sub-national level (e.g., the Welsh or Basque languages; Morgan, 2001; Valadez et al., 2015).

There are some further consistencies in the data concerning the intergenerational transmission and language learning contexts, regardless of the regional language, and these are reported in Tables A.6 and A.7 in Appendix A. For example, we consistently found that around 80% of the regional language users indicate learning the regional language in the home and around 20% outside the home (e.g., through friends, family, or people in the environment). There are some small fluctuations between the generations, but there is no clear shift in the main language learning context, which is consistently at home. The factors associated with parental language transmission of Frisian and Low Saxon are explored further in Chapter 3.

The main assumption behind the approach taken in this study is that language maintenance proceeds through the relationship between parents and children, but there is a slight caveat in how this was operationalized. The question posed to Lifelines informants was whether they spoke the regional language with their children, which does not account for scenarios of language attrition when a child moves to an area where the language is not used (e.g., for economic reasons; Kamwangamalu, 2013). Naturally, regional language users who do not have children cannot be included in this way either. Furthermore, there are also 'new speakers' of these regional languages (i.e., without community or home exposure), who may learn the language for specific goals (e.g., for work or to feel included). Consequently, there is some uncertainty surrounding this metric. This

uncertainty could have been minimized with further questions and metadata, but we expect that these scenarios are relatively uncommon and do not pose a significant problem to the approach we have taken here (particularly, given that we corrected for learning the regional language outside the home environment).

The overrepresentation of regional language users in our sample is regrettable, because the sample size is unique for a regional language questionnaire in the Netherlands. As mentioned before, the source of this sampling bias is unclear, but we argued that it is likely in part due to the form and content of the recruitment message. People who are monolingual or do not use a regional language were likely insufficiently encouraged to participate, which constitutes a type of non-response bias. There has been wide interest in avoiding excessive non-response bias, especially for online questionnaires, because they are widely used. Suggestions to improve this (see Manzo & Burke, 2012, for an overview) crucially rely on being able to control the exact presentation of the final questionnaire and contact with respondents. This was impossible in our case, because the Lifelines Cohort policy is that they handle these parts.

It may be beneficial to broaden the scope of the survey to topics that concern a larger audience if direct control over communication with respondents cannot be maintained. Schmeets and Cornips (2022) investigated regional languages in the entire Netherlands, but the relevant survey questions were embedded in a broader survey about social cohesion and well-being (Centraal Bureau voor de Statistiek, 2019). For such a survey, a specific subgroup of language users will not be discouraged from participating solely based on their self-reported language use. It is still possible that they may stop the online survey early once they reach the part about regional languages, but at least it is possible to estimate the size of the non-response bias in that scenario (i.e., people who stop are likely non-dialectal speakers or monolingual). One should take care when designing a longer survey, because the ideal survey length seems to be around 10 to 15 minutes (Galesic & Bosnjak, 2009; Revilla & Ochoa, 2017), and that was already the length of our questionnaire. Furthermore, if one part of the survey is (for example) about economic factors, respondents might relate their responses on another part about languages to economic contexts. Part of our survey concerned language attitudes, and respondents might then evaluate their regional languages as less economically viable (and consequently report less positive attitudes).

For our approach, we essentially estimated metrics for a younger generation based on data from an older parental generation. It is also possible to choose the

oldest generation in the data and to estimate the generational decline or stability for multiple generations consecutively, which may be seen as a type of survival analysis. Such an approach is not very attractive in our case, because the entire analysis depends on a single reference point in such a scenario. Furthermore, multiple reference points were available to us and our approach preserves more information. Unless one is certain that a single percentage is representative, which we could not be for the Low Saxon data, it is advantageous not to rely on a single reference point.

For the speaker count, we assumed that the estimates for the different age ranges provide a broadly accurate view of the speaker population. The total age range across the included generations was between 5 and 60 for Frisian and between 6 and 69 for Low Saxon. We have no accurate estimates for those outside this range. We could have extrapolated the estimated percentages to the ages younger than our youngest generation and older than our oldest generation, for example, by using the estimates of the closest generation. However, this would have been a simplified assumption, and we did not want to increase uncertainty around the speaker count.

There is a final caveat that may not generalize to all majority-minority language contexts, but it is worth mentioning in the Low Saxon context. Speakers of Low Saxon dialects often perceive their dialect as a variant of Dutch instead of Low Saxon. In our sample, around 87% of the Low Saxon speakers in Groningen and Drenthe indicated that they perceive their dialect as a variant of Dutch (as opposed to around 16% of the speakers of Frisian in Fryslân, likely those speaking a Frisian-Hollandic contact variety). This widespread view was part of the initial argumentation in the House of Representatives to not officially recognize Low Saxon under the European Charter for Regional Languages or Minority Languages (see Tweede Kamer der Staten-Generaal, 1995). One of the first questions in our survey was therefore whether the respondent speaks a variant of one of the regional languages in the Netherlands, which included explicit examples of better-known terms for regional varieties of Low Saxon (e.g., *Gronings* and *Drents*). The perception of Low Saxon dialects as dialects of Dutch should therefore only minimally impact our findings, but it is possible that some respondents indicated not speaking a variant of Low Saxon due to their perceived dialect classification (as it being a variant of Dutch).

Finally, it is important to remember that a speaker count only shows a partial impression of current regional language use. Rural parts of the regional language population may value the language differently and feel more inclined to trans-

mit the language. This seems to be the case if we inspect the Lifelines data (see Figures A.1 and A.2 in Appendix A)[7]. Intergenerational transmission of Frisian and Low Saxon is substantially lower in urbanized areas, although this seems more pronounced for Low Saxon. The higher prevalence of Frisian in rural areas is confirmed as well by Klinkenberg et al. (2018, p. 90). Standard Dutch is also increasingly acceptable in most regional language usage contexts, as opposed to (close personal) contexts being reserved for the regional language (Hinskens & Taeldeman, 2013, p. 5). Reporting regional language use averaged across the entire population may give a relatively pessimistic view of the sociolinguistic situation in rural areas, so assessing urbanized and rural areas separately is useful.

The key findings of this chapter confirm some contrasts between the Frisian and Low Saxon regional languages and their speaker populations. The findings also emphasize the challenges in obtaining reliable information about these regional languages. We are confident about the generalizability of our method and findings, but especially the Low Saxon estimates would be more reliable if there would have been no reason to doubt the estimates of the reference work we used: Bloemhoff (2005). It is advantageous to ensure and verify all aspects of the data collection beforehand rather than having to apply corrections to the data afterward, but it is unlikely that this issue can usually be completely avoided. To do so, control over the contact with informants likely has to be prioritized over obtaining a large sample size and this is increasingly rare (Ó Riagáin, 2018, p. 112). If control over the process can be secured, the study may first be piloted with smaller groups. One can then test whether the samples adequately represent the population and deploy a more reliable process for a larger participant pool afterward. Alternatively (or additionally), if future studies can determine a standardized methodology for language questionnaires that minimizes (the effects of) potential biases, it may again prove possible to relinquish control over contacting informants to external parties and therefore obtain a larger sample size.

[7] The Stellingwerven are included in Figure A.1 for the sake of completeness, because there were data available for them in the Lifelines dataset. Note that in some areas where Town Frisian varieties are spoken (e.g., in Het Bildt and Ameland) the Frisian transmission rate is high, but this does not mean that Frisian itself is more prevalent than Town Frisian or other varieties in these areas. The lower transmission rate for Frisian in towns could also partially be due to the presence of Frisian-Hollandic contact varieties in those areas, which results in more competition for language transmission than in Low Saxon areas.

2.6 Conclusion

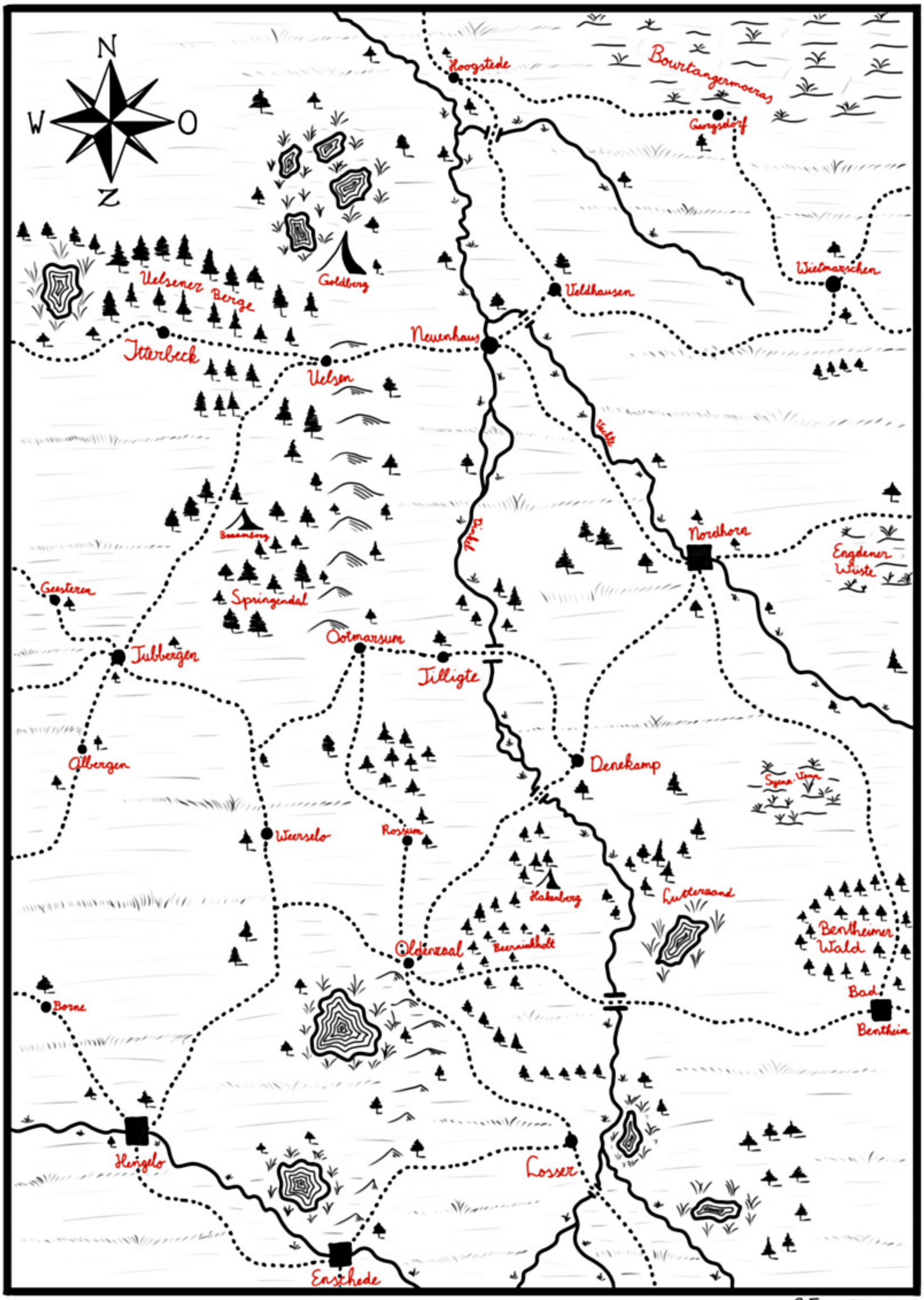
In this chapter, we attempted to derive reliable estimates and speaker counts for the regional languages Frisian and Low Saxon in the northern Netherlands. Our sample of 38,500 speakers suffered from sampling bias, because dialect speakers were overrepresented. We initially attempted to solve this problem by applying post-stratification, but this appeared insufficient. Instead, we devised a new method relying on earlier estimates of the use of these regional languages in combination with intergenerational transmission rates obtained from our sample to determine reliable estimates about the use of regional languages for younger generations.

The two metrics of interest in this chapter were self-reported speaking proficiency and self-reported language use at home, which offer two distinct views on the total use of these regional languages in the area of interest. We estimated that in 2021 around 62% of the population in Fryslân (excluding the municipalities of Weststellingwerf and Ooststellingwerf) was able to speak Frisian, and around 48% used the language at home. For Low Saxon in the provinces of Groningen and Drenthe, we estimated that around 41% of the population was able to speak Low Saxon, and around 14% used the language at home. These and earlier results show that the Frisian speaker population appears more stable over time than the Low Saxon one.

We evaluated our estimates by comparing them to other works covering these languages, but the questions surrounding language use and maintenance are not always consistently phrased in different questionnaires. This limited the number of works we could compare our results with. Devising a standardized methodology for questionnaires used for speaker counts would be highly useful for future studies, and it ensures that the size of minority language speaker populations can be consistently documented for language policy.

Tilligte, Overijssel

CHAPTER 3


N
W
O
Z
Hoogstede
Bourtangermoeras
Georgsdorf
Goldberg
Uelsener Berge
Wietmarschen
Veldhausen
Neuenhaus
Itterbeck
Uelsen
Nordhorn
Engdener Wüste
Geesteren
Springendal
Ootmarsum
Tubbergen
Tilligte
Albergen
Denekamp
Weerselo
Rossum
Hakenberg
Lutterzand
Bentheimer Wald
Oldenzaal
Borne
Bad Bentheim
Hengelo
Losser
Enschede

Voor Janine

FACTORS ASSOCIATED WITH REGIONAL LANGUAGE TRANSMISSION

Abstract

An important mechanism for language maintenance is transmission from parents to their children. In this chapter, we focused on parental language transmission of Frisian and Low Saxon and assessed which variables are associated with parental language transmission for both regional languages. For this purpose, we analyzed questionnaire responses from around 25,000 Frisian and Low Saxon speakers participating in the Lifelines Cohort Study (covering the three northern provinces in the Netherlands).

Language transmission was consistently higher for Frisian than for Low Saxon. Parental language transmission was strongly associated with whether their children's other parent speaks the same regional language and the frequency of language use in different social contexts. Other important factors included language attitudes and the degree of urbanization of the respondent's neighborhood. Taken together, these findings suggest that language maintenance for Frisian and Low Saxon could potentially be bolstered by adequately stimulating positive language attitudes and the use of the language in different social contexts by both the government and smaller societal organizations.[1]

3.1 Introduction

WITHOUT INTERVENTION, we may expect a five-fold increase in languages without native speakers by the end of the century (Bromham et al., 2021, p. 169), severely reducing the world's linguistic diversity. State protection is usually formally or informally reserved for majority languages, such as a national language (or languages). Majority languages often enjoy stable intergenerational transmission from parents to children, which is crucial for language maintenance (Wölck, 2004, p. 7). There is usually no stable intergenerational transmission for regional languages, such as Frisian and Low Saxon in the Netherlands. For Frisian, the transmission rate was stable at around 70% in 2021, while for Low Saxon the transmission rate went down from around 50%

[1]This chapter is adapted from: Buurke, R., Bartelds, M., Heeringa, W., Knooihuizen, R., & Wieling, M. (2024a). Intergenerational language transmission of Frisian and Low Saxon in the Netherlands. *Journal of Language and Social Psychology*, *43*(5-6), 1–18. https://doi.org/10.1177/0261927X241287765

to around 30% within two generations (see Table A.6 in Appendix A). Most Frisian and Low Saxon speakers indicate that they acquired the language mainly through their parents. We explore which factors are associated with Frisian and Low Saxon parental language transmission and address research question 2 this way (see Section 1.4). Our non-experimental study precludes us from determining cause and effect, but these associated variables may serve as a useful starting point for future studies, and for government initiatives to improve parental language transmission.

Mutual intelligibility between Frisian and Low Saxon, and between these languages and Dutch, is high (De Vries, 2011; Ter Denge, 2012; Belmar & Pinho, 2020). Frisian and Low Saxon have been protected in the Netherlands since the 1990s under the European Charter for Regional and Minority Languages (see Sections 1.1.2 and 1.1.3 in Chapter 1), although Frisian is protected according to Part II and III (as opposed to only Part II for Low Saxon) and therefore enjoys more financial support from the government. The teaching of Frisian in schools and its use in government correspondence are uniquely facilitated in the Netherlands, although language attitudes toward Frisian have not improved much (Hilton & Gooskens, 2013), and teaching in schools consistently falls short of set targets (Bayat et al., 2023). In short, the Frisian language currently has more fortunate circumstances than the Low Saxon language, but they are still far from ideal.

3.1.1 Factors potentially associated with language transmission — We rely on factors that can be extracted from the regional language questionnaire used for deriving speaker counts in Chapter 2. The questionnaire covers various variable groups that may be associated with parental language transmission behavior. These factors are summarized in the supplementary material for this chapter, which can be found at https://osf.io/jthvq/.

An important variable group concerns language use, although it can be difficult to distinguish language use, knowledge, and maintenance. We assume a parent is more likely to transmit their language when they frequently use it in different environments (e.g., at home, work, or church) or with people they see regularly (e.g., partners, parents, siblings, friends, or neighbors). We also incorporate whether someone started learning the language at a young age and whether they did so mainly through their parents. These factors reflect whether someone grew up in an environment where the regional language was embedded.

Our data also include information about educational attainment, which serves as a proxy for socioeconomic status. Regional language use is known to decrease monotonically from lower to higher social classes (Chambers & Trudgill, 1998, p. 58), which has also been observed for regional languages in the Netherlands (Driessen, 2005; Schmeets & Cornips, 2022). We also expect that language transmission decreases with increases in educational attainment in our data.

The number of (additional) foreign national languages (e.g., English, German, or French; see the supplementary material for details) a parent speaks may also influence their transmission probability, because they are likely more conscious of the associated benefits of multilingualism (e.g., economic or cognitive benefits Fan et al., 2015; Chibaka, 2018; Blom et al., 2017; Cockcroft et al., 2019). On the other hand, there is a common misconception that learning two languages simultaneously may be too confusing for children (Bialystok, 2011, p. 1249), although this has been disproven (Costa & Sebastián-Gallés, 2014). Given these conflicting ways of parental reasoning, we do not have a clear hypothesis about the effect of multilingualism on parental transmission rate.

As mentioned in Chapter 2, around 87% of the Low Saxon speakers in Groningen and Drenthe perceived their dialect as a variant of Dutch, while this was only the case for 16% of the Frisian speakers in Fryslân. In addition, dialect leveling occurs between Standard Dutch and these regional languages (e.g., Heeringa & Hinskens, 2015, and see also Chapters 5 and 6), and Standard Dutch itself has become less uniform (Grondelaers et al., 2016), which probably strengthens the perception of regional language varieties as dialects of Dutch. This perception pattern can have important consequences, because speakers may be less protective of the language as a part of their identity when they do not mentally separate the regional variety from the standard language (e.g., Groves, 2010, for the case of Cantonese). Given these findings, we predict that a parent's transmission probability decreases if they perceive their regional language (incorrectly) as a Dutch dialect.

Frisian and Low Saxon are reported to be more commonly used in rural than urban areas (Klinkenberg et al., 2018, and see also Figures A.1 and Figures A.2 in Appendix A), which is often the case for regional languages (Goeman & Jongenburger, 2009). Consequently, we investigate the relationship between language transmission and where someone grew up. We expect higher regional language transmission for parents who grew up in less urbanized areas or lived in such an area at the time of the questionnaire, because the regional language is more likely to be embedded in daily life in rural areas.

Speakers of minority languages maintaining positive attitudes about their language often use it more frequently (Wölck, 2004, p. 9), which we assume leads to increased language transmission (Bell, 2013). Language attitudes are affected by regional language attitudes in wider society (Dragojevic et al., 2021, p. 15), by whether there is a pragmatic purpose to the language (e.g., economically; Harbert, 1999; Gao, 2009), and by whether that language is part of someone's identity (Tseng, 2020). We assume that more positive language attitudes are associated with greater use and willingness to transmit the regional language.

3.2 Methods

3.2.1 Sample — To investigate the associated factors of language transmission for Frisian and Low Saxon speakers in the northern Netherlands, we rely on the participant pool of the Lifelines cohort study (Scholtens et al., 2015; Sijtsma et al., 2022) and the questionnaire that was distributed to it, as we did in Chapter 2. The obtained sample had around 38,500 respondents. We focus on the 25,606 respondents who indicated speaking Frisian or Low Saxon in their traditionally associated provinces: 10,116 Frisian and 1,143 Low Saxon speakers in Fryslân, 7,610 Low Saxon speakers in Groningen, and 6,737 Low Saxon speakers in Drenthe.

The large sample size is advantageous for the generalizability of our findings, but the Lifelines cohort study is not fully representative of the northern Dutch population (Klijs et al., 2015), as indicated in Section 2.2. Nonetheless, Klijs et al. (2015) concluded that the risk of selection bias in the Lifelines sample was low, because the sample broadly represents the northern Netherlands on socioeconomic factors (as well as diseases and general health). There is also no specific reason to assume that the data for specifically those speaking a regional language are biased (as the main problem in Chapter 2 was the ratio between the regional and non-regional speaker groups and not within groups), so we deem the Lifelines data generally suitable for regional language transmission analyses.

3.2.2 Variables — The exact question formulations used in the questionnaire can be found in the supplementary material. We used the question asking whether someone used their regional language with at least one of their children (i.e., as a binary variable) as our language transmission indicator. We excluded people who indicated that they did not have children (N = 4,467; approximately 17%),

because they cannot transmit the language through the mechanism we investigate.

3.2.2.1 The respondent's background. Educational attainment was included as a factor variable (low, middle, or high[2]) and as a binary variable later in the modeling procedure (contrasting high educational attainment with the other levels[3]). Participants were asked whether they speak a regional language or dialect spoken in the Netherlands, and more specifically whether they could participate in a simple conversation in that language. If this was the case, this was taken as evidence that someone could speak a language and someone was included in the study. Respondents were also asked to indicate which other national languages they could speak with the same conversational fluency threshold.

The respondents were additionally asked whether they perceived their dialect as a variant of Dutch or another language. They also filled in the exact age at which they started acquiring their regional language, which we coded as a binary variable: early acquisition in case someone filled in four or lower (i.e., primary school age in the Netherlands or earlier), and later acquisition for ages higher than four. Participants were furthermore asked whether they acquired their regional language primarily through their parents or in another way (e.g., their environment), which was encoded as a binary variable.

3.2.2.2 Language use. We also included in our language questionnaire whether individuals used their regional language with their children's other parent or guardian (i.e., as a binary variable). The specific question formulations can be consulted in the supplementary materials, but we ensured that various possible scenarios were covered. For example, we asked whether someone currently speaks with or used to speak with the other parent or guardian, and they could also indicate that the question was not applicable (e.g., in case there was no other parent or guardian). We quantified how often someone used their language, averaged over 12 specific environments, using a five-point scale ranging from 'never' to 'always'. Finally, a set of binary variables about language use with others in their close environment was included (e.g., parents, siblings, neighbors, or friends). Missing values were assigned when a variable did not apply to the respondent (e.g., in case of death of the other parent).

[2]See the Lifelines Wiki for details about this coding scheme: https://wiki-lifelines.web.rug.nl/.
[3]The binary variable appeared to be more statistically informative than the three-level contrast.

3.2.2.3 Geographical variability. The Lifelines cohort study includes the postal code of respondents at the time of the questionnaire by default, which we used to derive longitude and latitude coordinates and a factor variable representing their place of residence. Respondents additionally indicated in which Dutch province and geographical location they were (mainly) raised during the first twelve years of their lives from a drop-down list of known locations, which was included as a factor variable.

Statistics Netherlands also provides a measure of how urbanized an area is by counting the number of postal addresses in a square kilometer: *omgevingsadressendichtheid* 'domicile address density' (OAD; Centraal Bureau voor de Statistiek, 2023c). More urbanized areas have higher OAD values. We averaged the OAD values for each geographical location instead of the postal code, because affluent neighborhoods may have a relatively low OAD but may reside in a highly urbanized environment. This information was only available for a respondent's place of residence at the time of the questionnaire.

3.2.2.4 Language attitudes. We averaged how positive respondents were about their regional languages based on 13 statements, measured on a seven-point Likert scale ranging from 'very strongly disagree' to 'very strongly agree' (see the supplementary material for details). The topics included whether someone's regional language should remain in use, should be taught in schools, or is a part of someone's identity, but also whether it is advantageous to know the regional language in that region. We included three statements about prejudices toward the regional language. The average score of these three statements was included in addition to the general language attitude score based on all 13 statements.

3.2.3 Statistical modeling — Given the many interrelated factors, a structured statistical modeling approach was necessary to obtain an optimal model. We constructed logistic generalized additive mixed models (GAMMs) predicting language transmission in a forward stepwise procedure. GAMMs can incorporate both non-linear and linear relationships between the predictors and the dependent variable (Wood, 2017). They can also account for structural variability associated with random-effect factors in the data, such as the respondents' location of residence and upbringing. We followed the procedures proposed by Wieling (2018), comparing models that differ only by a single model term using the *itsadug* package for model comparison (Van Rij et al., 2022). We reported the explained deviance as an effect size indicator for individual predictors and the

final model, which is a generalization of the explained variance for non-Gaussian GAMs (Wood et al., 2016).

3.3 Results

Table 3.1: Coefficients for a logistic GAM predicting regional language transmission. Note that the reference level of the interactions is Frisian (FR) to which Low Saxon (LS) is compared.

	Estimate	SE	*z*-value	*p*-value	
(Intercept)	-3.88	0.20	-19.06	<0.001	***
Language use with the other parent (yes)	2.40	0.08	30.87	<0.001	***
Language use frequency in in different environments	1.27	0.04	35.59	<0.001	***
LS vs. FR language user	0.09	0.20	0.45	0.660	
High educational attainment (yes)	0.01	0.08	0.13	0.900	
Early age of acquisition (yes)	0.62	0.10	6.13	<0.001	***
Number of national languages the participant can speak, including Dutch	-0.21	0.03	-8.23	<0.001	***
Language use with the other parent (yes) × LS vs. FR language user	-0.67	0.10	-6.66	<0.001	***
High educational attainment (yes) × LS vs. FR language user	-0.32	0.10	-3.15	<0.010	**
Early age of acquisition (yes) × LS vs. FR language user	-0.39	0.13	-2.97	<0.010	**

The summary of the final model is provided in Tables 3.1 and 3.2. Marginal effect plots for the included predictors are shown in Figures 3.1 through 3.4. Note that the overall transmission rate across the sample for each language (Frisian: 66%, Low Saxon: 40%) cannot be derived from the model summary due to the interaction terms with the regional language contrast. A descriptive summary of all variables used in the model is given in the supplementary material.

The total explained deviance of the final model was 50.6%, and most of the model's predictive power came from only a few predictors. Regional language use with the other parent or guardian accounted for 35.0% of the explained deviance, and language use frequency accounted for a further 10.5% of the explained deviance. The other model terms accounted for the remaining 5.1% of

Table 3.2: Smooth terms for a logistic GAM predicting regional language transmission.

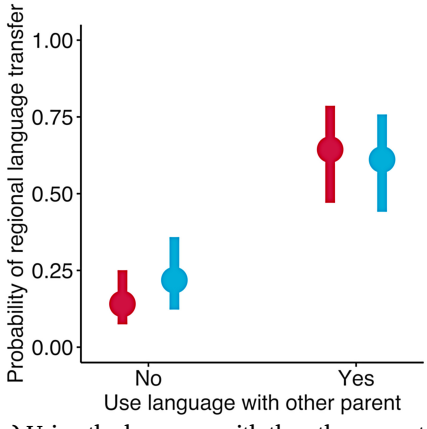
	Est. df	Ref. df	z-value	p-value	
Positive language attitude (FR)	2.83	3.56	73.95	<0.001	***
Positive language attitude (LS)	1.00	1.00	291.55	<0.001	***
OAD (FR)	2.89	3.36	22.62	<0.001	***
OAD (LS)	3.88	4.50	30.62	<0.001	***
Longitude $\times$ Latitude	11.37	14.55	99.40	<0.001	***
Location of residence (random intercept)	44.62	726.00	56.89	<0.010	**
Location growing up (random intercept)	97.81	1202.00	165.29	<0.001	***

the explained deviance, which is substantially less than the major effects of language use with the other parent or guardian and language use frequency, but the variables did improve the model fit sufficiently to be included in the final model.

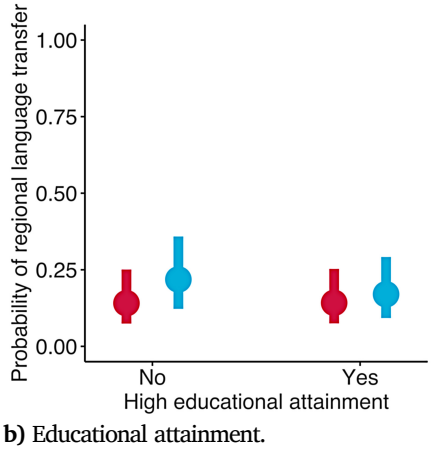
We checked for interactions between intergenerational transmission and the regional language, as we contrasted Frisian with Low Saxon speakers. The increase in estimated transmission rate was higher for Frisian than Low Saxon when the other parent also speaks the same language (see Figure 3.1a). There was no significant difference between people with different educational attainment levels for Frisian respondents, while Low Saxon respondents with lower educational attainment had a higher transmission probability (see Figure 3.1b). Additionally, the transmission probability increased for respondents who indicated an early age of acquisition, but this effect was weaker for Low Saxon speakers (see Figure 3.1c).

The transmission rate was higher for respondents residing in areas with a very low address density and also for respondents in areas with the highest address densities (see Figure 3.2a). The estimated transmission rate increase of Low Saxon was higher in areas with a low address density compared to the transmission rate increase of Frisian in low address density areas. Furthermore, respondents with a more positive language attitude were more likely to transmit their regional language (see Figure 3.2b). This effect was strongest for Low Saxon respondents.

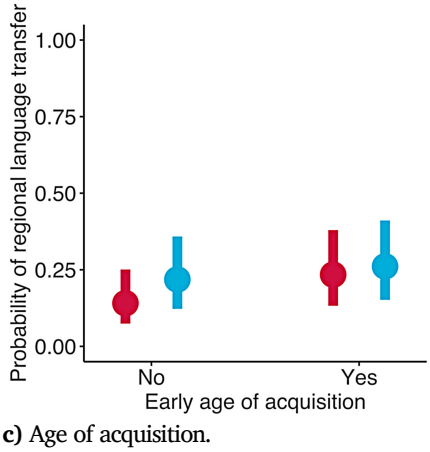
The contrast between Frisian and Low Saxon speakers was important in most cases, but there were two exceptions. The transmission probability increased with higher levels of language use frequency (see Figure 3.3a), and this effect was not statistically different for the two regional languages (but note that Frisian speakers have higher levels of language use frequency overall; see the model summary in the supplementary material). When respondents indicated speaking



a) Using the language with the other parent.



b) Educational attainment.

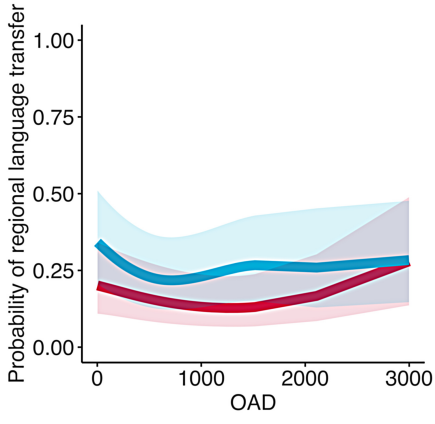


c) Age of acquisition.

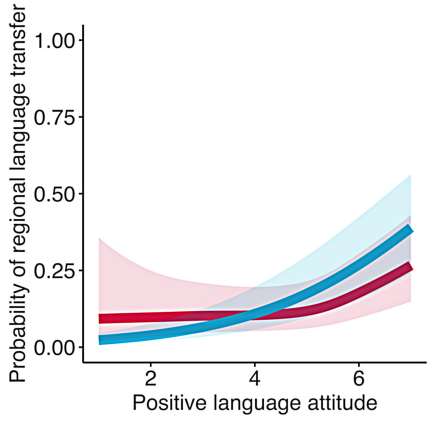
Figure 3.1: Marginal interaction effects for language transmission by use of the regional language with the other parent (top), educational attainment (bottom-left), and age of acquisition (bottom-right). Frisian is shown in red and Low Saxon in blue.

more national languages, this resulted in a lower regional language transmission probability (see Figure 3.3b), again without a statistical difference between the two regional languages.

There is one specific area where a relatively low rate of language transmission remained after accounting for all other variables: the northeast of the province of Groningen (see Figure 3.4). The lower transmission rate fanned out from that

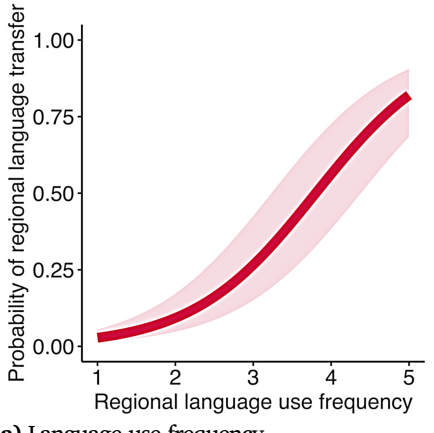


a) Environmental address density.

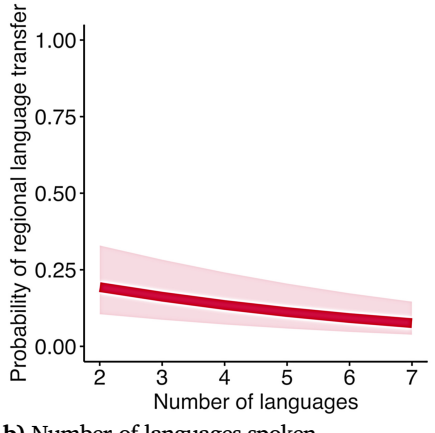


b) Positive language attitude.

Figure 3.2: Marginal interaction effects of environmental address density (OAD; left) and positive language attitude (right). Frisian is shown in red and Low Saxon in blue.



a) Language use frequency.



b) Number of languages spoken.

Figure 3.3: Marginal effects of regional language use frequency in different environments (left) and the number of languages a respondent speaks (right). Frisian is shown in red and Low Saxon in blue.

area and still covered a substantial part of the rest of the province.

We discuss these findings in Section 3.4, but we first explain why other variables are not included in the final model. Some variables (e.g., a respondent's perception of their language variety as a Dutch dialect or not) were not in-

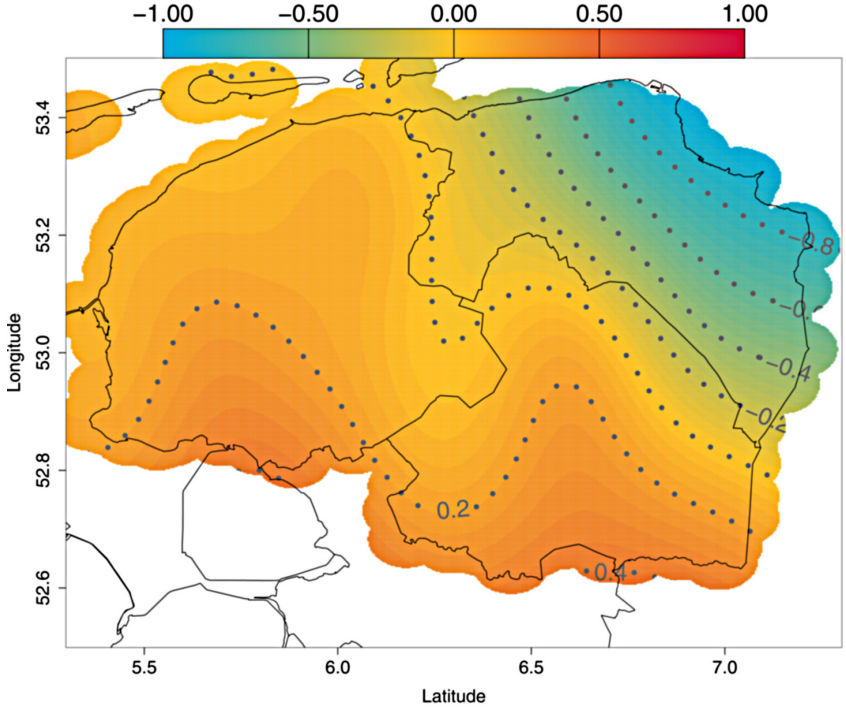


Figure 3.4: Marginal geographical effect smooth of the final model predicting regional language transmission. Redder colors indicate higher probabilities, whereas bluer colors indicate lower probabilities.

cluded, because other (statistically linked) variables had more explanatory power (e.g., language use frequency). There were also variables (e.g., early age of acquisition) that were strongly correlated with other more informative predictors (e.g., whether a respondent acquired the language through their parents), in which case the predictor with the greatest explanatory power was kept. Finally, some predictors were left out, because the direction of coefficients associated with other variables in the model reversed when these were included in the final model (e.g., specific statements about prejudices), which means that these were collinear.

3.4 Discussion

Stable and long-term survival of a language variety largely depends on language transmission from parents to their children, which Wölck (2004, p. 7) even named as the strongest candidate for a language maintenance universal. Using an exploratory procedure and a large-scale questionnaire in the northern Netherlands, we investigated which factors are associated with parental language transmission for Frisian and Low Saxon to address research question 2 (see Section 1.4).

We found that two variables, best described together as a reflection of the entrenchment of the regional language, strongly influenced transmission rates of Frisian and Low Saxon. A higher transmission rate was mainly associated with whether someone spoke their language with the other parent or guardian of their children. By definition, both caregivers can speak and transmit the regional language in this scenario. Furthermore, speakers who used the language more often showed considerably higher language transmission rates. However, the frequency of language use may be influenced by several processes. Even if speakers are internally motivated to use their language in many contexts, they are often increasingly expected to use the ubiquitous Standard Dutch (Hinskens & Taeldeman, 2013, p. 5) in nearly all social contexts (except close personal circles). Opportunities to use the regional language are also generally rarer outside of rural areas, which is particularly true for Low Saxon speakers (see Figures A.1 and A.2 in Appendix A). Regional language users are unlikely to move to a different environment to use their language, so this finding does not directly aid in potentially increasing language maintenance.

An early acquisition onset (i.e., before attending primary school) was associated with a higher transmission probability. This variable largely represents whether the language is a common mode of communication in a person's family, because there is a strong association with whether someone uses the language with their parents, siblings, and their close environment (see the supplementary material). The transmission probability gap between early and late acquisition is greater for Frisian than for Low Saxon (see Figure 3.1c), although more people learn Frisian at a later age through one of the more widely available Frisian language courses (see the supplementary material and also Kircher et al., 2023), which is relatively unlikely for Low Saxon. Furthermore, it may also simply reflect that more people learn Frisian at a young age.

More positive language attitudes were associated with higher language transmission rates, particularly for Low Saxon parents. Similarly, strongly negative

attitudes were also more strongly associated with lower transmission rates for Low Saxon parents than for Frisian parents. This finding suggests that fostering positive language attitudes for speakers of the lesser-maintained Low Saxon language may be beneficial to ensure its continued transmission and reduce its decline.

As expected, people with high educational attainment showed lower language transmission rates than those with lower or middle educational attainment. We first assessed a three-level operationalization, but the most important contrast was found between respondents with high attainment and those with lower educational attainments. The effect was non-significant for Frisian speakers, which may mean that the use of Frisian is less stigmatized than Low Saxon use for speakers from different educational backgrounds.

We also assessed whether transmission rates were higher in rural settings (as measured by a lower domicile address density), which Figure 3.2a shows is the case. However, the transmission probability did not decrease monotonically from more rural to more urban areas. The transmission probability was higher for both strongly rural and strongly urban areas, although this is not directly visible in the geographical distribution of regional language transmission (see Figures A.1 and A.2 in Appendix A). These effects are probably best understood in the context of the ongoing worldwide urbanization (Zhang, 2008), with many rural regional language speakers moving to larger towns and cities for better employment opportunities. In other words, many regional language users move to urban areas, but this does not mean that urban areas inherently benefit regional language transmission.

The transmission rates were lower in the northern and eastern parts of Groningen (see Figure 3.4). This residual geographical variation is striking, given the strong explanatory power of the variables already in the model. It is unclear why this area's remaining regional language transmission rate is lower. We tested whether the predictors used in the modeling procedure showed unique geographical patterns in this region (see the supplementary material). The age of acquisition in this region was on average around one year later than the average (approximately 2.6 years). This suggests that more speakers start acquiring Low Saxon in primary school than elsewhere and may consequently be less likely to transmit the language later in life. An in-depth follow-up study of this region is useful, as it may reveal information about this process and help to identify further variables relevant to regional language transmission.

We found evidence that speaking more national languages was associated

with a lower regional language transmission rate (see Figure 3.3b). People who speak more languages appear unlikely to also learn a regional language. This is possibly due to the lower prestige of regional languages, which causes the regional languages to be perceived as less useful than other national languages. Furthermore, positive language attitudes may be a relatively recent development in someone's life or society (e.g., regional languages have received more positive attention in recent decades; Goeman & Jongenburger, 2009; Slaats, 2020). If the parents had already decided not to transmit their language when they raised their children, positive attitudes in the present would not have influenced their behavior in the past. Language attitudes are also known to be dynamic (Lenz, 2009) and susceptible to external influences, such as language policies (Dragojevic et al., 2021). Regional languages are structurally subordinated to Standard Dutch due to language policies in the Netherlands (see Section 1.3.1 in Chapter 1 for details), so this may deter parents from transmitting their regional language.

Some of our findings may be used for fostering positive language attitudes (especially toward Low Saxon) from an institutional perspective, which fits into some recent governmental trends promoting regional languages (e.g., in teaching in secondary school; Prenger et al., 2023). Research on this topic rarely trickles down to language family policy (King et al., 2008, p. 913), so governmental organizations with a broad geographical scope are also advised to actively support parents in their multilingual child-rearing (as voiced by Wang & King, 2024). Studies focusing on parents with positive language attitudes (a substantial part of the population; see supplementary material) can assess how these parents can be encouraged to transmit their regional language and convince others to follow their example.

Finally, our results suggest that the differences between Frisian and Low Saxon can potentially be mitigated by giving Low Saxon an equal status to Frisian. Both languages have been protected in the Netherlands since the 1990s under the European Charter for Regional and Minority Languages, but Frisian is included under Parts II and III and Low Saxon only under Part II. An attempt to include Low Saxon also under Part III was rejected by the Dutch government in 2012 (see Parliamentary paper 2012Z05658[4]), but the political landscape frequently changes and new attempts are legally possible. By giving these languages equal status, the Dutch government can amend the incorrect perception of Low Saxon varieties as Dutch dialects. If this perception is successfully corrected,

[4] See https://www.tweedekamer.nl/kamerstukken/brieven_regering/detail?id=2012Z05658&did=2012D12185.

language attitudes will likely become more positive among Low Saxon speakers (e.g., due to a greater pride) and people who do not speak the language (e.g., due to more respect for the language and its users). People will likely also be more inclined to teach or learn the language, which is reflected by the greater number of Frisian respondents indicating they learned the language in other ways than through their parents (e.g., in school or a language course; see supplementary material and also Stefan, 2022). Ideally, the Dutch government regularly evaluates language attitudes towards regional languages after including Low Saxon under Part II to demonstrate the effectiveness of preserving linguistic diversity with government support to the international community.

3.4.1 Limitations — We only analyzed data from speakers in the northern provinces, but Low Saxon varieties are also widely spoken in the provinces of Overijssel and Gelderland (see Figure 1.2 in Chapter 1). Low Saxon is used by more speakers in Groningen and Drenthe than in other areas (Bloemhoff, 2005), and there may be other sociolinguistic differences between these areas. For example, the pressure from Standard Dutch on the local Low Saxon dialects seems to be greater for the southern Low Saxon areas (Wieling et al., 2011), which may strengthen perceptions of Low Saxon as a dialect of Dutch and influence language attitudes in those areas.

As mentioned in Chapter 2, a total of 87% of Low Saxon respondents in the Lifelines data perceived their dialect as a variant of Dutch, while only 16% of Frisian respondents indicated this. This state of mind may have serious repercussions, because reevaluation of varieties has been a powerful driver of political and sociolinguistic change in the past (Lenz, 2009). Furthermore, speakers who see their variant as a separate language may have a stronger tendency to support forming a written standard (see, for example, Groves, 2010). A detailed investigation of what this pattern could entail for interpretations of questionnaire responses is warranted, especially for questions comparing Dutch and regional language use. If someone puts these languages in the same category, the interpretations may differ accordingly.

Finally, children are also active agents in family language policy and also subject to external pressures (e.g., from school and peers; Dragojevic et al., 2021), which can lead to different preferences and resistance to learning a minority language (King et al., 2008; Fogle & King, 2013). Such preference mismatches are not unlikely for Low Saxon, given the strong decline in parental transmission in the two recent generations (see Table 2.2 in Chapter 2), and renewed interest

in dialects in the Netherlands since the end of the last century (Slaats, 2020). This could result in scenarios in which children may want to speak the regional language, while the parents discourage the use of the language.

3.5 Conclusion

We found that parental transmission of Frisian and Low Saxon in the Netherlands is primarily associated with environmental factors, such as frequent regional language use in different social contexts and the other parent or guardian speaking the same language. Other higher transmission contexts include when parents have positive language attitudes. The transmission rates can be high in both strongly urbanized and rural areas. The transmission rate is usually lower among parents with higher educational attainment. The methodology employed in this study cannot ascertain the direction of causality, so experimental intervention studies are necessary to explore how these different factors interact.

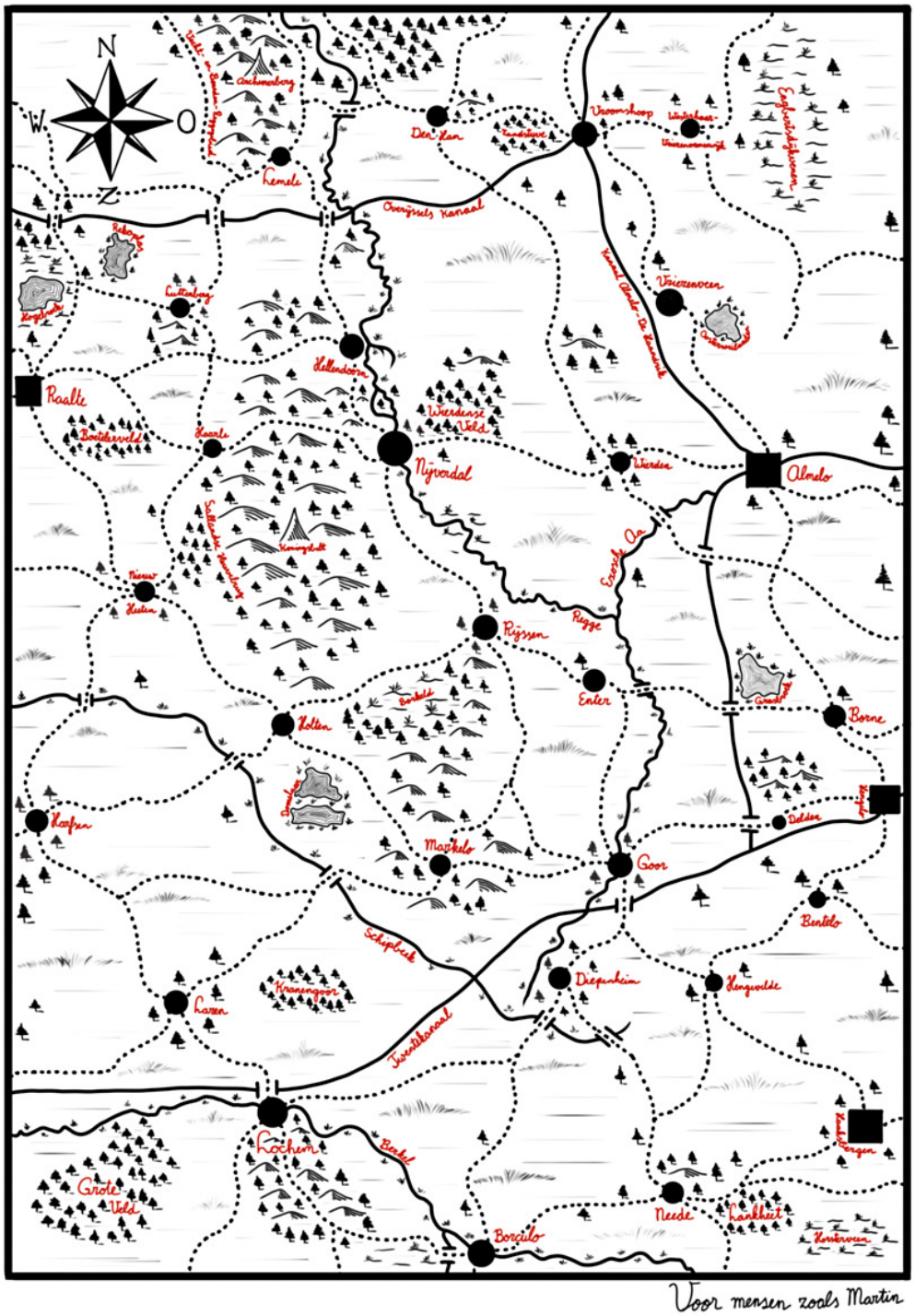
To stimulate regional language use, our results suggest that it would be beneficial if governmental organizations support Frisian and Low Saxon parents with multilingual child-rearing advice and practices. Furthermore, future studies may explore how parents with positive language attitudes can be activated to transmit their language to their children and encourage other parents to follow their example. These complementary approaches are necessary, because these regional languages are otherwise likely to disappear in favor of Standard Dutch, further increasing the substantial ongoing loss of linguistic diversity worldwide.

Part III

Pronunciation changes

Rijssen, Overijssel

CHAPTER 4



Voor mensen zoals Martin

REDUCING TRANSCRIBER VARIABILITY IN PHONETIC CORPORA

Abstract

Large phonetic corpora are frequently used to investigate pronunciation variation and change of regional varieties, but constructing them requires much effort. Many transcribers are often involved in making the phonetic transcriptions, usually resulting in inter-transcriber issues that may impact the reliability of analyses using these data. This problem is exacerbated when multiple phonetic corpora are compared when investigating real-time dialect change. In this chapter, we propose a method to automatically and iteratively merge phonetic symbols used in the transcriptions to obtain more coarse-grained phonetic transcriptions. Our approach is evaluated using two large phonetic dialect corpora focusing on dialect variation in the Netherlands and Flanders, which we use to estimate pronunciation change in this area in the 20th century.[1]

4.1 Introduction

To investigate pronunciation variation and change, researchers often rely on large data collections of transcribed speech. Depending on their specific interests, researchers consequently investigate a few linguistic variables in detail (the dialectological tradition; see Boberg et al., 2018) or many variables at the same time to obtain an aggregate view of dialect variation (the dialectometric tradition; see Nerbonne & Kretzschmar, 2013). The increase in computing power in recent decades has enabled large-scale quantitative analyses, which gave rise to novel dialectometric approaches (see Wieling & Nerbonne, 2015, for an overview). The dialectometric approach minimizes researcher bias due to not having to select a small set of specific linguistic variables, but relying on large data collections may still present some problems. Specifically, inter-transcriber variability may be an issue when many transcribers are involved in the data collection, especially when the transcribers adhere to different transcription processes in different projects.

In this chapter, we focus on research question 3 (see Section 1.4) and address whether phonetic corpora that differ in transcription practices can be meaningfully compared for pronunciation change analyses. To answer this research

[1] This chapter is adapted from: Buurke, R. S. S. J., & Wieling, M. (2023). Sound Change Estimation in Netherlandic Regional Languages: Reducing Inter-Transcriber Variability in Dialect Corpora. *Taal en Tongval*, *75*(1), 7–28. https://doi.org/10.5117/TET2023.1.002.BUUR

question, we specifically focus on two corpora that are often used for investigating Dutch dialect variation: the Reeks Nederlandse Dialectatlassen (RND; Blancquaert & Pée, 1930) and the Goeman-Taeldeman-van Reenen Project (GTRP; Taeldeman & Goeman, 1996). Goeman and Taeldeman indicated that their project, for which data were collected between 1980 and 1995, may be seen as a refinement of the RND, for which data were collected between 1922 and 1975.

Both datasets are described in more detail in Section 1.3.3 of Chapter 1, but it is worth noting that inter-transcriber issues have been pointed out for the GTRP before. Hinskens and Van Oostendorp (2006) analyzed nasal-plosive clusters (such as [nd]/[nt]) in the GTRP, and they noticed that these clusters were transcribed differently by transcribers from different areas. For example, transcriptions made by transcribers for Frisian varieties were structurally different from those made for Dutch varieties. In addition, the set of International Phonetic Alphabet (IPA) symbols used by the GTRP transcribers in the Netherlands is much larger than the set used by the transcribers in Flanders (Van Oostendorp, 2007; Wieling et al., 2007).

Inter-transcriber differences are not a novel phenomenon. They are a well-known problem for phonetic transcriptions and such differences are only expected to increase when transcribers have different linguistic backgrounds (see, e.g., how [l] and [r] are perceived differently by American and Japanese listeners; Miyawaki et al., 1975; Kuhl, 2004). Even among experienced transcribers, the inter-rater reliability is rarely higher than 80% (Amorosa et al., 1985), and it is even lower when narrow phonetic transcriptions are made as opposed to broader ones (Shriberg & Lof, 1991). Naturally, these transcriber effects do not devalue phonetic transcriptions, because a significant part of the pronunciation variation is likely still accounted for, but transcriber effects are detrimental to the reliability of an analysis and should therefore be minimized.

The degree to which inter-transcriber issues are a problem depends on the scale and the type of analysis. A synchronic study of a comparatively small geographical area is less likely to suffer extensively from such systematic inconsistencies than, for example, a diachronic (real-time) study of a large geographical area. In the latter scenario, more transcribers are usually involved, and the increased territory requires the transcriber to consider more linguistic variation. These factors substantially increase the task difficulty. We focus on this type of scenario and attempt to alleviate inter-transcriber issues in a relatively extreme case. If the approach proves effective, it can be fine-tuned for less extreme cases.

Specifically, in this study, we attempt to estimate the magnitude of pronunciation change across the Netherlands and Flanders, by comparing the GTRP and the RND datasets. To compare two datasets that differ in how narrow the phonetic transcriptions are (i.e., substantially more phonetic symbols were used for the GTRP transcriptions), we iteratively merge similar phonetic symbols until a shared phonetic symbol inventory remains. While information about variation is potentially lost before the final statistical analysis through this procedure, not all variation is equally meaningful. By merging symbols in the phonetic transcriptions that are most similar to each other, we effectively generate a less narrow phonetic transcription with likely fewer inter-transcriber inconsistencies. This approach is described in detail in Section 4.2.3, after which we explain how we estimate pronunciation change from the resulting transcriptions.

To evaluate the results after simplifying the phonetic complexity of the corpora, it is necessary to compare them to what is known about the regional languages within the geographical area. Varieties of three main language families are spoken in the Netherlands and Flanders: Frisian, Low Saxon, and Low Franconian (see Section 1.1 in Chapter 1 for details). Dialects of these families are influenced by a Dutch roof variety (i.e., Standard Dutch; Heeringa & Nerbonne, 2000; Heeringa & Hinskens, 2015), which is strongly embedded in all social contexts of daily life in the Netherlands and Belgium (although the standard languages differ between the Netherlands and Belgium; see Willemyns, 2003, for details). We expect this vertical convergence to be the main driver of pronunciation change (cf. Heeringa & Hinskens, 2015). Additionally, we assume that the rate of pronunciation change varies between and within the language varieties of interest, especially concerning Standard Dutch convergence. Specifically, language varieties with a relatively stable intergenerational transmission may show less convergence towards Standard Dutch.

For Frisian varieties, we expect pronunciation change to be relatively low, because the language is standardized and taught in school (Bayat et al., 2023), shows a relatively high intergenerational transmission (see Driessen, 2012, and also Chapters 2 and 3), and is protected politically (Hoekstra, 2003). Both Low Saxon and Limburgish are recognized as official regional languages in the Netherlands under the European Charter for Regional or Minority Languages (ECRML; see Chapter 1 for details about this Charter and the included languages). We expect a pattern similar to Frisian for Limburgish, due to its stable intergenerational transmission (Driessen, 2005), which is unique within the Low Franconian family (Versloot, 2021a). For Low Saxon, however, we expect a much larger change

(most likely toward Standard Dutch). While the regional language is recognized under the ECRML and positive attitudes toward the language exist to a certain degree (Ter Denge, 2012), the intergenerational transmission has declined rapidly (Versloot, 2021a, and also Chapters 2 and 3).

As the Low Franconian area is relatively large, we also provide predictions about dialect groups within that family (see Section 1.1.1 in Chapter 1). A small to moderate amount of change is expected for Brabantish and Zeelandic. These language varieties are not recognized under the ECRML and their intergenerational transmission is not very high (for Brabantish even very low; Driessen, 2012). However, they were more similar to Standard Dutch at the time of the RND and GTRP (e.g., due to substantial language contact), so the potential for pronunciation change is limited. We expect the rate of pronunciation change for varieties in Flanders[2] to be in between Low Saxon (comparatively much change) and Limburgish or Frisian (comparatively little change), because variable rates of dialect leveling have been noted for these varieties in recent decades (e.g., Taeldeman, 2013a, 2013b; Swanenberg & Van Hout, 2013). The West Flemish varieties in the area likely exhibit the least change and more eastern varieties more change, due to differences in intergenerational transmission (Taeldeman, 2013a, 2013b).

4.2 Methods

4.2.1 Data — We analyze the phonetic transcriptions of two large dialect corpora across the Netherlands and Flanders: the Reeks Nederlandse Dialectatlassen (RND) and the Goeman-Taeldeman-van Reenen Project (GTRP), which we introduced in Section 1.3.3. The RND was constructed during a much longer period (between 1923 and 1982) than the GTRP (between 1979 and 2001, with most recordings made between 1985 and 1989). The target informants of these corpora were non-mobile older rural males (NORMs), who were seen as desirably conservative language users (Chambers & Trudgill, 1998).

[2] We should note here that we do not necessarily expect convergence directly to the Standard Dutch variety in the Netherlands for these varieties, but instead to Belgian Standard Dutch, which had a relatively stable pronunciation in the time period of interest here (Willemyns, 2003). The varieties are relatively similar at the lexical level, so using Standard Dutch from the Netherlands as a reference point is not unreasonable. See also Vandekerckhove (2009) for a discussion on dialect-standard dynamics in Flanders.

The phonetic transcriptions for the RND sentences (used for eliciting dialect data) are partially available in a digitized form (see Gabmap; Leinonen et al., 2016). The available subset we use in this study consists of the phonetic transcriptions of 166 target words across 347 locations. For the GTRP, the target list for this project largely consisted of single words as opposed to sentences, although some sentences were still included.[3] In total, 1876 items were translated by informants in 613 locations. The translations were recorded and later transcribed phonetically. In this study, we use the GTRP subset extracted by Wieling et al. (2007, also available through Gabmap) which consists of phonetic transcriptions of 562 single words across the 613 locations.

We select locations and target words that overlap between these corpora to estimate the aggregated pronunciation change of regional varieties between the RND and GTRP time periods. The 192 overlapping locations are presented in Figure 4.1, and the 61 words that overlap are presented in Table 4.1. These data can be found in the supplementary materials for this chapter, which can be found at https://osf.io/v94tj/.

Table 4.1: Overlapping target words ($N = 61$) between the RND and GTRP.

bakken	dorsen	flauw	hooi	maart	saus	veel	zee
bier	dorst	gaan	kaas	melk	sneeuw	ver	zes
binden	drie	geld	komen	moe	spannen	vier	zijn
blauw	drinken	geweest	koud	nog	springen	vijf	zuur
brengen	droog	goed	krijgen	ook	stenen	voor	zwemmen
buigen	dun	gras	krom	op	tegen	vuur	
doen	duwen	groen	laten	potten	twee	weg	
dopen	eieren	hebben	licht	rijp	vader	wijn	

4.2.2 Levenshtein distance — To quantify the differences between the phonetic transcriptions, as well as to reduce the phonetic inventories of the corpora, we use a variant of the Levenshtein distance (Levenshtein, 1966) that has been optimized for linguistic purposes. The inputs of this algorithm are two phonetic strings, and the result is a count of how many binary operations are *minimally* necessary to turn one transcription into the other. Three possible operations are allowed to achieve this: insertions, deletions, or substitutions of two phonetic

[3]The full list is available on https://projecten.meertens.knaw.nl/mand/GTRPlijstgeheel.html.

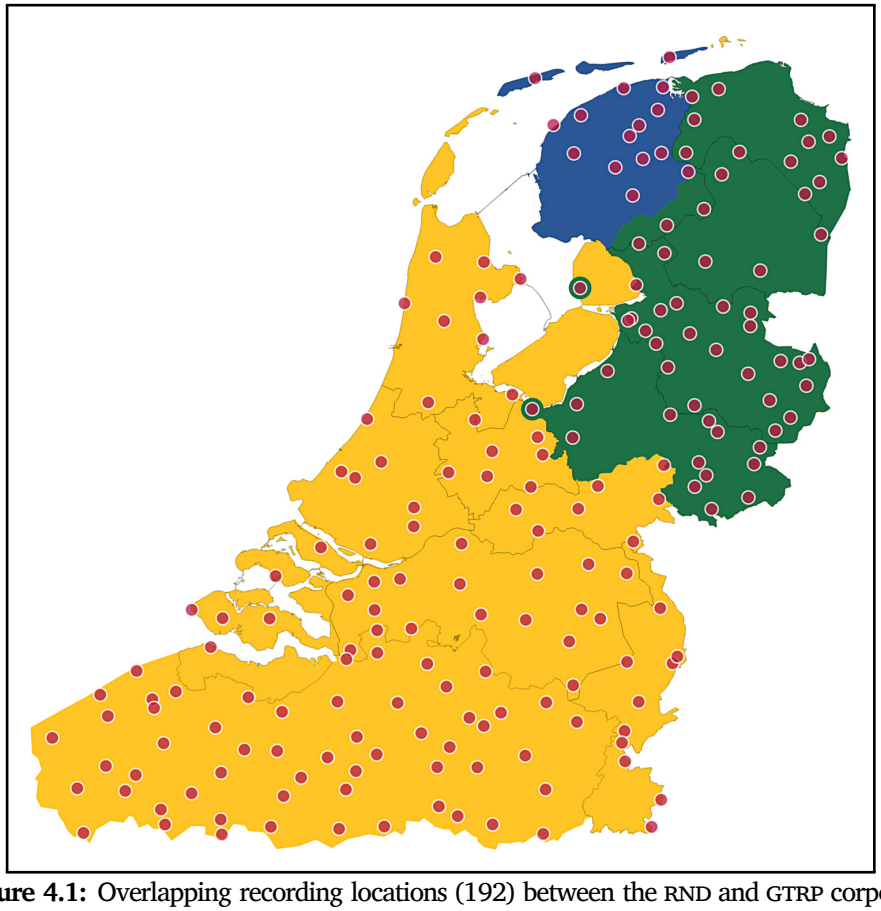


Figure 4.1: Overlapping recording locations (192) between the RND and GTRP corpora. The Frisian, Low Saxon, and Low Franconian areas are marked in blue, green, and yellow, respectively.

symbols simultaneously. These operations have an associated cost of 1 in the algorithm.

An example is given in Table 4.2, for which the transcriptions are taken from Heeringa and Hinskens (2015). The following operations are sufficient to transform [stʀodə] into [stʀɔət].[4] The first three phonetic symbols are already equal and therefore require no transformation. After this, [o] is substituted with [ɔ],

[4]These are two dialectal pronunciations of *straat* 'street' in the DIAREG corpus, which is not used in this chapter, but is described in Sections 1.3.3 in Chapter 1 and 5.2 in the next chapter.

Table 4.2: Levenshtein alignment between dialectal variants of Dutch *straat* 'street'.

	1	2	3	4	5	6	7
String 1	s	t	ʀ	o		d	ə
String 2	s	t	ʀ	ɔ	ə	t	
Operation	-	-	-	sub.	ins.	sub.	del.
Cost	0	0	0	1	1	1	1

the [ə] is inserted, and the [t] and [d] are substituted. Finally, the remaining [ə] is deleted. Four operations were required, which also represent the Levenshtein distance between these transcriptions. To correct for phonetic sequences of different lengths (longer sequences are less likely to be identical), we divide this distance by the alignment length, so the *normalized* Levenshtein distance is 4/7 (≈ 57%).

Note that it is possible to transform [strodə] into [strɔət] using three operations when vowels and consonants can substitute each other (i.e., if all phonetic symbols are substituted). This is not linguistically sensible, however, because consonants and vowels are different categories of sounds. We avoid such alignments by setting the associated cost of vowel-consonant substitutions to be very high. Vowel-consonant substitutions do not occur in that case, because only alignments with the lowest total operation cost are selected.

The Levenshtein distance can be optimized further for phonetic purposes. The binary weights used in the prior example treat all substitutions as equal, but it is sensible to penalize a substitution involving two phonetic symbols representing very different sounds (such as a substitution of [i] by [u]) more than a substitution involving phonetic symbols representing more similar sounds (such as a substitution of [i] by [ɪ]). In that case, the binary weights can be adjusted to a value between 0 and 1 (instead of 0 or 1) that reflects the distance between sounds in phonetic space. However, deriving reliable gradual weights is not a trivial problem, as is demonstrated in detail by Heeringa (2004, pp. 79–120).

We use the approach proposed by Wieling et al. (2012) to obtain gradual operation weights. This approach is based on the co-occurrence patterns of phonetic symbol segments in alignments (such as the one above) of phonetic transcriptions based on corpora, such as the RND and GTRP. The underlying idea is that phonetic symbols representing similar sounds will be more often substituted by each other in phonetic corpora covering language variation (Wieling et al., 2012). This data-driven approach can be used when the dataset from which the

alignments are generated is of sufficient size and has been shown to result in meaningful phonetic distances (Wieling et al., 2012). The Levenshtein distance algorithm incorporating these sensitive phonetic distances has been found to correlate well with the perception of pronunciation differences by listeners (Wieling et al., 2014b). In the approach of Wieling et al. (2012), the phonetic distances between phonetic symbols X and Y are determined via pointwise mutual information (PMI; Church and Hanks, 1990) according to Equation 4.1:

$$\text{PMI}(X, Y) = \log_2\left(\frac{\text{p}(X, Y)}{\text{p}(X)\text{p}(Y)}\right) \tag{4.1}$$

The numerator p(X, Y) reflects how often the relevant phonetic symbols are aligned *together* in Levenshtein alignments based on corpus data (where the pronunciations of each target word are compared between every pair of locations). The denominator is the multiplication of the probability of each phonetic symbol occurring *individually* in these Levenshtein alignments (i.e., the probability of the two phonetic symbols aligning simply due to chance). When phonetic symbols align more often than expected based on chance, the formula results in a positive value, and otherwise a negative value (0 indicates that they co-occur exactly as often as expected). More details about this algorithm can be found in Wieling et al. (2012).

To convert the PMI scores to phonetic distances, we first invert the PMI values according to Equation 4.2 and subsequently normalize the phonetic distances to values between 0 and 1 according to Equation 4.3.

$$\text{phonetic\_dist}_{(X, Y)} = 0 - \text{PMI}_{(X, Y)} \tag{4.2}$$

$$\text{normalized\_phonetic\_dist}_{(X, Y)} = \frac{\text{phonetic\_dist}_{(X, Y)} - \min(\text{phonetic\_dist})}{\max(\text{phonetic\_dist}) - \min(\text{phonetic\_dist})} \tag{4.3}$$

As a next step, a matrix can be constructed between all the phonetic symbols in the corpus, such as the hypothetical one in Table 4.3. The values in this matrix represent the normalized phonetic distances and can be used to weight the operations of the Levenshtein distance algorithm (cf. Wieling et al., 2012).

Table 4.3: Example of a PMI-based segment distance matrix.

	e	ɪ	ɛ	i	ø
e	0.000				
ɪ	0.018	0.000			
ɛ	0.019	0.022	0.000		
i	0.020	0.021	0.020	0.000	
ø	0.023	0.028	0.030	0.030	0.000

The distances obtained with this PMI approach correlate well with speaker perceptions. For example, Wieling et al. (2014a) tasked native speakers of American English with rating how 'nativelike' speech recordings of non-native speakers sounded, and the Levenshtein distance between transcriptions of these recordings compared to those of native American-English speakers correlated highly with the difference in native-likeness perception. In the same way, the PMI approach can be used to approximate perceptual differences between transcriptions of dialectal variants.

We apply the Levenshtein distance specifically on phonetically broad transcriptions, so we ignore any diacritics or suprasegmental information that already exists for the RND and GTRP transcriptions. This is mainly because narrower transcriptions are especially prone to transcriber errors (Amorosa et al., 1985; Shriberg & Lof, 1991), especially in the context of a compilation of transcriptions from many different transcribers (Hinskens & Van Oostendorp, 2006). We can also afford to rely on transcriptions that are less narrowly transcribed, because smaller distinctions between transcriptions at the word level usually disappear in an aggregate analysis that involves averaging over many items (see Wieling & Nerbonne, 2015), and we are mainly interested in the general patterns that occur across the Netherlands and Flanders.

4.2.3 Iteratively merging phonetic inventories — The PMI-based segment distance matrix can also be used to (iteratively) determine which phonetic segments should be merged (i.e., those with the smallest distances) to make two sets of phonetic symbol inventories more comparable. To achieve this, we first assess which phonetic symbols overlap and which are uniquely used in a particular phonetic symbol inventory. The overlapping phonetic symbols comprise the shared symbol set used for merging symbols. For the first inventory, we merge each phonetic symbol not occurring in the second inventory with its clos-

est alternative from the shared symbol set (based on the PMI-based segment distance matrix). After this procedure, we do the same for the phonetic symbols in the second inventory that do not occur in the first one. Note that the phonetic transcriptions are updated after each phonetic symbol merger (reflecting the merger), and a new PMI-based segment distance matrix is generated using these transcriptions. This procedure consequently yields a shared set of (partly merged) phonetic symbols and phonetic corpora that are updated according to this new shared inventory.

More specifically for the RND and GTRP, we follow a three-step approach to prepare the corpora for comparison. First, we merge the phonetic symbol inventories used for the Dutch and Belgian parts of the GTRP (as these differed substantially; Wieling et al., 2007).[5] Second, we merge the obtained GTRP phonetic symbol inventory and the RND phonetic symbol inventory. Third, as there may still be symbols in the resulting symbol inventory that occur very infrequently (which may indicate transcriber inconsistencies), we further merge those symbols with more frequently occurring symbols. For the third step, we require the minimum frequency of each phonetic symbol to be at least 1% of the total number of symbols occurring in all transcriptions. For example, if all transcriptions together contain 1000 phonetic symbols, any phonetic symbol that occurs fewer than 10 times across the combined corpus is merged with the most similar phonetic symbol that occurs sufficiently frequently and replaced in the phonetic transcriptions after every symbol merger.

4.2.4 Estimating pronunciation change — After iteratively merging the phonetic inventories according to the procedure illustrated above, we can determine the amount of pronunciation change by calculating the Levenshtein distance (with PMI-based data-driven phonetic distances) between the RND and the GTRP transcriptions (using the shared phonetic symbol inventory) for each word in every location. As the analysis focuses on the word level (and not the level of individual segments), values may exceed 1 as the operation costs are summed. Longer words have more phonetic symbol segments and consequently a higher potential for pronunciation change to occur. To ensure a more reliable comparison between the pronounced translations of Standard Dutch target words of different lengths, we normalize for different word lengths by dividing the Levenshtein distance by the length of the alignment (conform Section 4.2.2). As

[5] When we merge two symbols in the GTRP, we also update the RND transcriptions accordingly to prevent duplicate mergers.

a result, the pronunciation change value predicted for each RND-GTRP pair of transcriptions lies between 0 and 1.

We model pronunciation change based on geographical coordinates to visualize and evaluate the aggregate pronunciation change across the Netherlands and Flanders. More specifically, these coordinates are modeled as a two-dimensional (geographical) smooth in a generalized additive model (GAM; see Wieling et al., 2011). GAMs can be seen as an extension to linear regression, but with the capacity to deal with non-linear relationships and predictors (Wood, 2017). We use a GAM variant that incorporates beta regression, which is suitable for predicting values of the dependent variable in the interval (0,1), due to the bounded nature of the normalized pronunciation change variable. Pronunciation change values of exactly 0 or exactly 1 are increased or decreased by a small number (10^{-6}) to adhere to the requirements of the beta regression family. For that reason, we report the explained deviance of each model instead of the adjusted R^2, because this is better suited for non-Gaussian models (see Wood, 2017, p. 128, for the computation). Explained deviance may be seen as a generalization of R^2 and interpreted similarly: a higher percentage of deviance explained reflects a better fit of the model to the data.

4.3 Results

Table 4.4: IPA symbols that occur only in a subset of the GTRP and RND.

IPA	GTRP-NL	GTRP-BE	RND
c	✓		
h	✓		✓
ç	✓		
ð	✓		
ħ	✓		
ɒ	✓		✓
ɕ	✓		
ɡ	✓		✓
ɢ		✓	
ɣ	✓		
ɥ	✓		
ɦ	✓	✓	
ɨ	✓		
ɪ	✓		✓
ɬ	✓	✓	
ɱ	✓		
ɵ	✓	✓	
œ	✓		
ɷ	✓		
ɸ	✓		
ɹ	✓		
ɾ	✓		
ɾ	✓		
ʁ	✓		
ʉ	✓	✓	
ʊ	✓		✓
ʋ	✓		
ʌ	✓		
ʍ	✓		
ʎ	✓		
ʏ	✓		✓
ʝ	✓		
β	✓		
χ	✓		

4.3.1 Iteratively merging phonetic inventories — In Table 4.4, we present the 34 symbols that are absent in at least one of the phonetic corpora subsets: the Dutch part of the GTRP, the Belgian part of the GTRP (covering Flanders), or the RND (which did not show different transcription practices between the Netherlands and Flanders and is therefore not separated by country; see Wieling et al., 2007, for details). The check marks indicate in which subset an IPA symbol does occur. The overlapping RND-GTRP corpus (i.e., including the transcriptions from the 192 locations in Figure 4.1) has a total symbol inventory of 70 IPA symbols (of which 34 symbols are shared across all three sets). This symbol inventory is presented in Table B.1 in Appendix B.

Following the three-step procedure described above, we obtained a smaller and shared symbol inventory for the phonetic corpora subsets and iteratively replaced the phonetic symbols in the transcriptions. First, the GTRP-NL symbol inventory (69 symbols) was merged with the GTRP-BE symbol inventory (41 symbols) by identifying the symbols occurring in the GTRP-NL inventory which did not occur in the GTRP-BE inventory, and vice versa. A total of 30 symbols (29 symbols not occurring in the GTRP-BE and one symbol not occurring in the GTRP-NL) were iteratively merged with the most similar alternative symbol occurring in both inventories.

In the second step of the procedure, we merged the GTRP and RND symbol inventories obtained after the first step. As there were 38 overlapping symbols between the two sets, two symbols were merged with their most similar alternative occurring in both inventories. The resulting set of 38 symbols consisted of 14 vowels and 24 consonants.

As mentioned earlier, we continued merging symbols based on a frequency constraint in the third step of the procedure. If any phonetic symbol had a lower frequency than 1% of the total number of symbols occurring in all transcriptions after the second step of the procedure, it was replaced by its closest phonetic alternative within the iteratively merged GTRP-RND subset. After this procedure, the final phonetic inventory consisted of 29 symbols (12 vowels, 17 consonants) and is presented in Table 4.5. In sum, we reduced the original symbol inventory of 70 phonetic symbols (24 vowels and 46 consonants) to a combined symbol inventory of 29 symbols (11 vowels and 18 consonants). Consequently, the reduction also resulted in a greater decrease in phonetic symbols representing consonants than vowels.

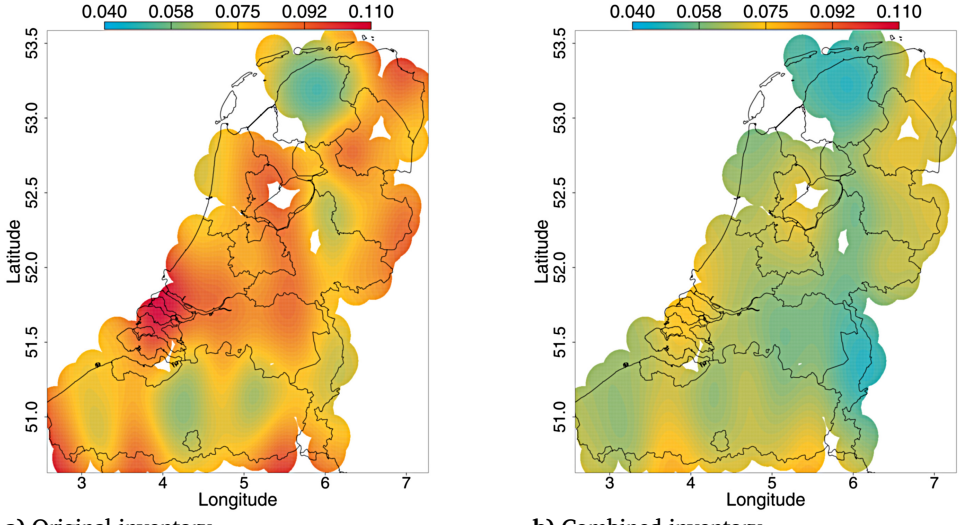
Table 4.5: Combined phonetic symbol inventory. The symbols that were merged due to low frequency are parenthesized.

Used symbol	← Merged symbol(s)						
a	a						
b	b	β					
d	d	ð					
e	e	ɪ					
f	f	ɸ					
i	i						
k	k	c	ɢ	(ʔ)			
l	l	ɬ	ʎ				
m	m						
n	n	ɱ	(ɲ)				
o	o	ɷ					
p	p						
r	r	ɹ	ɾ	ʁ	(ʀ)		
s	s	ʝ	ʈ	(ʃ)			
t	t	θ					
u	u	ʉ					
v	v						
w	w	ʍ	ʊ				
x	x	ç	χ	ɥ			
y	y						
z	z	(ʒ)					
ø	ø	œ					
ŋ	ŋ						
œ	œ	ʏ	ʌ	(ɵ)			
ɑ	ɑ	ɤ					
ɔ	ɔ	ʊ					
ə	ə	ɨ					
ɛ	ɛ	(æ)					
ɣ	ɣ	ħ	g	h	ɕ	(j)	(ɦ)

4.3.2 Aggregated pronunciation change estimation — We estimated pronunciation change in each location across the target words with the combined inventory. For comparison, we also estimated pronunciation change using the original transcriptions. The geographical distribution of pronunciation change was modeled using a generalized additive model (incorporating beta regression) including a two-dimensional smooth based on the longitude and latitude coordinates (see Wieling et al., 2011, for details). We included a random intercept for

each target word to account for the word-specific variability. Location-specific variability was modeled by the two-dimensional smooth and was not included separately as a random effect-factor. The model specification for both models was as follows (i.e., only the transcriptions are different for each model, and consequently the phonetic distances differ):[6]

```
Normalized_change ~ s(Longitude, Latitude, k = 50, m = 1)
                    + s(Word, bs = 're')
```

a) Original inventory. **b)** Combined inventory.

Figure 4.2: Proportions of pronunciation change predicted in the Netherlands and Flanders, based on a geographical smooth (using PMI-based weights). Red: more change. Blue: less change.

The visualizations of the geographical smooths are presented in Figure 4.2. The geographical smooth is significant in both models ($p < 0.01$), which indicates that the geographical distribution of change is not random. The explained deviance of the model based on the original transcriptions is 22%, whereas it is equal to 17.6% for the model based on the combined inventory.

Note that the values have been transformed from logits (the default link function used in beta regression) to proportions and should therefore be interpreted

[6]The parameters k and m define the number of basis dimensions for the smooth and a first-order derivative penalty, respectively, to avoid excessive extrapolation of the data. See https://cran.r-project.org/web/packages/mgcv/mgcv.pdf for further details.

as proportions of change, so (e.g.,) a value of 0.2 indicates an average change of 20% across all words in that area. The pronunciation change estimations are generally higher when the original transcriptions are used than when the combined transcriptions are used, which stems from an inherently higher probability of observing differences. The geographical smooths of Figures 4.2a and 4.2b should therefore not be compared according to their absolute values, but according to the patterns observable within the geographical smooth of each model.

There are several noticeable differences between the geographical patterns in the visualizations. For example, for the model based on the original inventory (Figure 4.2a), there is relatively much change around the border between the provinces of Gelderland and North Brabant. There is relatively little change in this area in the corresponding model based on the combined transcriptions (Figure 4.2b). Furthermore, the change in some areas has become slightly more homogeneous after iteratively merging the symbol inventories, such as in the Low Saxon areas around the Groningen-Drenthe and the Overijssel-Gelderland borders.

The patterns in Flanders are less affected by the combination of the symbol inventories, which was expected, because most phonetic symbol variation in the GTRP was found in the transcriptions of varieties spoken in the Netherlands. The main difference before and after iteratively merging the symbol inventories is the area in the southwest of Flanders. This was an area of relatively much change before the symbol combination procedure, but it is an area of little to moderate change after the procedure. The areas of stability in Flanders seem consistent before and after the procedure, because the areas of relatively much change around the border with Wallonia remain distinct.

The relative rate of pronunciation change in some areas is similar when comparing the two models. The relative stability of local varieties in the provinces of Fryslân and Limburg can still be observed after iteratively merging the phonetic inventory. Similarly, the relatively high rate of change in the Low Saxon area can be identified in both models. The local varieties in the southwest of the province of South Holland and a large part of the province of Zeeland consistently show a relatively high rate of change.

4.4 Discussion

It is difficult to imagine an aggregate-level study on the pronunciation change of dialects without relying on large corpora, such as the RND and GTRP. It is nearly impossible for a single researcher to construct corpora of this magnitude, because gathering the relevant data is costly (both financially and in terms of time). Consequently, such large corpora typically rely on many researchers contributing data, and potential transcriber inconsistencies often present a problem when many transcribers are involved (Amorosa et al., 1985; Shriberg & Lof, 1991). In this chapter, we addressed research question 3 (see Section 1.4) concerning whether it is possible to compare such corpora. For this reason, we presented a method that may mitigate the influence of inter-transcriber inconsistencies. This method relies on co-occurrence patterns of phonetic symbol segments in the corpora (building upon the work of Wieling et al., 2012) and was tested on data from the Netherlands and Flanders, but it may be applied to different geographical areas as well. We additionally posed several predictions specifically for the Netherlands and Flanders, which we used to evaluate the proposed method.

After iteratively merging the phonetic inventories, we created two geographical generalized additive models to interpret aggregate-level pronunciation change across the Netherlands and Flanders. In the following, we will discuss the patterns in the Netherlands and Flanders separately, as the patterns in the Netherlands are much more affected by iteratively merging the phonetic inventories than the patterns in Flanders (because these transcriptions were transcribed using a smaller symbol inventory in the GTRP corpus). When we observed the patterns in the Netherlands, most of the findings from the models were in line with what was expected. The Frisian and Limburgish varieties appeared relatively resistant to change, while the varieties spoken in the Low Saxon area seemed most prone to change. The observed pronunciation change in other regional language areas in the Netherlands mostly fell between these two extremes.

A noticeable exception to the general pattern in the Netherlands is the area located around the border of the provinces of Zeeland and South Holland, which showed a relatively high rate of change. We do not have a clear explanation for this observation. However, the relatively large difference in dates of collection between the RND and GTRP for this area (i.e., 48 years as compared to 29 years on average) may be a contributing factor, as there was more opportunity for change than in other areas. At the same time, the area in Fryslân that consistently showed little change was also an area with many years (i.e., 33 years)

between the RND and GTRP recording, so there are likely additional factors influencing these patterns.

The regional language areas in Flanders were much less affected by iteratively merging the phonetic symbol inventories than those in the Netherlands. One consistent finding comparing the two models was that the area in West Flanders showed relatively little change. The West Flemish varieties are subject to dialect loss, but it has been observed that they remain relatively conservative (Taeldeman, 2013b), which appears consistent with our findings. There is also a southern area in Flanders that showed relatively much change (i.e., near the rate of the Low Saxon area), which we did not express predictions about. It is striking that this change is located around the border with Wallonia (the French-speaking part of Belgium). This may potentially be explained by the linguistically tumultuous history of that border area in the 20th century, during which Flanders and Wallonia, respectively, became Dutch and French in the 1930s (see Willemyns, 2002, for details about directly identifiable border effects).

Comparing the results based on the original inventory to those based on the iteratively merged inventory reveals that the patterns are clearer (globally) and less noisy (locally) when using the PMI-based merger approach. This approach therefore seems capable of revealing the most important (and robust) areas of pronunciation change. In some further analyses (reported in Appendix B.2), we tested whether the patterns changed considerably after iteratively merging the inventories even further with stricter frequency constraints (e.g., a frequency constraint of 2.5% and 5% instead of 1%, as used in the analysis). The patterns remained stable until fewer than 10 symbols were left in the inventory, so the observed patterns of pronunciation change are quite robust. Overall, our findings indicate that the method can extract the most prominent patterns from transcribed pronunciation data while limiting the influence of inter-transcriber issues.

4.4.1 Limitations — It is important to stress that the proposed method of iteratively merging symbol inventories does not inherently reflect an ideal analysis. The approach is useful for reducing complexity in a meaningful and phonetically informed way, but determining the appropriate and optimal level of complexity is not trivial. The frequency constraint, for example, was chosen arbitrarily and could have been set at a different level. However, merging low-frequency symbols does not influence the results much at an aggregate level (see, e.g., Wieling et al., 2009). Nonetheless, it is also possible that the pronunciation variation

in a particular area is inherently noisy, and forcefully reducing the complexity may produce results that are further away from reality. There is no obvious reason to avoid the method altogether, but one should be informed of the existing pronunciation variation of the language area to avoid removing highly salient differences by accident.

It is important to note that the RND data collection spans many more years than the data collection of the GTRP. Consequently, in some locations, pronunciation change may be tracked during a much longer period (i.e., where the RND data were collected at the beginning of the project) than in others (i.e., where the RND data were collected toward the end of the project). Consequently, the observed patterns in different areas cannot be adequately compared.[7] For this reason, we return to investigating pronunciation change in Chapters 5 and 1 of this dissertation.

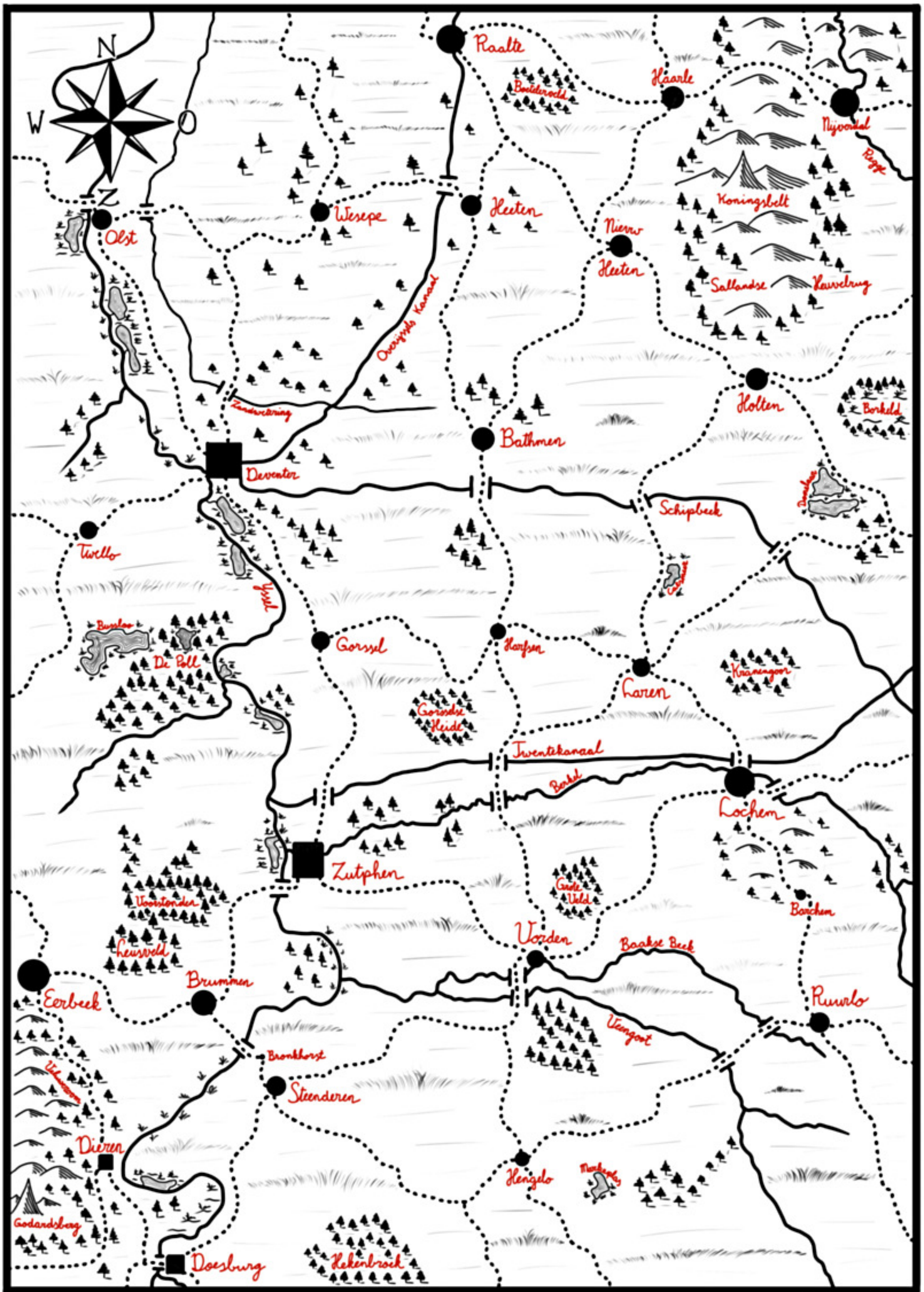
4.5 Conclusion

In this chapter, we have provided a potential solution for reducing transcriber-related differences and making different phonetically transcribed datasets more comparable, by iteratively merging the phonetic inventories and adjusting the phonetic transcriptions to use a shared set of phonetic symbols. Regardless of whether we used the original narrow transcriptions, or the coarser transcriptions with a reduced inventory of phonetic symbols, our analysis of pronunciation change in the Netherlands and Flanders in the 20th century showed the same pattern: the Frisian and Limburgish language areas were comparatively stable, while much change was observed for the Low Saxon area and the area around the province of Zeeland. Although the large differences between the years in which recordings were made in different regions make it difficult to draw reliable conclusions regarding pronunciation change, we have shown that our method of reducing transcriber-related variability and comparing different phonetically transcribed datasets can generate reliable results.

[7] At an earlier stage of the analysis, we attempted to adjust the distances by dividing them by the difference in recording years for each location. However, this resulted in disproportionately small distances for the locations with large differences. Another correction factor might have been applied, but there is no informed way of deriving this correction factor.

Laren, Gelderland

CHAPTER 5


N
W
O
Z
Raalte
Haarle
Nijverdal
Heeten
Nieuw
Heeten
Wesepe
Olst
Koningsbelt
Sallandse
Heuvelrug
Holten
Bathmen
Deventer
Schipbeek
Twello
Gorssel
Harfsen
Laren
Twentekanaal
Lochem
Zutphen
Vorden
Baakse Beek
Ruurlo
Barchem
Eerbeek
Brummen
Bronkhorst
Steenderen
Dieren
Doesburg
Hengelo

Voor Floor

ESTIMATING PRONUNCIATION CHANGE

Abstract

In this chapter, we investigate aggregated pronunciation change of Frisian, Town Frisian, and Low Saxon varieties in the northern and eastern Netherlands, which differ regarding key factors influencing dialect decline. Employing a real-time approach, we analyze phonetic corpora using dialectometric approaches to quantify change among older male dialect speakers across different time periods. We use a multidimensional variant of the Levenshtein distance to estimate how much dialect groups converged to and diverged from Standard Dutch in a period of about 20 years.

Our analyses indicate that pronunciation change is a slow process in this geographical area. The Frisian and Northern Low Saxon dialect groups appear most stable, while Westphalian Low Saxon varieties seem most prone to change. We offer possible explanations for our findings and discuss the shortcomings of the data and approach in detail.[1]

5.1 Introduction

STANDARD LANGUAGES coexist with non-standard varieties in many different constellations (see Auer, 2005, for an extensive typology of language communities in Europe). While there is typically a standardized spelling for the prestigious standardized variety, the pronunciation of standard language speech forms can be much more regionally conditioned (Pedersen, 2005). As mentioned in Chapter 1, pronunciation variation in Standard Dutch is nowadays acceptable to a certain extent (Smakman, 2006; Grondelaers et al., 2016) and regional variation in Standard Dutch is not necessarily perceived as inherently non-standard by native speakers (Grondelaers & Van Hout, 2010).

At the same time, the apparent regional flexibility of Standard Dutch speech does not mean there is no stigmatization of non-standard varieties. In fact, Vousten (1995) observed that especially the language attitudes of parents were a key factor for the decline of dialects in the 20th century in the Netherlands, due to both positive attitudes toward the standard (e.g., learning Standard Dutch was seen as being more useful and prestigious) and negative attitudes toward non-standard

[1] This chapter is adapted from: Buurke, R. S. S. J., Sekeres, H. G., Heeringa, W., Knooihuizen, R., & Wieling, M. (2022). Estimating the level and direction of aggregated sound change of dialects in the northern Netherlands. *Taal en Tongval*, *74*(2), 183–214. https://doi.org/10.5117/TET2022.2.002.BUUR

varieties and their users (e.g., associating regional language use with boorish behavior). The consequently rapid decline of dialects has been observed in Chapter 2 of this dissertation (and by others, e.g., Versloot, 2021a), although the situation of Frisian is considerably better than that of other regional languages in the Netherlands.

The regional varieties are not only spoken by fewer people than before, but they are also under considerable linguistic pressure. Dialect loss typically co-occurs with patterns of 'vertical' convergence (as Auer, 2018, terms convergence toward the standard language or 'roof' variety), and it may also co-occur with 'horizontal' convergence (i.e., convergence between neighboring varieties). The combination of these patterns can result in 'regiolectization', which is a process in which dialects and their neighboring varieties transform into varieties that occupy an intermediate space between the standard variety and the traditional local dialects (see Section 1.3.2 for examples and details). This has been observed for many areas in the Netherlands and Flanders, though mostly in the southern dialect groups (Vandekerckhove, 2009; Cornips, 2013; Swanenberg & Van Hout, 2013; Wilting et al., 2014). A new 'regiolect' may stabilize within a region (Ghyselen, 2015) and therefore, at least partially, preserve dialectal variation. As a consequence, newer generations of dialect speakers typically speak a mixture of geographically close dialects (see similar cases reported by, e.g., Leopold, 1959; Dorian, 1994), which themselves increasingly converge to the standard language in that country. We refer the reader to Auer (2005) for investigations of these dynamics in other European language areas.

In this chapter, we address research questions 4a through 4c concerning pronunciation changes in Frisian, Town Frisian, and Low Saxon: whether there is evidence for regiolect formation of these varieties, and whether they converge to or diverge from Standard Dutch (see Section 1.4). More specifically, we investigate to what degree the pronunciation of these regional languages is changing, and whether the change we observe is mostly toward Standard Dutch or away from it. Chiefly, we explore the usefulness of a phonetic transcription comparison method that analyzes multiple transcriptions simultaneously, because we think this can be particularly useful for the field. We aim to provide insight into whether the dialects of these regional languages are subject to regiolect formation, or are otherwise stable compared to what we know about them from the literature.

5.1.1 Linguistic groups of interest — In line with the research questions of this dissertation (see Section 1.4), we focus on Frisian, Town Frisian, and Low Saxon varieties and we leave the dialects of the Low Franconian group out of our analyses in this chapter (except for Standard Dutch, which also belongs to this group). The advantage of the more specific focus compared to Chapter 4 is that it becomes possible for a single transcriber to process all dialect recordings, preventing the need for approaches (such as the one illustrated in Chapter 4) to correct for transcriber-related variability.

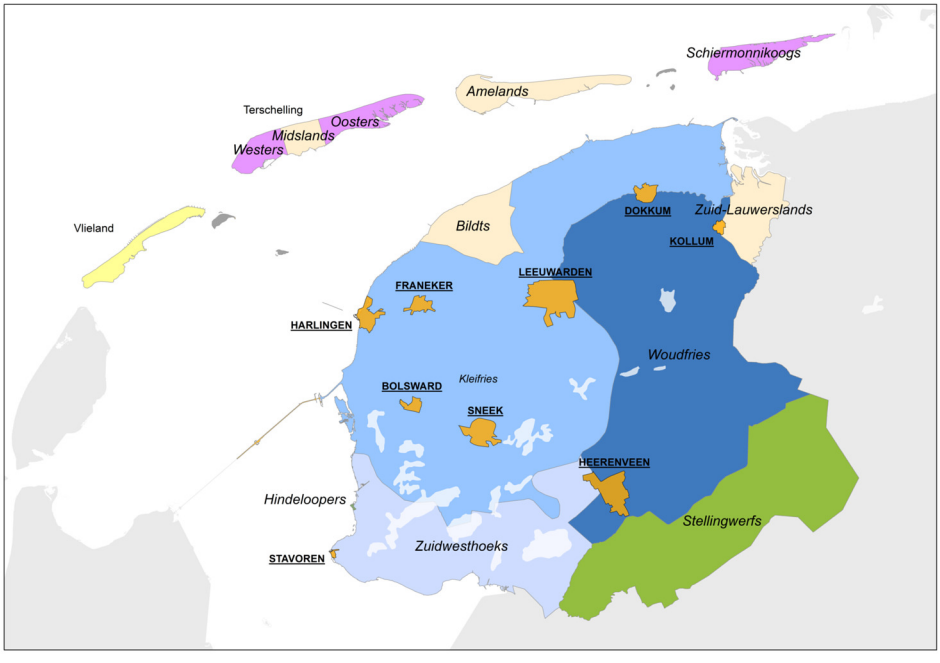


Figure 5.1: Traditional dialects in Fryslân (originally from Van de Velde, 2021). Major Low Saxon varieties are marked in green and major Frisian varieties in blue. Frisian-Hollandic contact varieties, including the towns, are marked in orange (indicating a Hollandic origin) and purple (Frisian origin).

The major regional language divisions in the Netherlands are shown in Figures 1.1 and 1.2 in Chapter 1. The Frisian varieties are shown in more detail in Figure 5.1 and their broad internal partitioning is detailed in Section 1.1.2 of Chapter 1. As mentioned in Chapter 1, we distinguish the Frisian-Hollandic contact varieties from the Frisian varieties and refer to them collectively as Town

Frisian (including Bildts, although it is typically seen as different from Town Frisian; cf. Duijff, 2002). These varieties have a largely Hollandic lexicon and cluster closely with Dutch (Gooskens & Heeringa, 2004; Van Sluis et al., 2016; Van de Velde et al., 2019). They also appear relatively conservative (Heeringa & Nerbonne, 2000; Versloot, 2021b), although their speaker populations are decreasing more rapidly than the surrounding Frisian speaker populations.

Figure 5.2: Dialect map of the Low Saxon area (scanned with permission from Bloemhoff et al., 2020).

The internal partitioning of the Low Saxon dialect continuum is less easily agreed upon than that of the Frisian group. We showed a broad division of the Low Saxon varieties in Figure 1.2 in Chapter 1, and a more narrow division is shown in Figure 5.2. While Bloemhoff et al. (2013a) provides a detailed de-

scription of the main differences between varieties, they also admit that existing classifications of the Low Saxon varieties need further work.

We explicitly aim to investigate differences between the Northern Low Saxon varieties and those part of the Westphalian Low Saxon group. As mentioned and detailed in Chapter 1, this distinction is based on linguistic phenomena and earlier dialectometric studies (Heeringa, 2004; Nerbonne & Heeringa, 2001; Bloemhoff et al., 2020, p. 47). We consider the dialects spoken in the province of Groningen, as well as *Veenkoloniaals* 'Peat colonies dialect', *Noord-Drents* 'Northern Drenthe dialect', and *Midden-Drents* 'Middle Drenthe dialect' in the province of Drenthe as Northern Low Saxon (see Figure 5.2). The other Low Saxon varieties are treated as Westphalian Low Saxon varieties.

5.1.2 Approach to investigating language change — There are two main approaches to investigating language change, and the choice for a particular framework generally depends on the type of data that is available or can be collected. Data can be collected and analyzed under the assumption that the apparent-time construct is valid (see, e.g., Tagliamonte, 2011), or one may opt for a real-time approach instead. Researchers working within the apparent-time paradigm compare speech data from younger generations to those of older ones, and the underlying assumption is that the younger generations use newer linguistic forms than the older generations. This assumption is based on the observation that individual language systems of speakers remain relatively stable after reaching adulthood. However, various studies provide evidence of post-adolescent language change (Blondeau, 2001; Ashby, 2001; Sankoff & Blondeau, 2007; Sankoff, 2019). If suitable data are available, the preferred approach is a real-time analysis, which samples the language at different points in time among speakers of approximately the same age. This approach therefore does not rely on the assumption that individual language systems remain stable for long periods. However, real-time studies are quite costly to conduct (Tillery & Bailey, 2003). Luckily, we benefit from the fact that two large dialect data collections have already covered our linguistic area of interest, which is approximately 20 years in between, and we can reuse these data. Consequently, we will proceed with a real-time approach in this chapter.

Similar to Chapter 4, we use a dialectometric approach to analyze pronunciation changes, which aims to study language varieties (synchronically or diachronically) by aggregating patterns from data (phonetic corpora in our case) across as many locations and words as possible. Dialectometric approaches have

been useful for detecting patterns of the Hollandic expansion (Kloeke, 1927; Wieling et al., 2011), reliably clustering dialectal varieties (Nerbonne & Heeringa, 1997; Heeringa, 2004; Wieling, 2007), and detecting change in dialects due to mutual influence (Heeringa & Hinskens, 2015). We can only estimate the general rate of pronunciation change of dialects at the group level when we sample a sufficiently large representative portion of its speakers and linguistic system.

We again rely on broad phonetic transcriptions for the analyses in this chapter, as we did in Chapter 4. We do acknowledge that comparing phonetic transcriptions abstracts away from potentially interesting features of pronunciation change, such as whether the changes are driven by production or perception (see Garrett, 2015, for an overview of key questions in phonetic change research). Furthermore, narrower phonetic transcriptions can also be useful when studying variation and change at a lower level of aggregation (e.g., at the individual level, Lee, 2016). However, we focus on patterns of change at a high level, namely that of the community, and therefore are interested in more general patterns. Narrow phonetic transcriptions are also relatively unreliable compared to broader ones, even if a single transcriber makes all transcriptions (Shriberg & Lof, 1991).

In a preliminary analysis of aggregated pronunciation change in the whole Netherlandic area, Buurke (2020a) found that the Low Saxon dialects appeared to have changed (irrespective of direction) more than the Frisian dialects in the 20th century (without distinguishing between Northern Low Saxon and Westphalian Low Saxon). This difference in pronunciation change was related to several key differences between the Low Saxon and Frisian situation in the Netherlands. Frisian speakers are more protective of their language than Low Saxon speakers (cf. Ytsma, 2007; Ter Denge, 2012) and the Frisian intergenerational transmission rate is more stable and higher (Driessen, 2012; Bloemhoff et al., 2013b, and see also Chapters 2 and 3). Frisian is also better protected by language policies, as it is protected under Parts II and III of the European Charter for Regional or Minority Languages, while Low Saxon is only protected under Part II (see Chapter 1 for details). These differences suggest that the Frisian varieties are likely to show most resistance to Standard Dutch influence (and thus convergence), while this is less likely for Low Saxon varieties. In line with the limited literature on this topic, we expect limited pronunciation change for the Town Frisian varieties.

We are interested in the rate of pronunciation change and the direction of pronunciation change, so we assess patterns of convergence and divergence between the regional varieties and Standard Dutch in this chapter. Several stud-

ies have shown convergence of specific dialects toward Standard Dutch (Smits, 2005; Giesbers, 2008), and convergence of dialects in the Netherlands as a whole toward Standard Dutch has also been demonstrated (for different periods between the 19th and 20th centuries; e.g., Heeringa & Nerbonne, 2000; Heeringa & Hinskens, 2015). However, these studies also note that the patterns are geographically variable. Convergence and divergence patterns may differ substantially between locations that are relatively close to each other. Overall, we expect to find the same general tendency for our data, with vertical convergence contributing more to language change than divergence.

We focus on four regional dialect groups: Frisian, Town Frisian, Northern Low Saxon, and Westphalian Low Saxon. We expect Frisian and the Town Frisian varieties to be relatively stable in our data due to the aforementioned key characteristics of their speaker populations (e.g., positive language attitudes, stable intergenerational transmission, and more protective language policies, especially for Frisian varieties). We expect that the Northern Low Saxon and Westphalian Low Saxon dialects converge more toward Standard Dutch, because their speaker populations decline rapidly (see Chapter 2). Additionally, the Low Saxon speaker population appears less actively interested in the preservation of regional varieties (Bloemhoff et al., 2013b), while the Frisian speaker population makes an active effort to promote Frisian (e.g., in school; Bayat et al., 2023).

5.2 Data

A real-time analysis requires multiple samples of the same variety at different points in time. Ideally, relevant background variables of the speaker samples are controlled for, such as age, gender, and socioeconomic status. In this chapter, we again rely on the Goeman-Taeldeman-Van Reenen project (GTRP; Taeldeman & Goeman, 1996), which has been described in detail in Section 1.3.3. We also rely on a newer dialect corpus introduced in Section 1.3.3: the *From Dialect to Regiolect* project (DIAREG; Heeringa & Hinskens, 2015).

The background of the speakers in both datasets is relatively similar, because non-mobile older rural males (often abbreviated to NORMs) were contacted for both dialect corpora. This approach is in line with traditional dialectological practices (Chambers & Trudgill, 1998). NORMs were the desired target group, as they are typically perceived to be the most conservative in their language use, have undergone little formal education, and are less influenced by urban

dynamics (where changes are induced through relatively sparse social networks involving diverse groups, whereas changes occur less frequently in the denser and more strongly reinforced social networks typically found in villages; Boberg et al., 2018, p. 11). The language patterns of these speakers are thought to reflect older language forms of their particular space in the dialect continuum. Speakers of the opposite type are mobile younger urban females, who are seen as the most innovative speakers and 'drivers' of linguistic change (Labov, 1990, 1994; Tagliamonte, 2011), although they can also be fluent traditional dialect speakers (Goeman, 2000). Irrespective of this, there may still be considerable individual variation across particular linguistic variables and pronunciation changes (Maclagan et al., 1999), or due to personal preferences and attitudes.

5.2.1 Differences between the phonetic corpora — The tasks used for eliciting dialect data differed between the GTRP and DIAREG, which requires further consideration if we want to compare these corpora. GTRP participants were visually presented with the base form of the Standard Dutch target words. They then translated these words into their local variant and their pronunciations were recorded and consequently transcribed for the dataset we use for this investigation. Similar to Chapter 4, we only analyze a subset of the transcriptions (selected by Wieling, 2007), which includes the base forms of nouns, adjectives, and verbs. This subset ignores most of the morphological variants of the target words, because our method of analysis (see Section 5.3.2) is not appropriate for dealing with morphological variation and would overestimate pronunciation change if morphological variants of the same target word are compared. This leaves 562 words per location, which remains a considerable amount of data.

The DIAREG project (Heeringa & Hinskens, 2014) offers a more recent phonetic dialect corpus collected between 2008 and 2011. Its primary aims were slightly different from the GTRP and the composition of the corresponding dataset therefore differs too. One of the aims of the project was to investigate whether there was aggregate-level evidence that dialects in the Netherlandic area were not only converging to Standard Dutch (Heeringa & Nerbonne, 2000; Heeringa & Hinskens, 2014), but also whether they formed regiolects. The DIAREG dataset contains more lexical variation than the GTRP, because the researchers were also interested in lexical change. This was reflected in the design of their experiment, as the researchers elicited running speech as opposed to word lists presented in Standard Dutch (which primes the Standard Dutch lexical variant). We ensure that these data can be used for our investigation by filtering out realizations that

are too distinct for direct comparison, such as phonetically reduced realizations (explained in further detail below).

Participants in the DIAREG study were presented with a silent movie in the form of stills and (written) narration. The presented story consisted of 23 relatively simple sentences with an average length of 7.6 words. The participants in the study worked in small groups of younger females or older males (usually in pairs, but up to four speakers in some cases), because Heeringa and Hinskens (2015) worked within the apparent-time paradigm. Each individual was first asked to write down their translated story version. Afterward, the participants compared their translations and made a new version everyone agreed on. This 'consensus' version was read aloud by the participants and transcribed. The more involved DIAREG approach avoids two potential pitfalls that typically occur in dialect research: the observer's paradox (i.e., altered behavior due to the presence of a researcher; Labov, 1972) and idiosyncratic language system differences between speakers obscuring the observed dialect variation (i.e., by using a 'consensus' version from multiple speakers). We obtained audio recordings of each group from the DIAREG authors (Heeringa and Hinskens). These recordings contained a subset of 13 sentences pronounced by all participants, yielding at most 125 target words per location (comprising 90 different word types).

Words pronounced in isolation, as was the case for the GTRP, are produced differently than words in running speech, as was the case for the DIAREG. Especially in running speech, the incentive to reduce the produced form was much higher. Clopper and Turnbull (2018) note that many potential factors can influence reduction, such as word frequency, semantic neighborhood density, and individual speaking style. The Levenshtein distance (which we use to quantify pronunciation change) analyzes phonetic transcriptions on a phonetic symbol-by-phonetic symbol basis, so phonetic reduction can greatly influence our results if left unaccounted for. For this reason, we manually annotate every transcription, indicating whether phonetic reduction occurred and leaving out transcriptions for which this is the case. Consequently, the overlap between the GTRP and DIAREG differs from location to location and the total number of comparable items decreases through this procedure.

5.2.2 Real-time comparison data — In line with our research questions (see Section 1.4), we are limiting the scope to the regional varieties spoken in the northern and eastern Netherlands in this chapter: the provinces of Fryslân, Groningen, Drenthe, Overijssel, and approximately the northern half of the province

of Gelderland (see the Low Saxon areas in Figure 5.2). We select recording locations and Standard Dutch target words that overlap between the GTRP and DIAREG, enabling a real-time comparison of pronunciation change. Because we analyze pronunciation change in real-time, we compare the GTRP speakers with the older men from the DIAREG.

There are 27 overlapping locations between the GTRP and the DIAREG (see Figure 5.3). We excluded two locations (Onstwedde and Nijverdal; marked as gray circles in Figure 5.3), because the older male DIAREG speakers of those locations were born earlier than the corresponding GTRP speakers, which means that the DIAREG speakers reflect an earlier form of the local dialect. Our dataset for analysis therefore comprises the recordings from 25 locations.

Instead of using the pre-existing transcriptions made by DIAREG and GTRP transcribers, a single transcriber (the second author of Buurke et al., 2022) made all relevant GTRP and DIAREG transcriptions anew. This ensures that there is no inter-transcriber variability, which is known to be a problem for the GTRP (Hinskens & Van Oostendorp, 2006). This also allows us to incorporate more Standard Dutch target words, as not all possible target words were transcribed by the GTRP contributors at the time. In total, 36 words overlapped between the two corpora (see Table 5.1).

Table 5.1: Overlapping target words ($N = 36$) between the GTRP and DIAREG corpora.

als	door	hem	houden	is	korte	loopt	naar	om	ruit	straat	vraagt
bij	glas	hij	huizen	juist	krijgen	lopen	niet	ook	steen	straten	wil
buiten	goed	hoek	in	klein	later	met	nu	op	stil	tijd	ziet

We have to account for lexical and morphological variation in the data, because the DIAREG corpus was also constructed to investigate change at these linguistic levels. If left unaccounted for, pronunciation change calculated based on these transcriptions would be inaccurate, because morphological or lexical change may drive the apparent pronunciation change instead. We therefore manually annotated the transcriptions of the comparisons. If the underlying cognates of a pair of GTRP-DIAREG transcriptions differed, these transcriptions were left out of the analyses. For example, Dutch *steen* 'stone' was on occasion translated by a DIAREG dialect speaker into a variant of *kei*, which has roughly the same meaning. Similarly, any variation due to diminutives, conjugation, or phonetically reduced forms was marked and removed as morphological mismatches. The number of target words per location that are used for the comparison after

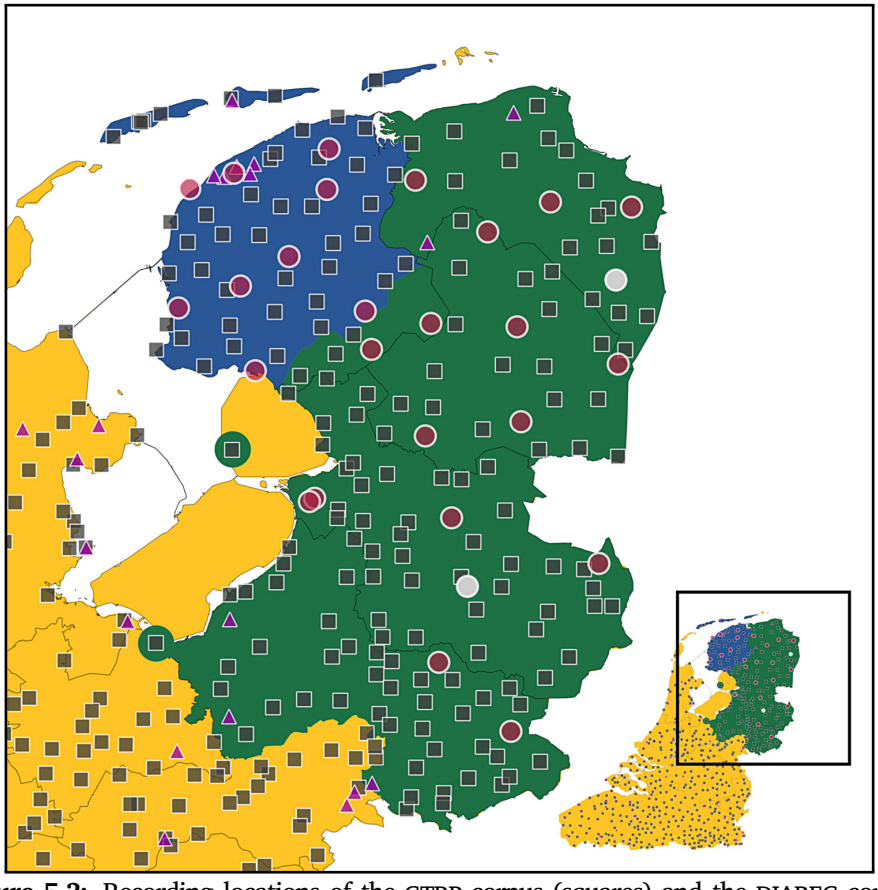


Figure 5.3: Recording locations of the GTRP corpus (squares) and the DIAREG corpus (triangles). The circles (27) indicate the overlapping locations in the north included in this study. Locations indicated by gray circles were left out of analyses, because the DIAREG speakers were older than the GTRP speakers. The Frisian, Low Saxon, and Low Franconian areas are marked in blue, green, and yellow, respectively.

filtering these variants is reported in Table 5.2.

There are no pairs of locations for which the full set of 36 possible overlapping Standard Dutch target words can be analyzed. This is mainly caused by the considerable phonetic reduction occurring for certain target words. For example, Dutch *hij* 'he' is pronounced as (a variant of the) phonetically reduced form [i] in 68% of the DIAREG cases. It is possible to replace the Standard Dutch target word with the reduced form, but this would still be problematic, because

Table 5.2: Number of available words per comparison after removal of lexical and morphological variation.

Location (number of words)			
Appelscha (22)	Grouw (24)	Noordwolde (27)	Tilligte (30)
Dokkum (21)	IJsselmuiden (32)	Ommen (25)	Veenwouden (31)
Eelde (28)	Jubbega (29)	Roswinkel (28)	Workum (29)
Finsterwolde (31)	Kampen (32)	Sexbierum (31)	Zwinderen (20)
Grijpskerk (26)	Koekange (33)	Sint Annaparochie (17)	
Groenlo (32)	Laren (23)	Slochteren (25)	
Grolloo (32)	Lemmer (30)	Sneek (16)	

the GTRP transcriptions then mismatch. The GTRP pronunciations were rarely reduced, because participants pronounced words in isolation. Consequently, we opted to exclude these data, because they cannot be compared adequately using our methodology.

In addition, some locations seem to have noticeably fewer words available for analysis. This applies in particular to Sint Annaparochie and Sneek. Upon closer inspection, it became clear that some GTRP transcriptions were unavailable for these locations. These Town Frisian speakers have a lexicon that is a unique mix of Dutch and Frisian, which may be the underlying cause of the lack of data (e.g., if only words that share the cognate with the Dutch target word are included in the GTRP). For example, *straat* 'street' is typically realized as [dik] (cf. Frisian *dyk*) in these dialects. We observe such Frisian lexemes in the DIAREG transcriptions only (again likely due to the task differences eliciting different kinds of speech). Comparisons between the GTRP and DIAREG are inappropriate in such cases, so these data are not used in the analyses.

The data for Appelscha were also problematic, because the GTRP speaker spoke Frisian, while the DIAREG speaker was a Low Saxon speaker. Appelscha is a Frisian-Low Saxon border town, and a mixture of the two speaker populations resides here. It turns out that the realizations represent the same cognate for 22 GTRP-DIAREG word pairs, so we could in principle include these data. However, we chose to err on the side of caution and leave out Appelscha from further analyses (yielding a total of 24 included locations). After these considerations, we are left with approximately 67% of the data (i.e., 652 word pairs) for our pronunciation change analysis.

Although all speakers in both corpora were NORMs, their exact ages differed,

so we need to take into account the distribution of age and recording year to establish the total time period across which we measure pronunciation change. These distributions are summarized with their mean and standard deviation for the two corpora (i.e., the subset of 24 locations) in Table 5.3. The metadata for the GTRP were incomplete, because speaker age data for Grolloo, Ommen, Sneek, and Workum were unavailable. The mean age of the DIAREG speakers is slightly higher than that of the GTRP speakers, but not problematically so, because a difference of six years is unlikely to yield substantially different patterns. Given that the ages for the comparison are roughly the same, we can assume that the time span of pronunciation change is approximately equal to the difference in recording years: on average approximately 20 years.

Table 5.3: Characteristics of speaker age and recording year per corpus (only including the 24 overlapping locations).

	GTRP (mean, standard dev.)	DIAREG (mean, standard dev.)
Speaker age	61 (8)	67 (6)
Recording year	1987 (2)	2009 (1)
Recording year diff.	22 (2)	

5.2.3 Representing Standard Dutch — We needed to select a representative reference speaker for Standard Dutch. We focused on national news presenters, because they are perceived as representative speakers of Standard Dutch (Smakman, 2006, p. 280). In 2021, a separate study was set up to determine the most 'standard' sounding news presenter out of 24 news presenters, including four speakers who were born in the northern Netherlands to introduce greater variation. These news presenters were between 27 and 54 at the time of their included recordings. Their recordings were selected from publicly available sources.

In an online questionnaire, respondents were presented with two audio samples of ten seconds each, read by a randomly selected news reader. We ensured that no locally identifiable terms occurred in the recordings, such as place names. Respondents were asked to rate which of the two news presenters sounded most 'standard' Dutch on a five-point scale ranging from the first speaker sounding 'much more standard' than the second speaker to the second speaker sounding 'much more standard' than the first speaker.

The questionnaire was distributed through social media and the website *Neerlandistiek*, which is an electronic journal for Dutch linguistics, literature, and

language proficiency in the Netherlands.[2] Respondents were only included if they indicated that they spoke Dutch as a first language (including bilinguals) and fully completed the questionnaire, reducing the initial 525 respondents to 271. Of these respondents, approximately 47% identified as a man and 53% as a woman. The respondents were between 16 and 85, and most respondents (71%) had a high educational attainment (i.e., higher professional or university education). Such a distribution of educational attainment is common for surveys (Christoffersen, 1987; Korkeila et al., 2001). Based on these 271 respondents, who rated 23 pairs of audio samples, Astrid Kersseboom was selected as the best representative of Standard Dutch speech (although there was little difference between the top-ranked speakers). After contacting her, she kindly agreed to record the words of our word list, and her pronunciations serve as the Standard Dutch reference points in this study.

We also included statements about the nature of Standard Dutch in the questionnaire, which was measured on a five-point Likert scale (ranging from 'strongly disagree' to 'strongly agree' with a neutral option). The results showed that most respondents thought Standard Dutch does change (91% agreed, 4% disagreed) and that most people understand Standard Dutch (83% agreed, 6% disagreed), but only a minority believed that everyone uses Standard Dutch (26% agreed, 61% disagreed). Most respondents agreed that Standard Dutch is the variant described in dictionaries (65% agreed, 17% disagreed), and many respondents considered Standard Dutch to be 'correct' Dutch (47% agreed, 23% disagreed). Many respondents also considered the comparison task difficult (48% agreed, 23% disagreed), likely because there was limited variation in the Standard Dutch pronunciations of the news presenters (except for the regional ones, which were rated noticeably lower). More details about the selection procedure, respondents, and results of this perceptual experiment are provided in the supplementary material of this chapter. The supplementary materials for this chapter can be found at https://osf.io/87zku/.

5.3 Methods

5.3.1 Levenshtein distance — We compute the degree of change using a variant of the Levenshtein distance (Levenshtein, 1966), which been adapted for comparing phonetic strings in dialectometry (Kessler, 1995; Nerbonne et al.,

[2] See https://neerlandistiek.nl.

1996) and historical linguistics (List et al., 2017). For a description of the basic steps involved in the algorithm, we refer to Chapter 4. Note that in the present chapter, we allow [ə] to be aligned with the sonorants [m, l, n, r, ŋ, j, w], because sonorant consonants have noticeably more acoustic energy than other consonants and are typically also voiced. The synchronic variation of Dutch *vier* 'four', [fiːr] vs. [fiːə] (see Heeringa, 2004, p. 125) also shows that sonorant consonants are likely to be structurally closer to vowels.[3] After determining the Levenshtein distance, we normalize it by dividing the distance by the length of the longest optimal alignment (in line with our approach in Chapter 4).

Similar to Chapter 4, we use linguistically sensitive costs for the operations, which range between 0 and 1 for each operation of the Levenshtein distance. For details about how these values can be derived, we refer to Section 4.2.2. For the analyses in this study, we used all available GTRP transcriptions to derive the PMI-based phonetic distances, because the total number of transcriptions in the GTRP-DIAREG dataset was too small to derive reliable phonetic distances.[4]

5.3.2 Measuring the direction of change — In addition to the rate of pronunciation change, we also analyze the direction of change by employing a three-dimensional (3D) version of the Levenshtein distance. The underlying principles of the two-dimensional Levenshtein distance (i.e., for comparing two sequences of phonetic symbols) can be extended to higher dimensions as well (Heeringa & Hinskens, 2015). Using a three-dimensional version of the algorithm, we can compute the difference between sequences of phonetic symbols (i.e., strings) while taking another string as a reference point, such as a standard language variant. An example (with binary costs for simplicity) is provided in Table 5.4.

Note that there are now seven possible operations instead of the usual three operations of the Levenshtein distance, and each operation constitutes a change in all strings. We can insert a phonetic symbol into one of the three strings (which implies deleting a phonetic symbol in the other two: three operations) and we can

[3] It is also possible to assume that the underlying structure is CVVC (with C: consonant, and V: vowel), instead of CVV/CVC, but the acoustic properties of sonorant consonants make the latter structure more likely.

[4] It is unclear from what sample size the PMI approach can reliably be applied, which is ideally explored in future work. However, the PMI distances based on the complete GTRP corpus can be used for the analysis in this chapter, because the symbols used in the (subset of transcriptions in the) DIAREG corpus overlap with those of the GTRP.

Table 5.4: 3D Levenshtein alignment example for dialectal variants of Dutch *straat* 'street'.

	1	2	3	4	5	6	7
Older variant	s	t	ʀ	o		d	ə
Newer variant	s	t	ʀ	ɔ	ə	t	
Standard variant	s	t	ʀ	a		t	
Older-standard operation	-	-	-	sub.	-	sub.	del.
Older-standard cost	0	0	0	1	0	1	1
Newer-standard operation	-	-	-	sub.	del.	-	-
Newer-standard cost	0	0	0	1	1	0	0
Direction of change	-	-	-	neutr.	div.	conv.	conv.

substitute phonetic symbols between each possible pair of strings (e.g., between strings 1 and 2, strings 1 and 3, and strings 2 and 3: three operations). We can also substitute a segment from each of the three strings at the same time (which is a single operation; cf. Heeringa & Hinskens, 2015).

From this single 3D alignment (for which we have used binary weights to facilitate interpretation in the examples), we can obtain the distances between the older variant and the standard, and between the newer variant and the standard.[5] For each segment of this alignment, we can then determine whether it reflects a pattern of convergence, divergence, neutral change, or stability. A particular segment is convergent (towards the standard) if the distance to the standard phonetic symbol is greater for the phonetic symbol in the *older* variant than for the phonetic symbol in the *newer* variant. If the distance to the standard is smaller for the phonetic symbol in the *older* variant instead, then the segment shows a divergent pattern. If these distances are equal, this segment concerns either neutral change or stability (in case the distance is zero).

Heeringa and Hinskens (2015) determined the type of change for each segment using binary weights, but we use gradual PMI weights. This makes it less likely to observe neutral change, because each distance between symbols has a gradual value between 0 and 1. Neutral change can only occur if two phonetic symbols occur with the same frequency in the transcriptions from which the PMI distances are derived. Given the example alignment in Table 5.4, this means that the direction of change for the fourth segment is either convergent (if the cost

[5] Note that the distance between the older and newer variant can also be obtained from this alignment, but for our analysis we are not interested in this particular distance.

of [o] ↔ [a] is greater than [ɔ] ↔ [a]) or divergent (if the cost of [o] ↔ [a] is less than [ɔ] ↔ [a]). We therefore use the following formula to determine the direction of change for each segment:

$$\text{directional\_change}_{(x,\,y,\,z)} = \text{distance}(y, z) - \text{distance}(x, z) \tag{5.1}$$

The phonetic transcriptions should always be compared in the same order when applying this formula, because a different order leads to a different interpretation. Following the above equation, we always make sure that x is the older variant (the GTRP transcription) and that y is the newer variant (the DIAREG transcription), while z is the standard variant. When this order is kept, a value greater than 0 indicates divergence for a segment and a value smaller than 0 indicates convergence.

We prefer the alignments generated by this 3D procedure over pairs of alignments generated by the typical two-dimensional (2D) procedure, because they better match our goal of investigating diachronic language change rather than synchronic variation. In principle, we can reach the 'same' goal by generating the typical 2D alignments between the GTRP–standard and between the DIAREG–standard transcriptions and then simply subtracting these two values. Indeed, the distances between the GTRP and DIAREG obtained by double-2D and single-3D alignment approaches are highly correlated ($r > 0.95, p < 0.01$). The 3D alignment, however, is conceptually better suited to investigating diachronic change, because a 3D alignment generates more possible segments of change. Given the example alignment in Table 5.4, we obtain a higher normalized distance with the double-2D Levenshtein than with the single-3D Levenshtein variant (because there is one more empty segment in the 3D alignment). We know that there was a segment after the final consonant ([d/t]) for the older variant, which was occupied by a [ə] at the time. We would ignore this historical property of the cognate if we did not use a 3D alignment, which is undesirable. The 3D version is therefore more suited than the 2D version to estimate diachronic differences, and we only carried out subsequent analyses with this algorithm.

We also hope to demonstrate in this chapter that using a multidimensional string comparison method can be a powerful and flexible tool for investigating pronunciation change. This is especially useful for diving deeper into the patterns of an aggregate-level analysis, because information can be attributed to and measured through individual segments. For example, one can investigate

the tendencies of directional changes in specific phonological classes using the 3D Levenshtein algorithm. Dialectologists often criticize the tendency of dialectometric studies to focus more on the description of patterns than on the drivers of the investigated patterns (Wieling & Nerbonne, 2015), but the methods used in this chapter can support the process toward more explanatory dialectometric studies.

5.3.3 Statistical analysis — In our statistical analysis, we model the distributions of language change for each dialect group separately. With this group-based analysis, we highlight differences in change between dialect groups which we expect to behave differently based on their linguistic history (e.g., we separate Town Frisian from the Frisian group, and we separate the Northern Low Saxon dialects from the Westphalian Low Saxon ones).

Similar to Chapter 4, we fit a generalized additive model that incorporates beta regression (suitable for modeling the bounded normalized pronunciation change variable). Note that the dependent variable is logit-scaled (i.e., the logarithm of the odds), because this is the link function used for beta regression. We account for the repeated nature of our data (i.e., multiple words pronounced by every individual) by including an appropriate random-effects structure. This enables us to obtain reliable results despite the different amounts of transcription data available per recording location.

To distinguish convergence and divergence in the data, we follow a similar procedure to Heeringa and Hinskens (2015), summing the costs of convergent and divergent segments *separately* for each alignment. For example, there are two convergent segments in Table 5.4, in addition to one divergent segment and one segment of neutral change. We normalize these values using the alignment length (to account for the different lengths of target words), so that we obtain a proportion of each type of change for each alignment, which serves as the dependent variable in our analyses (i.e., a value of 1 is the theoretical maximum value). Convergent and divergent change are analyzed in a single model, by including a binary predictor variable distinguishing between the two types of change for each data point (i.e., the GTRP–DIAREG–Standard Dutch triplet of transcriptions).

For the analysis, we use a categorical predictor distinguishing the four dialect groups (Frisian: FR, Town Frisian: TFR, Northern Low Saxon: NLS, and the Westphalian Low Saxon dialects: WLS). The locations associated with each dialect group are summarized in Table 5.5.

Table 5.5: Recording locations associated with each dialect group.

Dialect group (nr. of locations)	Locations
TFR (3)	Dokkum, Sint Annaparochie, Sneek
FR (7)	Grouw, Jubbega, Lemmer, Noordwolde, Sexbierum, Veenwouden, Workum
NLS (6)	Eelde, Finsterwolde, Grijpskerk, Grolloo, Roswinkel, Slochteren
WLS (8)	Groenlo, IJsselmuiden, Kampen, Koekange, Laren, Ommen, Tilligte, Zwinderen

5.4 Results

We created a generalized additive model that predicts change (separated into convergence and divergence using a binary factor 'Direction') based on the four previously defined linguistic groups (Frisian, Town Frisian, Northern Low Saxon, and Westphalian Low Saxon). A random intercept for each word and a random slope of direction per word (significantly improving the model fit, both p's < 0.01) were included in the model to account for by-word variation. We also tested whether including the age and birth year of the GTRP and DIAREG speakers improved the model, but this was not the case, so these variables were left out of the model.

When distinguishing the four different groups (FR, TFR, NLS, WLS), the model comparison reveals that adding this predictor did not offer an improvement over a (null) model without the group distinction. We discuss possible explanations for this finding in the discussion of this chapter.

Upon closer inspection of the data, a pattern of greater convergence (and less divergence) in the south compared to the north emerged. Given these pronunciation change differences between the north and the south in the area of interest, we therefore included a binary contrast between Westphalian Low Saxon (referred to as 'isWLS') and the other three language areas combined (i.e., FR, TFR, and NLS), which contrasts the most southern dialect group to the more northern ones. This resulted in a statistically improved model, and the results based on this optimal generalized additive model are shown in Tables 5.6. The interaction effect of the binary language group variable is shown in Figure 5.4. The

final model specification is as follows:

```
Normalized_change ~ isWLS * Direction
                    + s(Word, bs = 're')
                    + s(Direction, Word, bs = 're')
```

Table 5.6: Coefficients for a generalized additive model predicting change based on a binary distinction between Westphalian Low Saxon (WLS) and the other regional varieties. The direction of change is either convergence (conv.) or divergence (div.).

	Estimate	SE	*z*-value	*p*-value	
Intercept (non-WLS: conv.)	-3.76	0.05	-71.77	<0.001	***
non-WLS: div. vs. conv.	0.16	0.07	2.57	0.010	*
WLS vs. non-WLS: conv.	0.15	0.06	2.35	0.019	*
WLS: div. vs WLS: conv.	-0.21	0.09	-2.43	0.015	*

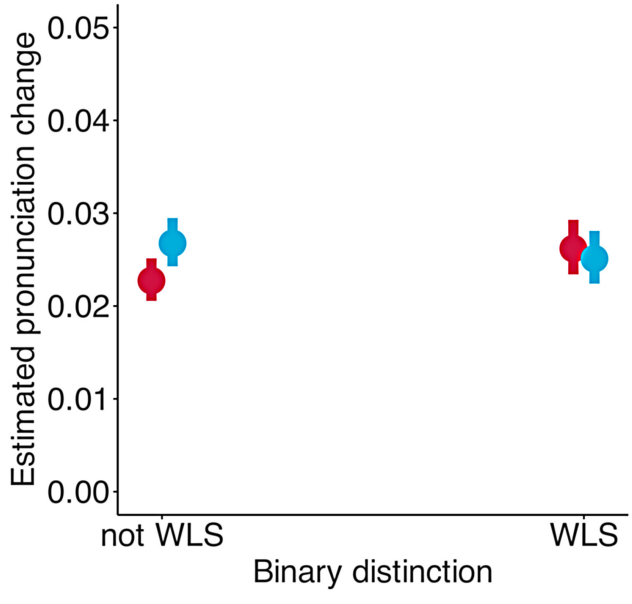


Figure 5.4: Estimated marginal interaction effect between change direction and a binary distinction between Westphalian Low Saxon (WLS) and the other regional varieties. Convergence is shown in red and divergence in blue. The *y*-axis has been transformed from logits into odds.

We can conclude from the model and Figure 5.4 that the rate of convergence toward Standard Dutch in the Westphalian Low Saxon group is significantly higher than in the other varieties (including Frisian and Frisian-Hollandic contact varieties). The divergence from Standard Dutch in the Westphalian Low Saxon group is lower than in the other language varieties. Divergence and convergence rates are about equal for the Westphalian Low Saxon group, whereas divergence rates are significantly higher than those of convergence for the other varieties. We turn to this in the discussion.

5.5 Discussion

We aimed to investigate the rate of pronunciation change in the northern and eastern Netherlands using a real-time paradigm, while accounting for the direction of these changes (addressing research questions 4a through 4c; see Section 1.4). We hypothesized that Frisian and Town Frisian dialects would be the most resistant to vertical convergence based on what is known from the literature about these dialect groups (e.g., a protective speaker population and greater political protection, in the case of Frisian). For Northern Low Saxon and Westphalian Low Saxon varieties, we expected to find more vertical convergence, because there is less intergenerational transmission within the population, the speaker population is less protective of their language variety, and the existing language policies protect Low Saxon to a lesser degree than the Frisian varieties. Overall, we expected to find more convergence than divergence, as previous studies found this as well (Heeringa & Nerbonne, 2000; Heeringa & Hinskens, 2015).

The results show that the overall rate of pronunciation change is low across a time span of approximately 20 years, at least for this geographical region and this particular time period. The percentages of change we observed differ from Heeringa and Hinskens (2015), who found an average percentage of pronunciation change of 13.3% across the whole Netherlandic area in an apparent-time analysis (6.8% convergence, 0.6% neutral change, 5.9% divergence). We observed only 3.4% of pronunciation change by summing the average convergence and divergence in our data (1.4% and 2.0%, respectively). This difference likely stems from the fact that we used PMI weights (which tend to be much smaller than 1, even after normalization) instead of binary weights. We obtain a total of 13.4% (5.9% convergence and 7.5% divergence) when we repeat the analysis binary

weights, which is similar to Heeringa and Hinskens (2015).[6] This indicates that the used weights influence the absolute levels of observed pronunciation change, so future studies may explore which weights are preferable. However, the relative pronunciation change between language varieties can be investigated using either type of weighting and these absolute numbers are difficult to interpret directly, and the choice in weighting is therefore not always relevant.

Within the relatively small range of pronunciation changes we observed, we find significantly more convergence for the Westphalian Low Saxon area than for the other dialect groups. These results are consistent with synchronic results obtained by Wieling et al. (2011), who related their results to the 'Hollandic expansion' (i.e., Hollandic speech norms expanding to peripheral provinces due to North and South Holland being the center of economic and political power; see Kloeke, 1927). At the same time, the geographical propagation of language change at an aggregate level is a contentious topic (see e.g., Nerbonne, 2010, for a discussion). Exploring the exact mechanism of pronunciation change in this particular geographical area, especially concerning Standard Dutch convergence, is a worthwhile topic for future studies.

The convergence toward Standard Dutch makes the Westphalian Low Saxon dialect group a candidate for regiolectization, if the strong convergence coincides with significant convergence between neighboring dialects. To better determine this, it would be necessary to analyze patterns of horizontal change together with the already investigated vertical change. In the future, it may therefore be interesting to explore methods that can, in addition to vertical change, estimate horizontal change, for instance by extending the Levenshtein distance to even more dimensions (see Heeringa & Hinskens, 2015). In Chapter 6, we explore other methods that aid in assessing horizontal convergence between the dialects spoken in the northern and eastern Netherlands and provide more comprehensive evidence of possible regiolect formation.

We did not expect a higher prevalence of divergent patterns compared to convergent ones in the Frisian, Town Frisian, and Low Saxon regions outside of the Westphalian Low Saxon area. However, due to the introduction of the binary distinction between the Westphalian Low Saxon group and the rest of the linguistic groups, it is difficult to interpret the patterns of the combined (Frisian, Town Frisian, and Northern Low Saxon) group. Future work may explore whether the

[6] We repeated our analyses using the younger female data from the DIAREG corpus to validate our effects. In this combined real-time/apparent-time analysis (reported in Appendix C.1), we find the same patterns of convergence within the Westphalian Low Saxon area.

divergent patterns in this time period are meaningful for the dialect groups in the long term, and what has driven these patterns. It would be interesting to explore whether this specific time period coincides with a (temporary) linguistic distancing from Standard Dutch in these areas, while these dialects may still converge toward Standard Dutch over a longer period.

One might wonder whether the convergent patterns occur in addition to the divergent ones, or whether they compensate for them. For example, convergence may theoretically occur specifically for vowels and divergence for consonants, leaving the dialects approximately equidistant from Standard Dutch. Our analysis allows quantifying both patterns, but these patterns are not inherently dependent on each other. Consequently, we checked whether there are classes of segments that are particularly prone to convergence or divergence. We therefore repeated the analyses, but with two new sets of transcriptions. All vowels were replaced by [ə] in one transcription set, which allowed us to investigate the pronunciation changes specific to consonants. In another set, all consonants were replaced by [x] to investigate the pronunciation changes specific to vowels. When focusing on change exclusive to the vowels, we observed that the Westphalian Low Saxon dialects converged significantly more and diverged significantly less than the other groups. This pattern was not observed for the consonants. In addition, for the consonants (and not the vowels) the Frisian and Northern Low Saxon dialects showed more divergence than convergence. These results suggest that the convergent and divergent patterns do not negate one another. More sophisticated tests of the interdependence of such effects are left to future studies, but the methods proposed in this chapter can also be used for such investigations.

5.5.1 Limitations — We now consider some caveats of our data and analyses. The most pressing of these is perhaps the fact that in some locations relatively few words could be used for analyses after our somewhat stringent inclusion criteria for the phonetic transcriptions. Our dataset comprised a maximum of 36 words across 24 locations, while Heeringa and Hinskens (2015) used (a maximum of) 125 words for each location. However, preliminary simulations in our lab have shown that when using only 25 randomly selected words out of a larger set of hundreds of words, the obtained aggregate dialect distances correlated at a level of $r = 0.90$ with the aggregate dialect distances based on the larger dataset (reported in Appendix C.2). Consequently, while our dataset is small, it does not

appear too small.[7]

Note that we have only used (a limited amount of) linguistic and spatial information in our analyses. Extralinguistic information within the same area and time period of interest could also be considered in future work. For example, we restricted our analysis to key characteristics that influence dialect decline at an aggregate level. We should, however, also operationalize factors that are thought to underlie dialect decline at lower levels (e.g., individual speaker attitudes and attitudes of the speaker population at large as determined through questionnaires). In Chapter 6, we therefore re-assess pronunciation change in the northern and eastern Netherlands and account for speaker characteristics (e.g., language attitudes and educational attainment) and linguistic information (e.g., word class and word frequency).

Upon closer inspection of the data from noticeably divergent locations, the divergence appears to stem from inserting or deleting a single phonetic symbol in various cases. Examples (from Sneek and Grijpskerk) of such minimally different pairs include [krɛiə] → [krɛiən] (Standard Dutch [krɛiɣə(n)] 'to get'), [la:tər] → [la:tə] (Standard Dutch [la:tər] 'later'), and [lo:pt] → [lo:p] (Standard Dutch [lo:pt] 'walks'), but also [dœ:r] → [dœ:] (Standard Dutch [do:r] 'through') and [lɔ:tər] → [lɔ:tr] (Standard Dutch [la:tər] 'later'). These minor differences may be caused by differences in task requirements between the GTRP and DIAREG, but in a more subtle manner than we accounted for. Recall that the DIAREG participants were asked to recount a story, whereas the GTRP participants translated isolated words and phrases. In running speech, the pressure to communicate efficiently causes speakers to 'simplify' their speech gestures (see Clopper & Turnbull, 2018, for an overview), and we may be observing lenition in our data as well (e.g., not pronouncing the word-final [n, t, r]). Transcribers may also fail to perceive [ə] if it is realized minimally, which is likely to occur in running speech. Finally, it is also possible that data derived from a single-word elicitation approach are relatively conservative and lag behind ongoing language changes (see Streckenbach, 2020), while data elicited in running speech may not be affected by this. Ideally, the phonetic data should be more comparable to avoid these uncertainties. We therefore constructed a new phonetic corpus based on new recordings in the 2020s, which is better comparable to the GTRP. This data collection and the accompanying analyses are described in Chapter 6.

One more general shortcoming concerns our artificial partitioning of lan-

[7] Additionally, the Cronbach's alpha of the transcription dataset was 0.85. A value of at least 0.80 is considered sufficiently reliable for dialectometric studies (see Heeringa, 2004, p. 170).

guage change into aggregated pronunciation change on the one hand, and change on different linguistic levels on the other hand, including simultaneous change at the lexical and phonetic or phonological levels due to lexical borrowing. This process is cross-linguistically widespread in contact situations (Durkin, 2015), and the potential interplay between change at these linguistic levels is perhaps more important than we give it credit for in this study. The longstanding language contact between, for example, the Low Saxon and Low Franconian areas has led to considerable lexical borrowing between the corresponding languages (though mostly from Standard Dutch into Low Saxon). However, a large amount of lexical borrowing does not entail an equal amount of pronunciation change. For example, the traditional Groningen dialect (part of the Northern Low Saxon area) word for ice-skating is *scheuvelen* [sxøvəɬn̩], but *schoatsen* [sxoːtsn̩] is commonly heard instead nowadays. This newer form is borrowed from Dutch *schaatsen* [sxaːtsə(n)], although it stays true to the regular correspondences between Standard Dutch and the Groningen dialect (i.e., [aː] in Standard Dutch cognates being pronounced as [oː] or [ɔː]). This illustrates how the linguistic levels concerning pronunciation may be more resistant to change than the lexical level, which Heeringa and Hinskens (2015) also found.

Finally, a point should be made about finding a suitable representative form of the standard variety. We mentioned that news presenters are typically considered representative speakers of Standard Dutch (Smakman, 2006, 2012), but we have only used pronunciation data from a single news presenter to represent Standard Dutch. The obvious shortcoming is that even highly experienced national news presenters with an intended wide coverage are unlikely to be entirely accentless. This is only natural, but since we only use a single speaker as a reference point, this makes it possible that our results are skewed toward this idiomatic version of the standard variety. At the same time, the average ratings given by respondents to our Standard Dutch questionnaire differed little between the news presenters, so the effect of a different news presenter on our analyses would likely have been limited. Using a single reference point is furthermore problematic when standard varieties also change. The Standard Dutch that is spoken nowadays is noticeably different from Standard Dutch spoken 50 years ago (see Smakman, 2006). Ideally, reference speakers should be chosen that represent the standard from around the average recording year for each dataset, but the data of these reference speakers should then have been collected around the same time as the other data for each dataset. Otherwise, these individuals may have changed their pronunciation (under the influence of the changing standard

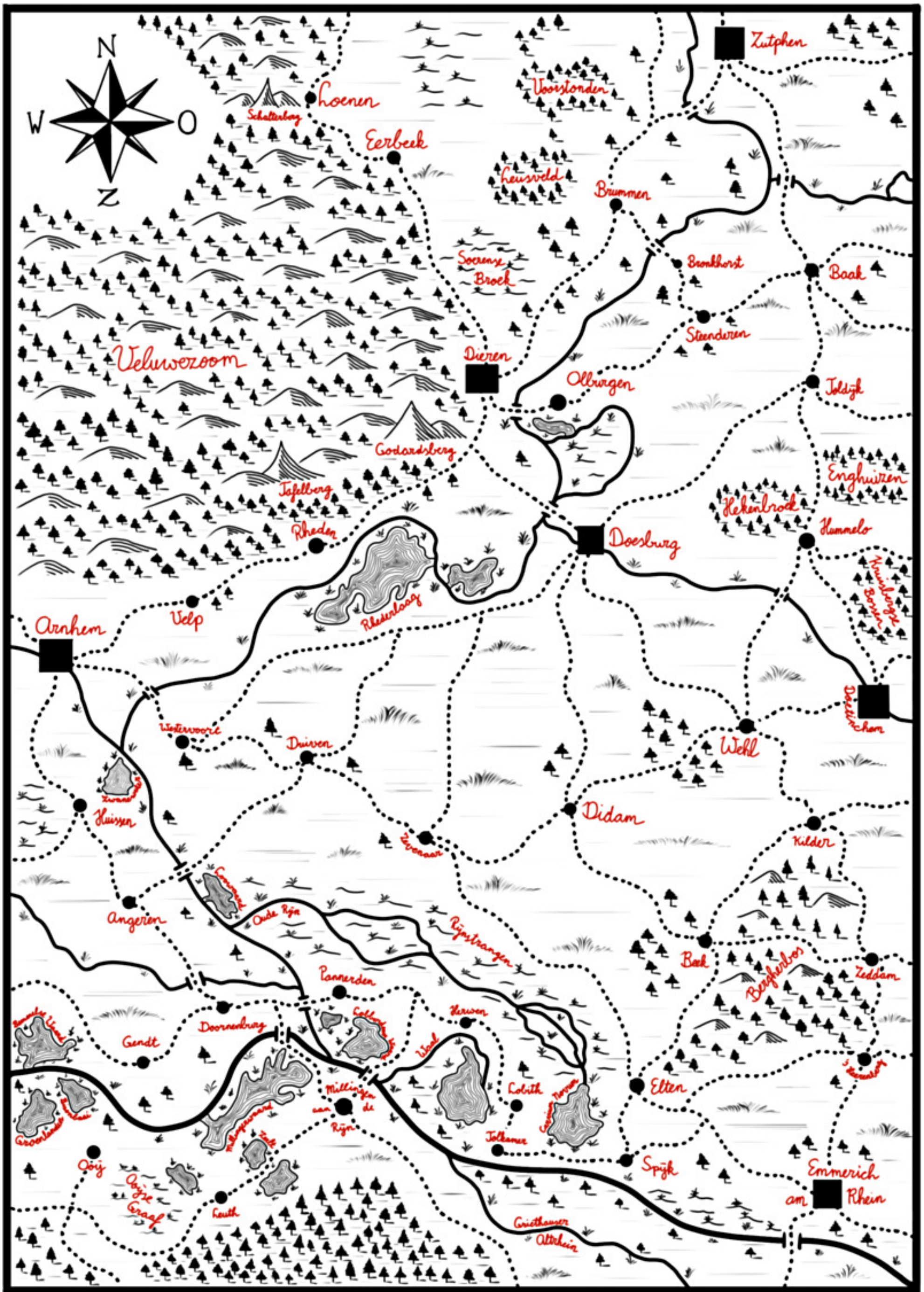
language) compared to when the dialect dataset was compiled (Blondeau, 2001; Ashby, 2001; Sankoff & Blondeau, 2007; Sankoff, 2019).

5.6 Conclusion

Using dialectometric analyses, we found that pronunciation change progresses slowly in the dialects of the regional languages spoken in the North of the Netherlands. There was more stability in our data than there was change, regardless of convergence to or divergence from Standard Dutch, but we observed different patterns between dialect areas within the ongoing changes. Regional varieties in areas further away from the economic center of the Netherlands converged to Standard Dutch to a lesser degree. Unfortunately, various limitations are associated with comparing the two dialect datasets we used, so our conclusion is only tentative. Consequently, in the following chapter, we aim to alleviate this issue by collecting new data to allow for a re-evaluation of pronunciation change in the Frisian, Town Frisian, and Low Saxon varieties.

Didam, Gelderland

CHAPTER 6


N
W
O
Z
Loenen
Eerbeek
Voorstonden
Leusveld
Zutphen
Brummen
Soerense Broek
Bronkhorst
Baak
Steenderen
Veluwezoom
Dieren
Olburgen
Toldijk
Godardsberg
Tafelberg
Enghuizen
Hekenbroek
Hummelo
Rheden
Doesburg
Rhederlaag
Velp
Arnhem
Westervoort
Duiven
Wehl
Doetinchem
Huissen
Didam
Zevenaar
Kilder
Angeren
Oude Rijn
Rijnstrangen
Pannerden
Beek
Bergherbos
Zeddam
Doornenburg
Gendt
Herwen
Waal
Elten
Millingen aan de Rijn
Tolkamer
Spijk
Emmerich am Rhein
Ooij
Ooijse Graaf
Leuth
Griethauser Altrhein

Voor Daan

RE-ESTIMATING PRONUNCIATION CHANGE

Abstract

To assess the possibility of regiolect formation, we estimated the pronunciation change of regional languages in the northern and eastern Netherlands in a period of about 35 years. We compared the pronunciation similarity between Frisian, Town Frisian, and Low Saxon varieties, and between these varieties and Standard Dutch. Specifically, we compared dialect recordings collected in 32 locations during the late 1980s to those collected during the 2020s. The recordings of dialect translations of a Standard Dutch word list were transcribed and consequently analyzed using the Levenshtein distance.

Our results showed evidence of regiolect formation in the Frisian and northern Low Saxon areas. Strong convergence to Standard Dutch was found in the areas bordering the Hollandic dialect area and in one eastern area in the province of Overijssel. An in-depth analysis showed that relatively low- and high-frequency words were most likely to change and that speaker characteristics interact with convergence and divergence patterns. Our findings confirm the importance of including multiple speakers per dialect location and accounting for speaker characteristics in dialectological research.[1]

6.1 Introduction

FRISIAN, TOWN FRISIAN, AND LOW SAXON are closely related West Germanic languages spoken in the northern and eastern Netherlands. Speakers of these regional varieties are (usually) also Standard Dutch speakers, which is the official language in the Netherlands. These regional languages and dialects have been converging to Standard Dutch on different linguistic levels since the middle of the 20th century (Heeringa & Hinskens, 2015), but there is also some tentative evidence of divergence from the standard language regarding the pronunciation level since the second half of the 1980s (see Chapter 5).

It is common for European regional languages to become more similar to their respective standard languages, usually by borrowing lexical and phonological features (Auer et al., 2005; Grant, 2015). This type of convergence mostly results from factors that stem from the relative power of standard languages,

[1] This chapter is adapted from: Buurke, R., Heeringa, W., Knooihuizen, R., & Wieling, M. (2024d). Recent changes in pronunciation patterns within and between Frisian, Town Frisian, and Low Saxon varieties [Submitted.].

such as the standard language's prestige, negative prejudices about regional languages and their speakers (possibly due to standard language ideology), and much stronger institutional support for the national standard language. Within such contexts, people may feel pressured to limit their regional language use to close interpersonal settings only, and they may be discouraged from transmitting this language to their children. At the same time, differences between traditional dialects are also decreasing in many areas. Such directional patterns of language change can indicate the formation of so-called 'regiolects' (Hoppenbrouwers, 1990), which are language varieties positioned between the traditional highly localized dialects and the standard language (see Section 1.3.2).

It is worth noting that there are different potential developmental paths for regiolects. One possible scenario is that (neighboring) dialects become more similar without any outside pressure of a standard language, but this is rare in Europe (Auer, 2018, p. 163). More often, dialects simultaneously converge toward the standard language and to neighboring dialects (e.g., Heeringa & Hinskens, 2015) due to the concurrent replacement of dialect features with standard language features. This widespread convergence often coincides with local divergence patterns for some varieties, especially around national borders (Smits, 2011). We assess evidence of regiolect formation in the Low Saxon and Frisian language areas, but we cannot make claims about a precise developmental path of potential regiolects, because our methods cannot ascertain the origin of the observed pronunciation changes.

The individual characteristics of speakers who contribute their recordings to dialect studies should be considered when investigating pronunciation change. Earlier dialect corpora covering Frisian, Town Frisian, and Low Saxon (such as the RND, GTRP, and DIAREG corpora; see Chapters 4 and 5) accounted for a speaker's age, gender, and location growing up, because these factors are usually associated with differing degrees of linguistic progressiveness (historically mainly leading to the inclusion of conservative non-mobile older rural male speakers; see Chambers & Trudgill, 1998). A speaker's educational attainment is also relevant (and therefore usually recorded, e.g., for the GTRP), as it is well known that educational background (or social class) interacts with the use of dialect features (i.e., dialect use is more prevalent among people with lower educational attainment or social class; Chambers & Trudgill, 1998; Driessen, 2005; Schmeets & Cornips, 2022, p. 58).

A speaker's resistance to linguistic change may also be influenced by their language attitudes toward their regional language. Strongly negative views of

someone's regional language (or strong standard language ideology) may lead to the disuse of the regional language and consequently negatively impact the reinforcement of nonstandard forms. In this case, a speaker is more likely to use the standard language or to be influenced by it. Conversely, strongly positive language attitudes may limit the impact of the influence of the standard language, because such speakers are more likely to value and adhere to conservative language use (especially if this is seen as a key part of preservation; Tulloch, 2006).

Pronunciation changes often do not affect the entire lexicon simultaneously or regularly (Phillips, 1984). Changes usually gradually spread out, not only across different words, but also across different geographical areas (Nerbonne, 2010). This lexical diffusion is partially driven by how often words are used (Bybee, 2002), and such word frequency effects are pervasive in the evolution of the world's languages (Pagel et al., 2007; Calude & Pagel, 2011). Phillips (1984) posits that frequency effects themselves also constitute a spectrum of processes, because it appears that low-frequency words may be more prone to the consequences of being 'forgotten' (e.g., being replaced more easily by higher-frequency words), while high-frequency words are usually consistently produced more economically (e.g., resulting in reduction, deletion or assimilation processes) and are usually more resistant to being borrowed or replaced (Monaghan & Roberts, 2019). It is therefore beneficial for language variation analyses to incorporate reliable word frequency measures.

Further evidence of lexical diffusion is reflected by the fact that some word categories are more resistant to change (e.g., nouns more than adjectives and verbs; Pagel et al., 2007; Wieling et al., 2011), although the literature on this pattern appears more scarce and the findings are less consistent. For example, Monaghan and Roberts (2019) provide evidence that open class word categories (i.e., categories that are open to new items, such as nouns, adjectives, and verbs) may be more easily borrowed than closed class word categories (e.g., numerals or determiners). While the effect of word frequency can more easily be linked with cognitive processes or processes of speech economy, there is no obvious reason why certain syntactically conditioned items should be more prone to change than others. Nonetheless, the fact that these effects are found in the literature suggests that word categories should also be accounted for when investigating language change.

We address research questions 4a through 4c again in this chapter (see 1.4), as we did in Chapter 5. We provide an aggregate and more in-depth view regard-

ing regiolect formation and pronunciation change in recent decades for Frisian, Town Frisian, and Low Saxon. The aggregate analysis directly addresses the research questions, as we assess whether the regional varieties have become more similar at an aggregate level and whether they have become more similar to Standard Dutch at an aggregate level. The in-depth analysis may provide insight into whether the newly recorded speakers or the chosen target words significantly influence our general findings. For this analysis, we directly incorporate the role of speaker characteristics and lexical covariates, because these characteristics may affect the convergence to and divergence from Standard Dutch.

6.2 Data

In this chapter, we attempt to improve upon earlier real-time studies of language change in Frisian, Town Frisian, and Low Saxon, including the one presented in Chapter 5. Chapters 4 and 5 demonstrated that existing dialect corpora are difficult to compare directly in a real-time approach. Similar to Chapters 4 and 5, we rely on the GTRP corpus (see Section 1.3.3 for details) and compare it with a set of newly collected dialect data (see Section 6.2.1 for details). Collecting new data allows us to control more strictly for speaker characteristics, such as a speaker's age, gender, and location of growing up. We specifically ensured that the speakers newly recorded for this project had the same gender as their GTRP reference speakers, were within three years of the sampling age, and grew up within ten kilometers of the location where the GTRP reference speakers grew up.[2]

We also focused on including more reference speakers for regional varieties for the newly collected data. For most GTRP locations, only a single speaker was recruited per location. The Standard Dutch elicitation list that GTRP speakers were asked to translate into their local dialect consisted of over 1800 items, so their language system is likely appropriately sampled. At the same time, it is unlikely that the single informant in each GTRP location was always the most representative speaker of the local dialect. Consequently, we tried to find at least two speakers per location for the newly collected corpus.

[2]Ideally the speakers would have grown up in the same location, but we could not always find participants from these locations. Limiting the range to ten kilometers ensured the dialects were still comparable. The metadata for each GTRP location is reported here: https://projecten.meertens.knaw.nl/mand/GTRPsprekersinfo.html.

We closely matched the data collection method of the GTRP, because mismatches between elicitation methods have been problematic before (see Chapter 5). Specifically, we ensured that the task underlying our new dialect corpus was comparable to that of the GTRP by presenting overlapping Standard Dutch target words in isolation, which we asked speakers to translate into their local dialect. This ensured a greater probability of detecting changes across this longer period than if (e.g.,) the DIAREG corpus had been the reference corpus.

6.2.1 The SPRAAKLAB corpus — We collected a new corpus with the mobile laboratory of the University of Groningen (Wieling et al., 2023), so we dub it the SPRAAKLAB corpus after the mobile laboratory. The mobile laboratory is equipped with professional directional microphones and a sound-dampened room. This ensured a high and consistent recording quality in a silent environment, a consistent experimental setup, and made it easier for dialect speakers to participate (Rebernik et al., 2024). See Figure 6.1 for an impression of the mobile laboratory.

The 32 GTRP reference locations for which new data were collected are shown in Figure 6.2, with relevant information per location in Table 6.1. We rendered Appelscha in gray in Figure 6.2, because the GTRP speaker spoke Frisian and the SPRAAKLAB speakers spoke Low Saxon, so the recordings from this location could not be used for estimating pronunciation change. Many Frisian speakers have migrated to this border area between Frisian and Low Saxon in the past century (Heeringa, 2005), so this might be why a Frisian speaker was selected for the GTRP in this traditionally Low Saxon area. Furthermore, the recordings in Workum were made with a laptop and high-quality head-mounted microphone due to the temporary unavailability of the mobile laboratory, but they were of high enough quality to be reliably transcribed and used for analysis.

The total number of SPRAAKLAB speakers whose data were included was 74. One GTRP speaker per location was included in this study (i.e., 31 in total), and a single Standard Dutch reference speaker, so there were 106 speakers in total. The data collection lasted from the autumn of 2022 to the autumn of 2023, which is a relatively short time period compared to the RND and GTRP corpora (see Chapter 4). This is beneficial, because it is challenging to disentangle pronunciation change from the difference in recording years between corpora (during which the varieties also change themselves) if the data collection process takes a long time to complete.

A variety of methods were used to find speakers, including posting letters to

Figure 6.1: Drone photo of the exterior of the mobile laboratory during recordings for the reference location Sexbierum, Fryslân.

people who participated in the study of Heeringa and Hinskens (2015), phoning companies in the reference locations, and referral sampling (i.e., a speaker asking other local dialect speakers they are familiar with). Post-initial communication primarily happened via e-mail, so speakers were familiar with general instructions before a recording session. We attempted to recruit at least two speakers in each reference location, which was successful for all locations except one (see Table 6.1). We found at least three speakers for over a third of the reference locations.[3]

A recording session generally consisted of a short introduction, during which participants gave written informed consent. This was followed by the instruction to translate the presented Standard Dutch target words into the speaker's dialect.

[3] Some speakers were not included in the real-time analysis, because they did not meet the inclusion criteria (i.e., matching the GTRP reference speaker), but their data are available for future research.

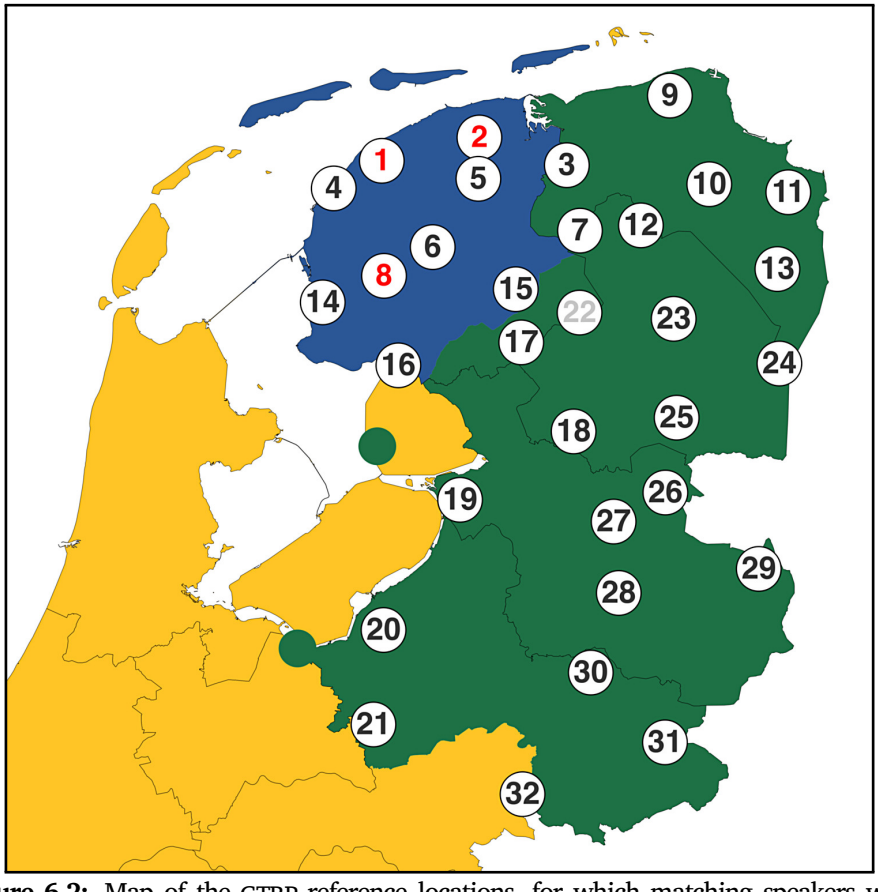


Figure 6.2: Map of the GTRP reference locations, for which matching speakers were found for the SPRAAKLAB corpus. The Frisian, Low Saxon, and Low Franconian areas are marked in blue, green, and yellow, respectively. Red numbers are used to indicate the Town Frisian varieties. The greyed-out number concerns Appelscha, of which the collected data could not be included in the analysis. The corresponding place names and metadata are provided in Table 6.1.

Participants were instructed to translate the words as they would pronounce them themselves (i.e., not another family member, who a participant may perceive as a better dialect speaker). They were also instructed to pronounce the translation only and not the target word.

The target words were individually presented in written form in Dutch in the center of a large screen for 2.5 seconds, after which a black screen was shown for

Table 6.1: Metadata per numbered reference location (see Figure 6.2): the language (Frisian: FR, Town Frisian: TFR, or Low Saxon: LS), province (Fryslân: FR, Groningen: GN, Drenthe: DR, Overijssel: OV, or Gelderland: GL), number of newly recorded speakers, gender (male: M, or female: F), and reference age.

	Name	Province	Language	Speakers	Gender	Reference age
1	Sint-Annaparochie	FR	TFR	3	M	55
2	Dokkum	FR	TFR	3	M	54
3	Grijpskerk	GN	LS	2	M	44
4	Sexbierum	FR	FR	2	M	64
5	Veenwouden	FR	FR	4	M	70
6	Grou	FR	FR	4	M	57
7	Zevenhuizen	GN	LS	2	F	68
8	Sneek	FR	TFR	2	M	60
9	Uithuizen	GN	LS	2	M	49
10	Slochteren	GN	LS	2	M	50
11	Finsterwolde	GN	LS	3	M	54
12	Eelde	DR	LS	2	M	69
13	Onstwedde	GN	LS	3	M	47
14	Workum	FR	FR	2	M	60
15	Jubbega	FR	FR	2	M	59
16	Lemmer	FR	FR	2	M	69
17	Noordwolde	FR	LS	2	F	72
18	Koekange	DR	LS	2	F	71
19	IJsselmuiden	OV	LS	3	M	68
20	Ermelo	GL	LS	3	F	64
21	Lunteren	GL	LS	2	F	55
22	Appelscha	FR	LS	3	M	64
23	Grolloo	DR	LS	2	M	49
24	Roswinkel	DR	LS	2	M	66
25	Zwinderen	DR	LS	2	M	70
26	Hardenberg	OV	LS	3	M	64
27	Ommen	OV	LS	1	F	56
28	Nijverdal	OV	LS	2	M	46
29	Tilligte	OV	LS	2	M	60
30	Laren	GL	LS	2	F	53
31	Groenlo	GL	LS	3	M	58
32	Didam	GL	LS	3	M	63

0.5 seconds. For example, if the Standard Dutch target word *huis* 'house' appeared on the screen, the speaker would translate and pronounce this word (typically into [hus] or [hys]) and then automatically move on to the next target word. Verbs

were presented with an underline to aid the speakers in differentiating between nouns and verbs, because many verbs and plural nouns are ambiguous in Dutch (e.g., *vissen* can both mean the plural noun 'fish' or the verb 'to fish').

The target word presentation cycle was repeated for all 150 target words of the task, of which 133 overlapped with the GTRP and were included in the analyses (see Table 6.2).[4] If a speaker failed to provide a pronunciation for a target word, the target was repeated at the end of the list. When a speaker indicated that no fitting translation for a particular target word existed in their dialect, they were instructed to remain silent, and the word was not used for the analysis.

Table 6.2: Standard Dutch target words ($N = 133$) for the word list translation task.

aarde	dreigen	huizen	maten	ribben	stil	wij hebben
achter	drie	jong	meid	rood	stof	wij kloppen
beginnen	dun	juist	moeten	rook	straat	wij krijgen
bij	durven	kamers	morgen	ruit	straten	wil
binden	eigen	kasten	nat	schade	suiker	zand
blauw	gewoon	katten	nieuw	schapen	tijd	zeep
blazen	glas	kinderen	noemen	scheel	trouwen	zeggen
bloeden	gras	klein	nu	scheppen	twee	zeilen
bouwen	grijs	kloppen	om	scherp	tweede	zelf
branden	groen	kneden	ons	schoenen	uilen	zes
breed	halen	koeien	ook	schrijven	vader	zetten
breken	hard	koken	op	schuiven	vals	zeventig
buiten	hebben	krijgen	paarden	slapen	vier	ziek
darmen	heet	krom	Pasen	smal	vraagt	ziet
deuren	hemel	kwaad	piepen	sparen	vrij	zilveren
dienen	hoek	later	planken	staan	vrijdag	zonder
donker	hol	lijken	ploegen	steen	vuur	zuur
door	hooi	loopt	raar	stenen	weten	zwaar
draaien	houden	lopen	redden	stijf	wij breken	zwemmen

Speakers filled in a short background questionnaire after the word list translation task, which included questions about the respondent's age, gender, and educational background. With nine possible values, the educational attainment

[4] The 17 extra target words overlapped with words in the target sentences from Heeringa and Hinskens (2015), which were collected for other potential analyses.

scale ranged from 'no education completed' to 'university education'. The complete background questionnaire is included in the supplementary materials. The supplementary materials for this chapter can be found at https://osf.io/hcavn/.

Additionally, 40 statements about regional identity adapted from the Swabian Orientation Index (Beaman, 2021, p. 107) were included in the questionnaire. These statements index how strongly someone identifies with their region in the linguistic and cultural sense, including questions about whether someone consumes media in their dialect, knows local folklore, and is proud of their dialect. The statements were adjusted for each dialect group to reflect terminology that is more widely in use by non-linguists (e.g., *Gronings* 'Groningen dialect' instead of *Nedersaksisch* 'Low Saxon'). The statements were presented on a five-point scale, ranging from 'strongly disagree' to 'strongly agree'. The average value was computed across all 40 statements after ensuring the scores for negatively framed statements were inverted, and was used to index the strength of someone's regional identity.

6.2.2 Including word frequency and word category metadata — There may be word frequency effects (Phillips, 1984; Bybee, 2002; Pagel et al., 2007) and word category effects (Wieling et al., 2011) when investigating language change, so we included this information for each Standard Dutch target word. There are several methods for estimating the frequency of specific words, but finding a suitable distribution that accounts for variation between individual language systems is challenging (Brysbaert et al., 2018). SUBTLEX-NL is a psycholinguistically motivated database for Standard Dutch based on film and television subtitles, which reflects how easily people recognize words (Keuleers et al., 2010). We therefore used the log-transformed frequencies of the target words in the SUBTLEX-NL database in the analysis.

The word category metadata was manually added, considering which interpretation was most plausible when a target word was presented in isolation (and always matching the verb distinction made during the experiment). Note that this approach is imperfect, because the word category of some target words remains ambiguous. For example, *bij* can mean both 'bee' and 'at' (or 'with', 'by', 'close to', or 'toward'). These words have the same pronunciation in Standard Dutch, but they have different pronunciations in some regional varieties. However, prepositional use is much more frequent, so this directs the speaker's interpretation in the translation task (in the absence of further context). The word list task was piloted several times, and after processing feedback from pilot

participants, the list in Table 6.2 minimized the risk of confusion for speakers. In total, 45 verbs, 39 nouns, 29 adjectives, 13 adverbs, five numerals, and two prepositions were included in the final list (a sufficient amount according to simulations of dialectometric studies; see Appendix C.2).

6.3 Method

6.3.1 Phonetic transcriptions — A single transcriber (the author of this dissertation) phonetically transcribed all target words for which valid recordings were available in the GTRP and SPRAAKLAB corpora, including transcriptions for Standard Dutch based on the pronunciations provided by a news reader (see Chapter 5 for details). We opted for broad phonetic transcriptions without suprasegmental information and diacritics (similar to Chapters 4 and 5), because they are often unreliable, even within transcriptions made by the same person (Shriberg & Lof, 1991). Phonetic distances obtained from transcriptions including such smaller distinctions also correlate strongly with phonetic distances obtained without these fine-grained distinctions, and fine-grained distinctions usually disappear in aggregated analyses involving many target words (Wieling & Nerbonne, 2015).

Note that the /r/ was always transcribed as [r]. Sebregts (2015) details how complex the nature of this phoneme is in Dutch, which can occur in at least ten variants in Standard Dutch (Van de Velde & Van Hout, 1999). Transcribing this variation would be highly time-consuming and likely unreliable, so the variation of this phoneme was simplified and is unlikely to affect the main conclusions of our aggregate-level analysis.

Finally, the transcriptions were made by listening to all recordings by target word (rather than participant or reference location), because the analyses also focus on differences at the level of words. The full set of transcriptions, also of the speakers not included in the real-time analysis of pronunciation change, is available in the supplementary materials. The 40 phonetic symbols (17 vowels, 23 consonants) used for transcribing all recordings (i.e., the GTRP and SPRAAKLAB recordings combined) are reported in Table 6.3.

6.3.2 Analyses — In this chapter, we use an aggregate and in-depth analysis to assess pronunciation change and regiolect formation. We again use the Levenshtein distance (for the aggregate analysis) and the 3D Levenshtein distance

Table 6.3: Phonetic symbols used in the combined corpus, ordered by place and manner of articulation according to the International Phonetic Alphabet (2005 version). Consonants are grouped into bilabials, labiodentals, (post)alveolars, velars, and glottals.

Vowels				Consonants				
Front	Central	Near-back	Back	Bilabial	Labiodental	(Post)alveolar	Velar	Glottal
i		ʊ	u	p	f	t	k	ʔ
y			o	b	v	d	g	h
ɪ			ʌ	m	ʋ	n	ŋ	
ʏ			ɔ			r	x	
e	ə		ɑ			ɾ	j	
ø						s		
ɛ						z		
œ						ʃ		
æ						ʒ		
a						l		

(for the in-depth analysis; see Chapter 5 for details) to quantify differences between the phonetic transcriptions for the realizations of Standard Dutch target words (i.e., always involving a triplet of transcriptions from the Standard Dutch speaker, the GTRP speaker, and a SPRAAKLAB speaker).

6.3.2.1 Aggregate analysis. For the aggregate analysis, we construct dialectometric maps based on the Levenshtein distance using Gabmap (Leinonen et al., 2016) to investigate pronunciation change between all regional varieties of the regional languages and Standard Dutch. These dialectometric maps include beam maps and multidimensional scaling maps (Nerbonne et al., 2011), which can be used to assess pronunciation variation (in the dialect corpora) at an aggregate level. We construct these maps for the GTRP and SPRAAKLAB corpora simultaneously and together with the Standard Dutch pronunciation, so we can assess whether existing dialect groups have become more internally similar over time, and also whether the distance of specific varieties to Standard Dutch has noticeably decreased.

We generate three maps using Gabmap: a beam map, a reference point map, and a multidimensional scaling (MDS) map. The beam map projects lines between the recording locations, which are colored according to the size of the linguistic distances. A beam map is useful for assessing whether groups of di-

alects have become more similar. The reference point map is also colored based on linguistic distances and used to see whether specific areas show an increased similarity to Standard Dutch (i.e., Standard Dutch is taken as the reference location in this map). To construct an MDS map, the high-dimensional distance data is reduced into three dimensions, which can then be mapped onto the three primary colors for interpretation. Areas with similar pronunciation are also colored more similarly, which is especially useful for visually assessing the dialect continuum. The reference point and MDS maps are partitioned into Voronoi tiles that differ in size according to the distribution of the recording locations, but the tile sizes themselves are not meaningful.

6.3.2.2 In-depth analysis. To obtain a more detailed view of the pronunciation change patterns, we construct a model to predict how many convergent, divergent, and neutral segments are expected in different dialect areas, while accounting for lexical variability and speaker characteristics. This approach is different from Chapters 4 and 5, in which we directly analyzed the PMI-based distances between phonetic transcriptions and normalized these according to the alignment length. While we lose the sensitive sound distances from the PMI approach, the present approach allows us to quantify the effect of neutral segments and it enables us to illustrate yet another approach for the dialectometrist's toolbox.

Given the nature of the predicted variable, we use a Poisson-based generalized additive mixed-effects regression model (GAMM; Wood, 2017), which is fitted using the *mgcv* library in R (Wood, 2000). This differs from the models fitted in Chapters 4 and 5, because the dependent variable is now a count of segments in the 3D Levenshtein alignments rather than a proportion of total pronunciation change per word. Note that using a Poisson-based model also means predictions are logit-scaled, similar to beta regression-based models.

The final model was constructed using an iterative modeling procedure in line with Wieling (2018) and Chapter 3, which ensured that the model explained as much of the variation in the data as possible using a minimal number of predictors. In line with the other models, we also assessed the inclusion of random intercepts and slopes (per target word). We initially only included the geographical effect (based on longitude and latitude coordinates), so it was unnecessary to include random effects per recording location.

Using the geographical effect model as a base model, we consecutively added new predictors, including lexical information (i.e., word frequency and word category) and speaker characteristics (i.e., gender, educational attainment, and

a measure of a speaker's regional identity strength). The more complex model was kept if the additional complexity was justified, which was assessed using the *compareML* function of the *itsadug* package in R (Van Rij et al., 2022). Maximum likelihood (ML) estimation was used when comparing models differing exclusively in the fixed effects, and the default fast restricted maximum likelihood (fREML) estimation method was used when models differed in their random effects. The fREML estimation method was used for the final model. Only significant predictors or interactions were retained in the final model.

6.4 Results

In the following sections, we showcase the findings from the aggregate analysis, which directly address the primary research inquiries in this chapter. Afterward, we present the in-depth analysis, which provides further insight into lexical and speaker effects in the overall observed pronunciation change across the geographical area. Note that the aggregate analysis focuses on both vertical and horizontal convergence patterns (i.e., between the regional language varieties and Standard Dutch, and also between the regional language varieties), while the in-depth analysis focuses only on vertical convergence (i.e., between the regional language varieties and Standard Dutch).

6.4.1 Aggregate analysis — The dialectometric maps are shown in Figure 6.3. The Frisian, Town Frisian, and Low Saxon language groups can be clearly distinguished at the aggregate level for both recording periods in all dialectometric maps.

The colors of the Frisian locations in the MDS space in Figure 6.3a remain virtually unchanged. Their similarity to Standard Dutch in Figure 6.3b seems stable, while the increased darkness of the lines in Figure 6.3c connecting the Frisian locations indicates that these varieties have become more similar in recent decades. The distance between the Town Frisian varieties appears stable in Figures 6.3a and 6.3c, and their similarity to Standard Dutch also seems stable in Figure 6.3b.

For the Low Saxon area, the three southwesternmost areas in the province of Gelderland bordering the Low Franconian area (i.e., Ermelo, Lunteren, and Didam; see Figure 6.2) appear to show a greater similarity to Standard Dutch over time (see Figures 6.3a and 6.3b), and these neighboring dialects remained rela-

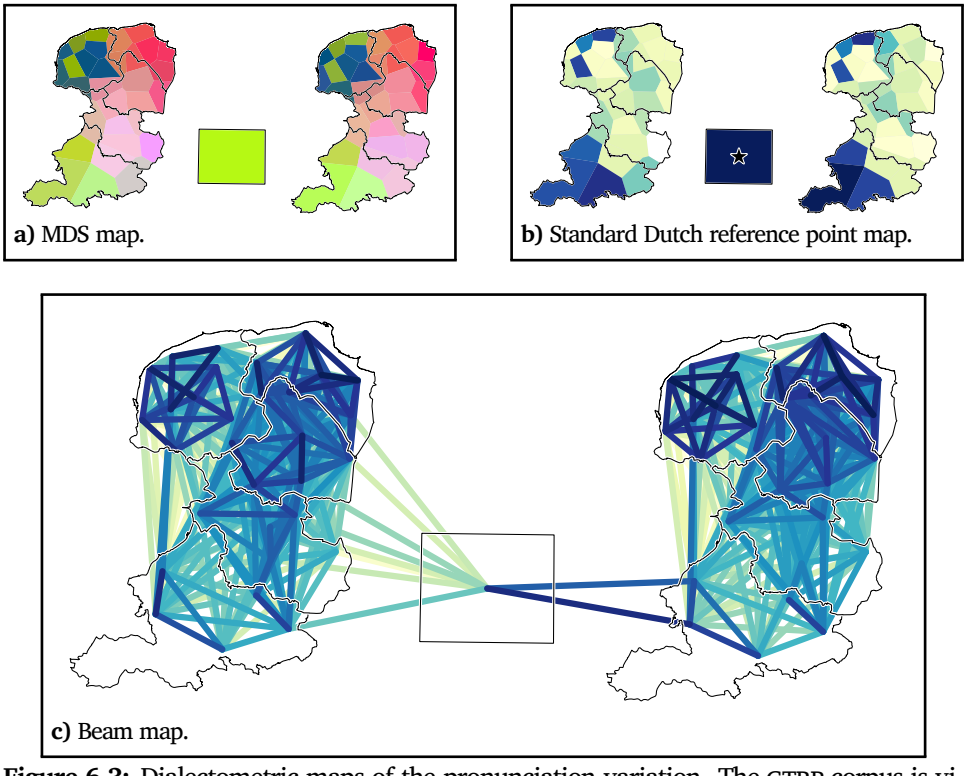


Figure 6.3: Dialectometric maps of the pronunciation variation. The GTRP corpus is visualized on the left, the SPRAAKLAB corpus on the right, and Standard Dutch is visualized via the rectangle in between. Note that darker colors in the beam map indicate greater similarity between locations, and darker colors indicate greater similarity to Standard Dutch in the reference point map. Similar colors in the MDS map indicate a more similar pronunciation in different areas.

tively similar to each other (see Figure 6.3c). These locations are also relatively distinct from the rest of the Low Saxon area, which is a consistent pattern across the two time periods (see Figure 6.3a). The eastern Twente and Achterhoek regions (i.e., close to locations 26 through 31 in Figure 6.2; see also the broader Low Saxon divisions in Figure 1.2 in Chapter 1) in the provinces of Overijssel and Gelderland have become more similar over time (see Figure 6.3a). The most substantial differences over time in these regions are found for the easternmost locations in Overijssel (i.e., Tilligte) and Gelderland (i.e., Groenlo), because of the increased similarity to surrounding areas (see also Figure 6.3b).

Furthermore, the northern Low Saxon areas in Groningen and the north of Drenthe appear to become more similar over time, although the increase in similarity seems less pronounced than for the Frisian locations (see Figure 6.3c). This pattern does not coincide with a substantially increased similarity to Standard Dutch in Figure 6.3b. The eastern locations in the provinces of Groningen and Drenthe remain relatively distinct from the other areas in Figure 6.3a, despite the increased similarity between other Low Saxon areas in Groningen and Drenthe.

6.4.2 In-depth analysis — We constructed a Poisson-based generalized additive mixed model predicting how many convergent, divergent, and neutral segments are expected given geographical variation, speaker characteristics, and lexical characteristics. We constructed a direction factor variable with three levels (abbreviated as 'Dir.') to distinguish between the different change directions with regard to Standard Dutch (i.e., convergence, divergence, or neutral). We also incorporated the length of each Levenshtein alignment in the model to correct for the fact that target words differ in length, which structurally affects the number of segments. The final model formula is as follows (note that the random intercept for word was not significant when including the by-word random slope for direction):

```
Count ~ Alignment length * Dir.
        + Regional identity strength * Dir.
        + Educational attainment * Dir.
        + s(Word frequency, by = Dir.)
        + s(Longitude, Latitude, by = Dir.)
        + s(Word, Dir., bs = 're')
```

The model summary for the parametric and non-parametric model terms are provided in Tables 6.4 and 6.5. The final model has an explained deviance of 78.1%.

As expected, a greater alignment length was associated with a greater pronunciation change for each direction (see Table 6.4), although the expected increase was smaller for neutral segments. Furthermore, while more convergence than divergence is predicted according to the model, neutral segments still account for most of the data (see Table 6.4). This is also confirmed when tallying the segments by direction, which shows that 84.8% of all segments were neutral,

Table 6.4: Parametric coefficients of a model predicting convergent (conv.), divergent (div.), and neutral (neutr.) segments.

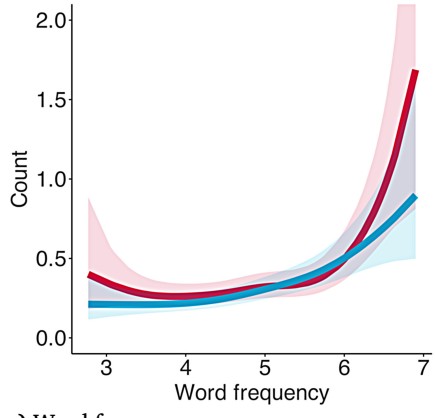
	Estimate	SE	*z*-value	*p*-value	
(Intercept)	-3.61	0.25	-14.47	<0.001	***
Alignment length	0.67	0.03	26.27	<0.001	***
Direction (div.)	-1.94	0.34	-5.65	<0.001	***
Direction (neutr.)	4.78	0.27	17.98	<0.001	***
Regional identity strength	-0.28	0.06	-4.87	<0.001	***
Educational attainment	0.02	0.01	1.35	0.176	
Alignment length × Dir. (div.)	0.18	0.04	5.13	<0.001	***
Alignment length × Dir. (neutr.)	-0.65	0.03	-23.52	<0.001	***
Dir. (div.) × Regional identity strength	0.41	0.08	5.28	<0.001	***
Dir. (neutr.) × Regional identity strength	0.29	0.06	4.96	<0.001	***
Dir. (div.) × Educational attainment	-0.07	0.02	-3.97	<0.001	***
Dir. (neutr.) × Educational attainment	-0.01	0.01	-1.02	0.308	

Table 6.5: Smooth coefficients and random slopes (sl.) of a model predicting convergent (conv.), divergent (div.), and neutral (neutr.) segments. The model includes a geographical smooth based on longitude and latitude coordinates.

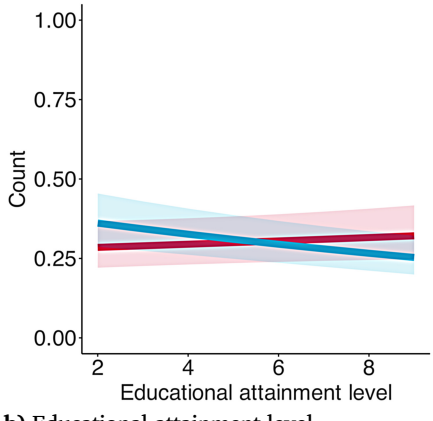
	Est. df	Ref. df	*z*-value	*p*-value	
Word frequency × Dir. (conv.)	3.71	3.80	32.88	<0.001	***
Word frequency × Dir. (div.)	2.37	2.44	28.81	<0.001	***
Word frequency × Dir. (neutr.)	1.16	1.16	3.26	0.114	
(Longitude × Latitude) × Dir. (conv.)	17.91	21.29	162.96	<0.001	***
(Longitude × Latitude) × Dir. (div.)	18.59	21.80	232.39	<0.001	***
(Longitude × Latitude) × Dir. (neutr.)	6.56	9.12	38.74	<0.001	***
Word, Dir. (random sl.)	360.28	394.00	3562.47	<0.001	***

7.5% of the segments were convergent, and 7.7% of the segments were divergent.

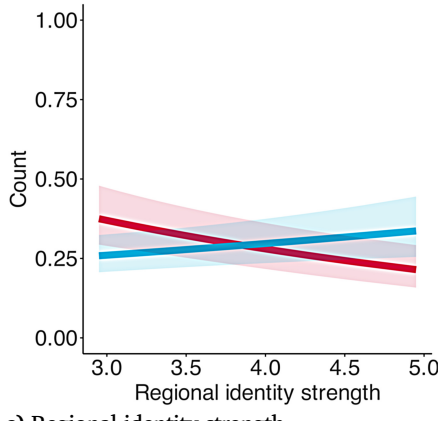
Due to the many interactions in the model, it is informative to look at the estimated marginal effects plots in Figure 6.4 and the marginal geographical effects in Figure 6.5. The estimates for neutral segments are left out of Figure 6.4, because we are interested in the convergence and divergence patterns, but the estimates for the neutral segments are reported in the supplementary material. The estimated marginal effect for word frequency in Figure 6.4a shows a slightly



a) Word frequency.



b) Educational attainment level.

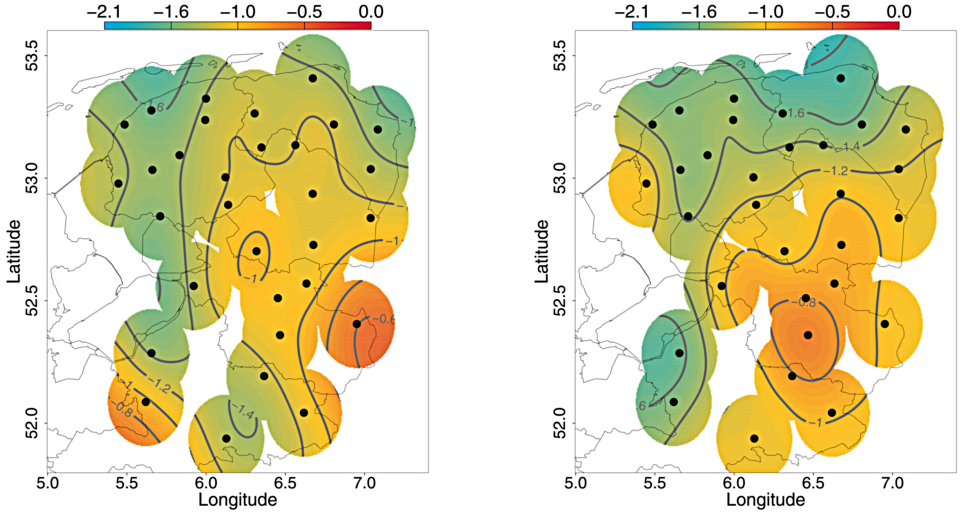


c) Regional identity strength.

Figure 6.4: Estimated marginal effects of the final model. The red line represents convergent segments, and the blue line represents divergent segments. The *y*-axis has been transformed from logits into odds.

higher amount of convergence for particularly low-frequency target words. The overall estimated pronunciation change is higher for higher-frequency words, but the increase occurs for both convergence and divergence. There was no statistically significant overall effect of word category in the final model, so this variable was not included.

The speaker characteristics visualized in Figures 6.4b and 6.4c significantly contributed to the model, but their overall effects are small and plotted with a



a) Convergence to Standard Dutch. **b)** Divergence from Standard Dutch.

Figure 6.5: Marginal geographical effects of the final model. Bluer colors indicate relatively less estimated change in that direction, and redder (or more yellow) colors indicate more change in that direction. The dots indicate the recording locations. The predicted values are logit-scaled.

smaller range of the *y*-axis for readability. The estimated divergence was higher for speakers with lower educational attainment, while the estimated convergence was higher for speakers with high educational attainment. Furthermore, the estimated divergence was somewhat higher for speakers with a stronger regional identity, while the estimated convergence was higher for speakers with a weaker regional identity.

The marginal geographical effect plot for convergence in Figure 6.5a is in line with the earlier Figure 6.3b, because it shows relatively strong convergence in the southwesternmost Low Saxon areas in the province of Gelderland and Tilligte in the province of Overijssel. More moderate amounts of convergence are also found for most other areas, with the least convergence occurring in the northeastern part of the province of Groningen and most of the province of Fryslân. The marginal effect of divergence in Figure 6.5b shows a relatively strong area of divergence in the eastern Twente and Achterhoek regions in the provinces of Overijssel and Gelderland, and relatively little divergence in other areas. The geographical variation of neutral segments is complementary to those already shown and therefore not reported here (see the supplementary material).

6.5 Discussion

Improving on the analysis of Chapter 5, we addressed research questions 4a through 4c (see Section 1.4) in this real-time study with closely controlled speaker parameters. We assessed whether there was evidence for regiolect formation for the Frisian, Town Frisian, and Low Saxon regional languages between the second half of the 1980s and the beginning of the 2020s. We estimated how much and in which direction these regional language varieties changed based on pronounced translations of 133 Standard Dutch target words. The aggregate analysis showed partial evidence of regiolect formation for Frisian and the northern Low Saxon region, because the local varieties in these areas became more similar over time. There was no clear evidence of increased similarity between neighboring dialects between the Town Frisian and southern Low Saxon varieties over time. Future studies may determine to which degree these potential regiolects are perceptually salient to speakers, and whether these regiolects replace the role of traditional dialects or instead add a potentially persistent extra layer of regional variation (as is the case in Germany; Kehrein, 2020).

A subsequent in-depth analysis showed that convergence to Standard Dutch was strongest around Lunteren and Ermelo, located in the southwestern Low Saxon border area with the Hollandic dialects. There was also a relatively strong area of convergence around Tilligte in the east of the province of Overijssel, although this could be due to the GTRP speaker being highly dissimilar from other varieties at the time (see Figure 6.3b). This speaker may not have been the best reference speaker for the local dialect, but that cannot retrospectively be ascertained either. More moderate convergence rates were found in most other areas, with the lowest predicted convergence rates in the province of Fryslân and the northeastern part of the province of Groningen. Divergence rates were predicted to be lower than convergence rates. The locations in the province of Overijssel (except IJsselmuiden) stood out as showing relatively more divergence from Standard Dutch, although there was also no indication of an increased similarity to Standard Dutch, so these dialects appear to fracture the dialect continuum.

The in-depth analysis showed that even when the recording period, reference location, and age range are relatively well controlled across time, other speaker characteristics can still influence the results to some degree. These characteristics included a speaker's regional identity strength and education attainment levels, although these effects were relatively minor. It is well known that educational background (or social class) and dialect use interact in the Netherlands

and elsewhere (Chambers & Trudgill, 1998; Driessen, 2005; Schmeets & Cornips, 2022, p. 58), although the effect is only statistically significant for divergence in our analyses. The observed regional identity effect occurs for both convergence and divergence, which indicates that this may play a more substantial role than educational backgrounds in the overall language change of regional varieties.

A driver of the observed divergence patterns at the speaker level could be a process in which hyperdialectisms (i.e., 'overdoing' dialectal features to express a regional identity) become entrenched in a speaker's language system, which has recently been observed for Brabantish and Limburgish (Van Spijk & Swanenberg, 2018; Doreleijers et al., 2021). This is not unexpected in the context of the substantial dialect loss that has been observed for Low Saxon (see Chapter 2), and it may explain the 'new' pronunciation variation observed in the province of Overijssel (showing both divergence from Standard Dutch and neighboring varieties). An example of this variation in Overijssel is the observed change from [ɑ] to in [ɔ] (e.g., [(h)ɑldn] →[hɔldn] 'to hold'), which is observed for all recording locations in that region. Furthermore, we observed changes for the target word *dreigen* 'to threaten' from [ɛi] to [i] (and vice versa), but also from [ɛi] to [a] and [i] to [ɛ]. These examples show increased vowel variability in this region, making the dialects simultaneously less similar to Standard Dutch and each other. It is also likely that such potential hyperdialectal behavior is more accepted among people with lower educational attainment, for example as an in-group solidarity marker (see, e.g., Doreleijers & Swanenberg, 2023b). Further studies can disentangle these identity effects more appropriately per individual speaker. Individual speaker effects can ostensibly be detected in aggregate analyses, confirming that regional identity strength should be accounted for in regional language change studies (see also Beaman, 2021).

Target words and their dialectal translations are also not equally likely to change. Low-frequency words (e.g., *uilen* 'owls') were particularly likely to converge to Standard Dutch, and high-frequency words (e.g., *nu* 'now') were more likely to change regardless of the direction. Low-frequency words are known to be prone to change (Bybee, 2002; Pagel et al., 2007), but the number of changing segments in high-frequency words appeared higher overall. These effects at the ends of the frequency spectrum probably have different underlying processes. Phillips (1984, p. 322) suggests that phonetic processes affect high-frequency words more than low-frequency words, while higher-level processes affect low-frequency words first. Numerous high-frequency words in our word list are short single-syllable words (e.g., *op*, *om*, *wil*, *nu*, *ook*, *bij*, and *door*; see Table 6.2), so

such words may change relatively easily. Bybee (2002, p. 269) suggests that especially high-frequency monosyllabic words are reduced more easily, even when compared to monosyllabic low-frequency words. However, we likely reduced this effect by presenting the words in isolation, so further work is necessary to ascertain whether this effect occurs in such an experimental setting. A potential underlying higher-level process, resulting in more change for low-frequency words, is that dialect words for infrequently used concepts are not regularly reinforced in a speaker's lexicon. Low-frequency words or features may more easily be replaced by high-frequency alternatives, for which the forms from Standard Dutch are prime candidates (due to the strong presence of Standard Dutch in all social contexts and institutions). Consequently, the regional language lexicon becomes smaller over time, even for speakers who are relatively proficient regional language users.

Comparing task differences between the GTRP and SPRAAKLAB recordings may be useful. The tasks were largely similar by design, but the task for the SPRAAKLAB corpus was more fast-paced than the GTRP task. Especially for low-frequency words, this may make it more likely that the Standard Dutch alternative is more readily activated. The pacing of the SPRAAKLAB task did not heavily bias the results with only Standard Dutch approximating forms, because there were clear cases where the non-Standard Dutch target words were pronounced instead. Variants of *tsjuster* (as opposed to Standard Dutch *donker* 'dark') were used by almost all Frisian speakers in the SPRAAKLAB recordings, whereas these variants were rare in the GTRP.

Another difference between the GTRP and SPRAAKLAB tasks is the degree to which the researcher was present during the recordings. The researcher's presence was minimized in the SPRAAKLAB recordings to avoid unintentionally influencing speakers to use more standard language forms (which was the variety the researcher used). This was not done for the GTRP corpus, although some field workers spoke in regional varieties instead of Standard Dutch, alleviating the problem of potential Standard Dutch accommodation. At the same time, the speakers may still use more Standard Dutch forms as an alternative if the regional varieties are not highly similar, because the speaker likely still experiences the field worker as a differently speaking outsider. A future study may discern whether the effects of higher-paced tasks negate the advantage of having no researcher present when eliciting dialect words, but this does not appear to have strongly impacted our findings.

One issue was identified only after the SPRAAKLAB recordings concluded. We

relied on the GTRP target words provided in Gabmap for the word list, which was assumed to include only the first-person plural of verbs (Wieling, 2007, p. 9). We discovered that the original GTRP list included the first-person plural form of only a few verbs and that most verb targets in the list were in the infinitive form instead. The confusion concerning these forms likely stems from the fact that these inflections have the same form and pronunciation in Standard Dutch and many dialects in the Netherlands (i.e., ending in [ə(n)]). The few targeted first-person plural verbs in the GTRP corpus are translated with a [t]-ending in numerous Low Saxon varieties (specifically, locations 17, 18, and 23 through 31 in Figure 6.2). All plural verb forms uniformly end in [t] for these Low Saxon locations.[5] However, we presented the underlined verb forms as first-person plural verbs to participants for the SPRAAKLAB recordings. Only three SPRAAKLAB speakers (a subset of speakers from locations 26 and 29) pronounced a [t]-ending for verbs, and only inconsistently so. The scarcity of the [t]-ending in the SPRAAKLAB corpus could be due to the choice to display the verbs without the preceding pronoun *wij* 'we', because we wanted to avoid speakers translating and pronouncing this part. This may have confused speakers, causing them to produce the infinitive form instead. At the same time, the scarcity of the [t]-ending can also result from an expected form of language change (i.e., a greater similarity to Standard Dutch). Given this complicated state of factors, we evaluated whether the patterns meaningfully changed when infinitives (i.e., in the GTRP) were compared to first-person plural endings (i.e., prompted in the SPRAAKLAB recordings), compared to excluding these items completely in our analysis. Fortunately, there was no meaningful difference in the results when the verbs were left out (see the supplementary material), despite the substantial sample size reduction (i.e., using data from 88 target words instead of 134).

Finally, it is attractive to think of ways to make obtaining relevant dialect data less time-consuming. For this chapter, a single transcriber (the author of this dissertation) made transcriptions for 106 speakers and 133 target words, which required listening to over 14,000 recordings several times. It is also possible to leverage neural acoustic models and automatically extract abstract numeric representations of the sound recordings. The differences between these representations can be quantified, similarly to how the Levenshtein distance can

[5] This is seen as a typical Saxon feature that is shared by many Low Saxon varieties in Germany (Bloemhoff et al., 2008a, p. 212), although many Low Saxon varieties have lost this feature nowadays (especially varieties that we included under the Northern Low Saxon group in Chapter 5; Bloemhoff et al., 2008b, p. 104).

be used to measure the difference between phonetic transcriptions (Bartelds et al., 2022; Bartelds, 2023). We also attempted this approach for our data, but the obtained distances were systematically disproportionately large between the GTRP and SPRAAKLAB recordings. This was surprising, because using this neural acoustic method for the SPRAAKLAB recordings proved to be successful for automatically quantifying regional language variation before (Buurke et al., 2024e), and likewise for the GTRP data (Bartelds & Wieling, 2022). This problem may be caused by the different sound recording methods of the corpora, because the GTRP was recorded on tape instead of with digital microphones.[6] This suggests that further investigations are necessary to assess the usefulness of the neural acoustic method in scenarios where different datasets are compared. This would be beneficial, as the acoustic method is likely more capable of fully processing all details in the sound signal. Human transcribers are limited by their perception, which can be problematic. For example, several Standard Dutch target words had word-initial /v/ or /z/. These sounds are involved in an ongoing language change of devoicing and (near-)merger with /f/ and /z/, respectively, which is most advanced in the northern Netherlands (Van de Velde et al., 1997; Pinget et al., 2016). A human Dutch transcriber (especially from the area of interest) likely has trouble distinguishing these sounds in these positions, because they are also subject to this ongoing change. Furthermore, it has been shown that the neural acoustic method aligns better with overall aggregated human perception than using the Levenshtein distance (Bartelds et al., 2022).

6.6 Conclusion

By analyzing phonetic corpora of regional languages in the northern and eastern Netherlands from the 1980s and 2020s, we found evidence for regiolect formation in these languages, especially in Frisian. The detection of regiolects in this area fits a more widely occurring pattern in the Netherlands and Flanders, although it is not yet clear whether the increased similarity between localized dialects has an active role in speaker perception. Furthermore, we found that convergence to Standard Dutch was strong in the border region between Low Saxon and the Hollandic dialect group and for an eastern village in the province of Overijssel. The divergence from Standard Dutch was also relatively strong in Overijssel, but this co-occurred with a divergence between dialects in this area, which appears

[6] See https://projecten.meertens.knaw.nl/mand/GTRPdatata.html.

to fracture the Low Saxon dialect landscape.

We managed to control for speaker characteristics when measuring community language change, which addressed shortcomings of previous studies in this area, although differences between data collection tasks remain difficult to avoid. At the same time, some speaker characteristics still presented statistically measurable effects, such as a speaker's educational background and regional identity strength. These findings indicate that it is essential not to rely on a single speaker in sampling locations, because inter-speaker differences can skew the results, although determining a reasonable minimum number of speakers remains a task for future studies.

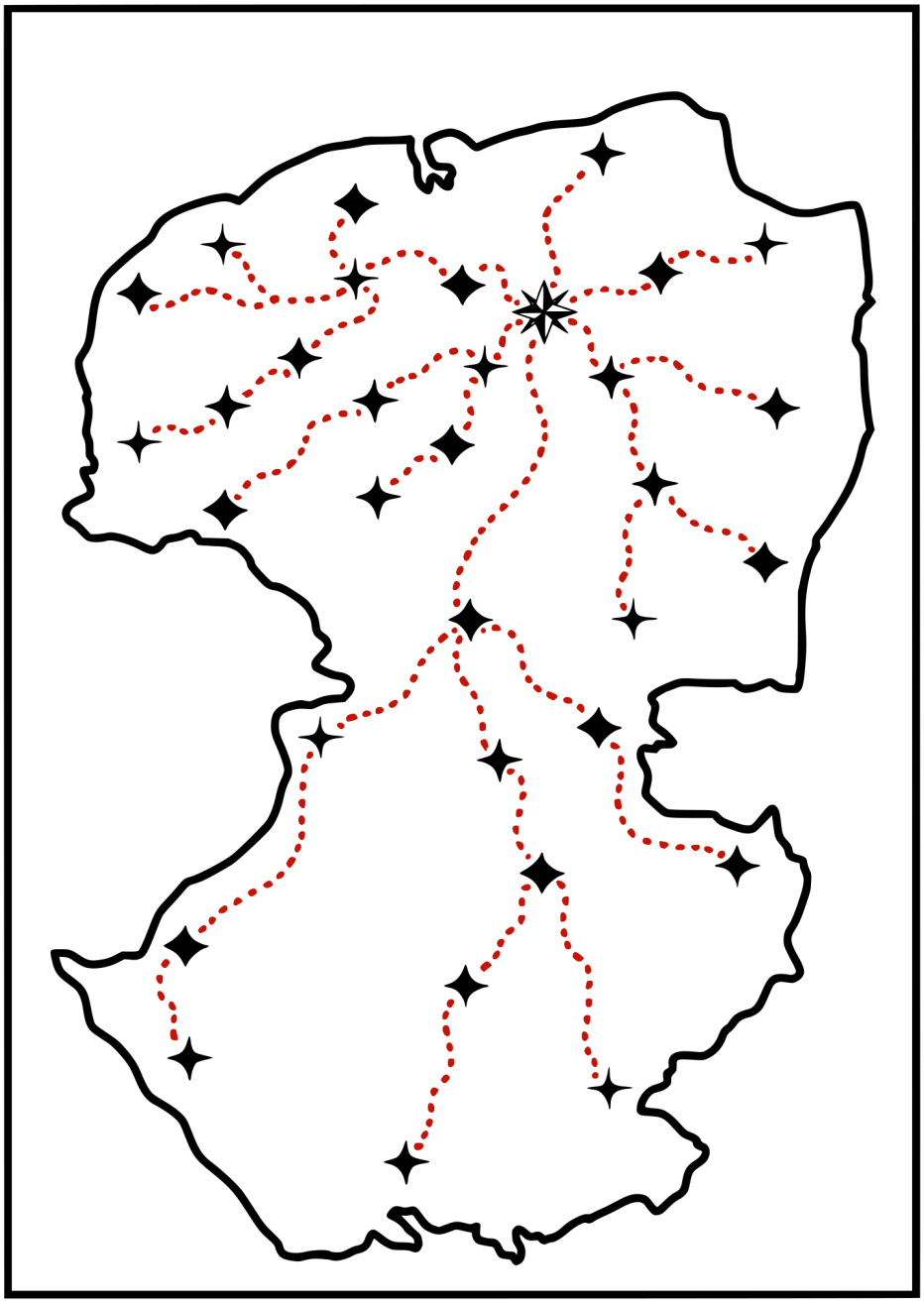
Further studies may investigate whether regiolect speakers also form a strong regional identity in a way that is similar to traditional dialect speakers. This may be used to strengthen language preservation efforts (e.g., with the formation of a regiolectal writing standard). Language preservationists might also attempt to explore the potential institutionalization or legal recognition of these regiolects. Exploring this concept is particularly interesting for Low Saxon, which enjoys fewer legal benefits than Frisian and whose speakers are spread across more provinces in the Netherlands.

Part IV

Conclusions

Lauwersoog, Groningen

CHAPTER 7



AN AMBISPECTIVE VIEW ON FRISIAN AND LOW SAXON

THE SUBSTANTIAL LINGUISTIC DIVERSITY in the northern and eastern parts of the Netherlands provides a fruitful context for research on minority languages and dialectology. The two major regional languages in this region are Frisian and Low Saxon, which are usually spoken alongside Standard Dutch. Varieties of these languages were historically spoken in a much larger area, but they have been subordinated to Standard Dutch in the Netherlands. Nowadays, these languages are under severe pressure from Standard Dutch due to a weaker political position and the fact that Standard Dutch has permeated all contexts of daily life. Consequently, the Frisian and Low Saxon speaker populations declined, and the languages have become substantially more similar to Standard Dutch. In this dissertation, we assessed how far these processes have progressed in recent decades. We also explored whether there is evidence for the formation of new regional language varieties, which have linguistic features that are less localized than those of traditional dialects.

The research questions, formulated in Chapter 1, and their answers based on this dissertation are discussed in the next section. These questions concerned changes in the current Frisian and Low Saxon speaker populations and pronunciation changes of varieties of both languages. While the comparison is relevant for those interested in Frisian and Low Saxon specifically, it also has more general value, because comparing the two languages may provide insights into how regional languages under pressure of the same standard language may develop along substantially different paths depending on specific differences.

7.1 Discussion

In Chapter 2 of this dissertation, we first assessed the size of the current speaker populations of Frisian and Low Saxon, specifically in the northern three provinces in the Netherlands. Through an innovative combination of information from earlier speaker counts with a new large-scale regional language questionnaire, we obtained percentages of how many people across three generations indicated that they (in their view) could speak the language, and whether they used the regional language at home. For Frisian, we found that between 59%

and 65% of the population in Fryslân could speak Frisian across different generations, while between 46% and 51% indicated using the language at home. There was a slight downward trend for self-reported speaking capability, while language use at home increased slightly over time. The language transmission from parents to children also increased for Frisian across generations (see Table A.7 in the Appendix), which is rare for European minority languages. By contrast, there were only downward trends for Low Saxon. The percentage of the population that could speak Low Saxon in Groningen decreased from approximately 51% to 28% across three generations, and the use of the language at home decreased from approximately 30% to 14%. For Drenthe, the percentage of Low Saxon speakers decreased from 55% to 39%, and regional language use at home decreased from approximately 37% to 23%. The intergenerational transmission also decreased from approximately 48% to 31% in Groningen and from 54% to 42% in Drenthe (see Table A.6 in the Appendix). These findings confirm our expectations based on earlier surveys (Bloemhoff, 2005; Klinkenberg et al., 2018; Versloot, 2021a), namely that the Frisian speaker population is stable (and might even grow in the next generation) and that the Low Saxon speaker population is sharply declining.

The different rates of intergenerational transmission of Frisian and Low Saxon are influenced by many factors simultaneously. In Chapter 3 we explored which factors are generally associated with transmitting Frisian or Low Saxon by parents who speak these languages. Many of the parents who did transmit their language indicated that the other parent or caregiver of their children spoke the same regional language, and these parents also frequently used their regional language in social contexts. These unsurprising findings indicate to which degree the regional language is embedded in the lives of these speakers and others around them. The positive effect on language transmission when both parents speak the regional language was stronger for Frisian than Low Saxon, which suggests that Frisian is more easily considered an alternative or addition to Standard Dutch in these families. Some other factors usually associated with regional language use were also associated with the parental transmission rate, such as someone's educational attainment and whether someone lived or grew up in a strongly urbanized environment, but these effects were smaller than the effect of the embeddedness of the regional language.

Perhaps the most characteristic difference between Low Saxon and Frisian language maintenance concerned the effect of a (positive) language attitude. Parents with a positive language attitude transmitted their language more often,

and the associated increase was larger for Low Saxon parents with a strongly positive attitude than for Frisian parents. Note that this does not mean that Frisian speakers do not have a positive language attitude. Instead, they had a more positive language attitude than Low Saxon parents overall. The Frisian language maintenance baseline was considerably higher than that of Low Saxon, so the potential positive impact of language attitude on language maintenance was also more limited. We also checked whether particular clusters of language attitude statements had a strong effect individually (e.g., statements about negative prejudices toward regional language use), but the explanatory power of the aggregated variable always exceeded that of smaller clusters of statements. These findings suggest that promoting a more positive language attitude regarding a broad range of topics among Low Saxon parents (e.g., emphasizing that the language is worth using and transmitting and that the language is not a form of improper Dutch) may be crucial for its continued maintenance, which has decreased strongly in recent decades. How proponents of Low Saxon can best proceed is, however, a question that requires future work, because language attitudes are extremely multi-faceted (Dragojevic et al., 2021). Nonetheless, the importance of language attitudes in this regional language context is clear.

The second part of this dissertation focused on measuring pronunciation change in Frisian, Town Frisian, and Low Saxon varieties. Many difficulties could and did arise in our investigation of this type of language change, as described in Chapters 4 through 6. The most important issue with existing dialect corpora in the Netherlands was that the existing transcriptions made by many different transcribers were inconsistent. This problem was addressed by merging phonetic symbols inventories (in Chapter 4) or by a single transcriber making all the transcriptions (as in Chapters 5 and 6). Another problematic aspect of comparing dialect corpora stemmed from the different elicitation tasks. In the case of the Goeman-Taeldeman-Van Reenen and From Dialect to Regiolect corpora (the GTRP and DIAREG corpora; Taeldeman & Goeman, 1996; Heeringa & Hinskens, 2015), this resulted in a relatively small subset of target words from both corpora that could be compared, because the pronunciation of target words was affected to a substantial extent by how they were elicited (i.e., as separate words in the GTRP corpus or in a sentence context in the DIAREG corpus). For the final chapter on language change (i.e., Chapter 6), we avoided these issues by collecting a new dataset overlapping as much as possible with the GTRP corpus in terms of elicited words, elicitation approach, and locations.

Focusing on the findings from Chapter 6, we can conclude that the overall

rate of pronunciation change was low (i.e., consistently less than 15% of the total measured sounds; see also Heeringa & Hinskens, 2015). When pronunciation change was observed, it often indicated convergence to Standard Dutch, although there was about as much divergence from Standard Dutch when tallying the convergent and divergent segments. It is unclear whether this means that convergence and divergence cancel each other out, but it is possible that some parts of the linguistic system are more prone to convergence (to Standard Dutch) and other parts are more likely to exhibit or reflect divergence. This is also reflected by (e.g.,) the word frequency effects we detected in our in-depth analysis in Chapter 6. A notable case of divergence was the consistent use of variants of *tjuster* 'dark' instead of variants of *donker* (its equivalent in Dutch) by the more recently interviewed speakers in the Frisian language area, while variants of *donker* were more commonly found in the older dialect corpus. However, such a substantial divergence was rare overall and likely reserved for a limited set of words.

Another area with relatively much divergence was the province of Overijssel, although the divergence mainly pertained to increased vowel variation instead of a consistent divergence from Standard Dutch pronunciation. There was no clear pattern in these data, which could result from speakers in these areas no longer acquiring the local dialects consistently. This is reminiscent of speakers of the Brabantish regiolect in the province of Brabant, who frequently produce dialectal variations that are not consistent with the traditional grammars, because these dialects are nowadays more commonly acquired as a second language instead of as a first language (Doreleijers & Swanenberg, 2023a). This explanation may be appropriate, but we cannot ascertain this, as the questionnaire underlying Chapter 2 only contained data from the provinces of Groningen, Fryslân, and Drenthe. Nonetheless, these linguistic patterns fit a situation in which there is renewed interest in dialects (possibly as a countermovement to globalization; Auer et al., 2005; Goeman & Jongenburger, 2009; Slaats, 2020) and more people wish to express their regional identity through dialect features despite having a low proficiency in the traditional dialect.

Finally, we assessed whether there was evidence for regiolect formation in Frisian, Town Frisian, and Low Saxon. We found clear evidence for increased similarity between Frisian varieties. This increased similarity was not necessarily a direct consequence of a strong convergence to Standard Dutch, because the distance to Standard Dutch appeared stable at an aggregate level. This indicates that this regional Frisian variety may not primarily be a consequence of conver-

gence to Standard Dutch, which is how regiolects usually form in the Netherlands and Flanders (such as *Tussentaal* in Flanders; Vandekerckhove, 1998). These varieties may converge to a separate Frisian roof variety instead, but future work is needed to evaluate this possibility. For the Low Saxon varieties, there was some evidence of regiolect formation in Groningen and the northern part of Drenthe. These varieties have become more similar, but the increase in similarity is less strong than it was for Frisian, and this seems to coincide more with convergence toward Standard Dutch, in contrast to the Frisian pattern. For the more southern Low Saxon and the Town Frisian varieties, there was no clear increase in similarity between these dialects (and even indications of increased dissimilarity, such as in Overijssel).

Our findings from Chapters 2 and 3 highlight how the development for most regional languages in Europe is one of steep decline, unless they are relatively strongly protected and have a speaker population that actively maintains the language. The willingness to protect and maintain a minority language appears related to a strong shared identity among its language users. Languages in a similar situation to Frisian include Irish in Ireland (Riagáin, 2001), Welsh in Wales (Azurmendi et al., 2001), and Basque in the Basque Country (Azurmendi et al., 2001). These languages are heavily intertwined with the identities of their speakers (Bourhis et al., 1973; Jones, 2008; Montaña, 2020), have had to survive along more powerful national languages, and are relatively well protected by law (i.e., they are included under Parts II and III of the ECRML). Nonetheless, these languages continue to struggle for survival, because their identity does not offer protection from domination by a more politically influential language (e.g., in the case of Irish). However, the advantage of a strong language-specific identity seems to put these languages in a different league than most European minority languages.

Low Saxon in the Netherlands lacks a strong shared identity among its speakers across the entire region (although stronger identities can exist more locally or regionally), so its proponents must rely more on strongly protective legislation instead. The notable changes in the Low Saxon speaker population and the linguistic systems may serve as a cautionary tale for proponents of other traditional regional languages without strong shared identities or strongly protective laws to help language preservation. Many European languages are in a comparable situation to Low Saxon, either because proponents of (closely related) minority languages act independently of each other instead of combining forces (e.g., Celtic languages in the United Kingdom; Mac Síthigh, 2018), or because

the corresponding nation is not inclined to recognize these regional languages (e.g., Belgium did not sign the ECRML and France did not ratify it; Hornsby, 2010; De Groot, 2018). Some languages in a similar situation as Low Saxon are included under the ECRML (or may still be included in the future). However, there are no judicial consequences for not adhering to the requirements stipulated in the ECRML (Dunbar, 2022). Consequently, nations cannot be forced to protect these languages actively even with the language being recognized under the ECRML. Our results suggest that these languages will eventually cease to exist due to the pressure of the accompanying national languages, albeit at different rates, unless their political protection is strengthened.

Our finding that the pronunciation of most Low Saxon varieties (especially south of Groningen) increasingly becomes more similar to Standard Dutch likely also hampers language maintenance, because it increases the probability of its speakers perceiving their language as a variant of the national language instead (e.g., Kirk, 2023). Additionally, increasingly similar language varieties are likely to compete for the same mental resources, because they share many cognates (Kirk et al., 2022). When a regional language starts to be perceived as a dialect of a standard language, the potential political benefits offered by the ECRML also become substantially more difficult to demand, because the ECRML only applies to languages that are not dialects of the state's official language(s). This fate may await numerous European languages that are no longer 'dissimilar enough' to their respective standard languages nowadays (e.g., English dialects and accents; Hughes et al., 2012). Even though language change due to language contact is natural and unstoppable (Aitchison, 2001), it may have negative consequences for minority languages in the current political climate in Europe.

Our findings may also be relevant to other minority languages. Specifically, the findings presented in this dissertation show how quickly relatively large languages can end up in a precarious situation. We found that only a few generations ago most inhabitants of the province of Groningen and Drenthe still had a relatively good command of the regional Low Saxon variety, but two generations later these speakers have become a small minority in these provinces. We cannot pinpoint which concrete factors most strongly contributed to this decline, but we can certainly observe the circumstances in which this occurred. Dutch citizens and the government do not appear actively hostile toward the language, but there does appear to be a lack of active undertakings to support the language. In the case of Frisian, this is compensated by a more actively supportive speaker population. This likely also influenced the apparently greater support of Frisian

by other Dutch citizens (e.g., a widespread perception of Frisian as a separate language) and the government. Our findings tell a cautionary tale: minority or heritage languages may at first only be perceived as less useful than a majority language or not as a separate language in need of support, but this may lead to the languages no longer being transmitted or their forms being replaced by those from another language. Consequently, the conditions that instigated the decline may be reinforced, resulting in a cyclical rapid decline in only a few generations.

7.2 Suggestions for future studies

Based on the work presented in this dissertation, there are several potential avenues for future research, which are discussed in detail in this section. A logical next step is to jointly investigate changes in the speaker population with pronunciation variation of Low Saxon in Germany (commonly referred to as Low German, *Platt*, or *Plattdeutsch*), where the language is spoken in a large area and included under Parts II and III of the ECRML instead of only under Part II. Reershemius (2011) reported that Low Saxon varieties are often used in advertising or to project a regional identity on social media, which is a relatively limited phenomenon in the Netherlands. It is worth assessing whether including Low Saxon under Part III (in addition to Part II) of the Charter has influenced the increased usage of Low Saxon in advertisements in Germany. Note that earlier studies reported that convergence to the overarching standard language occurs for Low Saxon varieties on the German side of the Dutch-German border as well (Smits, 2011). Consequently, convergence to a standard language does not seem to be avoided by the extended protection of Part III of the ECRML. However, it could be that the extended protection *limits* the degree of convergence to the standard language. A direct comparison of Low Saxon varieties across a larger area in the Netherlands and Germany would therefore be useful.

The approach taken in Chapter 2 to derive the speaker counts is reasonable, but using a single questionnaire to derive speaker counts is more reliable. The main problem with the questionnaire we sent out through Lifelines is that people who did not speak a regional language may have outright dismissed it, as they believed their participation in a questionnaire about regional language use would be irrelevant. This problem falls in the wider category of nonresponse biases, which are increasingly common for questionnaires (Manzo & Burke, 2012). However, the resulting bias is aggravated when it disproportionately affects one

subgroup (in this case, the underrepresentation of people who do not use the regional language). In future studies, it is advisable to use a multi-stage approach when distributing a questionnaire for speaker counts. The recruitment message should be tested in pilot studies by distributing different versions to subgroups to check whether any important subgroups are unlikely to respond. As long as enough pilot participants are reached, this approach also allows researchers to verify that the responses given to the survey questions are not heavily skewed, especially when someone is asked to indicate how proficient they are in a particular language. Language surveys are notoriously susceptible to different response patterns depending on the formulation of the questions (Duchêne & Humbert, 2018). The variability in response patterns to questions about self-reported speaking ability likely results from the fact that self-identification as a speaker can be difficult, and what is perceived as a language and what is not perceived as such can differ greatly between speakers. For example, most Low Saxon respondents (i.e., 87%) indicated speaking a Dutch dialect instead of a separate language.

We found a residual geographical effect in the model accounting for language maintenance in Chapter 3. Parents who spoke Low Saxon and lived in the province of Groningen were relatively unlikely to transmit their language compared to parents in Drenthe. This finding was initially surprising, because the most important factors were already accounted for in the statistical model. Upon further inspection, it appeared that the age of acquisition is around a year later on average in Groningen compared to Drenthe. This suggests that Low Saxon parents in Groningen are less likely to introduce their children to their regional language from birth or a very young age, so children of Low Saxon parents in this region may be more likely to start acquiring the regional language in primary school. There is no straightforward explanation for this finding, because it is not the case that language attitudes are particularly negative in the province of Groningen or that Low Saxon is uniquely facilitated or promoted in Drenthe. We focused on the differences between Frisian and Low Saxon in this dissertation, but inspecting the processes that underlie this language maintenance difference between Groningen and Drenthe may be worthwhile, as it could potentially also shed light on why Low Saxon transmission is so low compared to Frisian.

We always tried to find at least two speakers in our reference locations for the SPRAAKLAB corpus in Chapter 6. We succeeded for all reference locations except one and even made recordings with four or more speakers in some places. We advise that future studies investigating aggregated pronunciation variation or

changes in regional languages in the Netherlands also focus on contacting more speakers per reference location instead of finding an optimal single speaker, as was usually done for older dialect corpora. The continued influence of Standard Dutch causes an influx of Dutch features into traditional regional dialects, especially because people move around the country more than they did a century ago (Bek, 2022). Simultaneously, more speakers learn a local dialect as a second language instead of a first language (mainly in the case of Low Saxon), which may reduce the internal consistency of the local language and potentially hampers maintenance. These processes increase the variation between speakers of regional varieties, so relying on a single speaker is increasingly likely to result in a skewed sample of a local dialect. By relying on multiple speakers and analyzing the variation between them, dialectology becomes more similar to sociolinguistics than it has been historically, but bridging these fields is a worthwhile endeavor (and has been encouraged by others, such as Wieling & Nerbonne, 2015).

Returning to the question of regiolect formation, we want to stress that our evidence of regiolect formation for Frisian and the northern Low Saxon varieties warrants verification in future studies. The regional Frisian variety that appears to form may be due to a spoken Frisian roof variety or a spoken form of the standardized written variety, which we initially deemed unlikely. At the same time, the regional media in Fryslân actively promote and use Frisian, and the language is taught in schools, so perhaps this has resulted in a spoken Frisian standard. The question then arises whether this regional form is perceptually salient to speakers of Frisian, which should be assessed in a future study. We deem it less likely that the potential regiolect in the northern Low Saxon area is mentally represented as a separate variety that can be used for communication across a larger area. This should ideally be ascertained in a future study.

Finally, it may be worthwhile to explore to which degree alternative methods can be used instead of phonetic transcriptions to represent dialect pronunciations. Neural acoustic representations, for example, can be constructed if the recordings are consistent and of sufficiently high quality. These neural representations can subsequently be compared using dynamic time warping (Bartelds et al., 2022; Bartelds, 2023).[1] Such an approach has been successful for the GTRP before (Bartelds & Wieling, 2022), and also for the SPRAAKLAB recordings (Buurke et al., 2024e), but the approach failed when we tried comparing the

[1] A demonstration of the neural acoustic method can be found at https://github.com/rbuurke/single-word-comparisons-demo.

GTRP and SPRAAKLAB recordings directly using this method (see Chapter 6). This shows that further exploration is necessary to determine when this approach is appropriate. Nonetheless, we think the method has great potential as a tool in dialectology and language variation research, especially because it works as a better proxy for human perception than quantifying pronunciation variation using phonetic transcriptions (Bartelds & Wieling, 2022). Especially when the neural acoustic method is combined with automatic word segmentation (e.g., voice activity detection; Bredin, 2023), it can substantially reduce the amount of manual labor involved in collection and analyzing speech variations.

7.3 Conclusion

In this dissertation, through various methodological innovations, we found that the Frisian and Low Saxon languages and their speaker populations underwent different, but substantial changes in recent decades. Frisian and Low Saxon share a struggle for language maintenance and linguistic stability, but Frisian appears to fare better overall. To which degree the stability of these languages can be increased is a question for future studies, but our findings indicate that answers can be sought in various domains, such as language attitudes and further exploration of factors underlying language transmission in specific regional language areas (such as in the province of Groningen).

A new development for Frisian and Low Saxon is that an intermediate language variety seems to emerge. However, this is clearer for Frisian than the northern Low Saxon area. This development is in line with a general trend in the Netherlands and Flanders, but the potential regiolect that appears to form for Frisian seems to originate in an increased similarity between Frisian varieties, rather than convergence toward Standard Dutch. Convergence toward Standard Dutch does seem to underlie the formation of a potential regiolect for the northern Low Saxon area. These findings exemplify how Frisian develops in an apparently unique fashion compared to other regional languages in the Netherlands and Flanders. By contrast, Low Saxon seems to follow the general trend of European regional languages: it is becoming substantially more similar to the standard language (Dutch) and its speakers are increasingly likely to think of themselves as speakers of a dialect of the standard language rather than a separate language.

Our conclusions for Low Saxon may make it seem as if the language cannot be

maintained, but Low Saxon is not doomed yet. The findings regarding parental language transmission in particular show that if more positive language attitudes are fostered among Low Saxon speakers, language maintenance may follow suit. The Low Saxon speaker population itself must carry the responsibility for improving language attitudes, but the government should also be encouraged to follow through on its commitments to maintaining and promoting linguistic diversity and promoting positive language attitudes toward Low Saxon wherever possible. The more favorable situation of Frisian shows that this non-trivial task is achievable.

Part V

Appendices

APPENDIX A
SPEAKER COUNT ADDENDUM

A.1 Speaker count components

To obtain the estimates in Chapter 2, we used tables reported in reference works: Bloemhoff (2005) for the Low Saxon estimates, and Klinkenberg et al. (2018) for the Frisian estimates. We provide the relevant numbers from the Low Saxon reference work in Tables A.1 and A.2, and those from the Frisian reference work in Table A.5. The reference work numbers are summarized and adjusted for the provinces of Groningen and Drenthe in Tables A.3 and A.4. The adjustment components necessary for estimating the metrics for the next generation are based on the Lifelines questionnaire data and reported in Tables A.6 and A.7.

Table A.1: Percentages per level of Low Saxon speaking proficiency across different age groups, as reported in Bloemhoff (2005, p. 63).

Birth years	Age group	Low Saxon speaking proficiency				
		None	Bad	Reasonably	Well	Very well
1922 – 1942	61 +	21.8	8.1	14.4	18.4	37.3
1943 – 1963	40 – 60	21.9	11.2	14.5	20.2	32.2
1964 – 1985	18 – 39	22.7	12.9	25.9	18.2	20.3

Table A.2: Percentages of Low Saxon use at home across different age groups, as reported in Bloemhoff (2005, p. 61).

Birth years	Age group	Language used at home			
		Low Saxon	Dutch	Both	Something else
1922 – 1942	61 +	39.8	34.5	12.4	39.8
1943 – 1963	40 – 60	30.4	39.7	21.5	30.4
1964 – 1985	18 – 39	17.3	53.2	22.4	17.3

Table A.3: Adjusted percentages of Low Saxon speakers in Groningen and Drenthe, who indicate to speak Low Saxon reasonably well, well, or very well. See the supplementary material of Chapter 2 for details.

Birth years	Age group	Able to speak Low Saxon	
		Groningen	Drenthe
1922 – 1942	61 +	83	82
1943 – 1963	40 – 60	78	77
1964 – 1985	18 – 39	71	70

Table A.4: Adjusted percentages of Low Saxon speakers in Groningen and Drenthe, who indicate to speak Low Saxon at home or to do so together with Dutch.

Birth years	Age group	Using Low Saxon at home		Using Low Saxon and Dutch at home	
		Groningen	Drenthe	Groningen	Drenthe
1922 – 1942	61 +	48	36	55	42
1943 – 1963	40 – 60	47	28	55	32
1964 – 1985	18 – 39	36	16	42	18

Table A.5: Percentages of speakers living in Fryslân able to speak Frisian and using Frisian at home in Fryslân across different age groups. These data are based on Klinkenberg et al. (2018, p. 60).

Birth years	Age group	Able to speak Frisian	Using Frisian at home
1931 – 1951	65 +	69	49
1952 – 1966	50 – 64	69	54
1967 – 1986	30 – 49	72	62
1987 – 2007	$\leq$ 29	66	62

Table A.6: Adjustment components for Low Saxon speakers in Groningen and Drenthe based on the Lifelines data.

Birth years of child generation	Intergenerational transmission from parents to children (in %)		Rate of acquisition through parents	
	Groningen	Drenthe	Groningen	Drenthe
1952 – 1972	48	54	78	79
1973 – 1993	39	48	78	79
1994 – 2015	31	42	78	77

Table A.7: Adjustment components for Frisian speakers in Fryslân based on the Lifelines data.

Birth years of child generation	Intergenerational transmission from parents to children (in %)	Rate of acquisition through parents
1961 – 1981	74	78
1982 – 1996	71	82
1997 – 2016	69	84

A.2 Geographical distributions of language transmission

We present the geographical distributions of language transmission for Frisian in Figure A.1 and for Low Saxon (in Groningen, Drenthe, and the municipalities of Oost- and Westellingwerf) in Figure A.2.

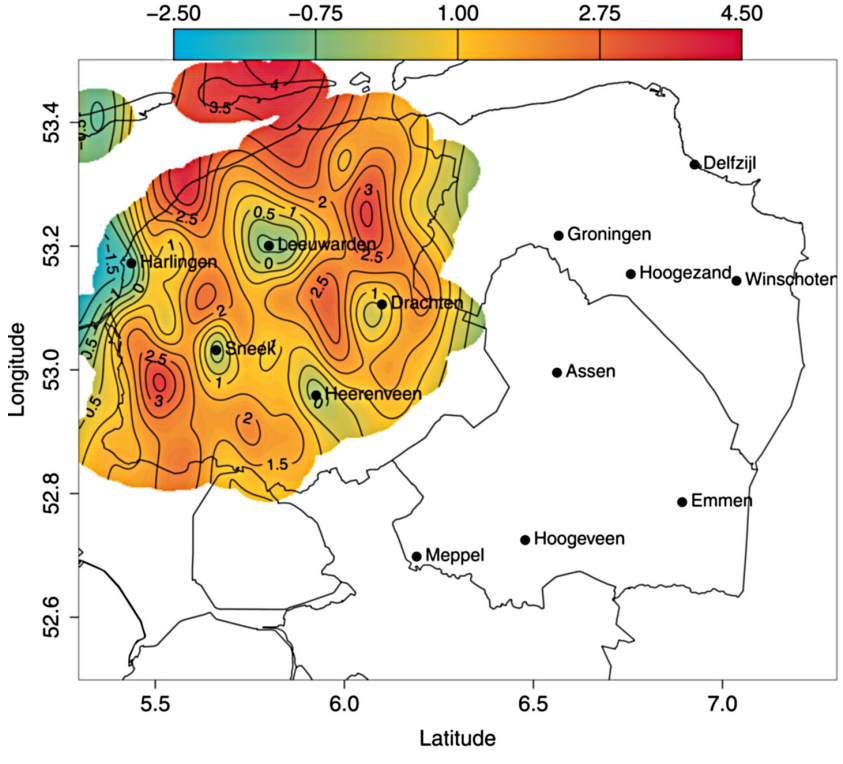


Figure A.1: Geographical distribution of the speakers who transmit Frisian to their children. Based on a logistic generalized additive model predicting the likelihood of intergenerational transmission based on a two-dimensional smooth from geographical coordinates. A value of -2, 0, or 2 indicates a transmission probability of respectively 12%, 50%, and 88%.

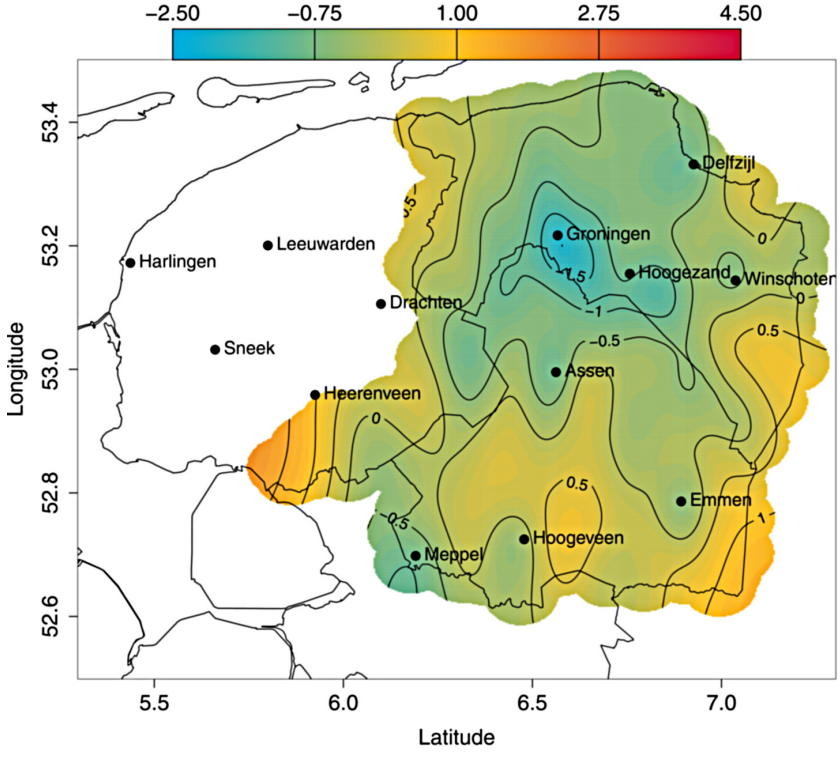


Figure A.2: Geographical distribution of the speakers who transmit Low Saxon to their children. Based on a logistic generalized additive model predicting the likelihood of intergenerational transmission based on a two-dimensional smooth from geographical coordinates. A value of -1, 0, or 1 indicates a transmission probability of respectively 27%, 50%, and 73%.

A

A.3 Estimates for the municipalities of Oost- and Weststellingwerf

We also added the estimates and adjustment components for the municipalities of Oost- and Weststellingwerf in Tables A.8 and A.9. The estimates are limited for these municipalities, as we only determined adjustment components (and resulting estimated percentages) when these could be based on data from more than 10 respondents. A cell is marked with an asterisk when there are insufficient data for a reliable estimate.

Table A.8: Adjustment components for Low Saxon speakers in the municipalities of Oost- and Weststellingwerf based on the Lifelines data.

Birth years of child generation	Intergenerational transmission from parents to children (in %)	Rate of acquisition through parents
1952 – 1972	44	*
1973 – 1993	49	83
1994 – 2015	*	68

Table A.9: Estimated percentages for speaking and using Low Saxon in the municipalities of Oost- and Weststellingwerf.

Birth years	Able to speak Low Saxon	Using Low Saxon and Dutch at home
1952 – 1972	*	*
1973 – 1993	37	31
1994 – 2015	*	*

APPENDIX B

REDUCING TRANSCRIBER VARIABILITY ADDENDUM

B.1 Phonetic symbols of the GTRP and RND corpora

Table B.1: Occurrence of all IPA symbols in the GTRP (Dutch: -NL, Flemish: -BE) and RND corpora. The check marks indicate in which subset an IPA symbol occurs.

IPA	GTRP-NL	GTRP-BE	RND
a	✓	✓	✓
b	✓	✓	✓
c	✓		
d	✓	✓	✓
e	✓	✓	✓
f	✓	✓	✓
h	✓		✓
i	✓	✓	✓
j	✓	✓	✓
k	✓	✓	✓
l	✓	✓	✓
m	✓	✓	✓
n	✓	✓	✓
o	✓	✓	✓
p	✓	✓	✓
r	✓	✓	✓
s	✓	✓	✓
t	✓	✓	✓
u	✓	✓	✓
v	✓	✓	✓
w	✓	✓	✓
x	✓	✓	✓
y	✓	✓	✓
z	✓	✓	✓
æ	✓	✓	✓
ç	✓		
ð	✓		
ø	✓	✓	✓

Continued on the next page.

IPA	GTRP-NL	GTRP-BE	RND
ħ	✓		
ŋ	✓	✓	✓
œ	✓	✓	✓
ɑ	✓	✓	✓
ɒ	✓		✓
ɔ	✓	✓	✓
ɕ	✓		
ə	✓	✓	✓
ɛ	✓	✓	✓
ɡ	✓		✓
ɢ		✓	
ɣ	✓	✓	✓
ɤ	✓		
ɥ	✓		
ɦ	✓	✓	
ɨ	✓		
ɪ	✓		✓
ɬ	✓	✓	
ɱ	✓		
ɲ	✓	✓	✓
ɵ	✓	✓	
ɶ	✓		
ɷ	✓		
ɸ	✓		
ɹ	✓		
ɻ	✓		
ɾ	✓		
ʁ	✓	✓	✓
ʁ	✓		
ʃ	✓	✓	✓
ʉ	✓	✓	
ʊ	✓		✓
ʋ	✓		
ʌ	✓		
ʍ	✓		
ʏ	✓		
ʏ	✓		✓
ʒ	✓	✓	✓
ʔ	✓	✓	✓

Continued on the next page.

IPA	GTRP-NL	GTRP-BE	RND
ɟ	✓		
β	✓		
χ	✓		

B.2 Using more reduced symbol inventories

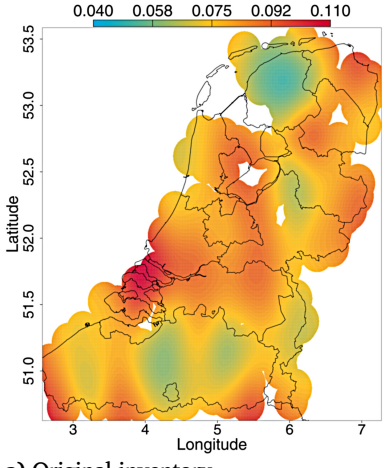
We used a frequency constraint as the final step in the procedure used to combine phonetic symbol inventories in Chapter 4. In this case, any phonetic symbols that occurred fewer than 1% of the times in the combined transcriptions were reduced. We can combine the symbol inventories of the RND and GTRP further to illustrate that the same general patterns are found with stricter frequency constraints, such as a threshold of 2.5% or 5%. The combined inventories for these analyses are presented in Tables B.2 and B.3, and the accompanying Figure B.1 shows the geographical distribution of pronunciation change with different frequency constraints.

Table B.2: Combined phonetic symbol inventory with a 2.5% frequency constraint (7 vowels, 13 consonants). The symbols that were merged due to low frequency are parenthesized.

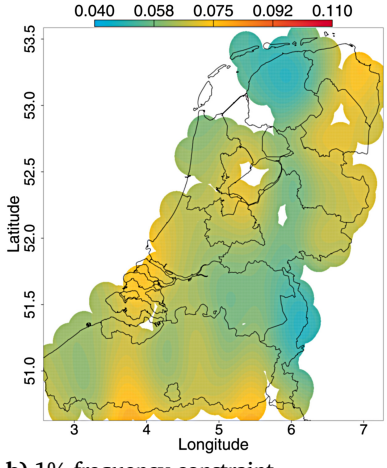
Used symbol	← Merged symbol(s)						
b	b	β	(p)				
d	d	ð					
e	e	ɪ					
f	f	ɸ	(v)				
i	i						
k	k	c	ɢ	(ʔ)			
l	l	ɬ	ʎ				
m	m						
n	n	ɱ	(ɲ)	(ŋ)			
r	r	ɹ	ɾ	ʁ	(ʀ)		
s	s	ʝ	ʈ	(ʃ)	(z)	(ʒ)	
t	t	θ					
u	u	ʉ	(y)	(o)	(ɷ)	(ø)	(œ)
w	w	ʌ	ʊ				
x	x	ç	χ	ɥ			
ɑ	ɑ	ɤ	(a)				
ɔ	ɔ	ʊ	(œ)	(ʏ)	(ʌ)	(ɵ)	
ə	ə	ɨ					
ɛ	ɛ	(æ)					
ɣ	ɣ	ħ	g	h	ɕ	(j)	(ɦ)

Table B.3: Combined phonetic symbol inventory with a 5% frequency constraint (4 vowels, 7 consonants). The symbols that were merged due to low frequency are parenthesized.

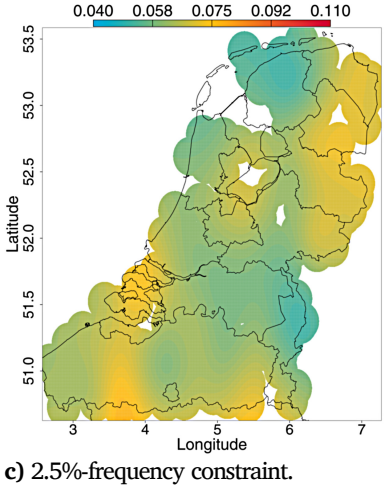
Used symbol	← Merged symbol(s)														
b	b	β	(p)	(m)	(l)	(ɬ)	(ʎ)								
f	f	ɸ	(v)	(w)	(ʍ)	(ʋ)									
i	i	(e)	(ɪ)	(ɛ)	(æ)										
n	n	ɱ	(ɲ)	(ŋ)											
r	r	ɹ	ɾ	ʁ	(ʀ)										
s	s	ʝ	ɽ	(ʃ)	(z)	(ʒ)	(k)	(c)	(ɢ)	(ʔ)					
t	t	θ	(d)	(ð)											
u															
ɔ	ɔ	ʊ	(œ)	(ʏ)	(ʌ)	(ɵ)	(ɑ)	(ɤ)	(a)	(u)	(ʉ)	(y)	(o)	(ɷ)	(ø)	(œ)
ə	ə	ɨ													
ɣ	ɣ	ħ	g	h	ɕ	(j)	(ɦ)	(x)	(ç)	(χ)	(ɥ)				



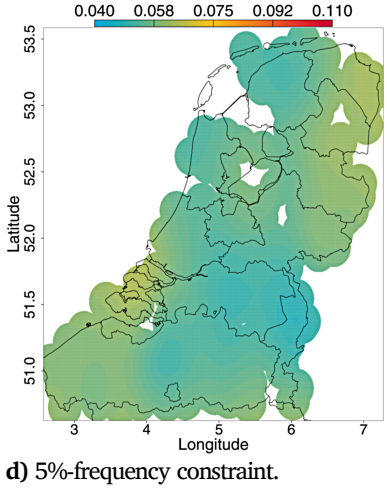
a) Original inventory.



b) 1%-frequency constraint.



c) 2.5%-frequency constraint.



d) 5%-frequency constraint.

Figure B.1: Proportions of pronunciation change predicted in the Netherlands and Flanders using different frequency constraints, based on a geographical smooth (using PMI-based weights). Red: more change. Blue: less change.

APPENDIX C

PRONUNCIATION CHANGE ESTIMATES ADDENDUM

C.1 Analysis using data from younger generation

In Chapter 5, we analyzed pronunciation change in the northern and eastern Netherlands using data from older men from the GTRP and DIAREG corpora. The DIAREG corpus also comprises transcriptions from younger women, to which we can apply the same analysis to assess whether we observe similar patterns.

The model specification for this analysis is the same as in Chapter 5, but now using the data from younger women. Both random intercept and slope were significant (p's < 0.01). The model summaries are presented in Table C.1 and the interaction effect is shown in Figure C.1. The observed pronunciation patterns are similar to those in Chapter 5, although there is no statistically significant difference between convergence and divergence for the non-Westphalian dialect group (including Frisian, Town Frisian, and Northern Low Saxon varieties).

Table C.1: Coefficients for a generalized additive model predicting change based on a binary distinction between the Westphalian Low Saxon (WLS) group and the other dialect groups (based on comparing data from GTRP speakers to younger women from the DIAREG). The direction of change is either convergence (conv.) or divergence (div.).

	Estimate	SE	z-value	p-value	
Intercept (non-WLS: conv.)	-3.734	0.051	-72.894	<0.001	***
non-WLS: div. vs. conv.	0.126	0.067	1.874	0.061	
non-WLS vs. WLS: conv.	0.132	0.057	2.319	0.021	*
non-WLS: div vs. WLS: conv.	-0.188	0.080	-2.337	0.196	*

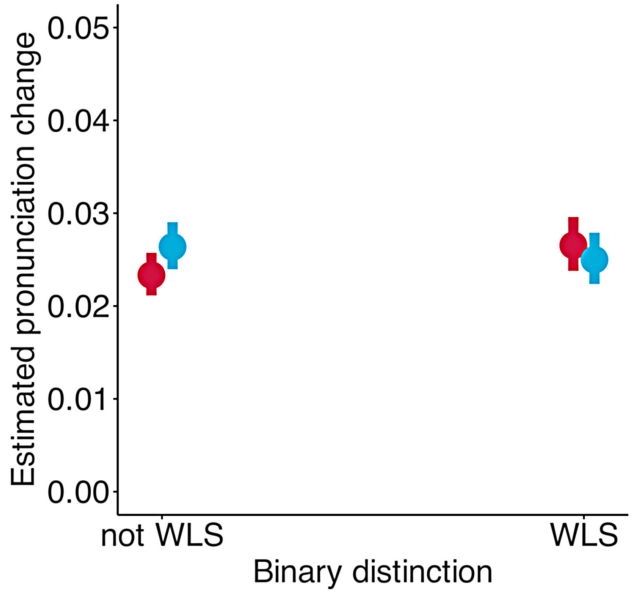


Figure C.1: Estimated marginal interaction effect between change direction a binary distinction between Westphalian Low Saxon (WLS) and the other regional varieties (based on comparing data from GTRP speakers to younger women from the DIAREG). Convergence is shown in red and divergence in blue. The *y*-axis has been transformed from logits into odds.

C.2 Simulation of different sample sizes

In Chapter 5, we stated that the total number of transcriptions used per recording location was limited. However, it is unclear what amount of data should be considered insufficient for dialectometric studies. We therefore simulated studies with different sample sizes to gain insight into what can be taken as a minimum sample size for this kind of study.

Using the GTRP corpus as a reference due to its substantial size, we drew random samples of the 562 target words available in the corpus, simulating studies using between 5 and 100 words (always including all 613 locations). We computed the correlation between the distance matrices obtained using all target words and the subset of the target words (using the PMI-based Levenshtein variant, as explained in Chapter 4). We repeated the process of drawing a random subset 1000 times for each sample size, and the corresponding distributions are plotted in Figure C.2.

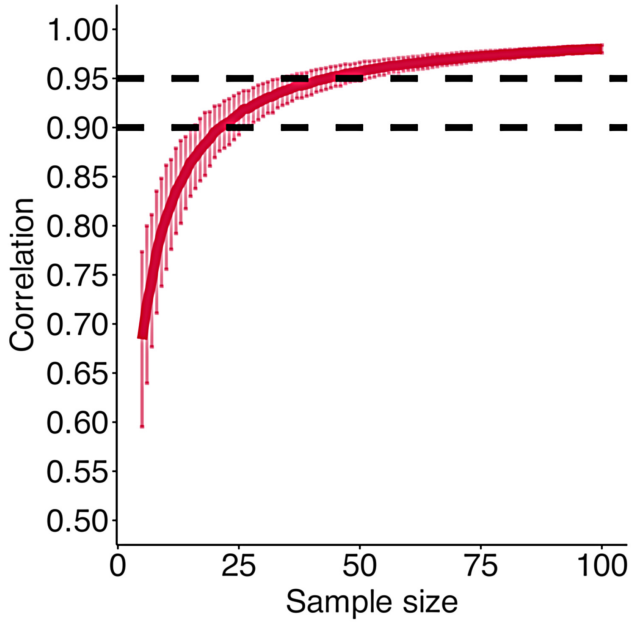


Figure C.2: Simulation of different target word sample sizes in dialectometric studies. The dashed lines indicate correlations of 0.90 and 0.95. Confidence intervals (at 95%) are added for each sample size.

The results are visualized in Figure C.2. Consistent correlations of over 0.90 are found from a sample size of approximately 25 target words, while around 45 words are necessary for consistent correlations of over 0.95. These findings indicate that the total number of target words used in the locations in Chapter 5 was generally sufficient for sampling a local dialect, and that dialects can be sampled reliably from as few as 25 target words. At the same time, the reliability can be increased further by recording translations of at least 45 target words with speakers from dialect locations, which we managed to do for the analyses in Chapter 6.

SUMMARIES IN ENGLISH AND DUTCH

Summary in English

The story of the Tower of Babel in the Book of Genesis, which attempts to explain why there are so many different languages in the world, suggests that people have long been interested in language variation. More recently, we have seen at least one language die every month. If no active countermeasures are taken, we can expect that about 80% of the world's languages will have disappeared by the end of this century. This language loss significantly reduces the world's cultural diversity and heritage and is primarily caused by placing certain languages above others. For example, official national languages are usually promoted over so-called minority languages.

In this dissertation, we focus on Frisian and Low Saxon. These languages have been spoken for centuries in what is now the northern and eastern Netherlands, but they have been subordinated to Dutch since the creation of the Kingdom in the early 19th century. They are recognized under the European Charter for Regional or Minority Languages by the Dutch government. However, Frisian is more extensively protected than Low Saxon and there is more financial support from the government for Frisian than Low Saxon. The Frisian-speaking population also appears more protective of its language, perhaps partly because of a strong shared identity, which the Low Saxon population seemingly lacks (possibly because they are spread over a larger geographical area).

By comparing these languages, we may gain insights that can be applied in similar majority-minority language contexts, especially if one regional language is doing better than others due to greater political protection or if the language benefits from a protective speaker population. In this dissertation, we assess the differences between Frisian and Low Saxon in two ways. In Part II we concern ourselves with the Frisian and Low Saxon speaker populations, focusing on the changing speaker population sizes and intergenerational transmission of these languages. In Part III we focus on language change as reflected by changing pronunciation patterns of Frisian and Low Saxon varieties, mainly under pressure from Dutch.

Speaker populations — Language counts for regional languages are scarce (especially for Low Saxon), so we derived a language count for Frisian and Low

Saxon, which is reported in Chapter 2. We distributed a regional language questionnaire through the Lifelines Biobank, a large-scale research project with a potential participant pool of over 130,000 people. We were able to reach many potential speakers (approximately 38,500 respondents) using this data source, although our geographical scope was consequently limited to the northern three provinces (Friesland, Groningen, and Drenthe).

When we inspected the sample it became clear that people who do not speak the regional language were underrepresented, so calculating the metrics of interest (i.e., how many people could speak the regional language and how many people use it at home) directly from the sample would be unreliable. We devised another method that relies on previous language counts that we assumed to be more representative, which reported language use and proficiency percentages for several generations. We consequently calculated estimates for the children of those generations by combining the prior language counts with the intergenerational transmission rates of these generations, which could be derived from the Lifelines data. Using this method, we estimated that about 41% of the population in the provinces of Groningen and Drenthe could speak Low Saxon in 2021, while about 17% used the regional language at home. For Frisian, we found that the percentages were around 62% and 48%, respectively. This means that Frisian is relatively much more in use in Fryslân than Low Saxon is in Groningen and Drenthe. We also found that the intergenerational transmission rate decreased substantially for Low Saxon (among parents born between 1922 and 1985), while it only slightly decreased for Frisian (for parents born between 1931 and 2007).

To determine which factors are associated with parental language transmission of Frisian and Low Saxon, we focused in Chapter 3 on those Lifelines respondents who reported using these regional languages. We found that parents were more likely to transmit Frisian or Low Saxon if the other involved caregiver (e.g., one's partner) spoke the same regional language, or if they reported using the language more often in different social contexts, which explained a substantial part of the variation between parents. The effects of the other associated factors in the final model were smaller, but still statistically significant. The language transmission rate was lower among parents with higher educational attainment in the case of Low Saxon (but not Frisian). The transmission rate was higher among parents who acquired their language at a young age and among parents living in either highly rural or highly urbanized areas. Finally, parents with a more positive attitude toward their regional language were more likely

to transmit their regional language. This effect was stronger for Low Saxon than for Frisian-speaking parents.

Pronunciation patterns — Previous studies found evidence that a substantial amount of lexical and pronunciation change in Frisian and Low Saxon varieties (and other regional varieties in the Netherlands) likely stems from Standard Dutch influence, mainly due to its institutionalized and protected role in Dutch society. This type of dialect leveling is commonly referred to as vertical convergence, distinguishing it from horizontal convergence (i.e., neighboring dialects becoming more similar). Vertical convergence, especially if it coincides with horizontal convergence, may lead to so-called 'regiolects'. These varieties often retain linguistic features from the standard language and traditional dialects, but they are usually less localized than traditional dialects. In this dissertation, we assess whether Frisian and Low Saxon varieties showed signs of Standard Dutch influence in their pronunciation several decades ago and whether this linguistic influence increased over time. We also investigate whether there is evidence of regiolect formation in the Frisian and Low Saxon areas (as has been evidenced for several other parts of the Netherlands and Flanders).

Dialectologists usually rely on phonetically transcribed dialect corpora, but phonetic transcriptions are often inconsistently transcribed by different people. In Chapter 4 we therefore devise a method to alleviate this common problem in dialect studies of pronunciation variation and change. We compare the *Reeks Nederlandse Dialectatlassen* (RND; 'Dutch Dialect Atlas Series') with the corpus of the Goeman-Taeldeman-Van Reenen project (GTRP), both of which cover pronunciation patterns of regional language variants in the Netherlands and Flanders. The RND corpus is an older corpus (recordings were made between 1923 and 1982) than the GTRP (recordings were made between 1979 and 2001, with most recordings collected in the period between 1985 and 1989). The GTRP transcriptions vary systematically between transcribers, which is reflected by the fact that about 70 symbols were used for the transcriptions for regional varieties in the Netherlands, while only about 40 symbols were used for the regional varieties in Flanders. Similarly, about 40 symbols were used for the RND transcriptions.

To reduce the impact of small transcription differences between GTRP transcribers on analyses of pronunciation change, we iteratively replaced phonetic symbols that were infrequent in the RND and GTRP corpora with phonetically similar but more frequently occurring symbols. This procedure yielded transcriptions that were better suited for comparison, because they reflected more

meaningful variation after grouping minor variations in phonetic space. We use an algorithm (the Levenshtein distance) to quantify how much two phonetic transcriptions differ from each other by counting how many binary symbol operations (specifically, insertions, deletions, or substitutions) are necessary to change one transcription into the other. When the transcriptions for these corpora were compared (i.e., RND and GTRP transcriptions for the same recording location and Dutch target word) using this algorithm, we found that the overall pronunciation change in Frisian varieties during the 20th century was limited and that Low Saxon varieties changed more substantially during this period. Nonetheless, this comparison was not ideal for answering the research questions regarding pronunciation change due to the considerable differences in recording years between the RND and the GTRP.

In Chapter 5 we focus on vertical convergence between Frisian and Low Saxon language varieties and Standard Dutch. We compare the GTRP with a newer corpus: the From Dialect to Regiolect project (DIAREG), which was also collected in a relatively short period (between 2008 and 2011). We extend the Levenshtein distance to concurrently process three transcriptions for each recording location and target word: the GTRP transcription, the DIAREG transcription, and the transcription for Standard Dutch (represented by a newsreader). A single transcriber made all relevant transcriptions to avoid inter-transcriber issues. We found that the Westphalian Low Saxon varieties (i.e., excluding the varieties in the province of Groningen and the northern part of Drenthe) showed more vertical convergence and less vertical divergence in this period than the other regional varieties. However, our analysis was again not ideal due to a relatively limited set of analyzable target words and differences between the GTRP and DIAREG tasks used for eliciting dialect forms.

We once more address the research questions regarding pronunciation change in Chapter 6, this time including both vertical and horizontal convergence. We follow the same methodology as in Chapter 5, but we use a new corpus designed to be as similar as possible to the GTRP. A subset of GTRP target words was presented to new speakers in the same locations as in Chapter 5, matching the GTRP reference speaker's age, gender, and location of growing up. Based on the maps resulting from our dialectometric analysis, we conclude that Frisian varieties became more similar over the intervening 35 years, indicating regiolect formation. Similar evidence of horizontal convergence was found for the Low Saxon varieties in the province of Groningen and the northern part of Drenthe, although this evidence was weaker than for the Frisian varieties. We saw no clear evidence of

regiolect formation in the other areas. In an in-depth analysis of vertical convergence and divergence, we also found that convergence toward Standard Dutch was stronger for relatively low-frequency and relatively high-frequency target words. We also observe more divergence from Standard Dutch for speakers with lower educational attainment, and we find that speakers with a strong regional identity showed less convergence toward Standard Dutch and more divergence away from it.

Conclusion — Summarizing the findings of Parts II and III, we can conclude that the situation of the Frisian language is more favorable than that of Low Saxon. Although these languages are both subordinated to Standard Dutch in the Netherlands, it is clear that Frisian is more actively maintained by its speakers and is more resistant to external linguistic pressure from Standard Dutch. The situation of Frisian seems strikingly fortunate in the context of European minority languages, almost all of which have declined significantly over the past century.

Pronunciation patterns of Low Saxon varieties appear consistently more influenced by Standard Dutch, although this effect is strongest in the more southern Low Saxon areas. Additionally, speaker characteristics (e.g., educational attainment and regional identity strength) interact with the variation observed in dialectometric studies, so it is advisable not to rely on only a single speaker when studying a local dialect. There is also evidence of regiolect formation in Frisian and the northern Low Saxon areas, but the extent to which they are actively perceived and used as such requires further investigation in future studies.

When we contrast the findings for Frisian and Low Saxon, it seems that the comparatively disadvantaged position of Low Saxon can be improved in several ways. Stimulating more positive attitudes toward the language may boost intergenerational transmission rates. Furthermore, fostering regional identities can contribute to a greater awareness of the language as a part of cultural heritage and a separate language from Dutch, which may lead to a greater willingness to either learn or maintain Low Saxon. These responsibilities are ideally carried by both the Low Saxon speaker population and the Dutch government, especially now there are still many speakers who can transfer the language.

Samenvatting in het Nederlands

Het verhaal over de Toren van Babel in het boek Genesis, dat probeert te verklaren waarom er zoveel verschillende talen zijn in de wereld, suggereert dat mensen al heel lang geïnteresseerd zijn in taalvariatie. Meer recentelijk hebben we elke maand minstens één taal zien uitsterven. Als er geen actieve tegenmaatregelen genomen worden, kunnen we verwachten dat ongeveer 80% van de talen in de wereld verdwenen zullen zijn tegen het einde van deze eeuw. Het verlies van deze talen, dat de culturele diversiteit en het culturele erfgoed van de wereld aanzienlijk vermindert, wordt voornamelijk veroorzaakt doordat bepaalde talen hiërarchisch boven andere talen geplaatst worden. Meestal worden officiële nationale talen bijvoorbeeld boven zogenaamde minderheidstalen geplaatst.

In dit proefschrift richten we ons op het Fries en het Nedersaksisch. Deze talen worden al eeuwenlang gesproken in wat nu Noord- en Oost-Nederland is, maar ze zijn ondergeschikt gemaakt aan het Nederlands sinds de stichting van het Koninkrijk in de vroege negentiende eeuw. Beide talen zijn door de Nederlandse overheid erkend onder het Europees Handvest voor regionale talen of talen van minderheden. Het Fries is echter uitgebreider beschermd dan het Nedersaksisch en er is meer financiële ondersteuning vanuit de overheid voor het Fries dan voor het Nedersaksisch. De Friestalige bevolking lijkt ook meer beschermingsgezind over haar taal, misschien gedeeltelijk vanwege een sterke gedeelde identiteit, die ogenschijnlijk ontbreekt onder de Nedersaksische bevolking (mogelijk vanwege haar verspreiding over een groter geografisch gebied).

Door deze talen te vergelijken kunnen we inzichten verwerven die toegepast kunnen worden in vergelijkbare situaties waarin een meerderheid van de bevolking een minderheidstaal spreekt, vooral als de ene streektaal het beter doet dan de andere vanwege betere politieke bescherming of als de taal profiteert van een beschermende sprekerspopulatie. In dit proefschrift bekijken we de verschillen tussen het Fries en het Nedersaksisch op twee manieren. In Deel II houden we ons bezig met de Friese en Nedersaksische sprekerspopulaties, waarbij we ons richten op het veranderende aantal sprekers en de intergenerationele overdracht van deze talen. In Deel III richten we ons op taalverandering zoals weerspiegeld in veranderende uitspraakpatronen van varianten van het Fries en Nedersaksisch, voornamelijk onder druk van het Nederlands.

Sprekerspopulaties — Taaltellingen voor streektalen zijn schaars (in het bijzonder voor het Nedersaskisch), dus we hebben een taaltelling gedaan voor het Fries en Nedersaksisch, waarover we rapporteren in Hoofdstuk 2. We hebben een streektaalvragenlijst verspreid via een grootschalig onderzoeksproject met een potentiële respondentenpool van meer dan 130.000 mensen: de *Lifelines*-biobank. Dankzij deze gegevensbron konden we veel potentiële sprekers bereiken (ongeveer 38.500 mensen), al werd onze geografische reikwijdte hierdoor wel beperkt tot de drie noordelijke provincies (Fryslân, Groningen en Drenthe).

Toen we de steekproef inspecteerden, werd het duidelijk dat mensen die geen streektaal spreken ondervertegenwoordigd waren. Het direct op basis van de steekprof berekenen van de statistieken waarin we geïnteresseerd waren (namelijk hoeveel mensen de streektaal konden spreken en hoeveel mensen die thuis gebruikten) zou daarom onbetrouwbaar zijn. We bedachten een andere methode die uitgaat van eerdere taaltellingen (waarvan we aannemen dat ze representatiever zijn), die deze statistieken rapporteerden voor verschillende generaties. We hebben vervolgens schattingen berekend voor de kinderen van deze generaties door de eerdere taaltellingen te combineren met de intergenerationele taaloverdracht van deze generaties, die we konden afleiden uit de Lifelines-data. Aan de hand van deze methode schatten we voor 2021 dat ongeveer 41% van de bevolking in Groningen en Drenthe Nedersaksisch kon spreken, terwijl ongeveer 17% de streektaal thuis gebruikte. Voor het Fries ontdekten we dat de percentages respectievelijk rond de 62% en 48% lagen, dus het Fries is relatief aanzienlijk meer in gebruik in Fryslân dan het Nedersaksisch in Groningen en Drenthe. Ook zagen we dat de mate van intergenerationele overdracht substantieel afnam voor het Nedersaksisch, terwijl dit slechts licht afnam voor het Fries.

Om te bepalen welke factoren geassocieerd zijn met de taaloverdracht van het Fries en het Nedersaksisch van ouders op kinderen, hebben we ons in Hoofdstuk 3 specifiek gericht op de Lifelines-respondenten die aangaven deze streektalen te gebruiken. We ontdekten dat ouders eerder het Fries of het Nedersaksisch overdroegen als de andere betrokken verzorger (bijvoorbeeld iemands partner) dezelfde streektaal sprak of wanneer ze aangaven de taal vaker te gebruiken in verschillende sociale contexten. Dit verklaarde een substantieel deel van de variatie tussen ouders. De effecten van de andere geassocieerde factoren in het uiteindelijke model waren kleiner, maar nog steeds statistisch significant. De mate van taaloverdracht was lager onder ouders met een hoger opleidingsniveau in het geval van het Nedersaksisch (en niet het Fries). De mate van taaloverdracht

was hoger onder ouders die hun taal op jonge leeftijd hadden verworven en ouders die in zeer landelijke of juist zeer verstedelijkte gebieden woonden. Tot slot gaven ouders met een positievere houding over hun streektaal eerder hun streektaal door. Dit laatste effect was sterker voor Nedersaksischtalige dan voor Friestalige ouders.

Uitspraakpatronen — Uit eerder onderzoek is gebleken dat een aanzienlijke hoeveelheid van de lexicale verandering en uitspraakverandering in Friese en Nedersaksische taalvarianten (en andere streektaalvarianten in Nederland) waarschijnlijk veroorzaakt wordt door invloed van het Standaardnederlands, voornamelijk vanwege zijn geïnstitutionaliseerde en beschermde rol in de Nederlandse samenleving. Dit type dialectnivellering wordt doorgaans verticale convergentie genoemd, ter onderscheid van horizontale convergentie (waarbij aangrenzende dialecten meer op elkaar gaan lijken). Verticale convergentie, vooral als het samen voorkomt met horizontale convergentie, kan leiden tot zogenaamde 'regiolecten'. Deze taalvarianten behouden vaak kenmerken van de standaardtaal en traditionele dialecten, maar ze zijn meestal minder lokaal gebonden dan traditionele dialecten. In dit proefschrift kijken we of Friese en Nedersaksische taalvarianten enkele decennia tekenen van Nederlandse invloed lieten zien in hun uitspraakpatronen of of deze invloed is toegenomen over de tijd heen. We onderzoeken ook of er bewijs is voor regiolectvorming in de Friese en Nedersaksische gebieden (zoals is aangetoond voor verschillende andere delen van Nederland en Vlaanderen).

Dialectologen vertrouwen vaak op fonetisch getranscribeerde dialectcorpora, maar fonetische transcripties worden vaak inconsistent getranscribeerd door verschillende mensen. In Hoofdstuk 4 bedenken we daarom een methode om dit veelvoorkomende probleem in dialectonderzoek over uitspraakvariatie en -verandering te verlichten. We vergelijken de Reeks Nederlandse Dialectatlassen (RND) met het corpus van het Goeman-Taeldeman-Van Reenen-project (GTRP), waarin in beide gevallen uitspraakpatronen van streektaalvarianten in Nederland en Vlaanderen zijn opgenomen. Het RND corpus is een ouder corpus (waarvoor de opnames tussen 1923 en 1982 gemaakt werden) dan de GTRP (waarvoor de opnames tussen 1979 en 2001 gemaakt werden, maar vooral tussen 1985 en 1989). De GTRP transcripties variëren systematisch tussen de transcribenten, wat gereflecteerd is in het feit dat ongeveer 70 symbolen gebruikt werden voor de transcripties voor streektaalvarianten in Nederland, terwijl maar ongeveer 40 symbolen werden gebruikt voor de streektaalvarianten in Vlaanderen. Tevens

werden er ongeveer 40 symbolen gebruikt voor de RND transcripties.

Om de impact van kleine transcriptieverschillen tussen GTRP-transcribenten te verminderen vervingen we iteratief fonetische symbolen die weinig voorkwamen in de RND- en GTRP-corpora door fonetisch vergelijkbare symbolen die vaker voorkwamen. Deze procedure leverde transcripties op die beter geschikt waren om te vergelijken, omdat ze meer betekenisvolle variatie reflecteerden na het groeperen van kleine variatie in de fonetische ruimte. We gebruiken een algoritme (de Levenshtein-afstand) om te kwantificeren hoeveel twee fonetische transcripties van elkaar verschillen door te tellen hoeveel binaire symboolaanpassingen (specifiek inserties, deleties of substituties) er nodig zijn om de ene transcriptie in de andere te veranderen. Door de transcripties voor deze corpora te vergelijken met dit algoritme (namelijk RND- en GTRP-transcripties voor dezelfde opnameplaats en voor hetzelfde Nederlandse doelwoord) ontdekten we dat de totale uitspraakverandering in Friese varianten in de twintigste eeuw beperkt was en dat Nedersaksische varianten in deze periode substantiëler veranderden. Deze vergelijking was echter niet ideaal voor het beantwoorden van de onderzoeksvragen met betrekking tot uitspraakverandering doordat de verschillen in opnamejaren tussen de RND en de GTRP aanzienlijk verschilden.

In Hoofdstuk 5 richten we ons op verticale convergentie tussen Friese en Nedersaksische taalvarianten enerzijds en het Standaardnederlands anderzijds. We vergelijken de GTRP met een nieuwer corpus: het *From Dialect to Regiolect*-project (DIAREG), dat ook in een relatief korte periode verzameld is (namelijk tussen 2008 en 2011). We breiden de Levenshtein-afstand uit om telkens tegelijkertijd drie transcripties te verwerken voor elke opnamelocatie en elk doelwoord: de GTRP-transcriptie, de DIAREG-transcriptie en de transcriptie voor het Standaardnederlands (vertegenwoordigd door een nieuwslezer). Alle relevante transcripties werden gemaakt door een enkele transcribent om problemen ten gevolge van transcribentverschillen te voorkomen. We ontdekten dat de Westfaalse Nedersaksische varianten (d.w.z. de varianten van het Nedersaksisch uitgezonderd die in Groningen en het noordelijke deel van Drenthe) meer verticale convergentie en minder verticale divergentie vertoonden in deze periode dan de andere streektaalvarianten. Onze analyse was echter opnieuw niet ideaal door de relatief beperkte set analyseerbare doelwoorden en de verschillen tussen de GTRP- en DIAREG-taken die gebruikt werden om dialectvormen te eliciteren.

We behandelen de onderzoeksvragen met betrekking tot uitspraakverandering opnieuw in Hoofdstuk 6, deze keer inclusief verticale en horizontale convergentie. We volgen dezelfde methodologie als in Hoofdstuk 5, maar we gebruiken

een nieuw corpus dat zoveel mogelijk overeenkomt met de GTRP. Een subset van de GTRP-doelwoorden werd aan sprekers gepresenteerd in dezelfde opnamelocaties, waarbij we ervoor zorgden dat de leeftijd, het geslacht en de opgroeilocatie van de referentiespreker in de GTRP overeenkwam. Op basis van de kaarten waarop we onze dialectometrische analyse hebben gevisualiseerd, concludeerden we dat varianten van het Fries meer op elkaar zijn gaan lijken in de 35 tussenliggende jaren, wat duidt op regiolectvorming. Vergelijkbaar bewijs voor horizontale convergentie werd ook gevonden voor de Nedersaksische varianten in Groningen en Noord-Drenthe, hoewel dit bewijs minder sterk was dan voor voor de Friese varianten. We zagen geen duidelijk bewijs voor regiolectvorming in de andere gebieden. In een verdiepende analyse van verticale convergentie en divergentie ontdekten we ook dat convergentie naar het Standaardnederlands sterker was voor relatief laagfrequente en relatief hoogfrequente doelwoorden. We observeerden ook meer divergentie weg van het Standaardnederlands voor sprekers met een lager opleidingsniveau, en we ontdekten dat sprekers met een sterke regionale identiteit minder convergentie naar het Standaardnederlands vertoonden en meer divergentie ervandaan.

Conclusie — Als we de bevindingen van Deel II en III samenvatten, kunnen we concluderen dat de situatie van de Friese taal gunstiger is dan die van het Nedersaksisch. Hoewel deze talen beide hiërarchisch ondergeschikt zijn aan het Standaardnederlands in Nederland, is het duidelijk dat het Fries actiever wordt onderhouden door de sprekers ervan en beter bestand is tegen externe talige druk van het Standaardnederlands. Het succes van de Friese situatie valt op in de context van Europese minderheidstalen, die in de afgelopen eeuw bijna allemaal sterk in verval geraakt zijn.

De uitspraakpatronen van Nedersaksische variëteiten lijken consistent meer onder invloed van het Standaardnederlands te staan, hoewel dit effect sterker is in de meer zuidelijke Nedersakische gebieden. Verder interacteren kenmerken van de spreker (zoals opleidingsniveau en sterkte van regionale identiteit) met de variatie die we zien in dialectometrisch onderzoek. Het is daarom raadzaam om niet te vertrouwen op een enkele spreker als men een lokaal dialect wil onderzoeken. Er is ook bewijs voor regiolectvorming in Friese en noordelijke Nedersaksische gebieden, maar de mate waarin deze actief zo wordt waargenomen moet in de toekomst verder onderzocht worden.

Als we de bevindingen voor het Fries en Nedersaksisch contrasteren, lijkt het erop dat de relatief achtergestelde positie van het Nedersaksisch op verschil-

lende manieren verbeterd kan worden. De intergenerationele overdracht kan versterkt worden door een positievere taalhouding te stimuleren. Verder kan het bevorderen van regionale identiteit bijdragen aan een groter bewustzijn van de taal als cultureel erfgoed en als een taal afzonderlijk van het Nederlands, wat kan leiden tot een grotere bereidwilligheid om het Nedersaksisch te leren of te onderhouden. Deze verantwoordelijkheden worden idealiter gedragen door zowel de Nedersaksische sprekerspopulatie en de Nederlandse overheid, zeker nu er nog veel sprekers zijn die de taal door kunnen geven.

ABOUT THE AUTHOR

Curriculum vitae

Raoul Buurke was born on the 3rd of April, 1995 in Groningen, Surinam. At the age of three, he was adopted by Dutch parents and has since resided in Groningen, the Netherlands. He graduated from the Willem Lodewijk Gymnasium (a high school offering pre-university education) in 2012 and had a broad interest in the humanities. He initially studied sociology, but switched to artificial intelligence after a year, and again switched a final time to linguistics after another year. Having developed an interest in statistics during his sociology studies, he was particularly interested in the linguistic fields that quantify language variation and change. Raoul obtained a Bachelor's degree and research Master's degree (with distinction) in linguistics, after which he pursued a PhD in computational linguistics.

In addition to his studies and research, Raoul had the opportunity to engage in science communication. His additions to the Dutch teaching method *Kern Nederlands* for high schools were a first stint in this direction, and there were many more opportunities to tell a broader public about regional language variation in the Netherlands after joining Speech Lab Groningen. He contributed to the Groningen dialect teaching program *Van Old noar Jong* 'From Old to Young', and he gave several guest lectures about Frisian and Low Saxon developments at other institutes or for other interested groups. He also developed a program called *Klankkaleidoscoop* 'Sound kaleidoscope', which roughly guesses where someone is from in the Netherlands based on their dialectal translations of single words (using automatic segmentation and quantifying pronunciation differences using machine learning methods). The most tangible result of Raoul's science communication efforts is the board game *Streektaalstrijd*, developed with colleagues from Speech Lab Groningen (and with support from the entire Netherlandic language area), which has been widely sold and appreciated.

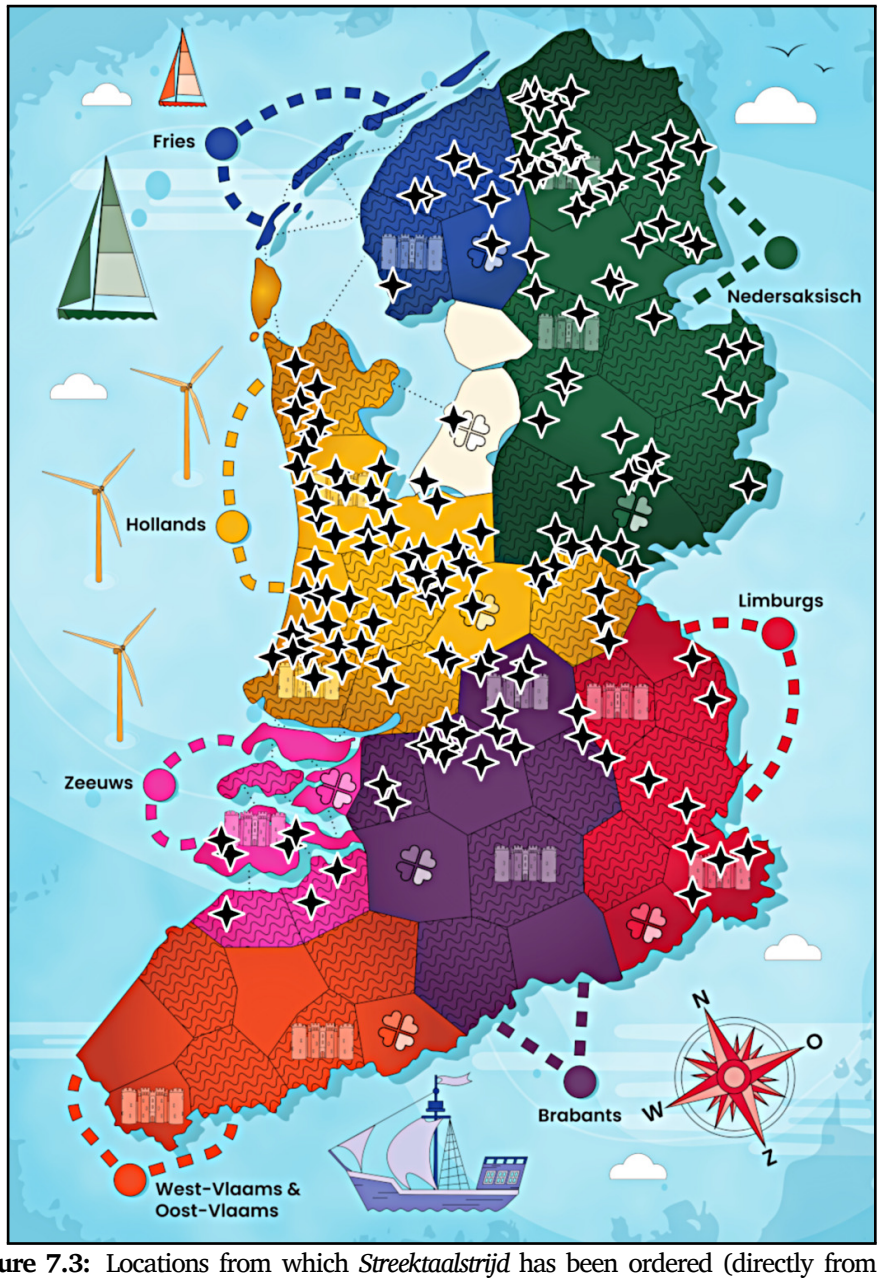


Figure 7.3: Locations from which *Streektaalstrijd* has been ordered (directly from the University Shop), as updated until the 22nd of March, 2024.

Publications

As first author:

Buurke, R. S. S. J., Sekeres, H. G., Heeringa, W., Knooihuizen, R., & Wieling, M. (2022). Estimating the level and direction of aggregated sound change of dialects in the northern Netherlands. *Taal en Tongval*, *74*(2), 183–214. https://doi.org/10.5117/TET2022.2.002.BUUR

Buurke, R. S. S. J., & Wieling, M. (2023). Sound Change Estimation in Netherlandic Regional Languages: Reducing Inter-Transcriber Variability in Dialect Corpora. *Taal en Tongval*, *75*(1), 7–28. https://doi.org/10.5117/TET2023.1.002.BUUR

Buurke, R., Sekeres, H., Visser, L., & Wieling, M. (2024f). *Streektaalstrijd: Een strategisch kennisspel over taal en dialect* (Tweede druk). Wink Games, Dongen

Buurke, R., Bartelds, M., Knooihuizen, R., & Wieling, M. (2024b). Multiple estimates of the Frisian and Low Saxon speaker population size in the Netherlands. *Linguistic Minorities in Europe Online*. https://doi.org/10.1515/lme.28672125

Buurke, R., Bartelds, M., Heeringa, W., Knooihuizen, R., & Wieling, M. (2024a). Intergenerational language transmission of Frisian and Low Saxon in the Netherlands. *Journal of Language and Social Psychology*, *43*(5-6), 1–18. https://doi.org/10.1177/0261927X241287765

Buurke, R., Heeringa, W., Knooihuizen, R., & Wieling, M. (2024d). Recent changes in pronunciation patterns within and between Frisian, Town Frisian, and Low Saxon varieties [Submitted.]

Buurke, R., Doreleijers, K., Farasyn, M., Hilte, L., Swanenberg, J., & Voeten, C. (Eds.). (2024c). Micro- and macrovariation: Diverse approaches in research on language variation and change [Special issue. In press.]. *Taal en Tongval*, *76*(2)

Co-authored publications:

Buurke, R. (2020b). Taaltechnologie. In *Kern Nederlands Taal & Cultuur* (pp. 46–49). Koninklijke Boom uitgevers

Brouwer, J., Buurke, R., Van den Berg, F., Knooihuizen, R., Loerts, H., Bartelds, M., Wieling, M., & Keijzer, M. (2024). Minority language happiness: The link between social inclusion, well-being, and speaking a regional language in the northern Netherlands. *Ampersand*, *12*(100173). https:

//doi.org/10.1016/j.amper.2024.100173

Tienkamp, T. B., Rebernik, T., Buurke, R., Polsterer, K., Van Son, R., Wieling, M., Witjes, M. J., De Visscher, S., & Abur, D. (2024). The Effect of Speaking Style on the Articulatory-Acoustic Vowel Space in Individuals with Tongue Cancer before and after Surgical Treatment. *Proceedings of the 13th International Seminar on Speech Production*, 65–68

Rebernik, T., Jacobi, J., Buurke, R., Tienkamp, T., Tsiwah, F., Oud, S., Reinders, A., Abur, D., & Wieling, M. (2024). SPRAAKLAB Mobile Laboratory for Acoustic and Articulatory Data Collection: Technical Specifications and Practical Experiences [Submitted.]

BIBLIOGRAPHY

Aitchison, J. (2001). *Language Change: Progress or Decay?* Cambridge University Press.

Amorosa, H., Von Benda, U., Wagner, E., & Keck, A. (1985). Transcribing phonetic detail in the speech of unintelligible children: A comparison of procedures. *International Journal of Language & Communication Disorders*, *20*(3), 281–287. https://doi.org/10.3109/13682828509012268

Ashby, W. (2001). Un nouveau regard sur la chute du *ne* en français parlé tourangeau: S'agit-il d'un changement en cours? *Journal of French Language Studies*, *11*(1), 1–22. https : / / doi . org / 10 . 1017 / S0959269501000114

Auer, P. (2005). Europe's sociolinguistic unity, or: A typology of European dialect/standard constellations. In N. Delbecque, J. Van der Auwera, & D. Geeraerts (Eds.), *Perspectives on Variation* (pp. 7–42). De Gruyter Mouton. https://doi.org/10.1515/9783110909579.7

Auer, P. (2018). Dialect change in Europe–Leveling and convergence. In C. Boberg, J. Nerbonne, & D. Watt (Eds.), *The Handbook of Dialectology* (pp. 159–176). Wiley Online Library.

Auer, P., Hinskens, F., & Kerswill, P. E. (2005). *Dialect Change: Convergence and Divergence in European Languages*. Cambridge University Press.

Austin, P., & Sallabank, J. (Eds.). (2011). *The Cambridge Handbook of Endangered Languages*. Cambridge University Press.

Azurmendi, M.-J., Bachoc, E., & Zabaleta, F. (2001). Reversing Language Shift: The Case of Basque. In J. A. Fishman (Ed.), *Can Threatened Languages be Saved?* (pp. 234–259). Multilingual Matters. https://doi.org/10.21832/9781853597060-012

Bakker, F., & Van Hout, R. (2017). De indeling van de dialecten in Noord-Limburg en het aangrenzende Duitse gebied. *Nederlandse Taalkunde*, *22*(3), 303–332. https://doi.org/https://doi.org/10.5117/NEDTAA2017.3.BAKK

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

Bartelds, M., De Vries, W., Sanal, F., Richter, C., Liberman, M., & Wieling, M. (2022). Neural representations for modeling variation in speech. *Journal of Phonetics*, *92*, 101137. https://doi.org/10.1016/j.wocn.2022.101137

Bartelds, M., & Wieling, M. (2022). Quantifying Language Variation Acoustically with Few Resources. *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 3735–3741. https://doi.org/10.18653/v1/2022.naacl-main.273

Bayat, Z., Kircher, R., & Van de Velde, H. (2023). Minority language rights to education in international, regional, and domestic regulations and practices: The case of Frisian in the Netherlands. *Current Issues in Language Planning*, *24*(1), 81–101. https://doi.org/10.1080/14664208.2022.2037291

Beaman, K. V. (2021). *Coherence in Real- and Apparent-Time: A sociolinguistic variationist investigation of language change in Swabia* [PhD Thesis]. University of London.

Bek, P. (2022). *No bicycle, no bus, no job: The making of workers' mobility in the Netherlands, 1920-1990*. Amsterdam University Press.

Bell, J. (2013). Language attitudes and language revival/survival. *Journal of Multilingual and Multicultural Development*, *34*(4), 399–410. https://doi.org/10.1080/01434632.2013.794812

Belmar, G., & Pinho, S. (2020). 'Kinsto it Frysk ferstean?': Intelligibility of West Frisian for Dutch native speakers. *Dutch Journal of Applied Linguistics*, *9*(1-2), 109–131. https://doi.org/10.1075/dujal.19034.bel

Berns, J., & Steusel, S. (2004). *Taal in stad en land: Noord-Hollands*. Sdu Uitgevers.

Bethlehem, J. G. (2009). *Applied survey methods: A statistical perspective*. Wiley.

Bialystok, E. (2011). Reshaping the mind: The benefits of bilingualism. *Canadian Journal of Experimental Psychology*, *65*(4), 229–235. https://doi.org/10.1037/a0025406

Blancquaert, E., & Pée, W. (1930). *Reeks Nederlandse Dialectatlassen*. De Sikkel.

Bloemhoff, H. (2005). *Taaltelling Nedersaksisch: Een enquête naar het gebruik en de beheersing van het Nedersaksisch in Nederland*. Nedersaksisch Instituut, Rijksuniversiteit Groningen.

Bloemhoff, H. (2008a). Stellingwerf. In *Handboek Nedersaksische Taal- en Letterkunde* (pp. 175–193). Koninklijke Van Gorcum.

Bloemhoff, H. (2008b). Taalsociologische aspecten. In *Handboek Nedersaksische Taal- en Letterkunde* (pp. 294–323). Koninklijke Van Gorcum.

Bloemhoff, H., Bloemhoff-de Bruijn, P., Nijen Twilhaar, J., Nijkeuter, H., & Scholtmeijer, H. (2019). Nedersaksisch in een notendop. Inleiding in de Nedersakische taal en literatuur.

Bloemhoff, H., Bloemhoff-de Bruijn, P., Nijen Twilhaar, J., Nijkeuter, H., & Scholtmeijer, H. (2020). *Introduction to Dutch Low Saxon: Language and literature*. Koninklijke Van Gorcum.

Bloemhoff, H., De Haan, G., & Versloot, A. (2010). Language varieties in the province of Fryslân. In F. Hinskens & J. Taeldeman (Eds.), *Language and Space: Dutch* (pp. 721–738). De Gruyter Mouton.

Bloemhoff, H., Niebaum, H., & Kocks, G. (2008a). Drenthe. In *Handboek Nedersaksische Taal- en Letterkunde* (pp. 194–220). Koninklijke Van Gorcum.

Bloemhoff, H., Niebaum, H., Nijen Twilhaar, J., & Scholtmeijer, H. (2013a). Low Saxon phonology. In F. Hinskens & J. Taeldeman (Eds.), *Dutch* (pp. 454–475). Walter de Gruyter.

Bloemhoff, H., Niebaum, H., Nijen Twilhaar, J., & Scholtmeijer, H. (2013b). Recent evolutions in the position and structure of Low Saxon. With an excursus on the new polders in Flevoland. In F. Hinskens & J. Taeldeman (Eds.), *Dutch* (pp. 495–511). Walter de Gruyter.

Bloemhoff, H., Van der Kooi, J., Niebaum, H., & Reker, S. (2008b). *Handboek Nedersaksische Taal- en Letterkunde*. Koninklijke Van Gorcum.

Blom, E., Boerma, T., Bosma, E., Cornips, L., & Everaert, E. (2017). Cognitive advantages of bilingual children in different sociolinguistic contexts. *Frontiers in Psychology*, *8*, 552. https://doi.org/10.3389/fpsyg.2017.00552

Blondeau, H. (2001). Real-time changes in the paradigm of personal pronouns in Montreal French. *Journal of Sociolinguistics*, *5*(4), 453–474. https://doi.org/10.1111/1467-9481.00160

Boberg, C., Nerbonne, J., & Watt, D. (Eds.). (2018). *The Handbook Of Dialectology*.

Bourhis, R. Y., Giles, H., & Tajfel, H. (1973). Language as a determinant of Welsh identity. *European Journal of Social Psychology*, *3*(4), 447–460. https://doi.org/10.1002/ejsp.2420030407

Braat, E. (2020). *Low Saxon in the Netherlands: Efforts put into protecting, promoting and stimulating the language after signing covenant 'Nedersaksisch'* [Master's thesis, Tilburg University].

Breder Birkenes, M., & Pheiff, J. (2022). The Classification of Dutch Dialects Revisited: A Dialectometric Approach to the "Kleiner Niederländischer

Sprachatlas" (KNSA). *Zeitschrift für Dialektologie und Linguistik*, *89*(2-3), 153–199. https://doi.org/10.25162/zdl-2022-0006

Bredin, H. (2023). Pyannote. audio 2.1 speaker diarization pipeline: Principle, benchmark, and recipe. *INTERSPEECH 2023*, 1983–1987. https://doi.org/10.21437/Interspeech.2023-105

Bremmer, R. H. (2008). North-Sea Germanic at the Cross-Roads: The Emergence of Frisian and Hollandish. *NOWELE. North-Western European Language Evolution*, *54-55*, 279–308. https://doi.org/10.1075/nowele.54-55.09bre

Bromham, L., Dinnage, R., Skirgård, H., Ritchie, A., Cardillo, M., Meakins, F., Greenhill, S., & Hua, X. (2021). Global predictors of language endangerment and the future of linguistic diversity. *Nature Ecology & Evolution*, *6*(2), 163–173. https://doi.org/10.1038/s41559-021-01604-y

Brouwer, J., Buurke, R., Van den Berg, F., Knooihuizen, R., Loerts, H., Bartelds, M., Wieling, M., & Keijzer, M. (2024). Minority language happiness: The link between social inclusion, well-being, and speaking a regional language in the northern Netherlands. *Ampersand*, *12*(100173). https://doi.org/10.1016/j.amper.2024.100173

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The Word Frequency Effect in Word Processing: An Updated Review. *Current Directions in Psychological Science*, *27*(1), 45–50. https://doi.org/10.1177/0963721417727521

Bubenik, V. (1993). Dialect contact and koineization: The case of Hellenistic Greek. *International Journal of the Sociology of Language*, *1993*(99), 9–24. https://doi.org/10.1515/ijsl.1993.99.9

Burke, P. (2004). *Languages and communities in early modern Europe*. Cambridge University Press.

Buurke, R. (2020a). *Dialects across time and space: Computational modeling of dialects in the Netherlandic language area* [Master's thesis, University of Groningen].

Buurke, R. (2020b). Taaltechnologie. In *Kern Nederlands Taal & Cultuur* (pp. 46–49). Koninklijke Boom uitgevers.

Buurke, R., Bartelds, M., Heeringa, W., Knooihuizen, R., & Wieling, M. (2024a). Intergenerational language transmission of Frisian and Low Saxon in the Netherlands. *Journal of Language and Social Psychology*, *43*(5-6), 1–18. https://doi.org/10.1177/0261927X241287765

Buurke, R., Bartelds, M., Knooihuizen, R., & Wieling, M. (2024b). Multiple estimates of the Frisian and Low Saxon speaker population size in the Ne-

therlands. *Linguistic Minorities in Europe Online*. https://doi.org/10.1515/lme.28672125

Buurke, R., Doreleijers, K., Farasyn, M., Hilte, L., Swanenberg, J., & Voeten, C. (Eds.). (2024c). Micro- and macrovariation: Diverse approaches in research on language variation and change. *Taal en Tongval*, *76*(2).

Buurke, R., Heeringa, W., Knooihuizen, R., & Wieling, M. (2024d). Recent changes in pronunciation patterns within and between Frisian, Town Frisian, and Low Saxon varieties.

Buurke, R., Knooihuizen, R., Heeringa, W., & Wieling, M. (2024e). Measuring sound differences using automatic quantitative procedures: Applied to Standard Dutch, Frisian, and Low Saxon [Presented at Methods in Dialectology XVIII.].

Buurke, R., Sekeres, H., Visser, L., & Wieling, M. (2024f). *Streektaalstrijd: Een strategisch kennisspel over taal en dialect* (Tweede druk). Wink Games, Dongen.

Buurke, R. S. S. J., Sekeres, H. G., Heeringa, W., Knooihuizen, R., & Wieling, M. (2022). Estimating the level and direction of aggregated sound change of dialects in the northern Netherlands. *Taal en Tongval*, *74*(2), 183–214. https://doi.org/10.5117/TET2022.2.002.BUUR

Buurke, R. S. S. J., & Wieling, M. (2023). Sound Change Estimation in Netherlandic Regional Languages: Reducing Inter-Transcriber Variability in Dialect Corpora. *Taal en Tongval*, *75*(1), 7–28. https://doi.org/10.5117/TET2023.1.002.BUUR

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, *14*(3), 261–290. https://doi.org/10.1017/S0954394502143018

Calude, A. S., & Pagel, M. (2011). How do we use language? Shared patterns in the frequency of word use across 17 world languages. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1567), 1101–1107. https://doi.org/10.1098/rstb.2010.0315

Centraal Bureau voor de Statistiek. (2019). Sociale samenhang en welzijn. Retrieved May 31, 2023, from https://www.cbs.nl/nl-nl/longread/statistische-trends/2021/talen-en-dialecten-in-nederland

Centraal Bureau voor de Statistiek. (2021). Kinderen krijgen. Retrieved April 18, 2023, from https://www.cbs.nl/nl-nl/visualisaties/dashboard-bevolking/levensloop/kinderen-krijgen

Centraal Bureau voor de Statistiek. (2023a). Bevolking 15 tot 75 jaar; opleidingsniveau, wijken en buurten, 2021. Retrieved November 23, 2023, from https://opendata.cbs.nl/statline/#/CBS/nl/dataset/85372NED/table?ts=1700743566965

Centraal Bureau voor de Statistiek. (2023b). Bevolkingspiramide. Retrieved March 19, 2023, from https://www.cbs.nl/nl-nl/visualisaties/dashboard-bevolking/bevolkingspiramide

Centraal Bureau voor de Statistiek. (2023c). Stedelijkheid (van een gebied). Retrieved November 23, 2023, from https://www.cbs.nl/nl-nl/onze-diensten/methoden/begrippen/stedelijkheid%E2%80%93van-een-gebied%E2%80%93

Chambers, J. K., & Trudgill, P. (1998). *Dialectology* (2nd ed). Cambridge University Press.

Chibaka, E. F. (2018). Advantages of bilingualism and multilingualism: Multidimensional research findings. In B. S. Chumbow (Ed.), *Multilingualism and Bilingualism* (pp. 15–36). InTech. https://doi.org/10.5772/intechopen.74625

Christoffersen, M. N. (1987). The educational bias of mail questionnaires. *Journal of Official Statistics*, *3*(4), 459–464.

Church, K. W., & Hanks, P. (1990). Word Association Norms, Mutual Information, and Lexicography. *Computational Linguistics*, *16*(1), 22–29.

Clopper, C., & Turnbull, R. (2018). Exploring variation in phonetic reduction: Linguistic, social, and cognitive factors. In F. Cangemi, M. Clayards, O. Niebuhr, B. Schuppler, & M. Zellers (Eds.), *Rethinking reduction: Interdisciplinary perspectives on conditions, mechanisms, and domains for phonetic variation* (pp. 25–72). Mouton de Gruyter.

Cockcroft, K., Wigdorowitz, M., & Liversage, L. (2019). A multilingual advantage in the components of working memory. *Bilingualism: Language and Cognition*, *22*(1), 15–29. https://doi.org/10.1017/S1366728917000475

Cornips, L. (2013). Recent developments in the Limburg dialect region. In F. Hinskens & J. Taeldeman (Eds.), *Dutch* (pp. 378–398). Walter de Gruyter.

Costa, A., & Sebastián-Gallés, N. (2014). How does the bilingual experience sculpt the brain? *Nature Reviews Neuroscience*, *15*(5), 336–345. https://doi.org/10.1038/nrn3709

Crystal, D. (2014). *Language Death* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139923477

De Caluwe, J., & Van Renterghem, E. (2011). Regiolectisering en de opkomst van tussentaal in Vlaanderen. *Taal en Tongval*, *63*(1), 61–77. https://doi.org/10.5117/TET2011.1.DECA

De Groot, G.-R. (2018). European Charter for Regional and Minority Languages. *Language variation. A factor of increasing complexity and a challenge for language policy within Europe (Contributions to the EFNIL Conference 2018 in Amsterdam)*, 115–128.

De Schutter, G. (2013). The dialects of the Brabant region: Phonological properties. In F. Hinskens & J. Taeldeman (Eds.), *Dutch* (pp. 277–296). Walter de Gruyter.

De Tier, V., Van de Velde, H., Schrier, M., Swanenberg, J., & Van de Wijngaard, T. (2008). Oe praet jie? Zeeuwen, Brabanders en Limburgers over hun eigen dialecten. *Zeeland. Tijdschrift van het Koninklijk Zeeuws Genootschap der Wetenschappen*, *17*, 53–61.

De Vries, T. (2011). De fersteanberens fan it Frysk foar dialektsprekkers. *Us Wurk. Tydskrift foar Frisistyk / Journal of Frisian Studies*, *59*(3-4), 132–157.

Delgado-Rodriguez, M. (2004). Bias. *Journal of Epidemiology & Community Health*, *58*(8), 635–641. https://doi.org/10.1136/jech.2003.008466

Doreleijers, K. (2021). Dialect in het onderwijs? Een verkennend onderzoek naar aandacht en animo voor dialecteducatie met een focus op Noord-Brabant. https://pure.uvt.nl/ws/portalfiles/portal/58119820/publieksrapport_dialecteducatie_2021.pdf

Doreleijers, K., Piepers, J., Backus, A., & Swanenberg, J. (2021). Language Variation in Dialect-standard Contact Situations. In G. Kristiansen, K. Franco, S. De Pascale, L. Rosseel, & W. Zhang (Eds.), *Cognitive Sociolinguistics Revisited* (pp. 175–185). De Gruyter. https://doi.org/10.1515/9783110733945-015

Doreleijers, K., & Swanenberg, J. (2023a). Hyperdialectisms revisited. *Linguistics in the Netherlands*, *40*, 39–54. https://doi.org/10.1075/avt.00078.dor

Doreleijers, K., & Swanenberg, J. (2023b). Putting local dialect in the mix: Indexicality and stylization in a TikTok challenge. *Language & Communication*, *92*, 1–14. https://doi.org/10.1016/j.langcom.2023.06.002

Dorian, N. (1994). Varieties of variation in a very small place: Social homogeneity, prestige norms, and linguistic variation. *Language*, *70*(4), 631–696.

Dragojevic, M., Fasoli, F., Cramer, J., & Rakić, T. (2021). Toward a Century of Language Attitudes Research: Looking Back and Moving Forward. *Journal*

of Language and Social Psychology, *40*(1), 60–79. https://doi.org/10.1177/0261927X20966714

Driessen, G. (2005). In Dutch? Usage of Dutch Regional Languages and Dialects. *Language, Culture and Curriculum*, *18*(3), 271–285. https://doi.org/10.1080/07908310508668747

Driessen, G. (2012). Ontwikkelingen in het gebruik van Fries, streektalen en dialecten in de periode 1995-2011, 5.

Duchêne, A., & Humbert, P. N. (2018). Surveying languages: The art of governing speakers with numbers. *International Journal of the Sociology of Language*, *2018*(252), 1–20. https://doi.org/10.1515/ijsl-2018-0012

Duijff, P. (2002). *Fries en stadfries*. Sdu Uitgevers.

Dunbar, R. (2022, November). Linguistic Human Rights in International Law. In T. Skutnabb-Kangas & R. Phillipson (Eds.), *The Handbook of Linguistic Human Rights* (1st ed., pp. 23–38). Wiley. https://doi.org/10.1002/9781119753926.ch2

Durkin, P. (2015). *Etymology* (J. R. Taylor, Ed.). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199641604.013.025

Edwards, J. (1992). Sociopolitical Aspects of Language Maintenance and Loss: Towards a Typology of Minority Language Situations. In W. Fase, K. Jaspaert, & S. Kroon (Eds.), *Studies in Bilingualism* (pp. 37–54, Vol. 1). John Benjamins Publishing Company.

European Commission. (2018). *Linguistic diversity in the European Union: The case of regional and minority languages*. Publications Office.

Fan, S. P., Liberman, Z., Keysar, B., & Kinzler, K. D. (2015). The exposure advantage: Early exposure to a multilingual environment promotes effective communication. *Psychological Science*, *26*(7), 1090–1097. https://doi.org/10.1177/0956797615574699

Feitsma, A. (1989). *Changes in the pronunciation of Frisian under the influence of Netherlandic*. Foris.

Fishman, J. A. (1997). *Reversing language shift: Theoretical and empirical foundations of assistance to threatened languages* (Reprinted). Multilingual Matters.

Fogle, L. W., & King, K. A. (2013). Child agency and language policy in transnational families. *Issues in Applied Linguistics*, *19*(2013), 1–25. https://doi.org/10.5070/L4190005288

Fokkema, K. (1937). *Het Stadsfries: Een bijdrage tot de geschiedenis en de grammatica van het dialect van Leeuwarden*. Van Gorcum.

Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly*, *73*(2), 349–360. https://doi.org/10.1093/poq/nfp031

Gao, F. (2009). Language and power: Korean–Chinese students' language attitude and practice. *Journal of Multilingual and Multicultural Development*, *30*(6), 525–534. https://doi.org/10.1080/01434630903147922

Garrett, A. (2015). Sound change. In C. Bowern & B. Evans (Eds.), *The Routledge Handbook of Historical Linguistics* (pp. 227–248). Routledge London; New York.

Ghyselen, A.-S. (2015). 'Stabilisering' van tussentaal?: Het taalrepertorium in de Westhoek als casus. *Taal en Tongval*, *67*(1), 43–95. https://doi.org/10.5117/TET2015.1.GHYS

Giesbers, C. (2008). *Dialecten op de grens van twee talen-een dialectologisch en sociolinguïstisch onderzoek in het Kleverlands dialectgebied* [PhD Thesis]. Radboud Universiteit Nijmegen.

Gledhill, C. (2020). Language diversity in the United Kingdom: A synthesis of the UK Census 2011 and the Language Rich Europe. In *Politique linguistique et société* (pp. 137–150). Presses universitaires de Valenciennes.

Goeman, A. (2000). Naast NORMs ook MYSFs in het veranderende dialectlandschap en het regiolect. *Taal en Tongval*, *52*, 87–99.

Goeman, T., & Jongenburger, W. (2009). Dimensions and determinants of dialect use in the Netherlands at the individual and regional levels at the end of the twentieth century. *International Journal of the Sociology of Language*, *2009*(196-197), 31–72. https://doi.org/0.1515/IJSL.2009.016

Gooskens, C., & Heeringa, W. (2004). The position of Frisian in the Germanic language area. In D. Gilbers, M. Schreuder, & N. Knevel (Eds.), *On the boundaries of phonology and phonetics* (pp. 61–87). University of Groningen.

Gorter, D. (2001). Extent and Position of West Frisian. In H. H. Munske, N. Århammar, V. F. Faltings, J. F. Hoekstra, O. Vries, A. G. Walker, & O. Wilts (Eds.), *Handbuch des Friesischen / Handbook of Frisian Studies* (pp. 73–83). De Gruyter. https://doi.org/10.1515/9783110946925.73

Goss, E. L., & Howell, R. B. (2006). Social and structural factors in the development of Dutch urban dialects in the Early Modern period. In T. D. Cravens (Ed.), *Current Issues in Linguistic Theory* (p. 59, Vol. 268). John Benjamins Publishing Company. https://doi.org/10.1075/cilt.268.04gos

Grant, A. (2015). Lexical Borrowing. In J. Taylor (Ed.), *The Oxford Handbook of the Word* (pp. 431–445). Oxford University Press.

Grondelaers, S., & Van Hout, R. (2010). Is Standard Dutch with a regional accent standard or not? Evidence from native speakers' attitudes. *Language Variation and Change*, *22*(2), 221–239. https://doi.org/10.1017/S0954394510000086

Grondelaers, S., Van Hout, R., & Van Gent, P. (2016). Destandardization is not destandardization: Revising standardness criteria in order to revisit standard language typologies in the Low Countries. *Taal & Tongval*, *68*(2), 119–149. https://doi.org/10.5117/TET2016.2.GRON

Groves, J. M. (2010). Language or dialect, topolect or regiolect? A comparative study of language attitudes towards the status of Cantonese in Hong Kong. *Journal of Multilingual and Multicultural Development*, *31*(6), 531–551. https://doi.org/10.1080/01434632.2010.509507

Groves, R. M., & Peytcheva, E. (2008). The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis. *Public Opinion Quarterly*, *72*(2), 167–189. https://doi.org/10.1093/poq/nfn011

Haeseryn, W. (2010). Belgian Dutch. In F. Hinskens & J. Taeldeman (Eds.), *Language and Space: Dutch* (pp. 700–720). De Gruyter Mouton.

Harbert, W. (1999). Endangered languages and economic development. In P. K. Austin & J. Sallabank (Eds.), *The Cambridge Handbook of Endangered Languages* (1st ed., pp. 403–422). Cambridge University Press. https://doi.org/10.1017/CBO9780511975981.020

Haugen, E. (1966). Dialect, Language, Nation. *American Anthropologist*, *68*(4), 922–935. https://doi.org/10.1525/aa.1966.68.4.02a00040

Heeringa, W. (2004). *Measuring dialect pronunciation differences using Levenshtein distance* [PhD Thesis]. University of Groningen.

Heeringa, W. (2005). Dialect variation in and around Frisia; classification and relationships. *Us Wurk*, *54*(3–4), 125–167.

Heeringa, W., & Hinskens, F. (2014). Convergence between dialect varieties and dialect groups in the Dutch language area. In B. Szmrecsanyi & B. Wälchli (Eds.), *Aggregating Dialectology, Typology, and Register Analysis* (pp. 26–52). De Gruyter. https://doi.org/10.1515/9783110317558.26

Heeringa, W., & Hinskens, F. (2015). Dialect change and its consequences for the Dutch dialect landscape. How much is due to the standard variety and how much is not? *Journal of Linguistic Geography*, *3*(1), 20–33. https://doi.org/10.1017/jlg.2015.2

Heeringa, W., & Nerbonne, J. (2000). Change, convergence and divergence among Dutch and Frisian. In P. Boersma, P. H. Breuker, L. G. Jansma, & J. Van der Vaart (Eds.), *Philologia Frisica Anno* (pp. 88–109).

Heeroma, K. H. (1953). De Westfaalse expansie. *Taal en Tongval*, *5*(1).

Hermans, B. (2013). Phonological features of Limburgian dialects. In F. Hinskens & J. Taeldeman (Eds.), *Dutch* (pp. 336–355). Walter de Gruyter.

Hilton, N. H., & Gooskens, C. (2013). Language policies and attitudes towards Frisian in the Netherlands. In C. Gooskens & R. Van Bezooijen (Eds.), *Phonetics in Europe: Perception and production* (pp. 139–157). Peter Lang.

Hinskens, F., & Taeldeman, J. (2013). *Dutch*. Berlin: De Gruyter Mouton.

Hinskens, F., & Van Oostendorp, M. (2006). De palatalisering en velarisering van coronale nasaal-plosief clusters in GTR. Talige, dialectgeografische en onderzoekerseffecten. *Taal en Tongval*, *58*, 103–122.

Hoekstra, E. (2003). Frisian: Standardization in progress of a language in decay. In A. Deumert & W. Vandenbussche (Eds.), *Germanic Standardizations: Past to Present* (pp. 193–209). John Benjamins Publishing Company.

Hof, J. J. (1933). *Friesche dialectgeographie*. Martinus Nijhoff.

Holt, D., & Smith, T. F. (1979). Post stratification. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *142*(1), 33–46. https://doi.org/10.2307/2344652

Hoppenbrouwers, C. (1990). *Het regiolect: Van dialect tot algemeen Nederlands*. Muiderberg: Coutinho.

Hornsby, M. (2010). From the periphery to the centre: Recent debates on the place of Breton (and other regional languages) in the French Republic. In *Marginal dialects: Scotland, Ireland and beyond* (pp. 171–197). Aberdeen: Forum for Research on the Languages of Scotland; Ireland.

Howell, R. B. (2006). Immigration and koineisation: The formation of early modern Dutch urban vernaculars. *Transactions of the Philological Society*, *104*(2), 207–227. https://doi.org/10.1111/j.1467-968X.2006.00169.x

Hughes, A., Trudgill, P., & Watt, D. J. L. (2012). *English accents & dialects: An introduction to social and regional varieties of English in the British Isles* (Fifth edition). Hodder Education.

Jones, M. (2008). Framing Welsh identity. *Textes & Contextes*, (1).

Jongbloed-Faber, L., Van de Velde, H., Van der Meer, C., & Klinkenberg, E. (2016). Language use of Frisian bilingual teenagers on social media. *Treballs de Sociolingüística Catalana*, *26*, 27–54.

Jongbloed-Faber, L., Van Loo, J., & Cornips, L. (2017). Regional languages on Twitter: A comparative study between Frisian and Limburgish. *Dutch Journal of Applied Linguistics*, *6*(2), 174–196. https://doi.org/10.1075/dujal.16017.jon

Jonkman, R., & Gorter, D. (1995). *Taal yn Fryslan: Op 'e nij besjoen*. Fryske Akademy.

Kamwangamalu, N. M. (2013). Rural-urban and south-north migrations and language maintenance and shift. *International Journal of the Sociology of Language*, *2013*(222), 33–49. https://doi.org/10.1515/ijsl-2013-0031

Kehrein, R. (2020). Vertical Language Change in Germany: Dialects, Regiolects, and Standard German. In S. D. Brunn & R. Kehrein (Eds.), *Handbook of the Changing World Language Map* (pp. 1357–1369). Springer International Publishing. https://doi.org/10.1007/978-3-030-02438-3_173

Kessler, B. (1995). Computational dialectology in Irish Gaelic. *Proceedings of the seventh conference on European chapter of the Association for Computational Linguistics* -, 60. https://doi.org/10.3115/976973.976983

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, *42*(3), 643–650. https://doi.org/10.3758/BRM.42.3.643

King, K. A., Fogle, L., & Logan-Terry, A. (2008). Family Language Policy. *Language and Linguistics Compass*, *2*(5), 907–922. https://doi.org/10.1111/j.1749-818X.2008.00076.x

Kircher, R., Kutlu, E., & Vellinga, M. (2023). Promoting minority language use to foster revitalisation: Insights from new speakers of West Frisian. *Applied Linguistics*, amad045. https://doi.org/10.1093/applin/amad045

Kirk, N. W. (2023). MIND your language(s): Recognizing Minority, Indigenous, Non-standard(ized), and Dialect variety usage in "monolinguals". *Applied Psycholinguistics*, *44*(3), 358–364. https://doi.org/10.1017/S0142716422000467

Kirk, N. W., Declerck, M., Kemp, R. J., & Kempe, V. (2022). Language control in regional dialect speakers–monolingual by name, bilingual by nature? *Bilingualism: Language and Cognition*, *25*(3), 511–520. https://doi.org/10.1017/S1366728921000973

Klijs, B., Scholtens, S., Mandemakers, J. J., Snieder, H., Stolk, R. P., & Smidt, N. (2015). Representativeness of the LifeLines cohort study (R. I. Ali, Ed.). *PLOS ONE*, *10*(9), e0137203. https://doi.org/10.1371/journal.pone.0137203

Klinkenberg, E., Jonkman, R., & Stefan, M. H. (2018). *Taal yn Fryslân. De folgjende generaasje*. Provinsje Fryslân.

Kloeke, G. (1927). *De Hollandsche Expansie in de zestiende en zeventiende eeuw en haar weerspiegeling in de hedendaagsche Nederlandse dialecten*. Gravenhage: Martinus Nijhoff.

Korkeila, K., Suominen, S., Ahvenainen, J., Ojanlatva, A., Rautava, P., Helenius, H., & Koskenvuo, M. (2001). Non-response and related factors in a nation-wide health survey. *European Journal of Epidemiology*, *17*, 991–999. https://doi.org/10.1023/a:1020016922473

Kruize, H. (2012). *Op Urk. Een kwalitatief onderzoek naar de plaatsidentiteit van een bijzonder dorp* [Master's thesis, University of Groningen].

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*(11), 831–843. https://doi.org/10.1038/nrn1533

Labov, W. (1972). *Sociolinguistic patterns*. University of Pennsylvania Press.

Labov, W. (1990). The intersection of sex and social class in the course of linguistic change. *Language Variation and Change*, *2*(2), 205–254. https://doi.org/10.1017/S0954394500000338

Labov, W. (1994). *Principles of linguistic change. 1: Internal factors* (repr). Wiley-Blackwell.

Langer, N. (2003). Low German. In A. Deumert & W. Vandenbussche (Eds.), *Germanic standardizations: Past to present*. J. Benjamins.

Lee, H. (2016). Parallel sound change between segmental and suprasegmental properties: An individual level observation. *Phonetics and Speech Sciences*, *8*(4), 23–29. https://doi.org/10.13064/KSSS.2016.8.4.023

Leinonen, T., Çöltekin, Ç., & Nerbonne, J. (2016). Using Gabmap. *Lingua*, *178*, 71–83. https://doi.org/10.1016/j.lingua.2015.02.004

Lenz, A. (2009). Emergence of varieties through restructuring and reevaluation. In P. Auer & J. E. Schmidt (Eds.), *Language and Space: An international handbook of linguistic variation, volume 3* (pp. 295–315). Walter de Gruyter. https://doi.org/10.1515/9783110220278.295

Leopold, W. (1959). The decline of German dialects. *Word*, *15*(1), 130–153. https://doi.org/10.1080/00437956.1959.11659689

Levenshtein, V. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, *10*(8), 707–710.

List, J.-M., Greenhill, S. J., & Gray, R. D. (2017). The Potential of Automatic Word Comparison for Historical Linguistics (R. C. Berwick, Ed.). *PLOS ONE*, *12*(1), e0170046. https://doi.org/10.1371/journal.pone.0170046

Mac Síthigh, D. (2018). Official status of languages in the United Kingdom and Ireland. *Common Law World Review*, *47*(1), 77–102. https://doi.org/10.1177/1473779518773642

Maclagan, M., Gordon, E., & Lewis, G. (1999). Women and sound change: Conservative and innovative behavior by the same speakers. *Language Variation and Change*, *11*(1), 19–41.

Manzo, A. N., & Burke, J. M. (2012). Increasing Response Rate in Web-Based/Internet Surveys. In L. Gideon (Ed.), *Handbook of Survey Methodology for the Social Sciences* (pp. 327–343). Springer New York. https://doi.org/10.1007/978-1-4614-3876-2_19

May, S. (2000). Accommodating and Resisting Minority Language Policy: The Case of Wales. *International Journal of Bilingual Education and Bilingualism*, *3*(2), 101–128. https://doi.org/10.1080/13670050008667702

Milroy, J. (2001). Language ideologies and the consequences of standardization. *Journal of Sociolinguistics*, *5*(4), 530–555. https://doi.org/10.1111/1467-9481.00163

Miyawaki, K., Jenkins, J. J., Strange, W., Liberman, A. M., Verbrugge, R., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics*, *18*(5), 331–340. https://doi.org/10.3758/BF03211209

Monaghan, P., & Roberts, S. G. (2019). Cognitive influences in language evolution: Psycholinguistic predictors of loan word borrowing. *Cognition*, *186*, 147–158. https://doi.org/10.1016/j.cognition.2019.02.007

Montaña, B. T. (2020). Language and Basque nationalism: Collective identity, social conflict and institutionalisation. In C. Mar-Molinero & A. Smith (Eds.), *Nationalism and the Nation in the Iberian Peninsula* (pp. 221–236). Routledge.

Montaquila, J. M., Brick, J. M., Hagedorn, M. C., Kennedy, C., & Keeter, S. (2008). Aspects of nonresponse bias in RDD telephone surveys. In J. M. Lepkowski, C. Tucker, J. M. Brick, E. D. De Leeuw, L. Japec, P. J. Lavrakas, M. W. Link, & R. L. Sangster (Eds.), *Advances in Telephone Survey Methodology* (pp. 561–586). Wiley Online Library.

Moore, R. E., Pietikäinen, S., & Blommaert, J. (2010). Counting the losses: Numbers as the language of language endangerment. *Sociolinguistic Studies*, *4*(1), 1–26. https://doi.org/10.1558/sols.v4i1.1

Morgan, G. (2001). Welsh: A European Case of Language Maintenance. In L. Hinton & K. Hale (Eds.), *The Green Book of Language Revitalization in Practice* (pp. 107–118). Brill. https://doi.org/10.1163/9789004261723_011

Munske, H. H., Århammar, N., Faltings, V. F., Hoekstra, J. F., Vries, O., Walker, A. G., & Wilts, O. (Eds.). (2001, December). *Handbuch des Friesischen / Handbook of Frisian Studies*. DE GRUYTER. https://doi.org/10.1515/9783110946925

Nerbonne, J., & Kretzschmar, W. A. (2013). Dialectometry++. *Literary and Linguistic Computing*, *28*(1), 2–12. https://doi.org/10.1093/llc/fqs062

Nerbonne, J. (2010). Measuring the diffusion of linguistic change. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1559), 3821–3828. https://doi.org/10.1098/rstb.2010.0048

Nerbonne, J., Colen, R., Gooskens, C., Kleiweg, P., & Leinonen, T. (2011). Gabmap – a web application for dialectology. *Dialectologia: Revista Electrònica*, *2*, 65–89.

Nerbonne, J., & Heeringa, W. (1997). Measuring dialect distance phonetically. In J. Coleman (Ed.), *Computational Phonology: Third Meeting of the ACL Special Interest Group in Computational Phonology* (pp. 11–18).

Nerbonne, J., & Heeringa, W. (2001). Computational comparison and classification of dialects. *Dialectologia et Geolinguistica, Journal of the International Society for Dialectology and Geolinguistics*, *9/2001*, 69–84. https://doi.org/10.1515/dig.2001.2001.9.69

Nerbonne, J., Heeringa, W., Van den Hout, E., Van der Kooi, P., Otten, S., & Van de Vis, W. (1996). Phonetic distance between Dutch dialects. In G. Durieux, W. Daelemans, & S. Gillis (Eds.), *CLIN VI: Proceedings of the Sixth CLIN meeting* (pp. 185–202).

Ó Riagáin, P. (2018). *Measures of Language Proficiency in Censuses and Surveys*. Springer International Publishing. https://doi.org/10.1007/978-3-319-72941-1

Østergaard, C. R., Timmermans, B., & Kristinsson, K. (2011). Does a different view create something new? The effect of employee diversity on innovation. *Research Policy*, *40*(3), 500–509. https://doi.org/10.1016/j.respol.2010.11.004

Pagel, M., Atkinson, Q. D., & Meade, A. (2007). Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature*, *449*(7163), 717–720. https://doi.org/10.1038/nature06176

Pedersen, I. (2005). Processes of standardisation in Scandinavia. In P. Auer, F. Hinskens, P. Kerswill, et al. (Eds.), *Dialect change: Convergence and divergence in European languages* (pp. 171–195). Cambridge University Press.

Phillips, B. S. (1984). Word frequency and the actuation of sound change. *Language*, *60*(2), 320–342. https://doi.org/10.2307/413643

Pietersen, L. (1969). *De Friezen en hun taal: Een onderzoek naar de lees-en spreekgewoonten in Friesland en naar de houding ten aanzien van het Fries*. Fryske Akademy.

Pinget, A.-F., Kager, R., & Van de Velde, H. (2016). Regional differences in the perception of a consonant change in progress. *Journal of Linguistic Geography*, *4*(2), 65–75. https://doi.org/10.1017/jlg.2016.13

Prenger, J., Pleumeekers, J., Van Zanten, M., Schmidt, V., Teunis, B., Brand, M., & Bron, J. (2023). Conceptkerndoelen Nederlands en rekenen en wiskunde (T. Wierda, Ed.). https://slo-kerndoelen.files.svdcdn.com/production/uploads/assets/updates/Informatiedocument_kerndoelen_SLO_V11_digitaal_2023-10-02-121848_tdkd.pdf?dm=1696249129

Provincie Fryslân. (2020). De Fryske Taalatlas 2020: Friese taal in beeld.

Rebernik, T., Jacobi, J., Buurke, R., Tienkamp, T., Tsiwah, F., Oud, S., Reinders, A., Abur, D., & Wieling, M. (2024). SPRAAKLAB Mobile Laboratory for Acoustic and Articulatory Data Collection: Technical Specifications and Practical Experiences.

Reershemius, G. (2011). Reconstructing the past? Low German and the creating of regional identity in public language display. *Journal of Multilingual and Multicultural development*, *32*(1), 33–54. https://doi.org/10.1080/01434632.2010.536238

Reker, S. (1984). *Groninger spelling: Handleiding voor het lezen en schrijven van Groninger teksten*. Stichting 't Grunneger Bouk.

Revilla, M., & Ochoa, C. (2017). Ideal and maximum length for a web survey. *International Journal of Market Research*, *59*(5). https://doi.org/10.2501/IJMR-2017-039

Riagáin, P. (2001). Irish Language Production and Reproduction 1981–1996. In J. A. Fishman (Ed.), *Can Threatened Languages be Saved?* (pp. 195–214).

Multilingual Matters. https://doi.org/10.21832/9781853597060-010

Rys, K., Heeringa, W., Rutten, J. S., Hinskens, F., De Caluwe, J., Balesar, U., & Doerga Misier, S. (2021). *Staat van het Nederlands - Onderzoeksrapport 2021* (tech. rep.). Nederlandse Taalunie.

Sankoff, G. (2019). Language change across the lifespan: Three trajectory types. *Language*, *95*(2), 197–229. https://doi.org/10.1353/lan.2019.0029

Sankoff, G., & Blondeau, H. (2007). Language change across the lifespan: /r/ in Montreal French. *Language*, *83*(3), 560–588. https://doi.org/10.1353/lan.2007.0106

Sassen, A. (1953). *Het Drents van Ruinen* [PhD Thesis]. University of Groningen.

Schmeets, H., & Cornips, L. (2022). Taaldiversiteit in Nederland: Language diversity in the Netherlands. *Taal en Tongval*, *74*(1), 75–106. https://doi.org/10.5117/TET2022.1.004.SCHM

Schmid, M. S., & Keijzer, M. (2009). First language attrition and reversion among older migrants. *International Journal of the Sociology of Language*, *2009*(200). https://doi.org/10.1515/IJSL.2009.046

Scholtens, S., Smidt, N., Swertz, M. A., Bakker, S. J., Dotinga, A., Vonk, J. M., Van Dijk, F., Van Zon, S. K., Wijmenga, C., Wolffenbuttel, B. H., & Stolk, R. P. (2015). Cohort Profile: LifeLines, a three-generation cohort study and biobank. *International Journal of Epidemiology*, *44*(4), 1172–1180. https://doi.org/10.1093/ije/dyu229

Sebregts, K. (2015). *The sociophonetics and phonology of Dutch r* [PhD Thesis]. Utrecht University.

Shriberg, L., & Lof, G. (1991). Reliability studies in broad and narrow phonetic transcription. *Clinical Linguistics & Phonetics*, *5*(3), 225–279. https://doi.org/10.3109/02699209108986113

Sijtsma, A., Rienks, J., Van der Harst, P., Navis, G., Rosmalen, J. G. M., & Dotinga, A. (2022). Cohort Profile Update: Lifelines, a three-generation cohort study and biobank. *International Journal of Epidemiology*, *51*(5), e295–e302. https://doi.org/10.1093/ije/dyab257

Silver, B. (2002). Nationality and language in the new censuses of the Baltic states. https://doi.org/10.13140/2.1.2366.4321

Slaats, I. (2020). *Codificering in Brabantse, Limburgse en Nedersaksische dialecten* [Master's thesis, Universiteit Utrecht].

Smakman, D. (2006). *Standard Dutch in the Netherlands: A sociolinguistic and phonetic description* [PhD Thesis]. Radboud University.

Smakman, D. (2012). The definition of the standard language: A survey in seven countries. *International Journal of the Sociology of Language*, *2012*(218). https://doi.org/10.1515/ijsl-2012-0058

Smits, T. (2005). Dialectverandering grensoverschrijdend bekeken: Methodologische overwegingen bij een socio-dialectologisch onderzoek in West-Münsterland (D) en de Achterhoek (NL). *Taal en Tongval*, *57*(1), 89–110.

Smits, T. (2011). Dialectverlies en dialectnivellering in Nederlands-Duitse grensdialecten. *Taal en Tongval*, *63*(1), 175–196.

Stefan, M. H. (2022). *Spoken Frisian: Language contact, variation and change*. LOT.

Stegeman, J. (2021). *Grote geschiedenis van de Nederlandse taal*. Amsterdam University Press.

Stevens, G. (1999). A century of US censuses and the language characteristics of immigrants. *Demography*, *36*, 387–397. https://doi.org/10.2307/2648061

Streckenbach, A. (2020). Zur Chronologisierung sprachlicher Daten am Beispiel der Realisierungsvarianten von nichts in Unterfranken und am Oberrhein. In M. Hahn, A. Kleene, R. Langhanke, & A. Schaufuß (Eds.), *Dynamik in den deutschen Regionalsprachen: Gebrauch und Wahrnehmung: Beiträge aus dem Forum Sprachvariation* (pp. 55–85). Georg Olms Verlag.

Stroop, J. (1998). *Poldernederlands: Waardoor het ABN verdwijnt*. B. Bakker.

Swanenberg, J. (2013). All dialects are equal, but some dialects are more equal than others. *Tilburg Papers in Culture Studies*, *43*, 12.

Swanenberg, J., & Van Hout, R. (2013). Recent developments in the mid southern dialects. In F. Hinskens & J. Taeldeman (Eds.), *Dutch* (pp. 319–335). Walter de Gruyter.

Taalunie. (2021). *Staat van het Nederlands - Publieksrapport 2021*.

Taeldeman, J. (2013a). Recent developments in the East Flemish dialect area. In F. Hinskens & J. Taeldeman (Eds.), *Dutch* (pp. 255–276). Walter de Gruyter.

Taeldeman, J. (2013b). The southwestern dialect area: Phonology. In F. Hinskens & J. Taeldeman (Eds.), *Dutch* (pp. 150–174). Walter de Gruyter.

Taeldeman, J., & Goeman, A. (1996). Fonologie en morfologie van de Nederlandse dialecten: Een nieuwe materiaalverzameling en twee nieuwe atlasprojecten. *Taal en Tongval*, *48*(1), 38–59.

Tagliamonte, S. (2011). *Variationist sociolinguistics: Change, observation, interpretation*. John Wiley & Sons.

Tagliamonte, S. A. (2016). *Making waves: The Story of Variationist Sociolinguistics* (1st ed). Wiley-Blackwell.

Ter Denge, G. (2012). *Linguae Intergermanica: The Use of Low Saxon, English, Dutch, German, and Receptive Multilingualism in Northern Dutch–Northern German Communication* [Master's thesis, Utrecht University].

Tienkamp, T. B., Rebernik, T., Buurke, R., Polsterer, K., Van Son, R., Wieling, M., Witjes, M. J., De Visscher, S., & Abur, D. (2024). The Effect of Speaking Style on the Articulatory-Acoustic Vowel Space in Individuals with Tongue Cancer before and after Surgical Treatment. *Proceedings of the 13th International Seminar on Speech Production*, 65–68.

Tillery, J., & Bailey, G. (2003). Approaches to real time in dialectology and sociolinguistics. *World Englishes*, *22*(4), 351–365. https://doi.org/10.1111/j.1467-971X.2003.00305.x

Trudgill, P. (1974). Linguistic change and diffusion: Description and explanation in sociolinguistic dialect geography. *Language in Society*, *3*(2), 215–246. https://doi.org/10.1017/S0047404500004358

Tseng, A. (2020). Identity in home-language maintenance. In *Handbook of Home Language Maintenance and Development* (pp. 109–129). De Gruyter Mouton Berlin New York.

Tulloch, S. (2006). Preserving dialects of an endangered language. *Current Issues in Language Planning*, *7*(2-3), 269–286. https://doi.org/10.2167/cilp098.0

Tweede Kamer der Staten-Generaal. (1995). Kamerstuk 24 092: Goedkeuring van het op 5 november 1992 te Straatsburg tot stand gekomen Europees Handvest voor streektalen of talen van minderheden (Trb. 1993, 1 en 199) [Parliamentary document 24 092: Approval of the European Charter for Regional or Minority Languages, drawn up in Strasbourg on 5 November 1992 (Trb. 1993, 1 and 199)]. Retrieved December 4, 2023, from https://zoek.officielebekendmakingen.nl/kst-24092-5.pdf

Valadez, C., Etxeberria, F., & Intxausti, N. (2015). Language revitalization and the normalization of Basque: A study of teacher perceptions and expectations in the Basque Country. *Current Issues in Language Planning*, *16*(1-2), 60–79. https://doi.org/10.1080/14664208.2014.947019

Van Bezooijen, R. (2009). The pronunciation of /r/ in Frisian: A comparative study with Dutch and Town Frisian. In J. N. Stanford & D. R. Preston (Eds.), *IMPACT: Studies in Language and Society* (pp. 299–318, Vol. 25).

John Benjamins Publishing Company. https://doi.org/10.1075/impact.25.15bez

Van Bezooijen, R., & Van den Berg, R. (1999). Word intelligibility of language varieties in the Netherlands and Flanders under minimal conditions. *Linguistics in the Netherlands*, *16*(1), 1–12. https://doi.org/10.1075/avt.16.03bez

Van Bree, C. (2001, December). "Stadtfriesisch" und andere nichtfriesische Dialekte der Provinz Fryslân. In H. H. Munske, N. Århammar, V. F. Faltings, J. F. Hoekstra, O. Vries, A. G. Walker, & O. Wilts (Eds.), *Handbuch des Friesischen / Handbook of Frisian Studies* (pp. 129–138). DE GRUYTER. https://doi.org/10.1515/9783110946925.129

Van Bree, C. (2004). *Taal in stad en land: Zuid-Hollands*. Sdu Uitgevers.

Van Bree, C. (2017). The Frisian Substrate beneath the Groningen Dialect. *Amsterdamer Beiträge zur älteren Germanistik*, *77*(1-2), 65–87. https://doi.org/10.1163/18756719-12340067

Van de Velde, H. (2021). It fertutearzjen fan it Frysk - De verwaarlozing van het Fries [Inaugural lecture.].

Van de Velde, H., Duijff, P., Dyk, S., Heeringa, W., & Hoekstra, E. (2019). *Fries-Nederlandse contactvariëteiten in Fryslân*. Fryske Akademy.

Van de Velde, H., Van de Wijngaard, T., Schrier, M., De Tier, V., & Swanenberg, J. (2008). Limburgs kalle, wie sjteit't d'rmit? *Jaarboek Veldeke*, *2007*.

Van de Velde, H., & Van Hout, R. (1999). The pronunciation of (r) in Standard Dutch. *Linguistics in the Netherlands*, *16*(1), 177–188. https://doi.org/10.1075/avt.l6.16van

Van de Velde, H., Van Hout, R., & Gerritsen, M. (1997). Watching Dutch Change: A Real Time Study of Variation and Change in Standard Dutch Pronunciation. *Journal of Sociolinguistics*, *1*(3), 361–391. https://doi.org/10.1111/1467-9481.00021

Van den Berg, C., & Van Oostendorp, M. (2012). *Dat is andere taal! streektalen en dialecten van Nederland*. Winkler Prins.

Van der Star, M. E., & Hochstenbach, C. (2022). Continuity among stayers: Levels, predictors and meanings of place attachment in rural shrinking regions. *Journal of Rural Studies*, *96*, 369–380. https://doi.org/10.1016/j.jrurstud.2022.11.010

Van der Velden, B. D. (2004). *'Waar gaan wij heen met het Fries?': Het gebruik van de Friese taal in het juridische en in het bestuurlijke verkeer in de laatste twee eeuwen*. Wolf Legal Publ.

Van Oostendorp, M. (2007). Kenmerkeconomie in de GTRP-database. *Taal en Tongval*, *18*(1), 148–159.

Van Rij, J., Wieling, M., Baayen, R. H., & Van Rijn, H. (2022). Itsadug: Interpreting time series and autocorrelated data using GAMMs. https://doi.org/10.32614/CRAN.package.itsadug

Van Sluis, P., Hoekstra, E., & Van de Velde, H. (2016). Bildts as a mixed language. *International Journal of the Sociology of Language*, *2016*(242), 59–80. https://doi.org/10.1515/ijsl-2016-0033

Van Spijk, Y., & Swanenberg, J. (2018). Dialect en standaardtaal in contact: De vele vormen van het dialect in de Langstraat. *Taal en Tongval*, *70*(1), 115–140. https://doi.org/10.5117/TET2018.1.SPIJ

Vandekerckhove, R. (1998). Code-Switching Between Dialect and Standard Language as a Graduator. *Zeitschrift für Dialektologie und Linguistiek*, *65*(3), 280.

Vandekerckhove, R. (2009). Dialect loss and dialect vitality in Flanders. *International Journal of the Sociology of Language*, *2009*(196-197). https://doi.org/10.1515/IJSL.2009.017

Varkevisser, N. A., & Walsweer, A. (2018). *It is mei sizzen net te dwaan: Inventarisatie naar de stand van zaken m.b.t. het (vak) Fries in het primair en voortgezet onderwijs, rapport taalplan Frysk* (M. d. Roo & G. Tigchelaar, Trans.). Lectoraat Taalgebruik en Leren.

Versloot, A. (2021a). Streektaaldood in de Lage Landen. *Taal en Tongval*, *72*(1), 7–16. https://doi.org/10.5117/TET2020.1.VERS

Versloot, A. (2021b). The volatile linguistic shape of 'Town Frisian'/'Town Hollandic'. In H. Van de Velde, N. H. Hilton, & R. Knooihuizen (Eds.), *Language Variation–European Perspectives VIII: Selected papers from the Tenth International Conference on Language Variation in Europe (ICLaVE 10), Leeuwarden, June 2019* (pp. 11–34, Vol. 25). John Benjamins Publishing Company. https://doi.org/10.1075/silv.25.01ver

Vousten, R. (1995). *Dialect als tweede taal: Linguïstische en extra-linguïstische aspecten van de verwerving van een Noordlimburgs dialect door standaardtalige jongeren* [PhD Thesis]. Amsterdam: Thesis Publishers.

Wang, L., & King, K. (2024). Language ideologies, language policies, and shifting regional dialect proficiencies in three Chinese cities. *Journal of Multilingual and Multicultural Development*, *45*(6), 2166–2182. https://doi.org/10.1080/01434632.2022.2044339

Wenker, G., & Wrede, F. (1889). *Der Sprachatlas des Deutschen Reichs*. Elwert (reprint by Kessinger Publishing, LLC).

Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics*, *70*(2018), 86–116. https://doi.org/10.1016/j.wocn.2018.03.002

Wieling, M., Bloem, J., Mignella, K., Timmermeister, M., & Nerbonne, J. (2014a). Measuring foreign accent strength in English: Validating Levenshtein distance as a measure. *Language Dynamics and Change*, *4*(2), 253–269. https://doi.org/10.1163/22105832-00402001

Wieling, M., Heeringa, W., & Nerbonne, J. (2007). An Aggregate Analysis of Pronunciation in the Goeman-Taeldeman-Van Reenen-Project Data.

Wieling, M., Margaretha, E., & Nerbonne, J. (2012). Inducing a measure of phonetic similarity from pronunciation variation. *Journal of Phonetics*, *40*(2), 307–314. https://doi.org/10.1016/j.wocn.2011.12.004

Wieling, M., Montemagni, S., Nerbonne, J., & Baayen, R. H. (2014b). Lexical differences between Tuscan dialects and standard Italian: Accounting for geographic and sociodemographic variation using generalized additive mixed modeling. *Language*, *90*(3), 669–692. https://doi.org/10.1353/lan.2014.0064

Wieling, M., & Nerbonne, J. (2015). Advances in Dialectometry. *Annual Review of Linguistics*, *1*(1), 243–264. https://doi.org/10.1146/annurev-linguist-030514-124930

Wieling, M., Nerbonne, J., & Baayen, R. H. (2011). Quantitative Social Dialectology: Explaining Linguistic Variation Geographically and Socially (M. Perc, Ed.). *PLOS ONE*, *6*(9), e23613. https://doi.org/10.1371/journal.pone.0023613

Wieling, M., Prokić, J., & Nerbonne, J. (2009). Evaluating the pairwise string alignment of pronunciations. *Proceedings of the EACL 2009 Workshop on Language Technology and Resources for Cultural Heritage, Social Sciences, Humanities, and Education - LaTeCH-SHELT&R '09*, 26–34. https://doi.org/10.3115/1642049.1642053

Wieling, M., Rebernik, T., & Jacobi, J. (2023). SPRAAKLAB: A mobile laboratory for collecting speech production data. *Proceedings of the 20th International Congress of Phonetic Sciences*, 2060–2064.

Wieling, M. B. (2007). *Comparison of Dutch dialects* [Master's thesis, University of Groningen].

Willemyns, R. (2002). The Dutch-French Language Border in Belgium. *Journal of Multilingual and Multicultural Development*, *23*(1-2), 36–49. https://doi.org/10.1080/01434630208666453

Willemyns, R. (2003). Dutch. In A. Deumert & W. Vandenbussche (Eds.), *Germanic standardizations: Past to present* (pp. 93–126). J. Benjamins.

Wilting, M., Van Hout, R., & Swanenberg, J. (2014). Regiolect verankerd: Een survey-onderzoek naar dialectgebruik van Eindhovense jongeren. *Taal en Tongval*, *66*(2), 143–171. https://doi.org/10.1557/TET2014.2.WILT

Winkler, J. (1874). *Algemeen Nederduitsch en Friesch dialecticon*. Nijhoff.

Wölck, W. (2004). Universals of Language Maintenance, Shift and Change. *Coll. Antropol.*, *28*, 5–12.

Wood, S. (2000). Mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation [Institution: Comprehensive R Archive Network Pages: 1.9-1]. https://doi.org/10.32614/CRAN.package.mgcv

Wood, S. N. (2017). *Generalized Additive Models*. Chapman; Hall/CRC.

Wood, S. N., Pya, N., & Säfken, B. (2016). Smoothing Parameter and Model Selection for General Smooth Models. *Journal of the American Statistical Association*, *111*(516), 1548–1563. https://doi.org/10.1080/01621459.2016.1180986

Ytsma, J. (2007). Language Use and Language Attitudes in Friesland. In D. Lasagabaster & Á. Huguet Canalís (Eds.), *Multilingualism in European Bilingual Contexts* (pp. 144–163). Multilingual Matters.

Zhang, W. (2008). A forecast analysis on world population and urbanization process. *Environment, Development and Sustainability*, *10*(6), 717–730. https://doi.org/10.1007/s10668-007-9081-8

GRONINGEN DISSERTATIONS IN LINGUISTICS (GRODIL)

1. Henriëtte de Swart (1991). *Adverbs of Quantification: A Generalized Quantifier Approach*.
2. Eric Hoekstra (1991). *Licensing Conditions on Phrase Structure*.
3. Dicky Gilbers (1992). *Phonological Networks. A Theory of Segment Representation*.
4. Helen de Hoop (1992). *Case Configuration and Noun Phrase Interpretation*.
5. Gosse Bouma (1993). *Nonmonotonicity and Categorial Unification Grammar*.
6. Peter I. Blok (1993). *The Interpretation of Focus*.
7. Roelien Bastiaanse (1993). *Studies in Aphasia*.
8. Bert Bos (1993). *Rapid User Interface Development with the Script Language Gist*.
9. Wim Kosmeijer (1993). *Barriers and Licensing*.
10. Jan-Wouter Zwart (1993). *Dutch Syntax: A Minimalist Approach*.
11. Mark Kas (1993). *Essays on Boolean Functions and Negative Polarity*.
12. Ton van der Wouden (1994). *Negative Contexts*.
13. Joop Houtman (1994). *Coordination and Constituency: A Study in Categorial Grammar*.
14. Petra Hendriks (1995). *Comparatives and Categorial Grammar*.
15. Maarten de Wind (1995). *Inversion in French*.
16. Jelly Julia de Jong (1996). *The Case of Bound Pronouns in Peripheral Romance*.
17. Sjoukje van der Wal (1996). *Negative Polarity Items and Negation: Tandem Acquisition*.
18. Anastasia Giannakidou (1997). *The Landscape of Polarity Items*.
19. Karen Lattewitz (1997). *Adjacency in Dutch and German*.
20. Edith Kaan (1997). *Processing Subject-Object Ambiguities in Dutch*.
21. Henny Klein (1997). *Adverbs of Degree in Dutch*.
22. Leonie Bosveld-de Smet (1998). *On Mass and Plural Quantification: The case of French 'des'/'du'-NPs*.
23. Rita Landeweerd (1998). *Discourse semantics of perspective and temporal structure*.
24. Mettina Veenstra (1998). *Formalizing the Minimalist Program*.
25. Roel Jonkers (1998). *Comprehension and Production of Verbs in aphasic Speakers*.
26. Erik F. Tjong Kim Sang (1998). *Machine Learning of Phonotactics*.
27. Paulien Rijkhoek (1998). *On Degree Phrases and Result Clauses*.
28. Jan de Jong (1999). *Specific Language Impairment in Dutch: Inflectional Morphology and Argument Structure*.
29. H. Wee (1999). *Definite Focus*.
30. Eun-Hee Lee (2000). *Dynamic and Stative Information in Temporal Reasoning: Korean tense and aspect in discourse*.
31. Ivilin P. Stoianov (2001). *Connectionist Lexical Processing*.

32. Klarien van der Linde (2001). *Sonority substitutions*.
33. Monique Lamers (2001). *Sentence processing: using syntactic, semantic, and thematic information*.
34. Shalom Zuckerman (2001). *The Acquisition of "Optional" Movement*.
35. Rob Koeling (2001). *Dialogue-Based Disambiguation: Using Dialogue Status to Improve Speech Understanding*.
36. Esther Ruigendijk (2002). *Case assignment in Agrammatism: a cross-linguistic study*.
37. Tony Mullen (2002). *An Investigation into Compositional Features and Feature Merging for Maximum Entropy-Based Parse Selection*.
38. Nanette Bienfait (2002). *Grammatica-onderwijs aan allochtone jongeren*.
39. Dirk-Bart den Ouden (2002). *Phonology in Aphasia: Syllables and segments in level-specific deficits*.
40. Rienk Withaar (2002). *The Role of the Phonological Loop in Sentence Comprehension*.
41. Kim Sauter (2002). *Transfer and Access to Universal Grammar in Adult Second Language Acquisition*.
42. Laura Sabourin (2003). *Grammatical Gender and Second Language Processing: An ERP Study*.
43. Hein van Schie (2003). *Visual Semantics*.
44. Lilia Schürcks-Grozeva (2003). *Binding and Bulgarian*.
45. Stasinos Konstantopoulos (2003). *Using ILP to Learn Local Linguistic Structures*.
46. Wilbert Heeringa (2004). *Measuring Dialect Pronunciation Differences using Levenshtein Distance*.
47. Wouter Jansen (2004). *Laryngeal Contrast and Phonetic Voicing: A Laboratory Phonology*.
48. Judith Rispens (2004). *Syntactic and phonological processing in developmental dyslexia*.
49. Danielle Bougaïré (2004). *L'approche communicative des campagnes de sensibilisation en santé publique au Burkina Faso: Les cas de la planification familiale, du sida et de l'excision*.
50. Tanja Gaustad (2004). *Linguistic Knowledge and Word Sense Disambiguation*.
51. Susanne Schoof (2004). *An HPSG Account of Nonfinite Verbal Complements in Latin*.
52. M. Begoña Villada Moirón (2005). *Data-driven identification of fixed expressions and their modifiability*.
53. Robbert Prins (2005). *Finite-State Pre-Processing for Natural Language Analysis*.
54. Leonoor van der Beek (2005) *Topics in Corpus-Based Dutch Syntax*.
55. Keiko Yoshioka (2005). *Linguistic and gestural introduction and tracking of referents in L1 and L2 discourse*.
56. Sible Andringa (2005). *Form-focused instruction and the development of second language proficiency*.
57. Joanneke Prenger (2005). *Taal telt! Een onderzoek naar de rol van taalvaardigheid en tekstbegrip in het realistisch wiskundeonderwijs*.
58. Neslihan Kansu-Yetkiner (2006). *Blood, Shame and Fear: Self-Presentation Strategies of Turkish Women's Talk about their Health and Sexuality*.

59. Mónika Z. Zempléni (2006). *Functional imaging of the hemispheric contribution to language processing.*
60. Maartje Schreuder (2006). *Prosodic Processes in Language and Music.*
61. Hidetoshi Shiraishi (2006). *Topics in Nivkh Phonology.*
62. Tamás Biró (2006). *Finding the Right Words: Implementing Optimality Theory with Simulated Annealing.*
63. Dieuwke de Goede (2006). *Verbs in Spoken Sentence Processing: Unraveling the Activation Pattern of the Matrix Verb.*
64. Eleonora Rossi (2007). *Clitic production in Italian agrammatism.*
65. Holger Hopp (2007). *Ultimate Attainment at the Interfaces in Second Language Acquisition: Grammar and Processing.*
66. Gerlof Bouma (2008). *Starting a Sentence in Dutch: A corpus study of subject- and object-fronting.*
67. Julia Klitsch (2008). *Open your eyes and listen carefully. Auditory and audiovisual speech perception and the McGurk effect in Dutch speakers with and without aphasia.*
68. Janneke ter Beek (2008). *Restructuring and Infinitival Complements in Dutch.*
69. Jori Mur (2008). *Off-line Answer Extraction for Question Answering.*
70. Lonneke van der Plas (2008). *Automatic Lexico-Semantic Acquisition for Question Answering.*
71. Arjen Versloot (2008). *Mechanisms of Language Change: Vowel reduction in 15th century West Frisian.*
72. Ismail Fahmi (2009). *Automatic term and Relation Extraction for Medical Question Answering System.*
73. Tuba Yarbay Duman (2009). *Turkish Agrammatic Aphasia: Word Order, Time Reference and Case.*
74. Maria Trofimova (2009). *Case Assignment by Prepositions in Russian Aphasia.*
75. Rasmus Steinkrauss (2009). *Frequency and Function in WH Question Acquisition. A Usage-Based Case Study of German L1 Acquisition.*
76. Marjolein Deunk (2009). *Discourse Practices in Preschool. Young Children's Participation in Everyday Classroom Activities.*
77. Sake Jager (2009). *Towards ICT-Integrated Language Learning: Developing an Implementation Framework in terms of Pedagogy, Technology and Environment.*
78. Francisco Dellatorre Borges (2010). *Parse Selection with Support Vector Machines.*
79. Geoffrey Andogah (2010). *Geographically Constrained Information Retrieval.*
80. Jacqueline van Kruiningen (2010). *Onderwijsontwerp als conversatie. Probleemoplossing in interprofessioneel overleg.*
81. Robert G. Shackleton (2010). *Quantitative Assessment of English-American Speech Relationships*.
82. Tim Van de Cruys (2010). *Mining for Meaning: The Extraction of Lexico-semantic Knowledge from Text.*
83. Therese Leinonen (2010). *An Acoustic Analysis of Vowel Pronunciation in Swedish Dialects.*

84. Erik-Jan Smits (2010). *Acquiring Quantification. How Children Use Semantics and Pragmatics to Constrain Meaning*.
85. Tal Caspi (2010). *A Dynamic Perspective on Second Language Development*.
86. Teodora Mehotcheva (2010). *After the fiesta is over. Foreign language attrition of Spanish in Dutch and German Erasmus Student*.
87. Xiaoyan Xu (2010). *English language attrition and retention in Chinese and Dutch university students*.
88. Jelena Prokić (2010). *Families and Resemblances*.
89. Radek Šimík (2011). *Modal existential wh-constructions*.
90. Katrien Colman (2011). *Behavioral and neuroimaging studies on language processing in Dutch speakers with Parkinson's disease*.
91. Siti Mina Tamah (2011). *A Study on Student Interaction in the Implementation of the Jigsaw Technique in Language Teaching*.
92. Aletta Kwant (2011). *Geraakt door prentenboeken: Effecten van het gebruik van prentenboeken op de sociaal-emotionele ontwikkeling van kleuters*.
93. Marlies Kluck (2011). *Sentence amalgamation*.
94. Anja Schüppert (2011). *Origin of asymmetry: Mutual intelligibility of spoken Danish and Swedish*.
95. Peter Nabende (2011). *Applying Dynamic Bayesian Networks in Transliteration Detection and Generation*.
96. Barbara Plank (2011). *Domain Adaptation for Parsing*.
97. Çagrı Çöltekin (2011). *Catching Words in a Stream of Speech: Computational simulations of segmenting transcribed child-directed speech*.
98. Dörte Hessler (2011). *Audiovisual Processing in Aphasic and Non-Brain-Damaged Listeners: The Whole is More than the Sum of its Parts*.
99. Herman Heringa (2012). *Appositional constructions*.
100. Diana Dimitrova (2012). *Neural Correlates of Prosody and Information Structure*.
101. Harwintha Anjarningsih (2012). *Time Reference in Standard Indonesian Agrammatic Aphasia*.
102. Myrte Gosen (2012). *Tracing learning in interaction. An analysis of shared reading of picture books at kindergarten*.
103. Martijn Wieling (2012). *A Quantitative Approach to Social and Geographical Dialect Variation*.
104. Gisi Cannizzaro (2012). *Early word order and animacy*.
105. Kostadin Cholakov (2012). *Lexical Acquisition for Computational Grammars. A Unified Model*.
106. Karin Beijering (2012). *Expressions of epistemic modality in Mainland Scandinavian. A study into the lexicalization-grammaticalization-pragmaticalization interface*.
107. Veerle Baaijen (2012). *The development of understanding through writing*.
108. Jacolien van Rij (2012). *Pronoun processing: Computational, behavioral, and psychophysiological studies in children and adults*.
109. Ankelien Schippers (2012). *Variation and change in Germanic long-distance depen-*

dencies.

110. Hanneke Loerts (2012). *Uncommon gender: Eyes and brains, native and second language learners, & grammatical gender*.
111. Marjoleine Sloos (2013). *Frequency and phonological grammar: An integrated approach. Evidence from German, Indonesian, and Japanese*.
112. Aysa Arylova. (2013) *Possession in the Russian clause. Towards dynamicity in syntax*.
113. Daniël de Kok (2013). *Reversible Stochastic Attribute-Value Grammars*.
114. Gideon Kotzé (2013). *Complementary approaches to tree alignment: Combining statistical and rule-based methods*.
115. Fridah Katushemererwe (2013). *Computational Morphology and Bantu Language Learning: an Implementation for Runyakitara*.
116. Ryan C. Taylor (2013). *Tracking Referents: Markedness, World Knowledge and Pronoun Resolution*.
117. Hana Smiskova-Gustafsson (2013). *Chunks in L2 Development: A Usage-based Perspective*.
118. Milada Walková (2013). *The aspectual function of particles in phrasal verbs*.
119. Tom O. Abuom (2013). *Verb and Word Order Deficits in Swahili-English bilingual agrammatic speakers*.
120. Gülsen Yýlmaz (2013). *Bilingual Language Development among the First Generation Turkish Immigrants in the Netherlands*.
121. Trevor Benjamin (2013). *Signaling Trouble: On the linguistic design of other-initiation of repair in English conversation*.
122. Nguyen Hong Thi Phuong (2013). *A Dynamic Usage-based Approach to Second Language Teaching*.
123. Harm Brouwer (2014). *The Electrophysiology of Language Comprehension: A Neurocomputational Model*.
124. Kendall Decker (2014). *Orthography Development for Creole Languages*.
125. Laura S. Bos (2015). *The Brain, Verbs, and the Past: Neurolinguistic Studies on Time Reference*.
126. Rimke Groenewold (2015). *Direct and indirect speech in aphasia: Studies of spoken discourse production and comprehension*.
127. Huiping Chan (2015). *A Dynamic Approach to the Development of Lexicon and Syntax in a Second Language*.
128. James Griffiths (2015). *On appositives*.
129. Pavel Rudnev (2015). *Dependency and discourse-configurationality: A study of Avar*.
130. Kirsten Kolstrup (2015). *Opportunities to speak. A qualitative study of a second language in use*.
131. Güliz Güneş (2015). *Deriving Prosodic structures*.
132. Cornelia Lahmann (2015). *Beyond barriers. Complexity, accuracy, and fluency in long-term L2 speakers' speech*.
133. Sri Wachyunni (2015). *Scaffolding and Cooperative Learning: Effects on Reading Comprehension and Vocabulary Knowledge in English as a Foreign Language*.

134. Albert Walsweer (2015). *Ruimte voor leren. Een etnografisch onderzoek naar het verloop van een interventie gericht op versterking van het taalgebruik in een knowledge building environment op kleine Friese basisscholen.*
135. Aleyda Lizeth Linares Calix (2015). *Raising Metacognitive Genre Awareness in L2 Academic Readers and Writers.*
136. Fathima Mufeeda Irshad (2015). *Second Language Development through the Lens of a Dynamic Usage-Based Approach.*
137. Oscar Strik (2015). *Modelling analogical change. A history of Swedish and Frisian verb inflection.*
138. He Sun (2015). *Predictors and stages of very young child EFL learners' English development in China.*
139. Marieke Haan (2015). *Mode Matters. Effects of survey modes on participation and answering behavior.*
140. Nienke Houtzager (2015). *Bilingual advantages in middle-aged and elderly populations.*
141. Noortje Joost Venhuizen (2015). *Projection in Discourse: A data-driven formal semantic analysis.*
142. Valerio Basile (2015). *From Logic to Language: Natural Language Generation from Logical Forms.*
143. Jinxing Yue (2016). *Tone-word Recognition in Mandarin Chinese: Influences of lexical-level representations.*
144. Seçkin Arslan (2016). *Neurolinguistic and Psycholinguistic Investigations on Evidentiality in Turkish.*
145. Rui Qin (2016). *Neurophysiological Studies of Reading Fluency. Towards Visual and Auditory Markers of Developmental Dyslexia.*
146. Kashmiri Stec (2016). *Visible Quotation: The Multimodal Expression of Viewpoint.*
147. Yinxing Jin (2016). *Foreign language classroom anxiety: A study of Chinese university students of Japanese and English over time.*
148. Joost Hurkmans (2016). *The Treatment of Apraxia of Speech. Speech and Music Therapy, an Innovative Joint Effort.*
149. Franziska Köder (2016). *Between direct and indirect speech: The acquisition of pronouns in reported speech.*
150. Femke Swarte (2016). *Predicting the mutual intelligibility of Germanic languages from linguistic and extra-linguistic factors.*
151. Sanne Kuijper (2016). *Communication abilities of children with ASD and ADHD. Production, comprehension, and cognitive mechanisms.*
152. Jelena Golubović (2016). *Mutual intelligibility in the Slavic language area.*
153. Nynke van der Schaaf (2016). *"Kijk eens wat ik kan!" Sociale praktijken in de interactie tussen kinderen van 4 tot 8 jaar in de buitenschoolse opvang.*
154. Simon Šuster (2016). *Empirical studies on word representations.*
155. Kilian Evang (2016). *Cross-lingual Semantic Parsing with Categorial Grammars.*
156. Miren Arantzeta Pérez (2017). *Sentence comprehension in monolingual and bilingual*

aphasia: Evidence from behavioral and eye-tracking methods.

157. Sana-e-Zehra Haidry (2017). *Assessment of Dyslexia in the Urdu Language.*
158. Srđan Popov (2017). *Auditory and Visual ERP Correlates of Gender Agreement Processing in Dutch and Italian.*
159. Molood Sadat Safavi (2017). *The Competition of Memory and Expectation in Resolving Long-Distance Dependencies: Psycholinguistic Evidence from Persian Complex Predicates.*
160. Christopher Bergmann (2017). *Facets of native-likeness: First-language attrition among German emigrants to Anglophone North America.*
161. Stefanie Keulen (2017). *Foreign Accent Syndrome: A Neurolinguistic Analysis.*
162. Franz Manni (2017). *Linguistic Probes into Human History.*
163. Margreet Vogelzang (2017). *Reference and cognition: Experimental and computational cognitive modeling studies on reference processing in Dutch and Italian.*
164. Johannes Bjerva (2017). *One Model to Rule them all. Multitask and Multilingual Modelling for Lexical Analysis: Multitask and Multilingual Modelling for Lexical Analysis.*
165. Dieke Oele (2018). *Automated translation with interlingual word representations*.
166. Lucas Seuren (2018). *The interactional accomplishment of action.*
167. Elisabeth Borleffs (2018). *Cracking the code - Towards understanding, diagnosing and remediating dyslexia in Standard Indonesian.*
168. Mirjam Günther-van der Meij (2018). *The impact of degree of bilingualism on L3 development English language development in early and later bilinguals in the Frisian context*.
169. Ruth Koops van 't Jagt (2018). *Show, don't just tell: Photo stories to support people with limited health literacy.*
170. Bernat Bardagil-Mas (2018). *Case and agreement in Panará*.
171. Jessica Overweg (2018). *Taking an alternative perspective on language in autism*.
172. Lennie Donné (2018). *Convincing through conversation: Unraveling the role of interpersonal health communication in health campaign effectiveness.*
173. Toivo Glatz (2018). *Serious games as a level playing field for early literacy: A behavioural and neurophysiological evaluation.*
174. Ellie van Setten (2019). *Neurolinguistic Profiles of Advanced Readers with Developmental Dyslexia*.
175. Anna Pot (2019). *Aging in multilingual Netherlands: Effects on cognition, wellbeing and health*.
176. Audrey Rousse-Malpat (2019). *Effectiveness of explicit vs. implicit L2 instruction: a longitudinal classroom study on oral and written skills*.
177. Rob van der Goot (2019). *Normalization and Parsing Algorithms for Uncertain Input*.
178. Azadeh Elmianvari (2019). *Multilingualism, Facebook and the Iranian diaspora*.
179. Joëlle Ooms (2019). *"Don't make my mistake": Narrative fear appeals in health communication.*
180. Annerose Willemsen (2019). *The floor is yours: A conversation analytic study of*

teachers' conduct facilitating whole-class discussions around texts.

181. Frans Hiddink (2019). *Early childhood problem-solving interaction: Young children's discourse during small-group work in primary school*.
182. Hessel Haagsma (2020). *A Bigger Fish to Fry: Scaling up the Automatic Understanding of Idiomatic Expressions*.
183. Juliana Andrade Feiden (2020). *The Influence of Conceptual Number in Coreference Establishing: An ERP Study on Brazilian and European Portuguese.*
184. Sirkku Lesonen (2020). *Valuing variability: Dynamic usage-based principles in the L2 development of four Finnish language learners*.
185. Nathaniel Lartey (2020). *A neurolinguistic approach to the processing of resumption in Akan focus constructions*.
186. Bernard Amadeus Jaya Jap (2020). *Syntactic Frequency and Sentence Processing in Standard Indonesian.*
187. Ting Huang (2020). *Learning an L2 and L3 at the same time: help or hinder?*.
188. Anke Herder (2020). *Peer talk in collaborative writing of primary school students: A conversation analytic study of student interaction in the context of inquiry learning*.
189. Ellen Schep (2020). *Attachment in interaction: A conversation analytic study on dinner conversations with adolescents in family-style group care*.
190. Yulia Akinina (2020*). Individual behavioural patterns and neural underpinnings of verb processing in aphasia*.
191. Camila Martinez Rebolledo (2020). *Comprehending the development of reading difficulties in children with SLI*.
192. Jakolien den Hollander (2021). *Distinguishing a phonological encoding disorder from Apraxia of Speech in individuals with aphasia by using EEG*.
193. Rik van Noord (2021). *Character-based Neural Semantic Parsing*.
194. Anna de Koster (2021). *Acting Individually or Together? An Investigation of Children's Development of Distributivity*.
195. Frank Tsiwah (2021). *Time, tone and the brain: Behavioral and neurophysiological studies on time reference and grammatical tone in Akan*.
196. Amélie la Roi (2021). *Idioms in the Aging Brain*.
197. Nienke Wolthuis (2021). *Language impairments and resting-state EEG in brain tumour patients: Revealing connections*.
198. Nienke Smit (2021). *Get it together: Exploring the dynamics of teacher-student interaction in English as a foreign language lessons*.
199. Svetlana Averina (2021). *Bilateral neural correlates of treatment-induced changes in chronic aphasia*.
200. Wilasinee Siriboonpipattana (2021). *Neurolinguistic studies on the linguistic expression of time reference in Thai*.
201. Irene Graafsma (2021). *Computer programming skills: a cognitive perspective*.
202. Pouran Seifi (2021). *Processing and comprehension of L2 English relative clauses by Farsi speakers*.
203. Hongying Peng (2021). *A Holistic Person-Centred Approach to Mobile-Assisted Lan-*

guage Learning.

204. Nermina Cordalija (2021). *Neurolinguistic and psycholinguistic approaches to studying tense, aspect, and unaccusativity*.
205. Aida Salčić (2021). *Agreement processing in Dutch adults with dyslexia*.
206. Eabele Tjepkema (2021). *Exploring content-based language teaching practices to stimulate language use in grades 7 and 8 of Frisian trilingual primary education*.
207. Liefke Reitsma (2021). *Bilingualism and contact-induced language change: Exploring variation in the Frisian verbal complex*.
208. Steven Gilbers (2021). *Ambitionz az a Ridah: 2Pac's changing accent and flow in light of regional variation in African-American English speech and hip-hop music*.
209. Leanne Nagels (2021). *From voice to speech: The perception of voice characteristics and speech in children with cochlear implants*.
210. Vasilisa Verkhodanova (2021). *More than words: Recognizing speech of people with Parkinson's disease*.
211. Liset Rouweler (2021). *The impact of dyslexia in higher education*.
212. Maaike Pulles (2021). *Dialogic reading practices: A conversation analytic study of peer talk in collaborative reading activities in primary school inquiry learning*.
213. Agnes M. Engbersen (2022). *Assisting independent seniors with morning care: How care workers and seniors negotiate physical cooperation through multimodal interaction*.
214. Ryssa Moffat (2022). *Recognition and cortical haemodynamics of vocal emotions — an fNIRS perspective*.
215. Diane Mézière (2022). *Using eye movements to develop an ecologically-valid AI measure of reading comprehension*.
216. Ann-Katrin Ohlerth (2022). *Improving preoperative nTMS with a dual-task protocol - the contribution of action naming to language mapping: The contribution of Action Naming to language mapping*.
217. Masha Medvedeva (2022). *Identification, Categorisation and Forecasting of Court Decisions*.
218. Jidde Jacobi (2022). *Coordination and Timing of Speech Gestures in Parkinson's Disease.*
219. Dorothée Hoppe (2022). *Linear Order in Language: An Error-Driven Learning Account.*
220. Janine Strandberg (2022). *Phonetic and lexical variation in Finland-Swedish.*
221. Wim Gombert (2022). *From "learning to use" towards "using to learn"? Long-term effects of structure-based versus dynamic usage-based programs for French.*
222. Marlon Ruwette (2023). *Reading promotion of fiction and nonfiction texts; Determinants of elementary school teachers' reading promotion behavior and their students' leisure time reading.*
223. Ahmet Üstün (2023). *Multilingual Learning and Adaptation for Neural Language Models.*
224. Suzan Dilara Tokaç (2023). *An investigation of heritage language speakers of Turkish: evidentiality processing and the effects of language experience.*

225. Marita K. Everhardt (2023). *Non-native prosody perception in cochlear implant-simulated speech.*
226. Jeroen Steenbakkers (2023). *Speelse toewijding. Een pedagogisch-didactisch onderzoek naar schrijfstijl- en formuleeronderwijs in klas 3 en 4 havo/vwo.*
227. Nienke Meulman (2023). *Second language processing: electrophysiological studies and data analysis methods.*
228. Antje Sien Lisanne Knol (2023). *Doing therapies: On the tangibility of theory in interactional practices.*
229. Sofia Bimpikou (2023). *Inside characters' minds: The role of reports in narrative perspective taking.*
230. Pauline Cuperus (2023). *Aphasia therapy software: Research, development, and implementation.*
231. Elizabeth Heredia Murillo (2023). *The semantics and the L1 acquisition of Spanish scalar (focus) particles: the case of (ni) siquiera and incluso.*
232. Mara van der Ploeg (2023). *Language learning never gets old: learning a new language in later life.*
233. Martijn Bartelds (2023). *Representing Low-Resource Languages and Dialects: Improved Neural Methods for Spoken Language Processing.*
234. Yingjie Wang (2023). *Unpacking the label "Made in China": A Critical Discourse Analysis of Chinese and US news discourse.*
235. Prajit Dhar (2024). *Improving Neural Machine Translation of Languages with Little Data and Rich Morphology.*
236. Effy Ntemou (2024). *Effects of linguistic properties in preoperative language mapping: Insights from VLSM and navigated TMS.*
237. Huiyuan Lai (2024). *Neural Text Rewriting: Style Transfer, Figurative Language, And Beyond.*
238. Chunliu Wang (2024). *Multilingual Semantic Parsing and Generation with Neural Models.*
239. Lukas Edman (2024). *The Little Data That Could: Making the Most of Low-Resource Natural Language Processing.*
240. Jelle Brouwer (2024). *Learning for life: Investigating the potential of language learning as a non-pharmacological intervention in older adults with (past) depression.*
241. Daan van Soeren (2024). *Chronicle of a Linguistic Form Foretold: How human efficiency and prosodic factors explain the distribution of labials and coronals in Spanish, English, and Dutch.*
242. Wietse de Vries (2024). *Evaluation and adaptation of neural language models for under-resourced languages.*
243. Suzanne Dekker (2024). *Multilingualism in Heads, Hearts, and Hands: Attitudes towards the development, implementation, and evaluation of multilingual approaches in Frisian primary schools.*
244. Cheyenne Svaldi (2024). *An investigation of language learning an processing in children with developmental language disorder and pediatric cerebellar tumor survivors.*

245. Gosse Minnema (2024). *Perspective matters: framing events in language and society.*
246. Solveig Castelli (2024). *The planning and preparation of words for language production.*
247. Luc Koning (2024). *Redzaamheidslezen.*
248. Lourens Visser (2024). *Adverbs of degree in early West Germanic.*
249. Floor van den Berg (2024). *Introducing bilingual experience as a healthy cognitive aging tool.*
250. Sara Shoghi Javan (2024). *The Impact of Visual Contextual Information and Verb Argument Structure on Syntactic Parsing Preferences: A Behavioral-Electrophysiological Investigation of Ambiguous Structures in Dutch.*
251. Andreas Hiemstra (2025). *Morphosyntactic Transfer from German and English into Dutch in Third Language Acquisition: A Study on Dutch Superlatives and Two-Verb Clusters.*
252. Joke van Balen (2025). *Making Room for Subjectification during Classroom Discussions. A Conversation Analytic Study of Interactional Practices in the Dutch Language Classroom.*
253. Raoul Buurke (2025). *Frisian and Low Saxon in flux.*

GRODIL
Center for Language and Cognition Groningen (CLCG)
P.O. Box 716
9700 AS Groningen
The Netherlands