



university of  
groningen



Minorities &  
Multilingualism

# Stimmen (fan Fryslân)

A Citizen Science Approach  
to Minority Languages

Nanna Haug Hilton  
[\(n.h.hilton@rug.nl\)](mailto:n.h.hilton@rug.nl)

# Incentive 1: Monolingual & Large Language Bias in Linguistics

- ±2,300 languages are not described beyond word lists or simple grammar sketches (Hammarström & Nordhoff, 2011)
- Embrace minority languages & multilingualism to extend linguistic theory (Ravindranath 2015; Stanford 2016)
- Underrepresentation of «multilingual» sociolinguistic studies cf. Nagy & Meyerhoff (2008)

# A Multilingual Context

- Context in the Netherlands with socio-politically embedded multilingualism:  
**Fryslân**



# Frisian to a Sociolinguist

- **Recognition:** through European Charter for Regional or Minority Languages, only minority in NL's ratification of FCNM
- **Robust speaker numbers:** ±75% of population (642,000) report proficiency of *spoken* Frisian (2011) ≈ 480,000 speakers
- “All” speakers of Frisian are **bilingual**
- Only 15% of province report ‘good’ **writing skills** in Frisian (Provincie Fryslân, 2015)

# Frisian as a Minority Language

De Haan (1996: 19) ‘The influence of Dutch on Frisian is a genuine case of levelling along the standard-dialect dimension, even though the languages do not stand in a standard-dialect relationship in a historical sense’

- Van Bezooijen and Gooskens (2005): number of cognate words (sample texts): 94%
  - Phonetic (Levenstein) distances: 10-20%
- Perfect testing ground for questions regarding multilingualism and language change: but where's data?

# Incentive 2: Test new Methodologies on Multilingual Minority Language Communities

- Smartphones have built-in high (enough) quality microphones for instrumental analyses (e.g. of F1; F2; pitch)
- Web-based applications means new possibilities for «citizen science»
  - Citizens as data collectors
  - Citizens as analysts
  - Citizens as educators
  - Citizens as formulators of research questions

(Bonney et al., 2009)



«New» methodologies  
for gathering large  
amounts of data

# Documentation Apps

- Aikuma (Bird et al., 2014); update 2018
  - Recording mode for audio and video
  - Translation mode



Picture from Bird et al. (2014:1)

# Games With a Purpose

- Gamification of data collection and analysis: ‘GWAP’: *games with a purpose*
  - Popular for development of NLP
  - User gives data/analyses by playing
  - Vannella, Jurgens, Scarfini, Toscani, Navigli (2014): *Gamers do better with simple analyses and sorting than volunteers, i.e. successful and efficient way of getting arduous tasks done*

# Crowd Sourcing through Gamification

RÉSOLVEZ L'ANALOGIE SUIVANTE :



30s

mettre un terme ici

Aucun terme proposé

Si vous ne savez pas répondre, il faut passer la partie. Si vous pensez qu'il n'y a pas de réponse possible, vous pouvez mettre \*\*\*. Vous pouvez supprimer un mot proposé en cliquant dessus dans la liste affichée à droite.

*... est à rire  
ce que  
triste est à gai*

OK



Diko



# ZOMBI LINGO



## BIENVENUE à TOI, JEUNE ZOMBIE !

JOUE POUR AIDER LES SCIENTIFIQUES !

SUIS MES RÈGLES,  
IDENTIFIE LES TÊTES  
ET MANGE-LES.

ATTENTION AUX PIÈGES, ILS SONT NOMBREUX !



SIGNÉ  
Prof. Frankenperrier.

*Pas de limite pour toi !  
Tu accèdes à toutes les options, bonus cachés !*

ESSAYER

*Cette version limitée va te rendre accro ! Mais tu ne pourras pas sauvegarder !*

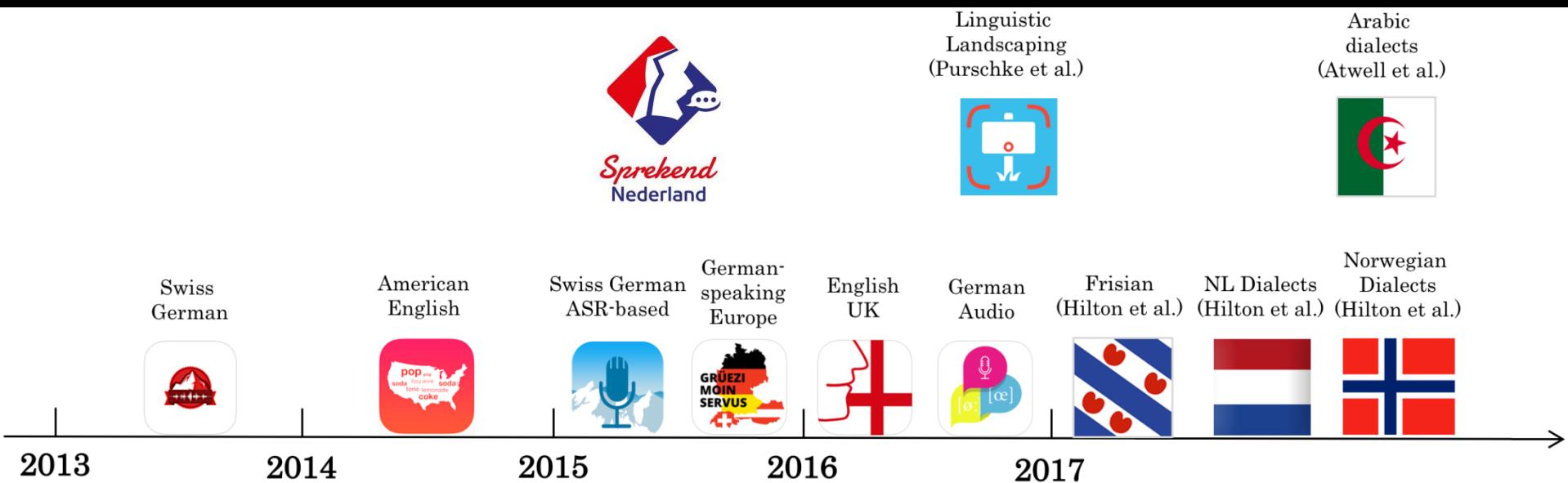
SEMAINE  
points

1 WELLINGTON : 2104723  
2 METHOS31 : 238239

3 CALYPSO : 0

MOIS  
TOTAL

# Sociolinguistic Crowd-Sourcing Tools





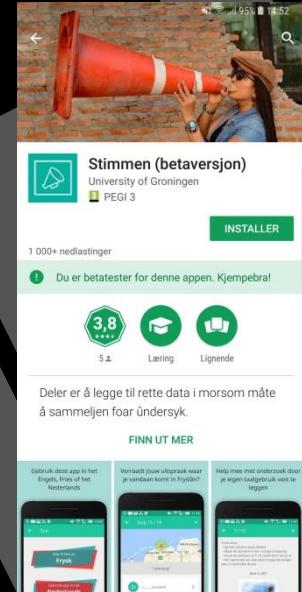
# Stimmen fan Fryslân



# Stimmen fan Fryslân

(Hilton, Gooskens, Leemann, Loerts, Jensma & Visser)

- Funded by ‘Lân fan taal’ programme for Leeuwarden as cultural capital 2018: Launch 20 September 2017
- Participant Goal: 10,000 informants
- ‘Citizen science’: data collection; education;
- Method: smart-phone (Android & iOS) and web-based application ‘stimmen’
- Linguistic Aims:
  - Frisian convergence with Dutch; Dialectology Frisian; Description of Dutch in Fryslân; Description of Frisian reading language
- + Speech Technology



## About you

Where are you from?

### Languages

Which language are you the most proficient in?

What is your gender?

Choose...

Choose...

Which languages do you actively use in your life?

What is your age bracket?

Choose...

Choose...

Tell us anything that you think is important to know about your dialect

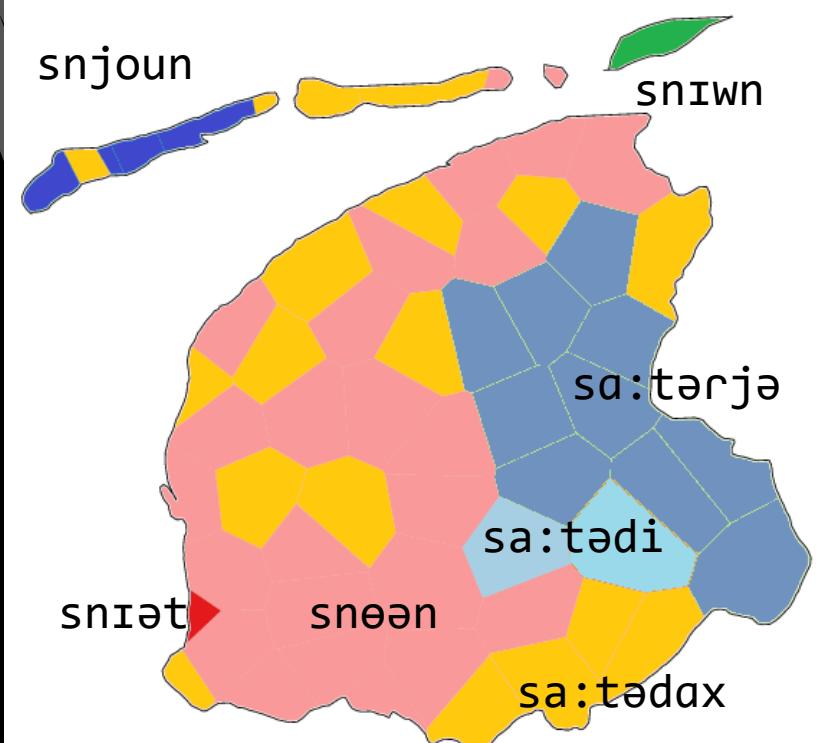
# Dialect Quiz

For Frisian, ‘city Frisian’ and Lower Saxon varieties in Fryslân

19 variables with 2-10 variants from GTRP corpus (1980s)

→ Initially, prediction of 58 locations on the basis of answers

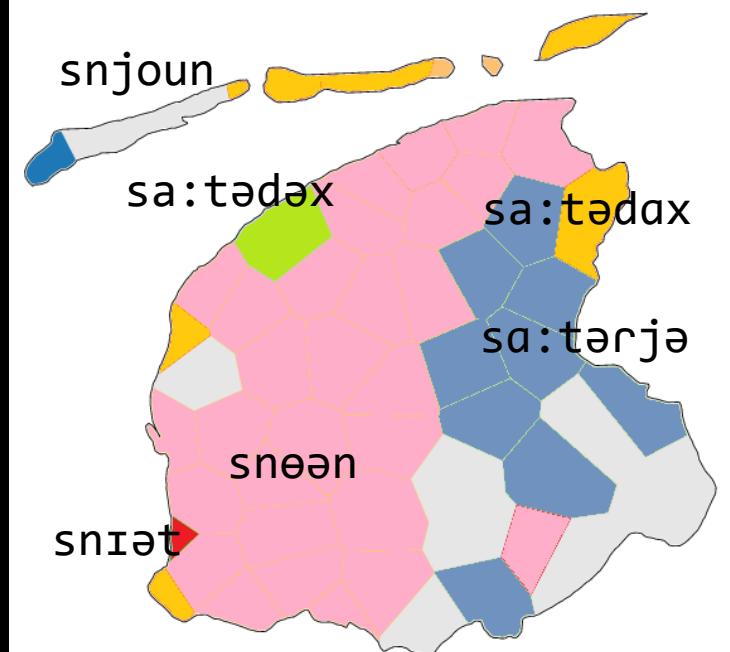
RQ: What are minority language users' self-reported dialect use?



1933



1980s



**2017-2018  
13,500 responses**

# The Picture Naming Task

- 88 pictures: made to elicit all Frisian phonemes
  - Allophones included
- User decides which language to record; i.e. can be **used on its own** by researchers elsewhere to record any language
  - Documentation purposes
  - Forced alignment training
- RQ: Is there convergence towards Dutch in Frisian phonology: (sk), (r), (brekking), (ei)? and what is Fryslân Dutch?

i:	iisbear (ijsbeer)		
u:	mûs (muis)	hûs (huis)	spûns (spons)
Ö:	reus (reus)	fleugel (vleugel)	reuzerêd (reuzerad)
ɛ:	blêd (blad)	gers (gras)	
a:	aap (aap)	banaan (banaan)	kano (kano)
ü:	druven (druiven)	duvel (duivel)	
e:	spegel (spiegel)	finster (venster)	grins (grens)
o:	mole (molen)	foto (foto)	
ɔ:	hân (hand)	bôge (boog)	



# Picture Task Recordings

- In the first 6 months:

≈ 42,000 recordings out of which

- ± 33,400 in Frisian
- ± 5,500 in Dutch
- ± 1100 in English
- ± 700 in Stadsfrys
- ± 600 in Bildts
- ± 200 in Lower Saxon

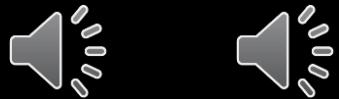
And the remaining others in: *Hielpers, Amelands, Schylgers, Stellingwarfs, Drents, Gronings, Spanish Castillian, German, Galician, French, North Frisian, Norwegian, Eastphalian, Russian, Papiamentu, Sarnami Hindustani & Berber Tarafit*

# Multilingual Speech Data

- Picture task data in more than one language:



- Male informant from Leuven; 21-30 yo: English most proficient language; actively uses German, Dutch and French



- Female Frisian-Dutch teenager from Arum: Dutch most proficient language; active user of Frisian and English too

# Speech Map

- Tell a #story/#joke/#recipe (...)
- Read a passage (North Wind & the Sun) or word list
- Data exhibited in the Language Exhibition in Leeuwarden (yellow pavilion on the water in the park)
- RQ: reading style in Frisian? (cf. Lamb, 2008)

*'I think it's nice that this app records how I speak because Frisian is predominantly a beautiful language to me because there are so many beautiful sounds in it: 'maaie'; 'gean'; 'slaaie': those kinds of words I think are more beautiful in Frisian than in any other language'*



Hitzum



Leeuwarden



Friese  
Surinamer

# Dissemination

- 1) Regional broadcaster Omrop Fryslân
  - Also national coverage (Radio 1)
- 2) School visits
  - <https://omrop.fr/JYGH>
- 3) Social media and own website
- 4) General ECC-programme:
  - public lecture series
  - language pavilion and exhibit in Leeuwarden

# Challenges

- Dialect quiz and word lists/pictures?
  - We also need other data!
  - Bilingualism research: more information about language background!

Mem berne yn Easternytsjerk En Heit in Damwâld.

Wij gebruiken het fries onderling tussen man en vrouw met onze kinderen en in het dagelijks leven en met anderen zoveel mogelijk Nederlands

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

Mooie app!

ik ben verhuisd van Ee naar Damwoude toen ik 8 was.

Ik ben Nederlands opgevoed omdat mijn moeder uit Groningen komt (vader uit Drachten) maar heb vanaf de basisschool Fries geleerd en ben dit toen later ook gaan praten met mijn vader.

tjsonge! ! útkomst is goed! ik kom út Ie, bûtenút tichteby Eastrum (oostrum)

Ik woon nu in Utrecht, maar thuis met familie, buren, etc altijd Fries. Ze kunnen amper anders.

ik bin bearne in frysland, mar fan myn 18e wenje ik buiten frysland en sprek ik in soad nederlands.

Ik hâld fan it Frysk, it sjongt mooi en hat prachtige wurden.

It is spytich dat hieltyd minder minsken Frysk prate. De skoallen moatte der in aktive rol yn spylje. Mar it begjit fansels mei de minsken sels. Dy moatte harren taal werdearje en der grutsk op wêze.

in het fries kan ik mij het beste uitdrukken. 1 woord stond 'mijn' uitspraak er niet bij

At it er op oan komt kin ik myn gefoelens it bêste ûterje yn it Frysk

It is in moaie sangerige taal. De woartels fan de taal lizze yn de kustkultuer om it Waad en de Noardwest hinne. It binne de Frysk sprekkenden sels dy't dizze âlde taal aktyf libjend hâlde kinne.

Spreek het liefst Fries, merk dat ik niet goed beheers omdat ik opgegroeid ben aan de Groninger kant van de grens. Zodra ik merk dat iemand ookFries spreekt schakel ik over. Woon 40 jaar in Winsum Gn.

Frysk moat bliuwe!

Ik vind het heel belangrijk om het Frysk der in te houden want ik ben zelf heel erg Frysk en ik zou niet andere willen

Ik skriuw, app en dream ek yn it Frysk

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!!

Frysk is de moaiste taal op ierde.

It sint my net dat wy better Hollânsk en Ingelsk skrieuwe (en sprekke) dan us eigen memmetaal..

At jo foar de Fryske taal altyd mar op twa gedachten hingjen bliuwe, bygelyks oer it kiezen fan plaknammen yn't frysk wol as net frysk boppe oan, dan wurde jo eins al net mear serieus nommen yn't lân

Ik ben trots dat ik fries spreek.

Ik soe It Frysk better behearskje wille.

De.memmetaal is fan en út it hert!

# More Challenges

- Dialect quiz and word lists/pictures?
  - We also need other data!
  - Bilingualism research: more information about language background!
- Data protection: Personal data & use by minors
- Validity and reliability of the data?

# More Challenges

- Dialect quiz and word lists/pictures?
  - We also need other data!
  - Bilingualism research: more information about language background!
- Data protection: Personal data & use by minors
- Validity and reliability of the data?
  - No real control over answers and recordings
  - Skewed data in mobile app-investigations: young & educated
- Recording quality? So far so good!

# Going forwards

- Throughout the year of 2018: public lectures and outreach activities
  - Making website-games with ‘game technology’ college students: ‘guess where the language is from!’
- Creating quizzes and app versions for Lower Saxon, Limburgish and Zeelandic
- Making the data publically available for download (with location, gender and age)
- Possibly creating an Australian version of the app with a new game function

# Going forwards

- Make (more) maps that reflect regional variation
- Continue transcribing picture task (with help from forced alignment - thanks FA & FAME!)
- Consider social effects and relationship variant use and reported language proficiencies
- Consider Dutch data

# Questions and Comments?

Stimmen for iPhone: [goo.gl/excxLv](http://goo.gl/excxLv)

Stimmen for Android: [goo.gl/nTKvDW](http://goo.gl/nTKvDW)

Please get in touch!

[n.h.hilton@rug.nl](mailto:n.h.hilton@rug.nl)