

Alle zeventiende-eeuwse kranten in Delpher getranscribeerd

# Oud nieuws voor nieuw onderzoek

**De transcriptie van een grote hoeveelheid kranten maakt het mogelijk om taal- en cultuurhistorische veranderingen in de Gouden Eeuw grootschalig te onderzoeken.**

*Nicoline van der Sijs*

Naar de ontwikkelingen van de Nederlandse Gouden Eeuw wordt veel onderzoek gedaan. Drukwerk uit de zeventiende eeuw vormt daarbij een belangrijke bron. En alhoewel er veel drukwerk is gepubliceerd en bewaard, bestond er tot nu toe geen aaneengesloten digitaal tekstcorpus waarmee taal- en cultuurhistorische veranderingen systematisch konden worden onderzocht.

## 20 miljoen woorden

Maar er is goed nieuws. Vrijwilligers hebben alle 17e-eeuwse kranten op Delpher - lopend van 1618 tot 1700 - getranscribeerd. Hiermee is het beschikbare digitale onderzoekscorpus van de zeventiende eeuw in één klap bijna verdubbeld. Het krantencorpus bestaat uit 6.184 verschil-

lende kranten die samen een kleine 20 miljoen woorden bevatten. Ter vergelijking: de DBNL-teksten voor deze eeuw bestaan uit circa 24 miljoen woorden. Door het transcriberen kunnen krantenteksten voor het eerst systematisch met de computer worden onderzocht. Tot nu toe was dat niet mogelijk omdat de optische tekenherkenning waarmee de teksten op Delpher waren gelezen, niet overweg kon met het gotische schrift en Oudnederlands. Medio 2020 komt het getranscribeerde krantencorpus beschikbaar via Delpher.

## Verrijking van data

Het Meertens Instituut werkt aan het verder verrijken van de digitale tekstbestanden. Zo worden de metadata opgeschoond en uitgebreid en worden afzonderlijke artikelen semi-automatisch afgesplitst en voorzien van informatie over de tekstsoort (zoals advertentie, binnenlands nieuws, officiële mededeling). Ook de geografische namen die in de krantenkoppen voorkomen, worden verrijkt en benut: aan iedere naam wordt de moderne spelling toegevoegd. Die moderne schrijfwijzen kunnen vervolgens worden ingevoerd in een kaartprogramma, dat week voor week visualiseert waar

het nieuws binnen en buiten Europa vandaan kwam, en hoe de geografische focus in de loop van de eeuw veranderde.

## Lacunes in kennis

De opgeschoonde en verrijkte krantenteksten komen in 2020 ook beschikbaar via een aparte interface. Dan kan iedereen zijn eigen onderzoeksvragen stellen, bijvoorbeeld naar maatschappelijke veranderingen of veranderingen in het taalgebruik. De teksten kunnen allerlei lacunes in kennis en gegevensbronnen aanvullen: zo ontdekte het Meertens Instituut al dat kranten een groot aantal woorden en spellingen bevatten die ontbreken in de bestaande historische lexica van het Nederlands. Het krantencorpus kan een proeftuin worden voor het testen van tools en modellen, zoals semantische vectoren en *topic modelling*. En de liefhebber kan natuurlijk ook gewoon het laatste nieuws van een bepaalde datum lezen.

[meertens.knaw.nl](http://meertens.knaw.nl)

*Nicoline van der Sijs is projectleider bij het Meertens Instituut. Heeft u vragen of suggesties, of wilt u meewerken aan dit project? Neem dan contact op: [post@nicolinevdsijs.nl](mailto:post@nicolinevdsijs.nl).*



**In plaats van woord voor woord lezen, kunnen onderzoekers de computer de getranscribeerde krantenteksten laten doorzoeken.**

*Credits: Detail uit Amsterdamsche Courant, 1684, via delpher.nl*

Artificiële Intelligentie achter de Liederbank

# Melodiegelijkenissen opsporen met algoritme

**In de Nederlandse Liederbank, waarin ruim 175.000 Nederlandse liederen zijn ontsloten, kun je zoeken naar melodiegelijkenissen. Peter van Kranenburg ontwikkelde het algoritme achter deze functie.**

*Mathilde Jansen*

Wie in de Nederlandse Liederbank 'Elf november is de dag' intypt, komt via 'vergelijkbare melodieën' terecht bij 'Daar was laatst een meisje loos'. Die mogelijkheid om naar melodiegelijkenissen te zoeken, is de verdienste van Peter van Kranenburg. Hij is computationeel musicoloog aan het Meertens Instituut en onderzoekt muziek aan de hand van computermodellen.

In 2010 promoveerde hij op een uitlijningsalgoritme. "Dat schrijft de melodieën zo onder elkaar, dat de overeenkomende noten precies onder elkaar komen te staan", legt Van Kranenburg uit. "Het algoritme zoekt uit op welke plekken ruimte toegevoegd moet worden, zodat de corresponderende delen onder elkaar staan. Hoe meer ruimte, hoe slechter de gelijkheid. Grof gezegd."

## Tune families

Als je een uitlijning maakt van een query-melodie met alle melodieën uit de Liederbank, en die sorteert, dan komen de meest gelijkende melodieën bovenaan. "Net als Google-resultaten", verduidelijkt Van Kra-

**In de Nederlandse Liederbank is het mogelijk om naar melodiegelijkenissen te zoeken. Een van de zoekresultaten bij 'Elf november is de dag' is 'Daar was laatst een meisje loos'. Credits: Meertens Instituut**

nenburg. Zo zie je welke melodieën varianten zijn van elkaar, en kun je ze onderverdelen in families, ook wel 'tune families' genoemd. Tegenwoordig is veel kunstmatige intelligen-

tie gebaseerd op neurale netwerken. Daarom onderzocht Van Kranenburg samen met collega's van het Meertens Instituut en de Universiteit Antwerpen of dit ook werkte bij melo-

diegelijkenissen. "Om het neurale netwerk te trainen, werden steeds twee melodieën aangeboden die wel op elkaar lijken en twee die niet op elkaar lijken. Als je dat lang genoeg doet, met heel veel verschillende melodieën, in ons geval zo'n zesduizend, dan hoop je dat zo'n netwerk op een gegeven moment leert wat het betekent dat twee melodieën op elkaar lijken." En dat lukte. Het model vond melodiegelijkenissen met een betrouwbaarheid van 70 tot 80 procent. Iets beter dan het uitlijningsalgoritme. "Nog geen grote verbetering, maar het laat wel zien dat het model werkt. En dat biedt perspectief voor de toekomst en vormt nieuwe uitdagingen. Want het neurale netwerkmodel is misschien wel intelligent, voor mensen is het soms moeilijk te interpreteren wat het allemaal doet. Het begrijpelijk maken van die netwerken is een belangrijk onderzoeksgebied. Daar willen we in een volgende stap aan bijdragen door te onderzoeken wat ons netwerk geleerd heeft over melodische gelijkheid."

[liederbank.nl](http://liederbank.nl)

## JONG TALENT

**'Een uitlijningsalgoritme schrijft de melodiën en overeenkomende noten precies onder elkaar'**



## Peter van Kranenburg

Van Kranenburg studeerde Musicologie aan de Universiteit Utrecht en Electrical Engineering aan de TU Delft. Hij promoveerde in 2010 aan de Universiteit Utrecht. Hij werkt als computationeel musicoloog bij het Meertens Instituut en de Universiteit Utrecht.