





# Addenda

Artikelen voor  
Ad Foolen



Sander Lestrade  
Peter de Swart  
Lotte Hogeweg  
(Red.)

# Addenda

Artikelen voor  
Ad Foolen

Sander Lestrade  
Peter de Swart  
Lotte Hogeweg  
(Red.)

Radboud Universiteit Nijmegen  
2015

Getypeset in  $\text{\LaTeX}$  ([www.latex-project.org](http://www.latex-project.org)) met de lettertypes Utopia en Computer Modern sans.

Omslagontwerp door Peter de Swart in *Processing* ([www.processing.org](http://www.processing.org)) met de lettertypes Coustard en Oxygen-Regular.

Gedrukt door Ipskamp drukkers ([www.ipskampdrukkers.nl](http://www.ipskampdrukkers.nl))

ISBN 978 90 902 9090 4

NUR 616

Deze uitgave is financieel mogelijk gemaakt door de afdeling Taalwetenschap, Faculteit der Letteren, Radboud Universiteit Nijmegen.

[www.ru.nl/taalwetenschap](http://www.ru.nl/taalwetenschap)

Copyright © 2015 by the authors. All rights reserved.

# Een DNA-test voor het bepalen van lexicale verwantschapspatronen

Nicoline van der Sijs

*Radboud Universiteit Nijmegen & Meertens Instituut, Amsterdam*

De moderne woordenschat van een taal is het resultaat van een lexicale *survival of the fittest*: sommige woorden zijn van marginaal – vaktaal, dialect, Bargoens – doorgedrongen tot het algemene gebruik, andere woorden zijn in de loop van eeuwen juist gemarginaliseerd. En dan zijn er natuurlijk in de loop van eeuwen veel nieuwe woorden gevormd of geleend, en in dezelfde tijd zijn er heel veel woorden uitgestorven.

Hiermee vertel ik niets nieuws. Evenzo is algemeen bekend dat voor-aanstaande taalgebruikers zich in de renaissance beijverden voor de vorming van een algemene, bovenregionale taal, en daarvoor een keuze maakten uit de taalvormen van het bestaande dialectlandschap. De consensus is dat zij zich vooral baseerden op wat op dat moment in de Hollandse steden gold als beschaafd taalgebruik. Dat beschaafde taalgebruik ging weer terug op oudere tradities die ontstaan waren in middeleeuwse schrijfcentra met grote uitstraling. Tot en met de twaalfde eeuw waren de Limburgse schrijfcentra dominant, in de dertiende en veertiende eeuw de Vlaamse, en in de vijftiende eeuw de Brabantse. Pas daarna verschoof het centrum naar het Hollands.

Dit zijn interessante algemene observaties over tendensen, maar ik zou dit graag eens gekoppeld willen zien aan een overzicht van de geografische herkomst van de concrete woordvormen in het moderne Standaardnederlands. Zijn de huidige woordvormen te herleiden tot specifieke Nederlandse dialecten? Hoeveel twaalfde-eeuws Limburgs bevat

---

Gepubliceerd in:

S. Lestrade, P. de Swart & L. Hogeweg (Red.) (2015).

*Addenda. Artikelen voor Ad Foolen* (pp. 401–407).

Nijmegen: Radboud Universiteit. ©2015 de auteur(s).

het Nederlands nog? En hoeveel vormen kunnen we terugvoeren tot een nauwverwant Nederduits of Hoogduits dialect? Dat er gedurende de gehele geschiedenis woordvormen zijn uitgewisseld tussen wat we nu met een modern construct Nederlandse en Duitse dialecten noemen, is zeker. De omvang van die uitwisseling is echter onbekend, en er bestaat geen consensus over: een van de bevindingen in mijn *Taal als mensenwerk: het ontstaan van het ABN* is dat die omvang wordt onderschat, en dat het Nederlands meer invloed van oostelijke dialecten heeft ondergaan dan algemeen wordt aangenomen. Als de geografische herkomst van moderne woordvormen bekend zou zijn, zou die stelling nader kunnen worden onderbouwd.

Maar hoe kunnen we achterhalen uit welk dialectgebied een bepaalde woordvorm afkomstig is? Daarvoor moet een methode worden ontworpen. In het onderstaande schets ik een mogelijk scenario. De kernvraag is: is het mogelijk een DNA-test te ontwikkelen die de genetische verwantschap bepaalt van een individueel woord met een bepaald dialectgebied?

Voor een dergelijke DNA-test is het allereerst nodig de genen van de dialecten (families) en van individuele woorden in kaart te brengen. Hoe groter de overeenkomst in genen is tussen families en individuen, hoe groter de kans dat ze verwant zijn, ofwel: hoe groter de kans dat een bepaalde woordvorm teruggaat op een bepaald dialect. De genen vinden we in de fonologische en morfologische structuur van de woorden.

Als ijkpunt stellen we om te beginnen de taalkundige DNA-profielen vast van de verschillende dialectgebieden van het Nederlands en het aangrenzende Neder- en Hoogduits. Die profielen bevatten informatie over de dialectverschijnselen (genen) die kenmerkend zijn voor de verschillende gebieden. Hiernaar is al veel onderzoek gedaan, waarop we ons kunnen baseren. Uitgangspunt is variatie in dialectverschijnselen te koppelen aan gebieden. Op die manier kunnen we DNA-profielen opstellen voor de verschillende dialectregio's.

Eerst inventariseren we welke variatie er momenteel bestaat. Bijvoorbeeld dat de vormen *strelen* en *wetering* voorkomen naast *stralen* en *water*, en dat *bleu*, *sneu* en *keuter* de varianten *blo*, *snood* en *kot* kennen. Als we de geografische verbreiding van de varianten bekijken, blijkt dat de vormen met *-e-* en *-eu-*, die het gevolg zijn vanumlaut, voorko-

men in oostelijke, Saksische dialecten. Het DNA-profiel van oostelijke dialecten krijgt dus de regel dat klinkers in deze dialecten in bepaalde omstandigheden door umlaut van kleur zijn veranderd. Een andere regel uit de DNA-profielen luidt dat in Saksische, Nederlandse en Hoogduitse dialecten de combinatie *-ald* of *-old* voorkomt die in andere dialecten is veranderd in *-oud*. De DNA-test zal dan *halt*, *kobold* en *waldhoorn* (naast Standaardnederlands *houd*, *kabout* en *woudhoorn*) als Hoogduits of Nederduits woord analyseren. Voorwaarde is dus wel dat er sprake is van variatie: zo zou een vorm als *alternantie* niet worden beschouwd als een oostelijke vorm, omdat er geen variant *outerantie* bestaat. De DNA-test bevat ook regels die *heiland* en *schlager* analyseert als Hoogduits op basis van de klank *ei* (naast Standaardnederlands *helen*) en *schl-* (naast Standaardnederlands *slager*).

Daarna vullen we de DNA-profielen aan met historische gegevens. Van DNA is bekend dat er mutaties kunnen optreden bij het doorgeven van de ene generatie op een volgende. Dergelijke mutaties kunnen ook plaatsvinden bij het doorgeven van dialectverschijnselen. We moeten dus niet alleen beschrijven welke dialectverschijnselen kenmerkend zijn voor de *huidige* dialectgebieden, maar ook hoe de situatie vroeger was. De dialecten zelf zijn in de loop van de tijd veranderd, en ook grenzen tussen individuele taalverschijnselen zijn verschoven. In de dertiende eeuw was bijvoorbeeld de vorm *leggen* voor *liggen* (*daar legt het*) gebruikelijk in het Hollands en ook langs de kust in Vlaanderen, terwijl in de rest van Vlaanderen en het Brabants *liggen* de gewone vorm was. Daar ‘ligt’ dus de oorsprong van de huidige standaardvorm *liggen*.

Voor het opstellen van de DNA-profielen van de dialectgebieden maken we gebruik van zoveel mogelijk exact gelokaliseerde en gedateerde dialectverschijnselen. Voor de huidige tijd, vanaf de negentiende eeuw, zijn veel taalkaarten en dialectgegevens beschikbaar. Voor de oudste periode is materiaal beschikbaar in het Oudnederlands en het Vroegmiddeleeuws Nederlands woordenboek, in de twee taalatlanten van het dertiende-eeuwse Nederlands, in het Corpus Gysseling van teksten uit de dertiende eeuw, in het CRM met veertiende-eeuwse oorkonden en in de atlas van het veertiende-eeuwse Nederlands. Een probleem van de middeleeuwse bronnen is wel dat de gegevens niet evenredig over het taalgebied zijn

verdeeld: er zijn veel meer gegevens beschikbaar uit Zuid-Nederlandse dialecten dan uit oostelijke of Hollandse dialecten. Erger nog is dat er een gat gaapt van de vijftiende tot de negentiende eeuw waaruit vrijwel geen dialectgegevens bekend zijn.

Ondanks de beperkingen van het materiaal zal het aan de hand van ijkpunten toch wel lukken om de DNA-profielen van dialectgebieden door de tijd heen te reconstrueren. Daarmee hebben we vergelijkingsmateriaal dat we kunnen gebruiken voor een eerste test van de huidige woordenschat: kunnen we uit de klank of woordstructuur van een moderne vorm opmaken uit welk dialectgebied deze vorm stamt? Daarvoor voeren we alle DNA-profielen in de computer in, zodat er één grote lexicaal DNA-databank ontstaat. Als een onderzoeker de herkomst van een individueel woord wil terugvinden, moet hij de verschillende hem bekende varianten intikken – voor de analyse is, net als bij biologisch DNA, altijd vergelijkingsmateriaal nodig. De computer analyseert de varianten en berekent automatisch met welk profiel ze de meeste verwantschap vertonen. Als je bijvoorbeeld de vormen *bracht* en *dacht* invoert, met als varianten (in dialecten en ouder Nederlands) *brocht* en *docht*, dan herleidt de computer de *a*-vormen tot Duitse dialecten. Evenzo berekent de computer dat *spiets* – naast *spies* – op basis van de klank *ts* waarschijnlijk afkomstig is uit een Hoogduits dialect. Maar bij *iets* geeft hij die analyse niet, want er bestaat geen vorm *ies*.

De computer kan vormen dus alleen analyseren als er meerdere varianten naast elkaar bestaan en de varianten klanken of woordstructuren bevatten waaruit het herkomstgebied kan worden afgeleid. Van lang niet alle Nederlandse woorden bestaan varianten naast elkaar, en in lang niet alle Nederlandse woorden zijn ondubbelzinnige DNA-sporen terug te vinden: van veel woorden zal het niet mogelijk zijn op grond van fonologische of morfologische kenmerken te bepalen in welk dialectgebied ze zijn ontstaan. Dat komt doordat het DNA-materiaal in de loop van de tijd wordt verdund. In taalkundige termen: de dialectkenmerken kunnen in de loop van de tijd worden uitgewist, doordat ze worden aangepast aan het Standaardnederlands. Aan de woorden *warenhuis* of *voorwoord* is niet te zien dat de oorsprong ervan ligt in het Duitse *Warenhaus* en *Vorwort*. Interne taalverschijnselen als klank- en woordstructuur zijn dus

onvoldoende om de genetische herkomst van individuele woorden te bepalen. Ze moeten worden aangevuld met externe informatie.

Die externe informatie kan op verschillende manieren worden gevonden. De eerste manier is het analyseren van expliciete uitspraken over de regionale herkomst van woorden in contemporaine bronnen. Zo voegen sommige oude woordenboeken streekaanduidingen toe achter een trefwoord. Het Latijns-Nederlandse woordenboek van Kiliaan uit 1599 noemt bijvoorbeeld *tillen* ‘Saksisch, Kleefs, Fries, Hollands’. Ook in literaire werken kunnen we informatie vinden over regionaal taalgebruik. Zo blijkt uit de *Spaanschen Brabander* van Bredero uit 1617 dat de uitspraak *j* voor intervocalische *d* kenmerkend was voor het Brabants, want Bredero legt Jerolimo *een goey meester, moeyer, vayer* in de mond. In de negentiende eeuw zijn veel verzamelingen aangelegd van (op dat moment als verderfelijk verworpen) Duitse leenwoorden, waarin bijvoorbeeld staat dat *afzet*, *bespreken*, *betwijfelen*, *bewonderen* en *inburgeren* germanismen zijn die dienen te worden vermeden. Aan de vorm van deze woorden kun je de Duitse herkomst nu niet meer direct zien, dus dergelijke lijsten zijn erg waardevol (al moeten ze kritisch worden gezien). Uit negentiende- en twintigste-eeuwse bronnen weten we ook dat woorden als *afzien* ‘lijden’, *eindejaarsuitkering*, *gezapig*, *nipt*, *peperkoek*, *prietpraat*, *stilaan*, *uitbater* en *op voorhand* zich vanuit het Zuid-Nederlandse taalgebied verspreiden over de gehele Nederlandse standaardtaal.

De tweede manier is individuele woorden systematisch terug te zoeken in gelokaliseerde en gedateerde corpora en achterhalen of ze ooit een beperkte regionale verspreiding hebben gekend. Dan blijkt bijvoorbeeld dat in de dertiende eeuw *zieden* en *huwen* beperkt waren tot het Vlaams en Brabants, terwijl *koken* en *trouwen* hun oorsprong hebben in Hollandse dialecten. Zoals eerder opgemerkt vertonen de gegevens uit de middeleeuwen en daarna veel lacunes, maar ook taalkaarten uit de negentiende en twintigste eeuw kunnen belangrijke informatie over de herkomst geven, omdat veel dialectgrenzen behoorlijk stabiel zijn. Uit dergelijke moderne taalkaarten blijkt bijvoorbeeld dat *ham* en *sturen* van oorsprong Hollands en Utrechts zijn, terwijl *hesp* en *zenden* uit Zuid-Nederlandse dialecten stammen en *schenk* en *schicken* in oostelijke dialecten voorkomen.



Via gelokaliseerde en gedateerde corpora wordt het wellicht mogelijk de precieze herkomst te bepalen van woorden waaraan wel te zien is dat ze uit een dialect afkomstig zijn, maar waarbij vooralsnog onbekend is uit welk dialect. Denk aan *achterdocht* en *besmeuren* – de verwachte standaardvormen zijn *achterdacht* en *besmeren*. Ook woorden als *kanen*, *knakker*, *knoeperd*, *meuren*, *pleuren*, *reuring* en *ruft* zullen oorspronkelijk dialectwoorden geweest zijn. Deze woorden zijn pas laat in het Standaardnederlands aangetroffen, en er zijn geen varianten van bekend. Hoewel ze erg jong zijn, hebben we geen idee uit welke regio ze stammen.

Het uiteindelijke doel is om de DNA-vingerafdruk van zoveel mogelijk woorden af te nemen, zodat we van een groot aantal woorden kunnen achterhalen uit welk dialectgebied ze afkomstig zijn. Zo kunnen we het complete genoom van de Standaardnederlandse woordenschat blootleggen. Dit zal leiden tot een groot aantal nieuwe inzichten in de historische patronen van ons lexicon. Als we weten wat het aandeel van de verschillende dialecten is in de huidige standaardtaal en tot hoe ver terug in de tijd dat aandeel per dialect gaat, kunnen we conclusies trekken over de status van dialecten en taalgebruikers in bepaalde periodes. Dat status al heel vroeg een beslissende rol speelde, blijkt uit het feit dat al omstreeks 1300 in het Vlaams de /eu/-uitspraak van bijvoorbeeld *koning* en *molen* als minder beschaafd gold en men daarom de spelling *kuening* en *muelen* in formele of hoogdravende teksten vermeed, hoewel dat de normale Vlaamse uitspraak was. Ook kunnen we onderzoeken wat er gebeurt als sprekers van verschillende dialecten met elkaar in contact komen: in bepaalde periodes trokken veel mensen uit een bepaalde regio naar een ander gebied (denk aan de Zuid-Nederlandse immigranten in de Republiek tijdens de Tachtigjarige Oorlog, of Duitse huurlingen in het zeventiende-eeuwse Nederlandse leger). Welk dialect kreeg in zo'n contactsituatie de overhand, en welke taalveranderingen traden er als gevolg van het taalcontact op? Verschilde dat per periode?

Kortom, we zijn dringend verlegen om een DNA-test voor het bepalen van lexicale verwantschap van standaardtalige vormen met de dialecten van vroeger en nu. Die test kan alleen worden samengesteld door een buitengewoon taalkundig onderzoeker met veel kennis van het Nederlands en het Duits die bovendien binnenkort alle tijd heeft voor onderzoek. Ik

denk dat ik wel iemand ken die in dit profiel past, Ad.

### Dankwoord

Ik dank Hans Beelen voor zijn nuttige suggesties.