Digitale Statenvertaling: vrijwilliger helpt wetenschap

PETER BOOT

Een groep van meer dan honderd vrijwilligers onder leiding van Nicoline van der Sijs digitaliseerde de eerste druk van de Statenbijbel (1637). Oud-KNAW-president Frits van Oostrom presenteerde op 22 mei het resultaat. Het project is nog niet voorbij: met deelprojectleider Hans Beelen werkt Van der Sijs nog aan een aantal andere bijbelversies. *e-data&research* sprak met haar over de rol van de vrijwilligers en het belang van het project voor de wetenschap.

NFP PHOTOGRAPHY

Nicoline van der Sijs met de Statenbijbel

'De motivatie van de vrijwilligers en de werkinzet is enorm, letterlijk onbetaalbaar', benadrukt de projectleidster. De inzet van de vrijwilligers kan volgens haar niet genoeg worden gewaardeerd. Daarbij is een goede organisatie van hun werk belangrijk. Van der Sijs toont zich vooral opgetogen over de snelheid waarmee is gewerkt: in minder dan een jaar zijn de 2.6 miljoen woorden overgetypt en nagekeken. 'Dat kan doordat de vrijwilligers zich serieus genomen voelen'. Een duidelijke streefdatum en heldere afspraken motiveren mensen om aan de slag te gaan. Dankzij die goede afspraken blijkt het mogelijk om niet alleen de Statenvertaling van 1637 te digitaliseren, maar daarnaast nog een hele serie andere vertalingen. Van der Sijs' organisatietalent werd tijdens de presentatie van het project geprezen door Herman Wiltink: 'Ze heeft alles goed doordacht en op de lange termijn gepland, met aandacht voor het allerkleinste detail'.

De groep van vrijwilligers die zich voor het project aanmeldde was enkele malen groter dan verwacht, zodat Van der Sijs zich de luxe van een wachtlijst kon permitteren. De meeste vrijwilligers zijn niet speciaal godsdienstig gemotiveerd, aldus van der Sijs, maar eerder door belangstelling voor taal en cultureel erfgoed. Qua beroep was de groep vrijwilligers zeer gemêleerd. Alleen voor de correctoren heeft Van der Sijs vrijwilligers gekozen die beroepsmatig of op een andere manier affiniteit hadden met het correctiewerk.

Inmiddels is behalve de Statenvertaling van 1637 ook de Delftse bijbel uit 1477 digitaal beschikbaar, net als de eerste Friese Bijbelvertaling (uit 1943). In september volgt waarschijnlijk de lutherse bijbel (1648), in december de zgn. Deux-Aesbijbel (1562), volgend jaar volgt nog een aantal. Zo ontstaat een corpus aan bijbelvertalingen dat voor taalkundig onderzoek van groot belang kan zijn.

In haar artikel 'Het boek der boeken nu ook digitaal' (link onderaan dit artikel) somt Van der Sijs een aantal mogelijkheden op. De doorzoekbaarheid van de teksten is van belang voor het onderzoek naar de ontwikkeling van de Nederlandse woordenschat. Het wordt ook mogelijk om teksten te vergelijken op stijl en woordgebruik: zowel de verschillende bijbelvertalingen, maar ook tussen de bijbeltekst en literaire teksten. Daarmee zou iets duidelijk kunnen worden over de manier waarop de bijbelvertaling het literair Nederlands heeft beïnvloed. Er kunnen ook praktische toepassingen zijn: de beschikbaarheid van een groot corpus in zestiende- en zeventiende-eeuwse spelling maakt het mogelijk OCR (tekstherkenning) te verbeteren, omdat programma's daarvoor gebruik maken van bekende spellingen.

Hoe kunnen onderzoekers bij de tekst? Op het moment zijn de gedigitaliseerde teksten beschikbaar op de sites van het Instituut voor Nederlandse Lexicologie (INL), de DBNL en het Nederlands Bijbelgenootschap. De tekst is doorzoekbaar via www.biblija.net. Daar kunnen ook verschillende bijbelversies worden vergeleken. Deze mogelijkheden zijn echter beperkt tot de eigenlijke Bijbeltekst. De Kanttekeningen van de vertalers blijven hier buiten beschouwing. Het INL werkt aan andere mogelijkheden voor onderzoekers, zoals een mogelijkheid om frequentielijsten af te leiden of concordanties te maken (lijsten van zoekwoorden in context). Onderzoekers die meer willen dan dat kunnen ook de hele tekst ter beschikking krijgen, maar moeten dan een contract tekenen dat ze de tekst alleen voor onderzoeksdoeleinden zullen gebruiken. De eerste belangstellenden hebben zich al gemeld.

http://www.inl.nl/images/stories/bijbels/statenvertaling1637/

TestweekLab: een collaboratory met psychologische testdata

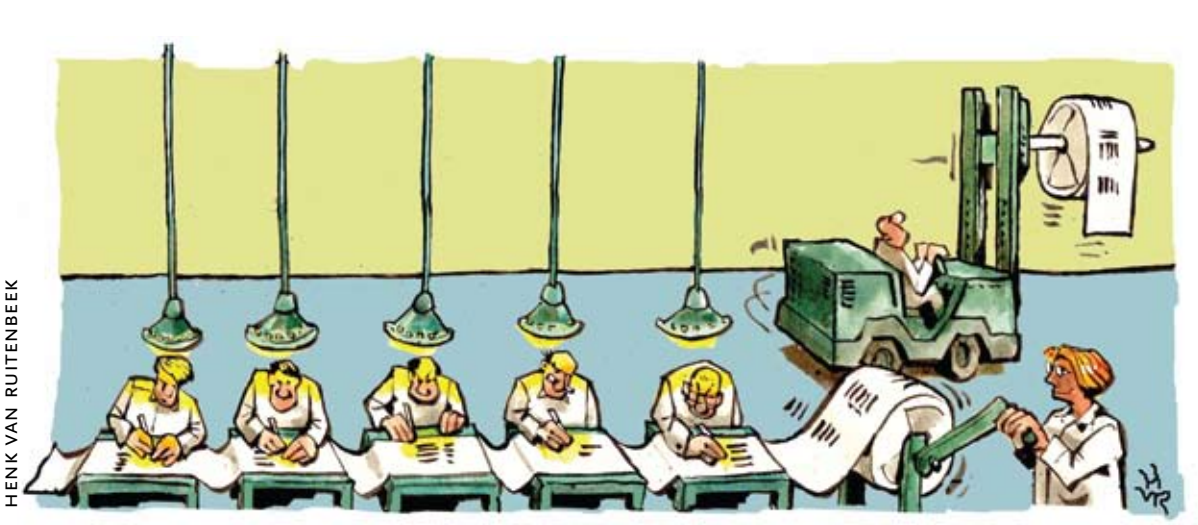
Al sinds veertig jaar organiseert de afdeling Psychologie van de Universiteit van Amsterdam (UvA) elk jaar de 'testweek'. Vorig jaar ging een project van start om de grote hoeveelheid data die dat opleverde te ontsluiten.

CASPAR TREIJTEL

De testweek is een propedeuseonderdeel van de opleiding Psychologie, waarin de psychologiestudenten worden getest. De naam dateert nog uit de tijd dat alle tests in één week werden gedaan; tegenwoordig vinden ze gedurende meerdere weken plaats.

Als methode van dataverzameling heeft hij twee unieke eigenschappen: er zijn zeer veel data beschikbaar en een aantal vragenlijsten wordt elk jaar afgenomen zodat van grote contuïteit kan worden gesproken.

In november 2007 is het project TestweekLab van start gegaan. Daarin is een begin gemaakt met het ontsluiten van data uit de testweek en het maken van een collaboratory daaromheen. Met de collaboratory worden twee doelen nagestreefd. Ten eerste wordt geprobeerd de heersende geslotenheid in het delen van elkaars datasets open te breken. Het online aanbieden van deze schat aan testdata moet in dat opzicht een voorbeeldfunctie vervullen. Het tweede doel is het ontsluiten van de grote hoeveelheid testdata, waarbij deze op een duurzame manier in het systeem worden opgeslagen.



HENK VAN RUITENBEEK

Actieve community

Het project TestweekLab is onderdeel van het thema collaboratories binnen het programma SURFshare. De collaboratory die is gebouwd, ook genaamd TestweekLab, heeft als uitgangspunten toegang te bieden tot elkaars testdata en samen te werken op het gebied van psychologisch onderzoek. In eerste instantie zal TestweekLab speciaal voor data uit de testweek worden gebruikt. In de toekomst zullen de opgeslagen testdata ook voor onderzoekers buiten de UvA toegankelijk zijn. Omdat TestweekLab een grote dataverzameling zal ontsluiten, waarmee zowel verticale (meerdere jaren testdata) als horizontale doorsnedes (meerdere tests in een jaar) gemaakt kunnen worden, wordt erop gerekend dat veel externe onderzoekers belangstelling zullen hebben voor het gebruik ervan.

Om het eerste doel te verwezenlijken – meer openheid rond data uit psychologisch onderzoek – is het belangrijk dat een actieve *community* ontstaat rondom de collaboratory. Een succesfactor voor een collaboratory is volgens Wikipedia de bereidheid van gebruikers om een gemeenschappelijk doel te verwezenlijken. Daarnaast is het cruciaal dat ze een *incentive* geboden krijgen voor het leveren van een bijdrage. In het geval van TestweekLab moeten de testdata als trekpleister dienen voor actieve deelname. TestweekLab biedt online functionaliteiten die samenwerking rond de data mogelijk maken, zoals emaillijsten, een wiki en gedeelde documenten. De afdeling Psychologie van de UvA zal een eerste vulling van deze onderdelen verzorgen, waarmee de meerwaarde van het online aanbieden van datasets expliciet gemaakt wordt.

Uitbouwen met meer testdata

De collaboratory is gebouwd met behulp van *Sakai* en *Fedora*. Sakai is een online *Collaboration and Learning Environment*. Het is een open source product dat een *plugin*-architectuur kent, wat wil zeggen dat externe partijen gemakkelijk kunnen meewerken aan uitbreiding van functionaliteit door modules te ontwikkelen (ook wel *Sakai tools* genaamd). Deze modules kunnen door beheerders van een Sakai-installatie optioneel geïnstalleerd worden. Voor TestweekLab wordt gebruik gemaakt van de functionaliteiten voor samenwerking die Sakai standaard aanbiedt. Daarnaast is speciaal voor TestweekLab een Sakai tool ontwikkeld die een interface biedt naar de Fedora repository, waarin de testdata worden opgeslagen. Deze worden in de vorm van SPSS-bestanden in de repository opgenomen. Met de Sakai tool kan de gebruiker zoeken naar testdata, browsen, en de metadata van de testen bekijken. Ook kan men testen downloaden en zelf testdata toevoegen.

In de nabije toekomst zal TestweekLab door de psychologen van de UvA verder uitgebreid worden met meer testdata. Voor de opslag van de testdata zal worden geprobeerd om aan het DANS Datakeurmerk te voldoen. Verder is de hoop erop gericht dat er onder psychologen voldoende belangstelling zal zijn om bij te dragen aan het online samenwerken naar aanleiding van psychologische testdata.