

Geografie en inwoneraantallen als verklarende factoren voor variatie in het Nederlandse dialectgebied

Wilbert Heeringa John Nerbonne René van Bezooijen Marco René Spruit
 Rijksuniversiteit Groningen Radboud Universiteit Meertens Instituut
 Groningen Nijmegen Amsterdam

Symposium kwantitatieve benaderingen in taal- en letterkundig onderzoek en elders in de geesteswetenschappen. Een kennismaking
 Meertens Instituut Amsterdam
 Donderdag 28 Jun 2007

1

Levenshtein-afstand

- Afstand tussen twee dialecten: gemiddelde Levenshtein-afstand voor een reeks woordenparen.
- Verfijning: gebruik graduele gewichten, namelijk de akrostische segmentafstanden.
- We staan alleen oplijningen toe waarin:
 - een linker correspondeert met een linker
 - een mederlinker correspondeert met een mederlinker
 - de [j] of [w] correspondeert met een linker
 - de [i] of [u] correspondeert met een mederlinker
 - de schwa correspondeert met een sonorant

7

Zwaartekrachtmodel

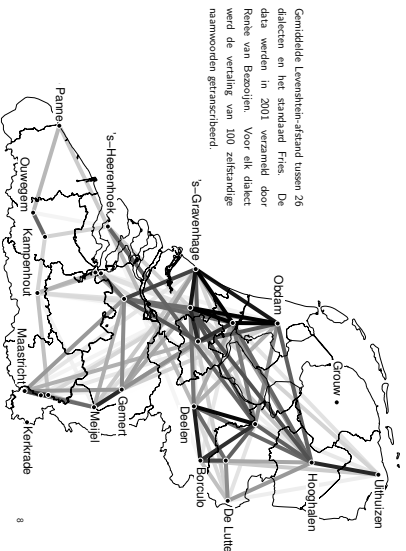
- Trudgill 1974, 1983 (ook Hinskens 1992, 1993) gebruiken het model als index voor taalkundige invloed tussen plaatsen.
- Geografie: hoe dichter plaatsen bij elkaar liggen, hoe meer contact. Kans dat een inwoner gaat naar een punt op een denkbeeldige cirkel rond zijn of haar woonplaats $1/r^2$.
- Inwoneraantallen: iedere inwoner uit de ene plaats kan in contact komen met iedere inwoner in de andere plaats: $P_1 \times P_2$
- G is de taalkundige verwantschap tussen beide plaatsen.

11

Levenshtein-afstand

- Nadtel methoden Seguy, Goelb en Hamner: twee items zijn of gelijk of ongelijk, geen gradualiteit.
- Nadtel methode geb. Hopperbroekers: methode is niet gevoelig voor volgorde van de segmenten in een woord, bijv. [kenin] en [knin] worden niet onderscheiden.
- In 1995 gebruikte Kessler de Levenshtein-afstand voor het meten van afstanden tussen lease dialecten.
- Geef graduele afstanden, is gevoelig voor volgorde van de segmenten in woorden.
- Ook toegepast op Nederlandse dialecten (Nerbonne et al., 1996, Heeringa 2004), Sardische dialecten (Bolognesi & Heeringa 2002), Noorse dialecten (Goskens & Heeringa 2004) en Duitse dialecten (Nerbonne & Stedje 2005).

5



Zwaartekrachtmodel

- In navolging van Nerbonne & Heeringa 2006 gebruiken we het model als index van sociaal contact.
- G vervalt: taalkundige verwantschap is de inverse van taalkundige afstand, dat laatste willen we juist verklaren met het model.
- Hypothesen:
 - De taalkundige afstand D is recht evenredig met de kwadratische afstand: $D \propto r^2$.
 - De taalkundige afstand D is omgekeerd evenredig met de producten van de inwoneraantallen: $D \propto \frac{1}{P_1 P_2}$
 - De taalkundige afstand D is omgekeerd evenredig met de mate van sociaal contact tussen twee plaatsen: $D \propto \frac{1}{P}$

12

Levenshtein-afstand

- Voorbeeld: *kinja* wordt uitgesproken als [ken:n] in het dialect van Amsterdam, en als [kn:n] in het dialect van Zwollekenspel. Hoe veranderen we de ene variant in de andere?
- Dit kan op meerdere manieren. Het Levenshtein-algoritme kiest de operaties zodanig dat de totale kosten minimaal zijn.

	k	ə	n	ɛ:	n	verwijder ə	1
	k	n	ɛ:	n		vervang ɛ: door i:	1
	k	n	i:	n		voeg toe ə	1
	k	n	i:	n	ə		3
	1	2	3	4	5	6	

- Deel de Levenshtein-afstand door de lengte van de oplijning: $3 / 6 = 0.5$. Als percentage: 50%.

6

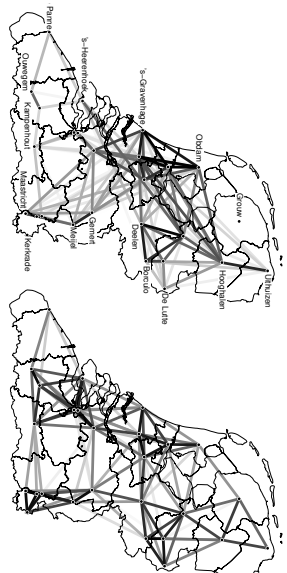
Zwaartekrachtmodel

- Zwaartekrachtmodel van Isaac Newton:

$$F = G \frac{m_1 \times m_2}{r^2}$$

F is de aantrekkingskracht
 m_1 en m_2 zijn de gewichten van beide objecten
 r is de afstand tussen beide objecten
 G is de kracht tussen twee objecten van elk 1 kg op 1 m afstand van elkaar

9



Links: gemiddelde Levenshtein-afstanden, rechts: henshloevde geografische afstanden. Correlatie $r=0.56$.

13

