

Vrijgebruiken in Nederland in kaart gebracht

Vrijers uit het verleden digitaal



Dankzij informanten is meer bekend over vrijers uit het verleden foto Chippix/Shutterstock

Afspraakjes maken, uitgaan en verkering krijgen. Voor velen, maar zeker voor historisch-demografen en gezinshistorici, interessante onderwerpen.

Douwe Zeldenrust en Hilde Bras

Al sinds 1931 stuurt het Meertens Instituut vragenlijsten naar informanten in Nederland en Vlaanderen. De vragenlijsten gaan over dialect, naamkunde en volkskunde.

Gemiddeld stuurden zo'n 1.000 mensen een ingevulde vragenlijst retour. Tegenwoordig worden de vragenlijsten digitaal verstuurd en ingevuld door het Meertens Panel bestaande uit ruim 5.000 informanten. De teller staat op 220 vragenlijsten. En deze worden de komende jaren gedigitaliseerd.

Op dit moment worden de antwoorden van vragenlijst nummer 40 uit 1971 gedigitali-

seerd. Lid van de Jonge Akademie van de KNAW en onderzoekster Hilde Bras: "Nummer 40 is een vragenlijst over vrijgebruiken. Het doel van deze vragenlijst was destijds om meer te weten te komen over verkering en verloving. De vragen gaan over de 'vrijer' en 'vrijster', de manier van kennismaken, de verkering en de verloving en bevat voor historisch-demografen en gezinshistorici zeer interessante gegevens over vrij- en verlovingpatronen. Dergelijke gegevens kunnen op geen enkele andere wijze op systematische manier verkregen worden."

Ook data wordt gekoppeld

"We koppelen de culturele gegevens uit nummer 40 aan recent beschikbaar gekomen grote databestanden. Voorbeelden zijn individuele gegevens over levenslopen (Historische Steekproef Nederlandse Bevolking) en sociaal-economische en demografische gegevens op gemeenteniveau (Historische Databank Nederlandse Gemeenten). Zo worden eerdere bezwaren en tekortkomingen van de vragenlijsten deels ondervangen. Hiermee ontstaat er in wetenschappelijke zin meerwaarde boven wat er tot op heden op grond van de vragenlijsten is gedaan."

Douwe Zeldenrust, coördinator van de onderzoekscollecties van het Meertens Instituut, voegt hieraan toe: "In dit kader zijn er nog meer interessante vragenlijsten. Na de vragenlijst over vrijgebruiken wilden de onderzoekers destijds meer inzicht krijgen in huwelijksgebruiken. Vragenlijst nummer 43 uit 1973 gaat over die huwelijksgebruiken en kan ook voor hedendaags onderzoek significante gegevens bevatten. We verwachten de eerste resultaten halverwege 2013."

The power of the family

De digitalisering van vragenlijst 40 is een samenwerking tussen de Radboud Universiteit en het Meertens Instituut (KNAW). De data die dit oplevert, worden door Hilde Bras geanalyseerd in het kader van haar project 'The power of the family', dat door NWO gefinancierd is met een Vidi-subsidie vanuit de Vernieuwingsimpuls. Wilt u meer weten over data over daten? Neem contact op met Douwe Zeldenrust (douwe.zeldenrust@meertens.knaw.nl) of Hilde Bras (h.bras@let.ru.nl).

www.ru.nl/hlcs/the_power_of_the_family

Online krantencorpus maakt vergelijkend onderzoek mogelijk

De toon in kranten is de laatste decennia informeler geworden, zo is de algemene opinie. Maar om dat met cijfers te kunnen onderbouwen, is een groot geannoteerd corpus nodig. En dat is nu online toegankelijk.

Erica Renckens

"Uit eerder onderzoek bleek al dat Amerikaanse en Britse kranten informeler zijn geworden, met name in de tweede helft van de 20e eeuw", vertelt Kirsten Vis. Vis werkt als postdoc aan de Universiteit Utrecht en promoveerde in 2011 aan de VU op een soortgelijk onderzoek naar Nederlandse kranten. "Toen ik begon met mijn onderzoek was er geen materiaal uit 1950 digitaal beschikbaar, dus heb ik zelf pagina's gedigitaliseerd." Uiteindelijk leidde dit tot het VU-DNC-corpus (Vrije Universiteit - Dutch Newspaper Corpus), met daarin bijna twee mil-

joen woorden uit vijf grote Nederlandse landelijke kranten uit de perioden 1950/1951 en 2002. Dit corpus is sinds eind 2012 beschikbaar via de CLARIN-infrastructuur.

Toename subjectiviteit

Het corpus verschilt op twee punten essentieel van het krantenproject van de Koninklijke Bibliotheek (KB) waarin Nederlandse dagbladen online beschikbaar worden gemaakt. "Het project van de KB heeft van de grote landelijke kranten alleen op grote schaal artikelen uit De Telegraaf overgenomen", legt Vis uit. "Alleen van de periode rond de Tweede Wereldoorlog zijn ook artikelen uit andere kranten opgenomen." Het VU-DNC-corpus bevat artikelen uit De Telegraaf, Algemeen Dagblad, Trouw, de Volkskrant en NRC Handelsblad. Daarnaast is het VU-DNC-corpus voorzien van verschillende annotatielagen. Vis: "Naast woordsoort-

en lemma-informatie bevatten die lagen ook annotaties met subjectieve elementen die de mening, houding of emotie van de schrijver of geciteerde aangeven." Dat zijn bijvoorbeeld modale elementen ('mogelijk', 'zeker', 'hopelijk'), intensiverende elementen ('nogal', 'erg'), uitroepen ('wat mooi!') en voornaamwoorden ('ik', 'mijn', 'jouw'). "Bovendien heb ik de directe citaten als zodanig gemarkeerd. Zo kan onderscheid worden gemaakt tussen de tekst waarin de journalist zijn eigen woorden gebruikt en die waarin de woorden van een ander worden geciteerd."

Met behulp van het VU-DNC-corpus kon Vis alsnog de gestelde onderzoeksvraag beantwoorden. "Ik vond dat er inderdaad meer meningen, houdingen en emoties worden uitgedrukt in kranten uit 2002 dan in kranten uit 1950/1951."

Maar deze toename aan subjectiviteit was vooral te danken aan de

woorden van de geciteerden. "De journalist was helemaal niet subjectiever gaan schrijven, maar het gebruik van directe citaten was in de tijd verdubbeld. En in de citaten vond ik wél een toename van subjectiviteit."

Ijkkpunt

Het VU-DNC-corpus biedt geesteswetenschappers niet alleen de mogelijkheid om vergelijkend kwantitatief onderzoek te doen naar landelijke dagbladen. Het corpus dient ook als ijkpunt voor OCR-postcorrectietools, waarmee spel fouten uit de automatische woordherkenning kunnen worden gecorrigeerd. "Na het digitaliseren zijn verschillende correcties uitgevoerd. De ongecorrigeerde en gecorrigeerde bestanden hebben we woord voor woord opgelijnd. Deze verzameling kan dienen voor toetsing en training van OCR-correctietools."

portal.clarin.inl.nl/vu-dnc/

KORT

KB lanceert de app 'Hier was het nieuws'

De KB lanceert de app 'Hier was het nieuws'. Met deze app kunt u overal in Nederland historische informatie oproepen over de plek waar u bent. De app van de KB is gebaseerd op de website Historische Kranten. Op deze website zijn miljoenen gedigitaliseerde krantenpagina's gratis toegankelijk. De app toont krantenartikelen tot 1940 en maakt gebruik van GPS. (HB)

www.hierwashetnieuws.nl

