

University of Groningen

Examining maintenance and shift in regional language populations

Buurke, Raoul Sergio Samuel Jan; Bartelds, Martijn; Knooihuizen, Remco; Wieling, Martijn

Published in:
Linguistic Minorities in Europe Online

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

Document Version
Early version, also known as pre-print

Publication date:
2023

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

Citation for published version (APA):

Buurke, R. S. S. J., Bartelds, M., Knooihuizen, R., & Wieling, M. (2023). Examining maintenance and shift in regional language populations: the case of Frisian and Low Saxon in the Netherlands. Manuscript submitted for publication.

Copyright

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

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

Take-down policy

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

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

Examining maintenance and shift in regional language populations: the case of Frisian and Low Saxon in the Netherlands

Raoul Buurke*
Rijksuniversiteit Groningen

Remco Knooihuizen
Rijksuniversiteit Groningen

Martijn Bartelds
Rijksuniversiteit Groningen

Martijn Wieling
Rijksuniversiteit Groningen

December 4, 2023

*Corresponding author: raoul.buurke@rug.nl.

Abstract

Language questionnaires can be used to determine the size of linguistic communities within an area, but these often suffer from bias. In this paper, we aim to investigate approaches to compensate for this bias. Specifically, we try to obtain reliable estimates of the speaker count of two regional languages in the Netherlands (i.e., Frisian and Low Saxon). We distributed a language questionnaire about a range of topics (e.g., language use, language proficiency, intergenerational transfer, and the language learning context of respondents) through an existing large-scale longitudinal study. This resulted in 38,500 respondents across the three northern provinces of the Netherlands (i.e., the provinces of Fryslân, Groningen, and Drenthe) where the two regional languages are spoken.

The relative prevalence of dialect speakers across the sample appeared to be unrealistically large, which indicated that the sample seemed to suffer from sampling bias. Initially, we applied post-stratification to account for differences between ratios in the sample and the population (e.g., for sex, age, domicile population density, and educational attainment). As this proved to be insufficient, we focused on an intergenerational transmission approach instead. Earlier usage estimates were used as reference points, and we estimated regional language use for the generations that followed the earlier generations.

The results show that the use of Low Saxon is declining with a total of about 350,000 speakers in 2021 (about 32% of the population), whereas Frisian use appears stable with a total of about 250,000 speakers (42% of the population). As these estimates seem plausible when compared to other speaker counts, we conclude that our intergenerational estimation approach may be a useful tool to obtain speaker estimates in case the necessary information is available.

Keywords— regional language use, intergenerational transmission, Frisian and Low Saxon, language questionnaire methodology, sampling bias

1 Introduction

Surveys and questionnaires are essential tools in academic and governmental research because they can offer valuable insights into societal patterns. Surveys conducted by governmental agencies can have especially substantial consequences for society, but ensuring accurate and representative results poses a significant challenge. This article delves 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, as it appears to be inadequate for the case study presented here. We propose a novel approach based on using prior surveys in combination with generational transmission to overcome the issue of sampling bias.

Ó Riagáin (2018: 5) notes that there are “no universally accepted survey measures of key sociolinguistic concepts such as language proficiency, language

use, or language attitudes". This 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 different interpretations of the questionnaire questions by the informants. Externally, a non-representative sample may be obtained due to a lack of guidelines on how to reach the relevant social groups.

The internal bias of a questionnaire can be minimized by asking questions most reliably, but this is notoriously difficult for language questionnaires ([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 speaker of a specific language, and they tend to be reluctant to classify themselves as one ([Moore, Pietikäinen & Blommaert 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). Moreover, 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](#): 52). A less ambiguous question is whether informants use the language at home, which is a closed question (i.e., yes or no). This is also likely to be a better determinant of language maintenance and shift, as it reflects the structural embedding of a language in everyday life. This metric may be conservative because people may use the language only in contexts other than the home (e.g., with friends or colleagues), especially if someone's partner does not use the regional language. In addition, the exact wording still matters, as participants may give different answers depending on whether they are thinking of speaking the language with their spouse or with their children. (cf. [Driessen 2012](#)).

Questions relating to self-perceived language proficiency and language use at home are both of interest and reported in this study because both are informative about regional language use. They differ in how reliably they are likely to be answered by participants, but they also differ in what they measure at a higher level. If a language is used at home, this suggests a structural embedding of the language in someone's life, but this is not necessarily the case when someone perceives themselves to be highly proficient in a particular language. The proficiency metrics indicate something about someone's knowledge of a language, but it is possible that someone can speak in a language and chooses not to do so. For example, the majority language may be strongly preferred as soon as someone who does not speak the minority language joins a conversation. In that case, the minority language is used much less than can be expected based on speaker numbers. Similarly, a person who can speak the regional language, potentially even very well, may have moved to a different area, resulting in a reduced usage of the regional language. It is therefore possible to see regional language use at home as a type of lower bound of in how many people's lives the regional language plays an active role, while self-reported speaking proficiency may be seen

as an upper bound.

In sum, regional language use can be difficult to estimate due to different interpretations of the most logical questions for respondents, as well as due to the many types of sociolinguistic situations that need to be covered. With this in mind, we want to explore whether it is possible to estimate regional language use based on a survey. Moreover, we want to find out which factors make such an effort more difficult and what can be done to remedy issues that arise.

2 Lifelines questionnaire

In this study, we explore to what extent issues of representativeness arise as we attempt to establish a regional language speaker count for regional languages in the northern Netherlands. This region is considered an ideal laboratory for multilingual research because there is considerable language variation in the three provinces of Fryslân, Groningen, and Drenthe, while only a relatively small part of the total Dutch population resides here (i.e., approximately 10%). In addition to the official Dutch standard language, these provinces are home to speakers of Low Saxon and Frisian (see Figure 0). Dialects that are a mix of Hollandic and Frisian are also in use in Fryslân, such as Bildts and Town Frisian.



Fig. 0: Illustrative map of main Frisian and Low Saxon areas in the Netherlands, with the names of relevant provinces. The blue area with diagonal lines and the green area with horizontal lines is where Frisian and Low Saxon are in use, respectively. The red squares indicate mixed dialects of Frisian and Low Franconian Hollandic.

In other scenarios where 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](#)). Alternatively, 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 biobank ([Scholtens et al. 2015](#)). Lifelines focuses on collecting data that can be used to study healthy aging, but it is also possible to use the existing infrastructure for specific add-on studies.

More than 167,000 participants have contributed their data to Lifelines, which translates to almost 10% of the northern population in 2022. The cohort has been followed over time and across different generations from 2006 onward, making it possible to estimate intergenerational transfer rates of the regional languages. We designed a questionnaire about regional language use (e.g., where it is used, self-perceived speaking proficiency, and regional language attitudes), which was added to Lifelines. The questionnaire was then implemented and distributed by the Lifelines team itself, which means that we did not have direct contact with the participants. After the data collection was finished by the Lifelines team, we obtained access to the data.

The data collection for our study was undertaken by the Lifelines team in November 2021. There were approximately 132,000 active participants in the Lifelines participant pool to which our questionnaire was sent out. Around 38,500 participants responded to the questionnaire, which means that the response rate was around 29%. About 14,500 respondents lived in Fryslân, about 12,000 in Groningen, and about 10,500 in Drenthe. The questionnaire was also filled in by respondents who no longer lived in the North (approximately 1,500). These participants were left out of analyses because our analysis focuses 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 participants presents a unique opportunity to reach many regional language users. The participant pool is not entirely representative of the northern population in the Netherlands, however. 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 the lower-educated and rural population ([Driessen 2005](#); [Goeman & Jongenburger 2009](#)).

On top of these general issues of representativeness in Lifelines, it became clear upon inspection of the collected data that the relative number of users of regional languages in the sample was disproportionately high. From the respondents, around 63% in Groningen, 65% in Drenthe, and 70% in Fryslân indicated they were able to speak the most prominent regional language (i.e., Low Saxon in Groningen and Drenthe and Frisian in Fryslân). These are likely overestimates

Original core part of recruitment message in Dutch:

Het spreken van twee talen kan op latere leeftijd positief effect hebben op het geheugen en concentratie. Onderzoeker [*naam van onderzoeker*] (*[naam van 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 researcher*] (*[name of university]*) wants to understand the cognitive benefits of bilingualism, especially 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.

Fig. 1: Core part of the recruitment message to potential Lifelines respondents.

when we compare these values to recent speaker counts, especially for Low Saxon ([Provincie Fryslân 2020](#); [Versloot 2021](#)). The sample is therefore not comparable to the population of the three northern provinces, and estimates based on the data need to be amended in some way.

While we cannot be certain about the cause of the overrepresentation bias of regional language users, we suspect that the balance of the information in the message to potential respondents discouraged self-perceived monolinguals and non-dialect speakers. The core part of the recruitment message without salutation and practical information (see [Figure 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, but this was likely insufficient to encourage this group of potential respondents to participate. If a potential respondent did not participate within 14 days, the Lifetimes team sent out a reminder. This probably ameliorated the issue, but insufficiently so because the provided information remained the same.

3 Post-stratification

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](#): 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., $50\% \div 40\%$). Likewise, the percentage for women should be multiplied by 0.83 (i.e., $50\% \div 60\%$). Estimates can be further improved when marginal distributions are available, for example of educational attainment by age group.

For the post-stratification approach, it is necessary to obtain reliable population information for the provinces of Fryslân, Groningen, and Drenthe. Statistics Netherlands is a governmental agency that collects such population statistics and provides them online. We apply post-stratification here for a few key factors that are available from Statistics Netherlands and possibly influence regional language use. These include the age, sex, domicile population density (i.e., how urban the location is where a respondent lives), and educational attainment of respondents. We adjust our sample ratios accordingly to assess whether this yields more realistic estimates of regional language use.

For the age variable, we divided the data into categories of ages 0 to 30, 31 to 60, and 61 to 90. The scale of domicile population density is divided into five levels, and it is based on how many postal codes there are in a square kilometer (i.e., a higher density indicating a more urban area; [Centraal Bureau voor de Statistiek 2023c](#)). 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).

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. 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 1](#). The other sample ratio multiplication factors can be found in the supplementary material. The variation around the unadjusted sample means is at most 2% when averaging the deviation of the variables sex, domicile population density, and educational attainment for each province.

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). Given that these percentages are still too high in the most extreme case (i.e., adjusting only for age) when compared with other published usage estimates, we need to accept that estimates that rely on the sample ratios between regional language users and non-users are untrustworthy for population inference. There are no marginal distributions available from Statistics Netherlands across all variables of interest, so we cannot apply post-stratification across all variables simultaneously. Even if these marginal distributions were available, however, the estimates would only become higher and

therefore even less realistic.

Table 1: Multiplication factors for the age variable to match the sample ratios to population ratios in 2021, as reported by Statistics Netherlands.

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

4 Reference works and intergenerational transmission

4.1 Method

The post-stratification approach was shown to be insufficient for obtaining realistic estimates of regional language use. Alternatively, it may also be possible to combine data from different sources, alleviating the bias present in the current sample. For example, if a suitable and more reliable reference work is available, and information is present about intergenerational transmission within a sample of regional language speakers, it is possible to estimate how many children acquired the language in the next generations (i.e., the generations after the ones reported in the reference work).

4.1.1 Reference works

In this study, we take [Bloemhoff \(2005\)](#) as a reference work for Low Saxon, and [Klinkenberg, Jonkman & Stefan \(2018\)](#) as a reference work for Frisian. The regional language use (i.e., self-perceived 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.¹ 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 \(2021: 9\)](#) expressed doubts about the Low Saxon reference work, because the reported percentages seem high. We share these doubts, but we also acknowledge that there is no good alternative speaker count available. Moreover, the sampling procedure of the reference work seems to be trustworthy (i.e., random phone number calling by an external company) and we cannot explore in detail what could have caused relatively high estimates. For this reason, we think the most reasonable

¹While regional language speakers are overrepresented in our sample, there is no reason to assume the regional language speakers of a given age represent a biased sample.

point of departure for this method is still [Bloemhoff \(2005\)](#). For the reference work of Frisian, we assume that the estimates are accurate because the sampling procedure seems sound as well: 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, Jonkman & Stefan \(2018\)](#) acknowledge that there are some representation issues (because younger people were underrepresented), but they adjusted their estimates using post-stratification.

As noted before, the exact formulation of questions can influence the interpretation of participants. Fortunately, the exact formulations are available for the different studies, which we translate to English here 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 as presented in the Appendix (Table 6). For the province of Groningen, the regional language variant was “Gronings or Westerkwartier-Gronings”, whereas for Drenthe it was “Drents or Stellingwerfs-Drents”. [Klinkenberg, Jonkman & Stefan \(2018\)](#) used the following formulation (translated from Frisian): “Can you speak Frisian?”. The range of options (i.e., very easily to completely not) is similar to [Bloemhoff \(2005\)](#), but they did not provide the names of more local and regional variants of Frisian.

[Bloemhoff \(2005\)](#) used the following formulation for the use of Low Saxon at home: “What do you mainly speak at home?”, with the options again presented in the Appendix (Table 7). For the use of Frisian at home, [Klinkenberg, Jonkman & Stefan \(2018\)](#) 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 did not give such a specification of with whom the regional language is used at home. The percentages 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 percentages that were reported for Low Saxon about language use at home were intended and likely interpreted as language use with someone’s spouse as well. This interpretation seems the most likely one, as there are no further specifications in the question. The Frisian reference data (for both questions discussed in this section) are presented in Table 11 in the Appendix.

Additional remarks on Low Saxon

For the Low Saxon reference work, a few caveats need to be addressed first. 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, we are interested in (1) the self-perceived speaking proficiency of Low Saxon and (2) the use of mainly Low Saxon or Dutch, or both, at home. These reference values are described for the entire Low Saxon area across age groups in Tables 6 and 7 in the Appendix. As the data from Lifelines only cover the Low Saxon provinces of Groningen and Drenthe², we would need the age-based usage values for both provinces separately. While these are not available in Bloemhoff (2005), usage data for both provinces as a whole is present. From this information, it is clear that a higher proportion of speakers 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 upwards. We do this by comparing the average (not separated by age group) percentages across the whole Low Saxon area for each individual level of the variables to the corresponding percentage of each individual region. For example, if the average percentage of ‘speaking Low Saxon very well’ is 20% for the whole Low Saxon area, but it is 25% for Groningen, then the age-group percentages for the whole Low Saxon area for ‘speaking Low Saxon very well’ are multiplied by 1.25 (i.e., $25 \div 20$) to yield the age-group percentages for Groningen. By correcting the metric levels this way (i.e., also for all levels of self-perceived speaking proficiency and language use at home), we obtain adjusted age-group estimates for both Groningen and Drenthe from the Low Saxon reference work. We do not provide all these calculated percentages here, but they can be found in the supplementary material. It is important to 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.

4.1.2 Using 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. As indicated, we rely on Bloemhoff (2005) as reference work for Low Saxon, and Klinkenberg, Jonkman & Stefan (2018) as reference work for Frisian. From these reference works, we use the reported (or computed, see above) percentages from 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 necessary main component in our method is the intergenerational transmission rate of the generations for which speaker count information is reported in the reference works. Because of the large (apparent) time depth in both the Lifelines data and the reference works, it is possible to match the information from both sources by the birth year of generations of interest. The Life-

²We have left out the Low-Saxon-speaking municipalities of Weststellingwerf and Ooststellingwerf located in Fryslân from the data from Lifelines because only the provinces of Groningen and Drenthe are specifically covered by the reference works.

lines participants were prompted for intergenerational transmission using the following formulation: “Do you speak (or did you speak) your regional language/dialect with your child(ren)?”. If they indicated that they spoke the regional language with at least one child, this was taken 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 (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., $100\% \div 80\%$) 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 (i.e., “How did you learn your regional language/dialect?”), and use this to adjust the transmission rate upward.

These components can be 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 the division of 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 age at the first childbirth of women in the Netherlands in recent decades ([Centraal Bureau voor de Statistiek 2021b](#)).

$$\begin{aligned} \text{metric estimate}_{(x,y)} = & \text{metric estimate in reference work}_{(x+z, y+z)} \\ & \times \text{parental transmission rate}_{(x+z, y+z)} \\ & \times \frac{100\%}{\text{rate of acquisition through parents}_{(x,y)}} \end{aligned} \quad (1)$$

As an example, consider the 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 then look at the data from that generation in Lifelines. There is a difference of 18 years between the data collection for the reference work and our Lifelines questionnaire, so we look at the transmission rate of the speakers aged between 48 and 68 in our Lifelines dataset. Their children are approximately 30 years younger, and so we also look at the speakers between 18 and 38 in Lifelines that indicate that they have acquired the regional language outside the home. If we assume a parental transmission rate of 50% and an adjustment ratio of 1.25 for language learning context (i.e., 80% learned it at home, but 20% learned it elsewhere), we then obtain an estimate of about 44% (i.e., $0.70 \times 0.50 \times 1.25 \times 100\%$) for the children of the reference generation in the reference work (i.e., those currently aged between 18 and 38).

4.1.3 Validation of results

Of course, the calculated estimates need to be evaluated. For this comparison, 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 \(2012\)](#) who investigated the language in which parents and their children interacted, and an extra dataset obtained from Statistics Netherlands (SN; *Centraal Bureau voor de Statistiek*; CBS) based on a regional language usage study carried out in 2019 ([Centraal Bureau voor de Statistiek 2019](#); [Schmeets & Cornips 2022](#)).

As the data collection periods differed for the individual datasets, it is first necessary to ensure all age groups are converted to birth year ranges. In case the birth year range in the dataset we compare with differs from that of our study, we estimate the usage in the birth year range used in our study by proportionally combining information from multiple birth year ranges. As an example, consider evaluating the usage estimate we obtained in our study for the people born between 1973 and 1993. For this purpose, we could compare our estimate to data from Statistics Netherlands collected in 2019. However, this dataset only contains usage data for people born between 1980 and 2004 (for which the regional language use was 16.9%), and people born between 1960 and 1979 (a usage of 29.7%). Out of the 21 years covered by our age group, the 1980-2004 group overlaps 14 years (1980-1993), whereas the 1960-1979 group overlaps seven years (1973-1979). We then estimate the regional language usage percentage in the data of Statistics Netherlands for 1973-1993 by multiplying the usage percentage of the younger group with $2/3$ ($16.9\% \times 2/3$), and the older group with $1/3$ ($29.7\% \times 1/3$) and adding these two values together ($\approx 21\%$). In case data from multiple groups cannot be combined (e.g., due to an age cut-off in the dataset), we use the data of the group with the largest overlap.

4.2 Results

4.2.1 Groningen and Drenthe

It is now possible to construct estimates based on the Lifelines data and the reference works. The data from the Low Saxon reference work were collected around 2003, so the age groups in the reference work were 18 years older at the time of the Lifelines survey in 2021. The components necessary for the estimates for Low Saxon are reported in Table 10 of the Appendix. Using these adjustments, we obtain our final estimates of the percentages of people who indicate speaking Low Saxon and speaking both Low Saxon and Dutch at home in 2021. These results are shown in Table 2. The established estimates are evaluated by comparing them to other estimates. For the self-perceived speaking metric, we use the percentages from [Bloemhoff \(2005\)](#), while we compare our findings with those of [Driessen \(2012\)](#) and Statistics Netherlands for the speaking at home metric.

The percentages reported in the works we use for comparison, adjusted in line with Section 4.1.3, are presented in Table 3. Note that for the evaluation, the reported percentages from the reference work and Statistics Netherlands are

averaged across Groningen and Drenthe³, while the percentages from [Driessen \(2012\)](#) are for the entire Low Saxon region, and therefore likely an underestimate for Groningen and Drenthe (as the regional language usage in those provinces was higher compared to the others; [Bloemhoff 2005](#)). The birth years 1994 to 2015 were not available in the reference work of [Bloemhoff \(2005\)](#) and [Driessen \(2012\)](#), but these years were present in the data from Statistics Netherlands.

Table 2: Estimated percentages for speaking and using Low Saxon (at home), by age group, for Groningen (Gron.) and Drenthe.

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

Table 3: Reference percentages for speaking and using Low Saxon (at home). Italicized values are adjusted based on multiple generations in the Low Saxon reference work (see Section 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
1994 - 2015	-	-	17	-	-
1973 - 1993	71	53	21	39	15
1952 - 1972	75	58	33	46	23

There is a substantial gap between the estimates and what would be expected based on the Low Saxon reference work (see Tables 8 and 9 in the Appendix), although the differences are most pronounced for the comparisons with the Low Saxon reference work itself (see Table 3). In that case, the differences for different generations range from 20% to 32% for the self-perceived speaking proficiency metric, from 30% to 39% for the language use at home metric, and from 6% to 16% for the regional language use at home together with Dutch metric. The differences compared to the other works used for comparison are between 2% and 11% (i.e., compared to Statistics Netherlands) and 7% and 15% (i.e., compared to [Driessen 2012](#)), specifically for the metrics of regional language use at home and regional language use at home together with Dutch. Moreover, except for the values obtained from [Driessen \(2012\)](#), the estimated percentages obtained

³The percentages from the reference work were not split across the two regions, as these are highly similar as shown in Table 8 in the Appendix.

through our method are lower than those reported in the works used for comparison. Possible explanations for these patterns are explored in Section 5.

4.2.2 Fryslân

For Frisian, we rely on [Klinkenberg, Jonkman & Stefan \(2018\)](#) as a reference work (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 Fryslân. The municipalities of Weststellingwerf and Ooststellingwerf in Fryslân are not included in the reference work and the Lifelines data because these areas are traditionally Low Saxon areas instead of Frisian ones. The procedure to obtain estimates and make comparisons is the same as for Low Saxon: the reference percentages are multiplied by the components obtained from the Lifelines data for the next generation of Frisian speakers. The estimates for Frisian are presented in Tables 4 and 5.

Table 4: Estimated percentages for speaking and using Frisian (at home) in Fryslân (excluding Oost- and Weststellingwerf).

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

Table 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 4.1.3).

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

It is immediately clear that Frisian language use declines to a lesser degree than that of Low Saxon, and it seems to be the case that the use of Frisian at home is even increasing slightly. The difference in percentages between the estimates and the works used for comparison is also less substantial for the Frisian case than the Low Saxon one. In the Frisian case, however, the differences between the estimates 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, Jonkman & Stefan \(2018\)](#) report noticeably higher percentages of Frisian language use at home than [Driessen \(2012\)](#) and Statistics Netherlands, but the opposite is the case when the estimates are compared to the reported percentages of the reference work.

It is now possible to calculate absolute speaker counts. Statistics Netherlands provides the population count for each age in 2021 ([Centraal Bureau voor de Statistiek 2021a](#)). By summing the population count for each age range in each generation in Tables 2 and 4, we can obtain the absolute number of language users by multiplying the population counts with the estimated percentages. We do not provide these calculations here, but they can be found in the supplementary material. For Groningen, this means that we estimate that there are about 183,000 people (out of a total of 589,762 people; 31%) who indicate being able to speak Low Saxon and about 65,000 (11%) who indicate mainly using the regional language at home. For Drenthe, these values are, respectively, about 180,000 (out of a total of 497,162 people; 36%) and 75,000 (15%). In Fryslân (excluding the municipalities of Weststellingwerf and Ooststellingwerf), there are about 251,000 people (out of a total of 597,019 people; 42%) who indicate being able to speak Frisian and about 195,000 people (33%) who indicate using the language at home.

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 therefore different types of biases may be introduced. In this study, we set out to estimate the number of speakers of Frisian and Low Saxon in the northern Netherlands, but there was an overrepresentation of regional language users in our sample that did not align with the established literature. As a solution, we instead estimated the generational decline in the users of these languages by combining data from existing sources and intergenerational transmission data from a new large-scale questionnaire and evaluated this metric. Our findings align with patterns in other studies ([Driessen 2012](#); [Versloot 2021](#)), 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 to be able to speak the regional language (from reasonably to very well) and whether they used the language at home (in addition to or together with Standard Dutch). The former type of question has the potential to be confusing for informants ([Duchêne & Humbert 2018: 9](#)), which might explain why the differences between the estimated values differ substantially from what is reported in other works. For Low Saxon, these differences between the estimated values and the reference work can be up to 32 percentage points for the self-perceived speaking metric and up to 18 percentage points for the speaking at home metric. For Frisian, the differences are overall smaller, with differences up to 15 percentage points for the self-perceived speaking metric and at most 8 percentage points for the speaking at home metric.

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

The intergenerational transfer 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 \(2012\)](#). The overall level remains higher for Frisian, however. Frisian also seems to be used at home slightly more often in younger generations than older ones. This is clear from both [Klinkenberg, Jonkman & Stefan \(2018\)](#) and the data from Statistics Netherlands. It contrasts with the number of self-perceived speakers of Frisian in Fryslân, as that seems to be decreasing slightly. This is not a universally established pattern (cf. [Driessen 2012](#)), but Frisian does seem to be 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, although the Dutch government has also been criticized for its lackluster support of teaching Frisian ([Bayat, Kircher & Van de Velde 2023](#)). Moreover, people who move to Fryslân also tend to learn Frisian ([Kircher, Kutlu & Vellinga 2023](#)). If Frisian is indeed becoming more common again in the home, this may indicate a rare case where the regional language decline has halted, and it may even become more widespread again over time. This is in stark contrast with many other regional languages in the Netherlands and elsewhere in Europe ([Versloot 2021](#); [Auer 2005](#)), 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, Etxebarria & Intxausti 2015](#)).

There are some further consistencies in the data concerning intergenerational transfer and language learning contexts, regardless of the regional language, and these are reported in Tables 10 and 12 in the Appendix. For example, the adjustment ratio between learners within and outside the home seems to be consistently around 1.20 to 1.25. This is equal to around 80% of people indicating to learn the language in the home and around 20% indicating to learn the language outside the home (e.g., through friends, family, or people in the envi-

ronment). There are some small fluctuations between the generations, but there is no clear shift in the main language learning context, which is at home.

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. Moreover, 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). There is therefore some uncertainty surrounding this metric, which could have been resolved with further questions and metadata, but we expect that these scenarios are uncommon enough to 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 speaker count 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 participants. This could not be done in our case, because these parts were handled by the external Lifelines team.

One way to preserve the possibility of distributing a survey without direct control over the communication with participants is to broaden the scope of the survey to topics that concern a larger audience. [Schmeets & Cornips \(2022\)](#) investigated regional languages in the entire Netherlands, but the relevant survey questions were embedded in a broader survey about social cohesion and wellbeing ([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-perceived 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 add-on questionnaire. Moreover, if one part of the survey is about economic factors, then respondents might relate their responses on another part to economic contexts. Part of our survey concerned language attitudes, and respondents might then evaluate their regional languages as less

economically viable (and therefore as less positive).

We use the given percentages of different generations from a single reference work and consequently estimate the metrics for the children of these generations. It is also possible to choose the oldest generation in the data and to estimate the generational decline for multiple generations consecutively, which may be seen as a type of survival analysis (e.g., agent-based modeling of a multilingual community; [Civico 2019](#)). Such an approach is not very attractive in our case, because the entire analysis then depends on a single reference point. Moreover, multiple reference points were available to us and our approach preserves more information. Unless one is certain that a single percentage is properly representative, which we could not be for the Low Saxon data, it is advantageous to not be too reliant on a single reference point.

There is a final caveat that may not generalize to all majority-minority language contexts, but which 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). 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 the survey was 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 variants of Low Saxon (e.g., Gronings and Drents). This 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 Low Saxon due to their perceived dialect classification.

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 transfer the language. This seems to be the case if we inspect the Lifelines data (see [Figures 2 and 3](#) in the Appendix), which show that intergenerational transfer is substantially lower in urban areas for both Low Saxon and Frisian. For Frisian, this could also partially be due to the presence of mixed dialects in towns, which results in more competition for language transfer than in Low Saxon areas. The higher prevalence of Frisian in rural areas is confirmed as well by [Klinkenberg, Jonkman & Stefan \(2018: 90\)](#). Standard Dutch is also increasingly acceptable in all social contexts of both regional languages, as opposed to (close personal) contexts being reserved for the regional language ([Hinskens & Taeldeman 2013: 5](#)). Reporting regional language use averaged across the entire population may give a relatively pessimistic view of the sociolinguistic situation in rural areas, and it is likely useful to treat different geographical areas separately.

The key findings of this work confirm some contrasts between the Frisian and Low Saxon regional languages and their users, but they also highlight how difficult it is to obtain reliable information about these regional languages. While

we are confident about the generalizability of our method and our findings, it is clear that for Low Saxon the estimates would be more reliable if there was no reason to doubt the estimates of the reference work we used (Bloemhoff 2005). It is more 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 be completely avoided. To do so, control over the contact with informants likely has to be prioritized over obtaining a large sample size. 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.

References

- Auer, P. (2005), Europe's sociolinguistic unity, or: A typology of European dialect/standard constellations, in N. Delbecq, J. Van der Auwera & D. Geeraerts, eds, 'Perspectives on Variation', De Gruyter Mouton, Berlin, pp. 7–42.
- 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.
- Belmar, G., Van Boven, C. & Pinho, S. (2020), 'Why do adults decide to learn a minority language? A study of the motivation(s) of potential new speakers of West Frisian', *Linguistic Minorities in Europe Online*.
- Bethlehem, J. G. (2009), *Applied survey methods: a statistical perspective*, Wiley Series in Survey Methodology, Wiley, Hoboken, NJ.
- Bloemhoff, H. (2005), *Taaltelling Nedersaksisch. Een enquête naar het gebruik en de beheersing van het Nedersaksisch in Nederland [Language count Low Saxon. A survey into the use and proficiency of Low Saxon in the Netherlands]*, Sasland.
- Centraal Bureau voor de Statistiek (2019), 'Sociale samenhang en welzijn. [Social cohesion and well-being.]'. Retrieved from <https://www.cbs.nl/nl-nl/longread/statistische-trends/2021/talen-en-dialecten-in-nederland> on 31st of May, 2023.
- Centraal Bureau voor de Statistiek (2021a), 'Bevolking op 1 januari en gemiddeld; geslacht, leeftijd en region [Population on January 1 and on average; gender, age and region]'. Retrieved from <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/70072NED/table?fromstatweb> on 21st of November, 2023.
- Centraal Bureau voor de Statistiek (2021b), 'Kinderen krijgen. [Having children.]'. Retrieved from <https://www.cbs.nl/nl-nl/visualisaties/dashboard-bevolking/levensloop/kinderen-krijgen> on 18th of April, 2023.
- Centraal Bureau voor de Statistiek (2023a), 'Bevolking 15 tot 75 jaar; opleidingsniveau, wijken en buurten, 2021. [Population 15 to 75 years;

- education level, districts and neighborhoods, 2021.]. Retrieved from <https://opendata.cbs.nl/statline/#/CBS/nl/dataset/85372NED/table?ts=1700743566965> on 23rd of November, 2023.
- Centraal Bureau voor de Statistiek (2023b), 'Bevolkingspiramide. [Population pyramid.]'. Retrieved from <https://www.cbs.nl/nl-nl/visualisaties/dashboard-bevolking/bevolkingspiramide> on 19th of April, 2023.
- Centraal Bureau voor de Statistiek (2023c), 'Stedelijkheid (van een gebied). [Urbanity (of an area).]'. Retrieved from <https://www.cbs.nl/nl-nl/onze-diensten/methoden/begrippen/stedelijkheid--van-een-gebied--> on 23rd of November, 2023.
- Christoffersen, M. N. (1987), 'The educational bias of mail questionnaires', *Journal of Official Statistics* **3**(4), 459–464.
- Civico, M. (2019), 'The dynamics of language minorities: Evidence from an agent-based model of language contact', *Journal of Artificial Societies and Social Simulation* **22**(4).
- Delgado-Rodriguez, M. (2004), 'Bias', *Journal of Epidemiology & Community Health* **58**(8), 635–641.
- Driessen, G. (2005), 'In Dutch? usage of Dutch regional languages and dialects', *Language, Culture and Curriculum* **18**(3), 271–285.
- Driessen, G. (2012), Ontwikkelingen in het gebruik van Fries, streektalen en dialecten in de periode 1995-2011 [Developments in the use of Frisian, regional languages and dialects in the period 1995-2011]. Nijmegen: ITS.
- 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.
- 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.
- 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.
- Gorter, D. & Walker, A. G. (2023), 'Frisian as a European Minority language', *Linguistic Minorities in Europe Online*.
- Hinskens, F. & Taeldeman, J. (2013), *Dutch*, Berlin: De Gruyter Mouton.
- Holt, D. & Smith, T. F. (1979), 'Post stratification', *Journal of the Royal Statistical Society Series A: Statistics in Society* **142**(1), 33–46.
- 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.
- Kircher, R., Kutlu, E. & Vellinga, M. (2023), 'Promoting minority language use to foster revitalisation: Insights from new speakers of West Frisian', *Applied Linguistics* p. amad045.
- Klijns, B., Scholtens, S., Mandemakers, J. J., Snieder, H., Stolk, R. P. & Smidt, N. (2015), 'Representativeness of the LifeLines cohort study', *PLOS ONE*

- 10(9), e0137203.
- Klinkenberg, E., Jonkman, R. & Stefan, N. (2018), 'Taal in Fryslân: De volgende generasje [Language in Fryslân: The next generation]'.
 Korkeila, K., Suominen, S., Ahvenainen, J., Ojanlatva, A., Rautava, P., Helenius, H. & Koskenvuo, M. (2001), 'Non-response and related factors in a nationwide health survey', *European Journal of Epidemiology* 17, 991–999.
 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', Springer, pp. 327–343.
 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. Japac, P. J. Lavrakas, M. W. Link & R. L. Sangster, eds, 'Advances in Telephone Survey Methodology', Wiley Online Library, pp. 561–586.
 Moore, R. E., Pietikäinen, S. & Blommaert, J. (2010), 'Counting the losses: numbers as the language of language endangerment', *Sociolinguistic Studies* 4(1), 1–26.
 Morgan, G. (2001), Welsh: A European case of language maintenance, in L. Hinton & K. Hale, eds, 'The Green Book of Language Revitalization in Practice', Brill, pp. 107–118.
 Ó Riagáin, P. (2018), *Measures of Language Proficiency in Censuses and Surveys*, Springer International Publishing.
 Provincie Fryslân (2020), 'De Fryske Taalatlas 2020: Friese taal in beeld [The Frisian Language Atlas 2020: Frisian language in pictures]'.
 Revilla, M. & Ochoa, C. (2017), 'Ideal and maximum length for a web survey', *International Journal of Market Research* 59(5).
 Schmeets, H. & Cornips, L. (2022), 'Taaldiversiteit in Nederland [Language diversity in the Netherlands]', *Taal en Tongval*, 74(1), 75–106.
 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.
 Stefan, M. (2022), *Spoken Frisian: Language contact, variation and change*, LOT.
 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 from <https://zoek.officiëlebezoekingen.nl/kst-24092-5.html> on 24th of November, 2023.
 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.
 Versloot, A. (2021), 'Streektaaldood in de Lage Landen [Regional language death in the Low Countries]', *Taal en Tongval* 72(1), 7–16.

Related articles/sources in LME

Research articles

Appendix

Table 6: Percentages per level of Low Saxon speaking proficiency by age group, as reported in [Bloemhoff \(2005\)](#).

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

Table 7: Percentages of Low Saxon use at home by age group, as reported in [Bloemhoff \(2005\)](#).

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

Table 8: Adjusted percentages of Low Saxon speakers in Groningen and Drenthe, who indicate to speak Low Saxon reasonably well, well, or very well. The data are based on [Bloemhoff \(2005\)](#).

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

Table 9: 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. The data are based on [Bloemhoff \(2005\)](#).

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

Table 10: Adjustment components for Low Saxon speakers in Groningen and Drenthe. These data are based on the Lifelines data.

Birth years of child generation	Generational transfer from parental generation (in %)		Adjustment ratio for children in target generation learning language outside the home instead of in the home	
	Groningen	Drenthe	Groningen	Drenthe
1994 - 2015	31	42	1.28	1.32
1973 - 1993	39	48	1.28	1.25
1952 - 1972	48	54	1.28	1.25

Table 11: Percentages of speaking Frisian and speaking Frisian at home in Fryslân (excluding Oost- and Weststellingwerf). These data are based on [Klinkenberg, Jonkman & Stefan \(2018\)](#).

Birth years	Age group	Speaking Frisian	Speaking Frisian at home
1987 - 2007	≤ 29	66	62
1967 - 1986	30 - 49	72	62
1952 - 1966	50 - 64	69	54
1931 - 1951	65+	69	49

Table 12: Adjustment components for Frisian speakers in Fryslân. These data are based on the Lifelines data.

Birth years of child generation	Generational transfer from parental generation (in %)	Adjustment ratio for children in target generation learning language outside the home instead of in the home
1997 - 2016	69	1.19
1982 - 1996	71	1.22
1961 - 1981	74	1.28

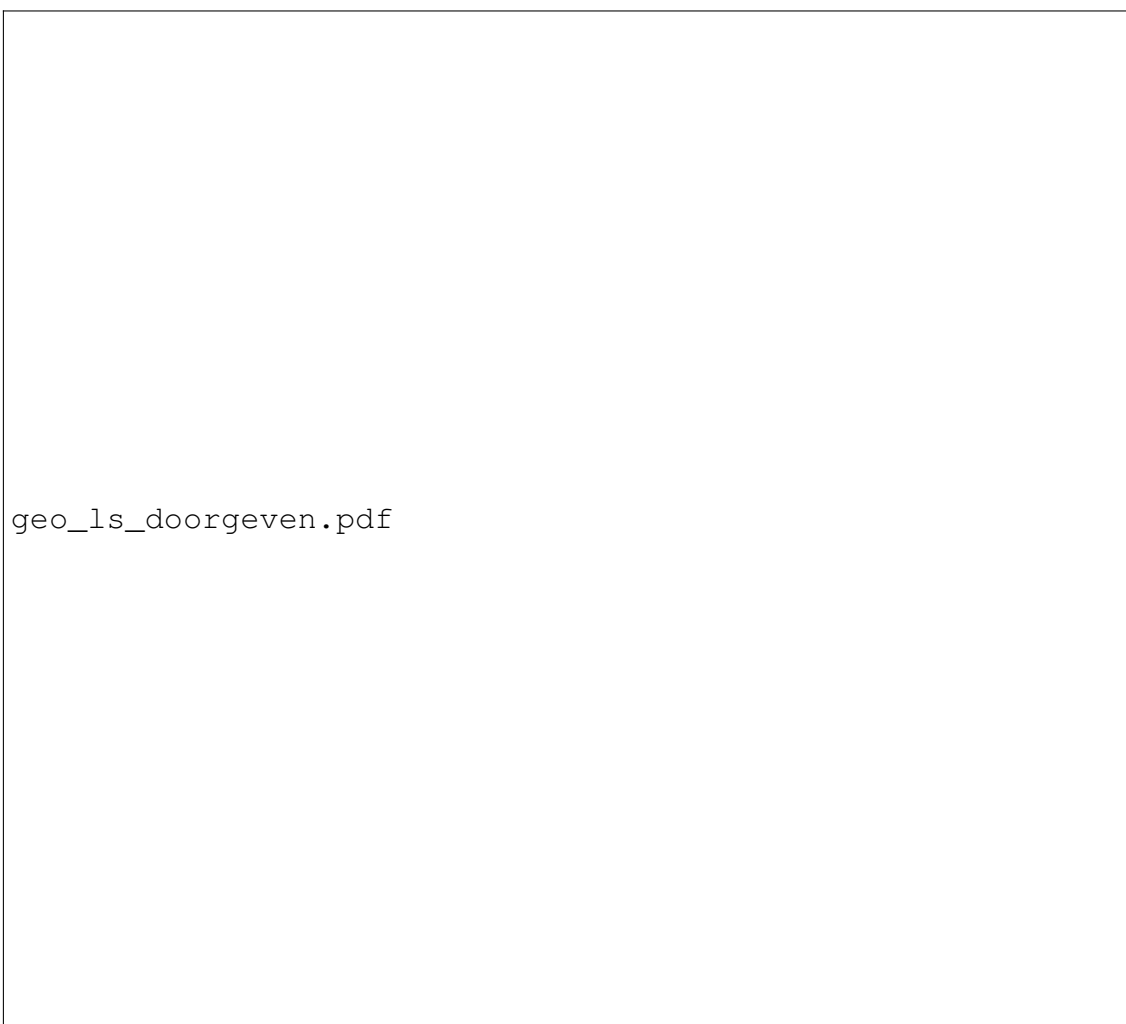


Fig. 2: The geographical distribution of the speakers who transfer Low Saxon to their children. Based on a logistic generalized additive model predicting the likelihood of intergenerational transfer based on a two-dimensional smooth from geographical coordinates. A value of -2, 0, or 2 indicates a probability of respectively 12%, 50%, and 88%.

geo_fr_doorgeven.pdf

Fig. 3: The geographical distribution of the speakers who transfer Frisian to their children. Based on a logistic generalized additive model predicting the likelihood of intergenerational transfer based on a two-dimensional smooth from geographical coordinates. A value of -2, 0, or 2 indicates a probability of respectively 12%, 50%, and 88%.