

# Twiqs.nl: Analysis of Dutch tweets

Erik Tjong Kim Sang

netherlands

eScience center

by SURF & NWO

Probeer ook eens de [nieuwe snelle versie van twiqs.nl](#). [Feedback](#) is van harte welkom!

[twiqs.nl](#) [FAQ](#) [Blog](#) [Ngrams](#)

Naam:

Wachtwoord:

Inloggen

[Registreren](#)

## Zoek in tweets:

Zoeken

26 : september : 2017 : - : : : : :

Op twiqs.nl kan je in Nederlandstalige tweets vanaf december 2010 zoeken. De gevonden tweets kan je dan op Twitter bekijken. De zoekdatabase is niet compleet. Contactpersoon voor deze website is Erik Tjong Kim Sang <erikt(at)xs4all.nl>

### Voorbeelden:

- ["mh17" op donderdag 17 juli 2014](#) (grafiek)
- ["koningsdag" op zaterdag 26 april 2014](#) (kaart)
- ["ajax.pcc" op zondag 20 april 2014](#) (grafiek)
- ["nagels" in september 2014](#) (gebruikersinformatie)
- ["hedde" in juli-september 2014](#) (kaart voor dialectwoord)
- ["zwarte piet" in september 2014](#) (woordenwolk)

Created at: 26 September 2017 13:33:10. erikt(at)xs4all.nl ([clusterstatus](#))

# Twiqs.nl: available views

**Graph: number of tweets per topic vs time**

**Map: number of tweets per topic vs position**

**Pie: user age and gender distribution per topic**

**Cloud: frequent words and frequent users per topic**

**Emoji: average sentiment per topic**

**Ngrams: word frequencies per month**



# Twiqs.nl stats

**Database with 3.4 billion Dutch tweets**

**70 unique IP visits per day**

**Dozens of papers citing twiqs.nl**



# Studying dialects on Twitter

The language Dutch has dozens of regional variants

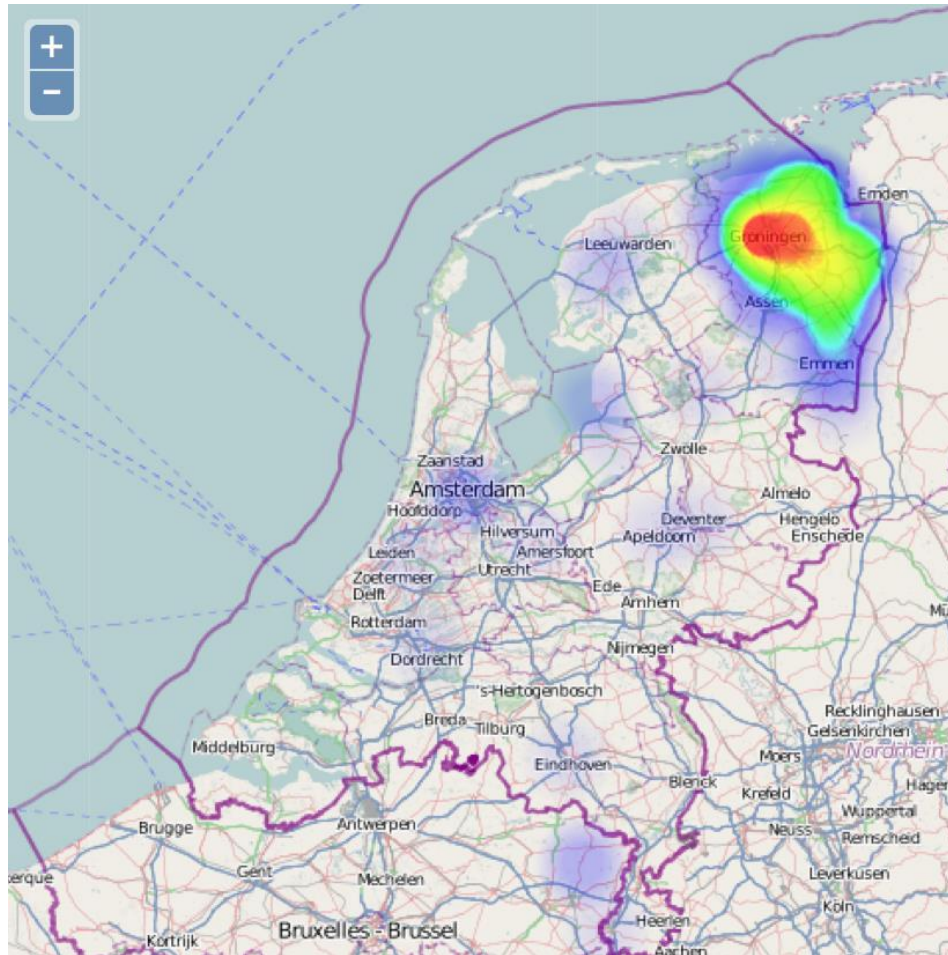
Regional language variance can be studied on Twitter:

- Search for phrases
- Make a map

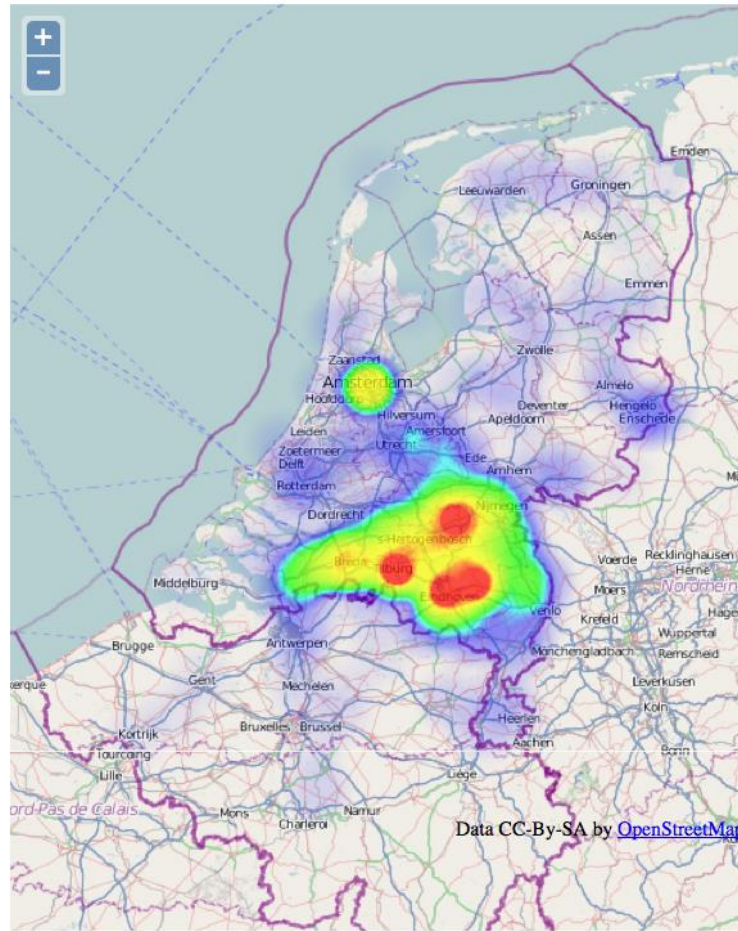
**Note:** choose a large time frame, like one month



**Nait**  
NL: Niet  
EN: Not



**Hedde**  
NL: Heb je  
EN: Do you have



# Particle research (Ton van der Wouden)

Dutch uses small words without meaning: particles:

- Je wil **zeker wel** iets te drinken, **hè?**  
(You want **surely** something to drink, **don't you?**)

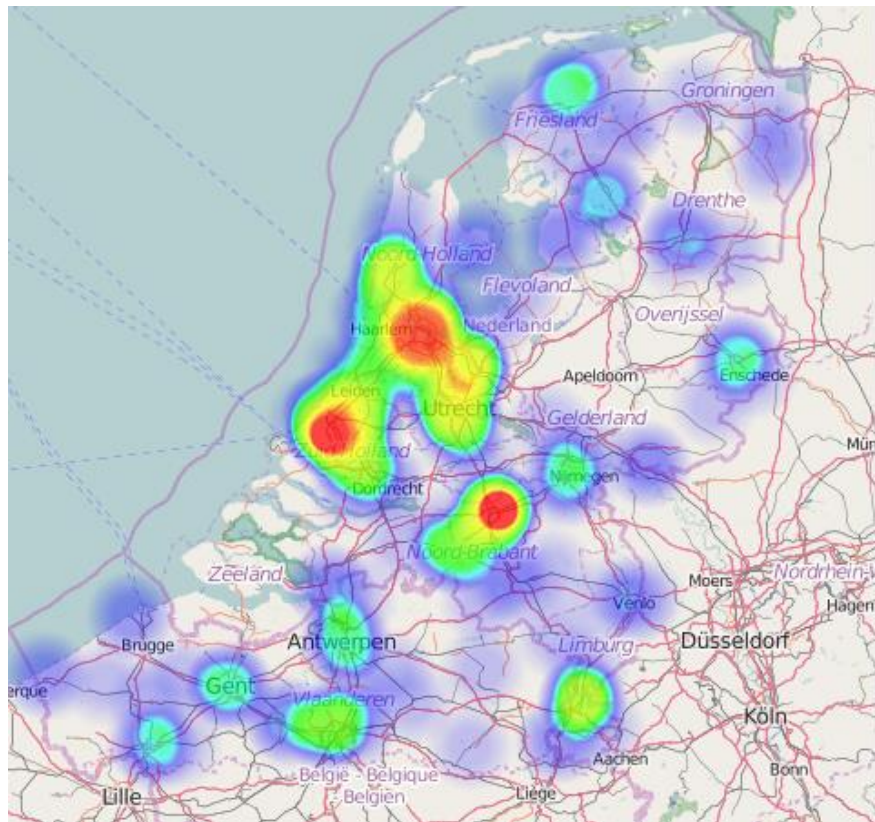
There is regional variance in particle usage:

**alweer vs weeral**

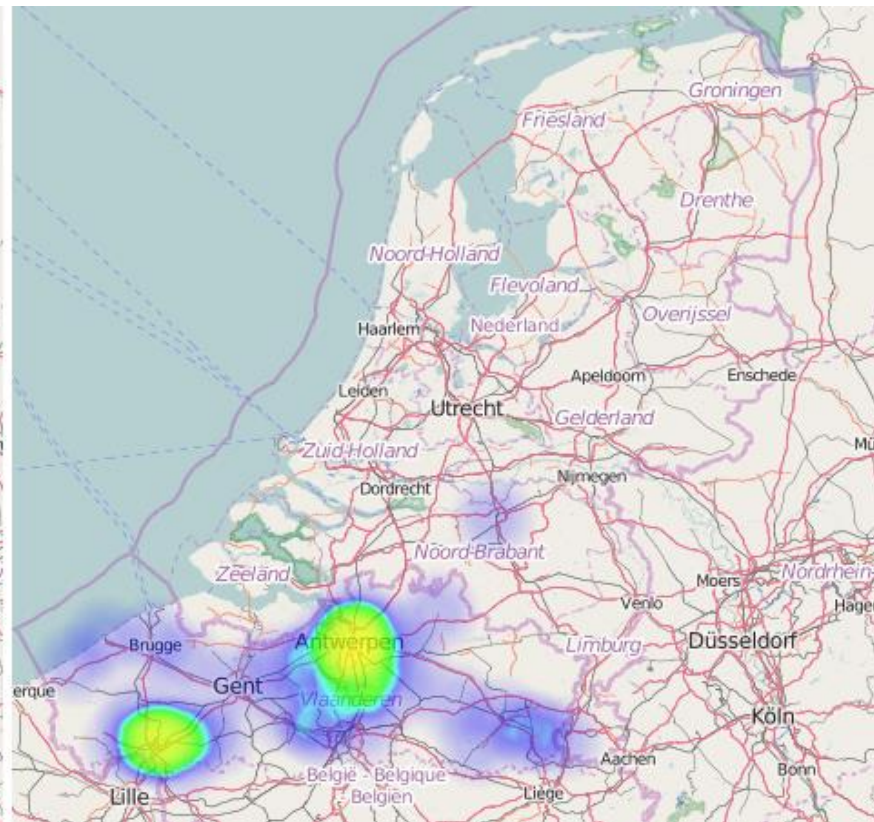




alweer



weeral



# Predicting election results (with J. Bos)

**Dutch provincial/Senate elections of 2011:**

- 12 provinces
- 12 parties
- 15 million Dutch tweets (1 week of data)

**Challenge: too little data for provincial predictions!**

**Filters: topic: 64,000 – geo: 640 – province: 53 – party: 4**



# Predictions based on tweet counts

**Test prediction based on party counts on Twitter: in Senate seats compared with poll results:**

**PVV: 18 seats (+6), VVD: 13 (-2), CDA: 12 (+2), PvdA: 9 (-4), SP: 7, GL: 6 (+2), D66: 5, CU: 2 (-1), PvdD: 1, SGP: 1 (-1), OSF: 1, 50+: 0 (-2)**

**Two different problems in predictions PVV and 50+**



# Improving tweet count prediction

**Problem: negative tweets boost party score**

**Solution: remove negative tweets**

**Problem: unusual age distribution on Twitter**

**Solution: normalize predictions with poll results**



# Final predictions based on tweets

**VVD: 14 (-2), PvdA: 16 (+2), PVV (10), CDA: 8 (-3), D66: 8 (+3), SP: 6 (-2), GL: 3 (-2), CU: 3 (+1), 50+: 2 (+1), SGP: 2 (+1) , PvdD: 2 (+1), OSF 1**

## Comparison with polls:

- **Politieke Barometer: 14 seats off**
- **Maurice de Hond: 14 seats off**
- **Twitter: 18 seats off**



## More information

**Ton van der Wouden, Zit je nu alweer te twitteren? In: De kaartenbank: over taal en cultuur, AUP, 2014.**

**Erik Tjong Kim Sang and Johan Bos, Predicting the 2011 Dutch Senate elections with Twitter. In: Proceedings of SASN 2012, ACL, Avignon, 2012.**

