Nanna Haug Hilton*

Stimmen: A citizen science approach to minority language sociolinguistics

<https://doi.org/10.1515/lingvan-2019-0017>

Abstract: This paper presents the project *Stimmen fan Fryslân* ‘Voices of Fryslân’. The project relies on a smartphone application developed to involve local communities in the creation of speech corpora, particularly of lesser used languages. This paper lays out the scientific and societal context of the project, showcases the smartphone application and gives an overview of the results from the project that attracted more than 15,000 users. Some key methodological issues are considered, and the paper discusses the role of smartphone technology for citizen science in minority language areas while also showing new maps with distributions of lexical and phonological variation in Frisian.

Keywords: citizen science, minority languages, language documentation, smartphones

1 Introduction and background

This paper presents the project *Stimmen fan Fryslân* ‘Voices of Fryslân’, referred to as *Stimmen* for short. The project started in 2017 as a citizen science initiative, with the aim of increasing the availability of lesser-used languages, and multilingual, data for the purposes of linguistic research. Of the estimated 5000–7000 languages of the world at least a third lack proper description (Hammarström and Nordhoff 2011); more than a thousand varieties are only described with a wordlist, a text collection, or not at all. An issue for linguistics as a discipline is that the largest national languages of the world, and particularly the written versions of these (cf. Linell 2005), have formed the empirical foundations upon which we have built most of our theories. Publications within linguistics largely represent research into (speakers of) large languages, and especially English. Nagy and Meyerhoff (2008) conclude, for example, that in 449 publications and presentations in four key outlets for research in sociolinguistics 50–70% of all output concerns the English language, depending on the publication type (see also Smakman 2015; Stanford 2016). Another empirical bias in linguistics is the reliance on data from monolingual populations (cf. Ricento 2013). Nagy and Meyerhoff (2008) find that output in sociolinguistics journals considering more than one language make up between 10 and 30% of all studies, when most estimates of the number of bilinguals across the globe are over 50% (Fromkin, Rodman and Hyams 2018).

The bias in empirical foundations is restrictive, for theoretical development, but also for applications of the scientific work. The creation of teaching material, the development of language technology, as well as forensic (linguistic) practice, all rely on the documentation of languages found across the globe. Currently a challenge for work on multilingual communities, and under-resourced languages, is the lack of available data, and particularly of speech data. While collections of speech can often be found on the Internet, it may not be in a format that lends itself well for research, without annotation and translations, for example, or of poor quality due to recordings made in noisy environments.

A way to increase representation of under-resourced language communities is to engage in “citizen science”, and rely on participation of the general public instead of on scientific researchers only. Citizen science means engaging communities in the scientific endeavour either in the preparatory stages of a project, for data collection, for data processing, for analyses, or for dissemination of results (cf. Bonney et al. 2016). While the label “citizen science” is relatively recent, the practice itself is not. In linguistics the contribution of

*Corresponding author: Nanna Haug Hilton, Centre for Language and Cognition Groningen, University of Groningen, Oude Kijk i/t Jatstraat 26, Groningen, 9712 EK, Netherlands, E-mail: n.h.hilton@rug.nl

the general public as providers of empirical data has a history that goes back centuries. This is particularly true for dialectological research (Wenker 1881) for which speakers of localised varieties have always provided data for the analysis of regional variation in language. However, recent technological developments have made public involvement in linguistic research substantially easier to facilitate. Examples of dialectological surveys being conducted using the Internet are Vaux (2003) and Möller and Elspass (2015). However, smartphone use offers further functionalities and large amounts of language samples as well as observations can be collected. Applications for iPhones or Android have so far been used to collect speech for development of language technology (De Vries et al. 2014), to make high-quality recordings for acoustic phonetic research (De Decker and Nycz 2011), or to collect reported language use for creation of new dialect maps (Leemann and Kolly 2013).

Stimmen was particularly inspired by the language documentation application for smartphones *Aikuma* (Bird et al. 2014) that gives the public a chance to collect speech recordings and translations. Documenting a language with *Aikuma* is made easy by providing speakers with the option of recording themselves or others, and of translating the recordings. Nonetheless, the creation of data is relatively time-consuming for the user. To ensure that *Stimmen* is of added value alongside *Aikuma*, its development focussed on removing the need for translations and written language; reaching a larger audience; and collecting smaller amounts of data per individual. The recording function in the application in *Stimmen* is therefore offered as a gamified picture naming task, rather than a speech-and-translation tool, that can be used by speakers of any variety to record any variety (as long as the researcher adds the particular variety to the list of possible languages), and multiple times, thus allowing for monolingual as well as multilingual recordings.

2 Geographic and sociolinguistic context of *Stimmen*

The *Stimmen* project is funded by the programme *Lân fan Taal* in the European Capital of Culture project for Leeuwarden 2018. Leeuwarden is capital of the multilingual province of Fryslân in the Netherlands, a region in which several (regional and migrant) minority languages are used. Frisian-Dutch bilingualism is widespread in the province: 75% of the inhabitants of the region report being able to speak Frisian (Fryslân 2015), which corresponds to some 485,000 speakers (Centraal Bureau voor de Statistiek 2018). These speakers are all presumed bilingual as secondary schooling is taught partially, or only, in Dutch. The province Fryslân is also home to bilingual mixed languages (results of long-term contact between Dutch and Frisian) ‘Bildts’ and ‘City Frisian’ (Bree 1994; Hoekstra and van Koppen 2000), as well as to varieties of another West-Germanic language family, Low Saxon, spoken along the southern border of the province of Fryslân.

To engage the communities in Fryslân, *Stimmen* includes an additional element of gamification in the smartphone application: a ‘dialect quiz’, made specifically for the Frisian, bilingual mixed languages, and Low Saxon speaking audiences in the region. The Dialect Quiz is a self-reporting dialect task in which the app guesses where someone is from (within the Netherlands) on the basis of answers about the user’s own dialect. Such ‘dialect quizzes’ have previously been employed in sociolinguistic projects (e.g. Leemann and Kolly 2013) and have been highly successful in reaching large audiences.

3 The *Stimmen* smartphone application

The *Stimmen* application consists of the following components: a start screen with a choice of interface language (English, Dutch, or Frisian in 2019. Low Saxon does not have one written standard and was therefore not included as an interface language), a tutorial, a root menu with choices of the components picture game, free speech recording, speech map and dialect quiz, and an ‘about’ section.

3.1 Language options and tutorial

The initial screen in *Stimmen* is a choice of interface language (Figure 1). While there are translations of the interface into other varieties, as of now the three choices in the published version are ‘English’, ‘Dutch’ and ‘Frisian’. For the first-time user the initial screens in the application are a tutorial (Figure 2). This tutorial explains the purpose of the project, as well as giving the user the notification that upon continuing their speech samples will be made publically available and their data can be used for research. Next, the user is directed to a root menu with four options in Sections 3.2, 3.3, 3.4, and 3.6 below.

3.2 Picture naming task

The picture naming task consists of 87 hand-drawn images of everyday objects in the Netherlands. The scientific purpose of this module is to collect data that can be used to document phonological and phonetic patterns for any of the languages recorded. However, Germanic languages were used as a starting point and the 87 pictures were chosen to represent all the phonemes and their allophonic realisations in the localised minority languages in the north of the Netherlands.

After the creation of the picture task the images were all tested for their nameability in 10 online surveys (with nine pictures in each), created in SurveyGizmo and distributed through social media. The surveys asked the informants to name the pictures and type the first word that occurred to them in an open answer text box under each picture. The survey response rate was between 19 and 63 ($M = 35.8$). A picture was kept if >90% of respondents used the term the researchers had intended to elicit with the picture. Two pictures were discarded (‘dyke’ and ‘sea’) and one was tested twice, after making changes to the picture (‘mouth’) (Figure 3).

When opening the picture task from the root menu a prompt informs the participants that the challenge is to name as many pictures as possible, and in as many languages as they can. Next, participants are asked to

Figure 1: Interface language choice in *Stimmen*.

Figure 2: Page two of tutorial in *Stimmen*.

Figure 3: Screenshot of the picture task in *Stimmen*, showing the final version of the picture 'mouth'.

provide meta-data (Figure 4) that includes where they are from (to indicate this on a map), their gender, their age bracket, which languages they are most fluent in, which languages they actively use in their life, and to answer the open question whether there is anything they would like to share with the researchers about their language or dialect. Further social data was not collected due to privacy concerns, as the speech recordings are publically available alongside the social information collected.

Participants are then asked which language they would like to record in. The current list of options includes 38 different languages but more can be added by contacting the *Stimmen* team (contact details in the ‘About’ section of the app). After choosing a language (Figure 5), a randomised selection of pictures is made and shown to the user. They can then press the screen to record, and thereafter send the recording to the researchers. A prompt asks whether the user is sure they want to share the recording publically. After naming 10 pictures the user can choose to go on, or go to the gallery of named pictures to check their progress (Figure 6).

3.3 Free speech and reading module

The free speech and reading module is in the root menu of the *Stimmen* app under ‘Record’. After providing meta-data (the app asks “Is this still you” and shows the meta-data stored from the previous session) the user is encouraged to record themselves telling their own story, a joke, a story from their childhood, a recipe, or reading a text (*The North Wind and the Sun* in Frisian, Dutch or English) (Figure 7). Another option includes giving the user’s own word for ‘potato’, as that particular word was deemed too difficult to draw in a way that would unanimously elicit ‘potato’ as a response for the picture naming task. Again, the user has to indicate which language they would like to record in, after which they can record and choose to submit their own voice to the database.

The image shows a mobile app interface titled 'Stimmen'. The screen displays a questionnaire under the heading 'About you'. The first question is 'Where are you from?' with a location pin icon. The next two questions are 'What is your gender?' and 'What is your age bracket?', both with 'Choose...' dropdown menus. The 'Languages' section follows, with 'Which language are you the most proficient in?' and 'Which languages do you actively use in your life?' also using 'Choose...' dropdowns. At the bottom, there is an open text field with the prompt 'Tell us anything that you think is important to know about your dialect'.

Figure 4: The meta-data questionnaire used for all tasks in *Stimmen*.

Figure 5: Some of the possible recording languages in *Stimmen*.

Figure 6: Word Gallery after naming four pictures in *Stimmen*.

Figure 7: Menu in the free speech task in *Stimmen*.

3.4 Speech map

The recordings of the Picture Naming Task and the Free Speech and Reading Module are all instantly visible on the website¹, as well as under the option ‘Speech Map’ in the root menu of the smartphone application. Figure 8 is a screenshot of the north of the Netherlands in the speech map on the website taken 1 February 2019.

Additionally, the recordings will be made available on the website in a downloadable format. The data will be kept for the coming 10 years on the server of The University of Groningen.

3.5 Data protection and privacy in *Stimmen*

Although the application asks the user whether they are sure they want to share their speech recording publically before they can go on to share their data, the application users have the right to have their data removed. The application includes an ‘About’ page (Figure 9) under which users can indicate that they want their data removed. If they press this option they send a simple web form to the researchers who can then get in touch and remove the specific data in question. By the beginning of 2019 only one user had requested this.

3.6 Dialect quiz

The gamified reported dialect use task Dialect Quiz was based on Leemann and Kolly (2013) and their subsequent applications, and was programmed by the developing studio *iBros*. Data gathered with this module

¹ stimmen.nl.

Figure 8: Speech Map in *Stimmen* and on *stimmen.nl*: a clickable map where the public can listen to the recordings made in the project.

Figure 9: The 'About' page in *Stimmen* with the button to press to have speech data removed.

provides researchers with the public's knowledge of traditional linguistic variants, and regional and social (age, gender) variation in reported dialect use can be analysed.

To develop a Dialect Quiz, an existing corpus of speech samples is needed. A prediction about where a person hails from is made on the basis of answers about dialect knowledge. The prediction in *Stimmen* was created based on the GTRP database (Goeman et al. n.d.) with data from 58 informants from the 1980s in Fryslân. The GTRP thus allowed for creation of a multilingual prediction for 58 different locations in Fryslân in

which varieties of Frisian, Low Saxon and bilingual mixed languages are spoken (some words, like ‘fish’ have the same variant used in more than one language).

Stimmen’s Dialect Quiz asks “How do you say ...”, i.e. what the app user’s own local variant is for 19 Standard Dutch words (see Table 1), as all inhabitants of Fryslân are believed to be fluent in Dutch. 2–10 possible variants are available to the user to listen to, and to choose between. After giving their personal variant to all 19 words, the app makes three guesses of where a user could be from. The user can then indicate whether this prediction is correct or not, and fill in the meta-data questionnaire, as used for Sections 3.2 and 3.3 above.

4 Community involvement

To get in touch with the community and engage the general public to participate in *Stimmen*, three main paths were used. First, collaboration with the regional broadcasting corporation *Omrop Fryslân* was sought, resulting in a number of feature segments on local radio and television about the project. The project leader was able to address the need for more recordings of the local minority languages in these segments, and the corporation shared the application on their webpages and Facebook page.²

Secondly, 12 local secondary schools³ were visited in the project period with a short teaching programme about multilingualism, that allowed use of the application in the last 15 min of class. An estimated 2000 informants were reached in this way. These educational visits gave the researchers an opportunity to share findings with the general public, and, crucially, for the public to approach the researchers to post questions about language.

Table 1: Words in *Stimmen*’s dialect quiz.

Standard Dutch prompt	Translation	No. of variants
<i>armen</i>	‘arms’	9
<i>avond</i>	‘evening’	9
<i>Bij</i>	‘bee’	10
<i>Blad</i>	‘leaf’	3
<i>borst</i>	‘breast’	8
<i>Dag</i>	‘day’	4
<i>deurtje</i>	‘door’ (+ diminutive)	12
<i>geel</i>	‘yellow’	5
<i>(ik ben) gegaan</i>	‘I have gone’	11
<i>(ik heb) gezet</i>	‘I have placed’	6
<i>heel</i>	‘whole’	5
<i>kaas</i>	‘cheese’	7
<i>koken</i>	‘cook’ (infinitive)	11
<i>oog</i>	‘eye’	7
<i>(toe)sprak</i>	‘spoke’	8
<i>tand</i>	‘tooth’	8
<i>trein</i>	‘train’	5
<i>vis</i>	‘fish’	2
<i>zaterdag</i>	‘Saturday’	7

² <https://www.omropfryslan.nl/nijis/756952-rug-en-omrop-dogge-undersyk-nei-de-stimmen-fan-fryslan#videoplayer756999>.

³ SG Liudger, Dockinga College, Inspecteur Boelensschool, CSG Anna Maria van Schurman, CSG Comenius, OSG Piter Jelles, Leeuwarder Lyceum, CSG Bogerman, Lauwers College, OSG Singelland, Het Drachtster Lyceum, OSG Singelland.

Third, a number of online and offline activities were held. On social media the public could comment or post questions to the researchers, or partake in the quiz or download the application. The application and its speech map were displayed in a public exhibition in a park in Leeuwarden, and a lecture series was organised there to share knowledge about multilingualism and minority languages in the north of the Netherlands.

5 Preliminary findings

The *Stimmen* application has been downloaded 6039 times (iOS: 2989; Android: 3050). The dialect quiz has been used a total of 15,131 times. The discrepancy between the two figures can be explained by the fact that the Dialect Quiz is also available as an easily shared web application on stimmen.nl.

In the rest of this paper, we give descriptive statistics of the data collected, with some chosen examples to illustrate the functionality of the application for research purposes. Publications with linguistic analyses of the recorded data, as well as analyses of the dialect predictions, are forthcoming.

5.1 Findings speech recordings

Of the 6039 downloads, 1925 distinct individuals have used the picture naming task, creating 41,553 recordings of words. 24,214 of these were created by female speakers, 17,028 by male speakers and 311 by speakers identifying as ‘other’. The age distribution of the recordings is found in Figure 10.

The top 10 languages recorded in the picture naming task are shown below (Table 2). Nearly three quarters of the recordings are made in Frisian, or varieties spoken within Fryslân.

[Examples 1–4: Recordings of ‘mouth’ in bilingual mixed language ‘Bildts’]

So far only 11 informants have made recordings in more than one language. This means that there is, as of yet, no substantial data set on which to conduct intra-individual analyses of bilingual speech patterns.

Three hundred and eight recordings have been made in the free speech module, out of which 261 in Frisian; 67 in Dutch; nine in City Frisian; three in Bildts; 12 in English; seven in German and French. There are 146 recordings of ‘potato’, 77 recordings of reading passages; 31 recordings of jokes; 30 recordings of people telling their own story and 11 recordings of people’s favourite recipes. Below are examples of recordings of the word ‘potato’ in Frisian from the corpus. Hoekstra et al. (1994) find some 21 variants of the word within Fryslân, and 16 of these are also found in the *Stimmen* recordings.

[Examples 5–9: Recordings of ‘potato’ in varieties of Frisian]

Figure 10: Distribution of recordings from the picture task in *Stimmen* across age groups.

Table 2: Number of Recordings Made by Languages in *Stimmen* Picture Task.

Language	No. of recordings	Language	No. of recordings
Frisian	31,580	English	1246
Dutch	5858	German	207
City Frisian	914	French	168
Bildts	540	Others	1040

5.2 Dialect quiz data

Of the 15,131 times the dialect quiz has been used the survey was filled in 3340 times: 1688 times by a female user; 1633 by a male user and 19 times by a user identifying as ‘other’. This rather low proportion (some 22%) of respondents filling in meta-data could have to do with the fact that the survey was easily skipped in the popular web version of the Quiz (it was the very last component and the respondents had already received their outcome). The age distribution of those using the Dialect Quiz can only be looked at in those filling in the survey and is given in Figure 11.

With the survey data one can easily consider regional and social (age, gender) variation in answers. To give an example, one regional marker between dialect areas in Fryslân is the word for ‘Saturday’ which has competing forms, particularly *saterje/saterdei* or *sneon* for Frisian speakers (other variants include *saterdech* for Stellingwerf Low Saxon speakers and *snjoun* on the island Ameland). The *sneoun* – *saterje* isogloss separates the western *Klaaifrysk* ‘Clay Frisian’ from eastern *Wâldfrysk* ‘Forest Frisian’. The isogloss as drawn by Hof (1933) some 90 years ago is in Figure 12. The interactive Figure 13 shows the different responses by informants to variants in the *Stimmen* Dialect Quiz corpus, grouped by neighbourhood within municipalities, as identified by the Dutch National Bureau for Statistics *Centraal Bureau voor de Statistiek* (CBS).

The distinction between Forest Frisian and Clay Frisian remains when we look at Figure 14 with the isogloss from Hof (1933) superposed upon the distribution map of the ‘*sneon*’ variants in the *Stimmen* quiz data. It is clear that the variant ‘*sneon*’ is used throughout the province, apart from in the Low Saxon speaking area in the south east. While there is still a clear East – West divide within the province, it must be noted that 20–40% of respondents note they use ‘*sneon*’ also in the East of the province. It could be that the spread of ‘*sneon*’ for ‘Saturday’ in the quiz is due to purist reasons. Typical Eastern variants ‘*saterje*’ and ‘*saterdei*’ are close to the Dutch form ‘*zaterdag*’, while ‘*sneon*’ is stereotypically Frisian.

6 Answers to the open question in the survey

Of the 5265 completed meta-data surveys (1925 in the picture naming task + 3340 in the dialect quiz) 490 informants filled in the open question “is there something else you would like to tell us about your language?”. Responses were separated into six categories: answers about A) details of the informants’ language learning history, or their own or parents’ background (345); B) responses about the functionality of the application (72); C) responses that professed positive attitudes towards Frisian (33); D) sharing observations about language use (19); E) the posting of actual questions to the researchers (4); or F) nonce responses (17). Examples of the first five categories can be found below.

A. “Fries geleerd van mijn ouders, komen beide uit verschillende streken. Mijn fries is een mengvorm daarvan.”

Learnt Frisian from my parents, both came from different regions. My Frisian is a mixture of that.

B. “Bij sommige woorden gebruik ik meerdere opties in mijn dagelijks leven dus heb ik de meest gebruikte aangetikt”

With some words I use several options in my daily life so I checked the one I use the most

Figure 11: Distribution of survey responses in the Dialect Quiz in *Stimmen* across ages.

Figure 12: Isogloss between *saterdei* and *sneon* in Frisian in the 1930s, from Hof (1933).

Figure 13: The distribution of the variants for 'Saturday' in the *Stimmen* Dialect Quiz responses: screenshot of the variants for 'sneon' (interactive map by Herbert Kruitbosch available at <http://stimmen.nl/uitspraakkaart-zaterdag/>).

Figure 14: The isogloss from Hof (1933) superposed on the ‘sneon’ – ‘saterje’ isogloss in the *Stimmen* Dialect Quiz responses.

C. “Koesterje de Fryske taal. Soarchje foar goed lesmateriaal en lit bern wer mear in it Frysk dwaan op skoalle. It is een fabel dat bern dan net goed it hollânsk leare. Myn generaazje kin it ek!!”

Cherish the Frisian language, ensure there is good education material and allow children to do more in school in Frisian. It is a myth that children won’t learn Dutch well then. My generation knows it too!

D. “Goedendag, ik wil dat jullie weten dat mensen misschien wel thuis Fries praten maar niet tegen vrienden. Sommige praten het een beetje door elkaar heen. Liefs x”

Hello, I would like you to know that people may speak Frisian at home but not to their friends. Some people speak it interchangeably. Love, x”

E. “Praten er veel mensen Harlingen?”

Do many people speak Harlingen [dialect]?

While answers in groups A and B provide helpful comments to the researchers, answers in C–E indicate an interest and emotional connection to the language that the users wish to share with the researchers. The range of attitudes and language ideologies provided in these answers shows how a fairly straightforward smartphone application can be used to collect rich qualitative data about language. One possible use for this data is to inform us further of the relationship between a regional or national identity and language change in minority language areas. One line of investigation is, for instance, whether respondents who utter strong feelings of local solidarity also use more localised linguistic variants in their speech.

7 Discussion and conclusions

The primary goal of *Stimmen* was to create a citizen science project for minority language users with the aid of a smartphone application. *Stimmen* has been successful in engaging a minority language community. The participation rate within the population of Fryslân was high, with at least 15,000 users (in a population of 645,000), and that community responded well to the project team’s plea for speech recordings to help further

knowledge of minority languages. Large amounts of spoken Frisian data have been provided – but there is no bilingual data to speak of. Future endeavours within the *Stimmen* project will emphasise the possibility within the application to record in several languages to see whether the number of bilingual recordings can be increased.

Overall, *Stimmen* is a successful attempt in gathering large amounts of speech data from minority language users, but primarily in the north of the Netherlands. Its success in this region shows how important satellite events and activities are for smartphones projects. The data collected in our project is predominantly from younger users, but is well-balanced in terms of gender distribution. The large amount of responses from young participants is another indication that the educational programme the *Stimmen* project ran made a real difference in the amount of data collected.

The Dialect Quiz has more users than the other tasks in the application. There may be two reasons for this. One is the fact that this task gives participants an incentive to participate through its gamified character. Gamification of tasks for scientific research has led to highly successful linguistic projects in recent years, and so-called *GWAPS*, games with a purpose, seem a reliable and efficient way of gathering information, if the user is sufficiently clear about their role and mission (Lafourcade et al. 2015). It is also of importance that the Dialect Quiz existed as a web application in addition to being available in the smartphone application. To reach larger audiences, future projects should consider using web applications instead of a smartphone application. As long as sound recording is made possible through support in the browser, a web application can have similar benefits to a smartphone application, especially when still used in browsers on tablets and smartphones with microphones that make high quality recordings. The benefit of web applications lies predominantly in the fact that the user does not need to download anything. The *Stimmen* project group has recently released a dialect quiz for Low Saxon spoken in the east of the Netherlands (Hilton et al. 2019), in which users can record themselves in the web application.

While the social background information gathered from our informants is quite limited for the purposes of sociolinguistic variation and change, the answers collected to the open question in the app provides a rich source of qualitative data. Furthermore, the dialect knowledge data in the Dialect Quiz lends itself well for making new maps of Frisian, bilingual mixed languages and Low Saxon within Fryslân, providing us with new information about the variation and change in progress in languages in the north of the Netherlands.

Acknowledgements: The author would like to thank the following individuals for their invaluable contributions to the *Stimmen* project: Dani Wanitsch, Jelske Dijkstra, Charlotte Gooskens, Hanneke Loerts, Willem Visser, Adrian Leemann, Jolene Roffel, Jordy van Dijken; Renee van Bezooijen, Vincent van Heuven, Piet van Reenen, Jan Pieter Kunst, Herbert Kruitbosch and Martijn Wieling.

References

- Bird, Steven, Florian R. Hanke, Oliver Adams & Haejoong Lee. 2014. Aikuma: A mobile app for collaborative language documentation. In Jeff Good, JuliaLeeHirschberg & Owen Rambow (eds.), *Proceedings of the 2014 Workshop on the Use of Computational Methods in the Study of Endangered Languages*, 1–5.
- Bonney, Rick, Tina B. Phillips, Heidi L. Ballard & Jody W. Enck. 2016. Can citizen science enhance public understanding of science? *Public Understanding of Science* 25. 2–16.
- Bree, Cor van, 1994. Het probleem van het ontstaan van het ‘Stadsfries’ in verband met nieuwe talen in contact-theorieën [The problem of the evolution of “Stadsfries” in connection with new languages in contact theories]. In Stanislaw Predota (ed.), *Handelingen Regionaal Colloquium Neerlandicum*, 43–66. Wrocław.
- Centraal Bureau voor de Statistiek, 2018. CBS StatLine – Regionale kerncijfers Nederland [CBS Statline – Regional core numbers the Netherlands]. Available at: <https://statline.cbs.nl/Statweb/publication/?DM=SLNL&PA=70072ned&D1=0-88,292-293&D2=6,20-22&D3=21-23&VW=T> (07 February 2019).
- De Decker, Paul, Jennifer Nycz. 2011. For the record: Which digital media can be used for sociophonetic analysis? *University of Pennsylvania Working Papers in Linguistics* 17: article 7.

- De Vries, Nic J, Mariele H. Davel, Jaco Badenhorst, Willem D. Basson, Febe De Wet, Etienne Barnard & Alta De Waal. 2014. A smartphone-based ASR data collection tool for under-resourced languages. *Speech Communication* 56. 119–131.
- Fromkin, Victoria, Robert Rodman & Nina Hyams. 2018. *An introduction to language*. South Melbourne: Cengage Learning.
- Goeman, Ton, Johan Tældeman, Piet Van Reenen. n.d. MAND/FAND/GTRP-database. <http://www.meertens.knaw.nl/mand/database/> (07 February 2019).
- Hammarström, Harald, Sebastian Nordhoff. 2011. How many languages have so far been described. *Presented at the NWO Endangered Languages Programme Conference, Leiden*.
- Hilton, Nanna H, Daniel Wanitsch & Martijn Wieling. 2019. Stemmen. <https://woordwaark.nl/stemmen/> (20 December 2019).
- Hoekstra, Eric & Marjo van Koppen. 2000. In D. Boutkan & A. Quak (eds.), *Language contact. Substratum, superstratum, adstratum in Germanic languages*. *Amsterdamer Beiträge zur Alteren Germanistik*. Band 54. 89–106. Amsterdam, Atlanta: Rodopi.
- Hoekstra, Jarich F, Klaas F. van der Veen, Durk H. Veenstra & Willem Visser. 1994. De ierappel op'e souwe. De dialektgeografyske en fonologyske ûntjouwing fan it wurd'ierappel'yn it Frysk. [“De ierappel op'e souwe” The dialectological and phonological development of the word for potato in Frisian]. *It Beaken* 56. 281–298.
- Hof, J.J. 1933. *Friese Dialectgeografie*. [Frisian dialect geography]. Den Haag: Nijhoff.
- Lafourcade, Mathieu, Alain Joubert, Nathalie Le Brun. 2015. *Games with a purpose (GWAPS)*. John Wiley & Sons.
- Leemann, Adrian & Marie-Jose Kolly. 2013. Dialäkt App, apps. apple.com/us/app/dialäkt-app/id606559705.
- Linell, Per. 2005. *The written language bias in linguistics: Its nature, origins and transformations*. Abingdon: Routledge.
- Möller, Robert & Stephan Elspaß. 2015. Atlas zur deutschen Alltagssprache. [Atlas of everyday German] In Roland Kehrein, Alfred Lameli & Stefan Rabanus (eds.), *Regionale Variation des Deutschen – Projekte und Perspektiven*. 519–540. Berlin, Boston: de Gruyter.
- Nagy, Naomi & Miriam Meyerhoff. 2008. The love that dare not speak its name – The fascination with monolingual speech communities in sociolinguistics. https://projects.chass.utoronto.ca/ngn/pdf/NN&MM_monolingual_poster.pdf.
- Provinsje Fryslân. 2015. *De Fryske taalatlas 2015, Fryske taal yn byld*. [The Frisian language atlas 2015, Frisian language in focus]. Leeuwarden.
- Ricento, Thomas. 2013. Language policy, ideology, and attitudes in English-dominant countries. In Robert Bailey, Richard Cameron & Ceil Lucas. (ed.), *The Oxford handbook of sociolinguistics*. Oxford: Oxford University Press. <https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199744084.001.0001/oxfordhb-9780199744084-e-26>.
- Smakman, Dick. 2015. The westernising mechanisms in Sociolinguistics. In Dick Smakman, & Patrick Heinrich (eds.), *Globalising sociolinguistics. Challenging and expanding theory*. Routledge: London.
- Stanford, James N. 2016. A call for more diverse sources of data: Variationist approaches in non-English contexts. *Journal of Sociolinguistics* 20. 525–541.
- Vaux, Bert. 2003. American dialects. Extended version of the article published in Let's Go USA 2004. https://www.choicesmagazine.org/UserFiles/file/article_115.pdf.
- Wenker, Georg. 1881. *Sprach-Atlas von Nord- und Mitteledeutschland*. Karl J. Trübner: Strassburg.

Refer to the *Stimmen* application in the following way:

Hilton, N. H., H. Loerts, W. Visser, G. Jensma, C. Gooskens, D. Wanitsch & A. Leemann. *Stimmen: A citizen science app for languages*. Released 20-Sep-2017 for iTunes and Google Play.