
TREFWOORD

1996

Sdu Uitgevers, Den Haag 1997

11

Mogelijkheden en onmogelijkheden van de cd-rom: een praktijkvoorbeeld

door **Nicoline van der Sijs**

Iedereen die ooit een dag met een groot woordenboekbestand op cd-rom heeft gespeeld, weet dat dit medium ongekeerde mogelijkheden biedt, maar ook ongekeerde *on*mogelijkheden. Toen ik eens in een algemeen gezelschap uitlegde dat ik Nederlandse ontleningen in het Engels zocht en daarvoor de cd-rom van de *OED* gebruikte, reageerde een van de aanwezigen wat verontwaardigd met: 'Dus dan hoef je zelf niets te doen, de computer doet alles voor je!' Nee dus.

Het formuleren van een criterium dat alle verlangde woorden zou selecteren, kostte me een dikke week. Toen had ik een bestand dat voor een derde bestond uit de gezochte woorden, en voor twee derde uit 'ruis', die handmatig verwijderd moest worden (een paar dagen werk). De *OED* geeft namelijk in twee gevallen Nederlandse vormen: bij Engelse woorden die ontleend zijn aan het Nederlands, en bij erfwoorden om de verwanten binnen het Germaans te tonen. Idealiter zijn die twee gevallen terminologisch duidelijk van elkaar onderscheiden (en kunnen dus alleen de leenwoorden geselecteerd worden). Maar van die ideale situatie is (nog) geen sprake, omdat het woordenboek geschreven werd in het pre-computertijdperk.

Selectieproblemen

Ter illustratie zal ik een paar voorbeelden geven van problemen die zich tijdens de selectie voordeden.

Bij een beetje vraagstelling kun je niet buiten de zogenaamde 'Booleaanse operatoren' EN, OF, NIET, EN NIET: selecteer a EN b EN c EN NIET d. Bij verschillende selecties kreeg ik echter telkens een foutmelding. Het heeft me een dag gekost om de oorzaak te achterhalen. Om u die dag te besparen zal ik de oplossing verklappen: als een van de selecties nul resultaten heeft, gooit de cd-rom er het bijltje bij neer. De handleiding waarschuwt daar niet voor. Bij een foutmelding moeten dus alle deelselecties apart uitgevoerd worden, totdat blijkt welke nul resultaten heeft. Als deze deelselectie verwijderd wordt, werkt de selectie wel.

Een ander probleem was, dat het niet mogelijk bleek bepaalde codes te zoeken die vóór het trefwoord staan. Ze worden op de cd-rom wel getoond, maar zijn niet selecteerbaar. Het ging mij om de code voor 'Obs.' (*obsolete*, dus verouderd; weergegeven met een *dagger* ofwel overlidensteken voor het trefwoord) en de code voor 'alien, not naturalized' (dus: in ongewijzigde vorm geleend, bijvoorbeeld *advocaat* 'drankje' Dutch *advocaat*, dit wordt weerge-

geven met twee verticale strepen voor het trefwoord). Verouderde woorden zijn wel op te vragen door de afkorting 'Obs.' te zoeken, maar dan krijg je als resultaat alle woorden waarin *ergens in het trefwoord* de afkorting 'Obs.' gebruikt wordt, dus zowel verouderde trefwoorden (wat ik zocht) als trefwoorden met een of meer verouderde betekenissen (wat ik niet zocht). De code voor 'alien' is helemaal niet op te vragen.

Tot zover de onmogelijkheden van de cd-rom. Voordat sceptici gesterkt worden in hun idee dat een cd-rom alleen maar extra werk oplevert en niet leidt tot nieuwe gegevens, gaan we gauw over naar de gevonden resultaten.

Vondsten

Hoewel de selectiemogelijkheden van de huidige cd-roms nog niet perfect zijn (in de toekomst zullen ze overigens steeds beter worden), leiden ze ook nu al tot vondsten die zonder de cd-rom niet mogelijk zijn – behalve wanneer iemand de hele *OED* doorleest, en wie heeft daar de tijd en het geduld voor? Ik heb in het *OED*-bestand een aantal etymologieën van Nederlandse woorden gevonden die mij niet bekend waren uit Nederlandse bronnen. Ik geef een paar voorbeelden.

Wervelwind is volgens de *OED* een leenwoord uit het Oudnoords. De Nederlandse etymologische woordenboeken noemen weliswaar de Oudnoordse vorm, maar 'vergelijken' deze slechts met de Nederlandse vorm, ze geven geen ontleningsrelatie. Toch acht ik deze zeer waarschijnlijk, omdat het woord in het Nederlands voor het eerst in de zestiende eeuw opduikt, het Oudnoords meer zeevaarttermen geleverd heeft (zoals de visnamen *dors* 'jonge kabeljauw' en *haai*), en Vondel volgens het *WNT* in 1671 schrijft: 'Wervelwint. Hier zegt men *dwarrelwint*.' De contacten tussen Nederland en de Scandinavische landen werkten in deze periode overigens beide kanten op: het woord *maelstrom* is door de Scandinavische talen, het Engels en het Duits uit het Nederlands geleend (Deens, Noors *malstrøm*, Zweeds *malström*, Engels *maelstrom*, Duits *Mahlstrom*). *Maelstrom* wordt, wederom volgens het *WNT*, in 1595 voor het eerst in het Nederlands genoemd, en wel op de atlas van Mercator op de kaart van Noorwegen, ter aanduiding van een draaikolk bij de Lofoteneilanden voor de noordwestelijke kust van Noorwegen.

Verder blijkt uit de *OED* dat een aantal wetenschappelijke termen door Nederlandse geleerden gevormd is van Latijnse of Griekse elementen, en vervolgens door andere talen is overgenomen. Zo is in het Engels *adstratum* ontleend aan Nederlands *adstraat*, dat gevormd is van Latijn *ad* 'naar' en *stratum* 'uitgespreid', in navolging van de taalkundige termen *substratum/substraat* 'oudste taal van een land, later door een ingevoerde taal vervangen' en *superstratum/superstraat* 'taalkundige bovenlaag'.

Adstraat is volgens de *OED* gevormd door Marius Valkhoff, en voor het eerst gebruikt in zijn inaugurele rede *Latijn, Romaans, Romeens* (1932:17). En inderdaad wordt het woord hier vermeld en uitgelegd. Het duidt op 'taalelementen of eigenschappen die wijzigingen veroorzaken in een taal die verder dominant is'. Het woord *adstratum* komt voor het eerst in 1939 in het Engels voor.

Aprosexia is in het Engels ontleend aan Nederlands *aproxia*, een woord dat voor het eerst door A.A.G. Guye in 1887 in het *Nederlandsch Tijdschrift voor Geneeskunde* (XXIII:381) gebruikt wordt en dat gevormd is van Grieks *aproxia*, van ontkenkend *a-* en *prosexis* 'aandacht'. *Aproxia* betekent 'abnormaal verminderd vermogen om de aandacht te richten'. Het komt in het Engels in 1889 voor het eerst voor, in een Engels artikel van dezelfde Guye.

Het Engels heeft *elaioplast* 'soort plastide, vooral gevonden in sommige angiospermen of bedektzadigen, dat olie bevat of dit afscheidt' ontleend aan Nederlands *elaioplast*. J.H. Wakker noemde dit elaioplast als eerste in 1887 in het *Maandblad voor Natuurwetenschappen* (XIII:110). Hij vormde het woord van Grieks *elaiion* 'olie' en *plastos* 'gevormd'; in het Nederlands wordt het daarom ook wel 'vetvormer' genoemd. Al een jaar later, in 1888, komt het in het Engels voor.

Veelgebruikte woorden zijn *supergeleidend*, *supergeleider* of *suprageleidend*, *suprageleider*. Deze woorden zijn dan ook in alle Nederlandse woordenboeken opgenomen, maar zonder verklaring van de herkomst. Het Engels gebruikt *superconducting*, *superconductor*, zelden *supraconducting*, *supraconductor*. Omdat in deze eeuw de Nederlandse wetenschap zwaar beïnvloed is door de Engelse en Amerikaanse en de internationale wetenschappelijke voertaal Engels is, is men geneigd te veronderstellen dat de Nederlandse woorden wel vertalingen van de Engelse zullen zijn. Het was voor mij dan ook een verrassing in de *OED* te lezen dat het precies andersom is en dat de Engelse woorden vertalingen van het Nederlands zijn. Nu had ik ook niet direct paraat dat de supergeleiding in 1911 ontdekt was door de Nederlander Heike Kamerlingh Onnes, die daarvoor in 1913 de Nobelprijs kreeg. De *OED* vermeldt zelfs in welke publicatie Kamerlingh Onnes voor het eerst de woorden *suprageleidend* en *suprageleider* gebruikt: in de *Verlagen van de gewone Vergadering der Wis- en Natuurkundige Afdeling van de Akademie van Wetenschappen te Amsterdam* (1912-1913, XXI:1390). *Superconductor* wordt vanaf 1931 in het Engels gebruikt, *superconducting* vanaf 1935.

Tot besluit het Engelse *undation theory* 'theorie dat selectieve interne verwarming van de aardmantel grote golfvormige plooiën in de aardkorst veroorzaakt'. Deze geologische term is volgens de *OED* een vertaling van het Nederlandse woord *undatietheorie*, dat in 1932 door R.W. van Bemmelen in het *Natuurkundig Tijdschrift Nederlandsch-Indië* (XCII:93) gemunt is. Het komt van Latijn *undare* 'golven', en is vanaf 1950 in het Engels geattesteerd.

Mijn conclusie is dat de tijd die nu nog verloren gaat door de nog niet perfecte selectiemogelijkheden van de cd-rom, ruimschoots vergoed wordt door de vondsten. Binnen afzienbare tijd zal geen enkele serieuze lexicograaf nog buiten het hulpmiddel van cd-roms van woordenboeken en tekstbestanden kunnen.

Cd-romversie van het WNT

De cd-romversie van het *WNT* heeft noodgedwongen minder zoekmogelijkheden dan die van de *OED*, doordat het *WNT* een veel vrijere en gevarieerdere lemmastructuur heeft. Daardoor bestaat er bijvoorbeeld geen apart 'etymologieveld' in het *WNT*. Maar desondanks is ons een enorme rijkdom in de schoot geworpen, nu het *WNT* ontsloten is. Een reusachtig woord- en citatenbestand is plotseling toegankelijk geworden, ieder woord of iedere combinatie van woorden kan in het gehele *WNT*-bestand gezocht worden, en de toepassingen voor taal- en letterkundigen zijn legio. Sterker nog: de toepassingen moeten nog verzonnen worden, terwijl de mogelijkheden er al zijn. Naarmate het aantal gebruikers van de cd-romversie toeneemt, zal zowel het aantal toepassingen als het aantal wensen toenemen. Je raakt wel snel aan de luxe gewend: ik werk nu ruim een half jaar met het *WNT* op cd-rom, en ik kan me al bijna niet meer voorstellen hoe ik het 'vroeger' deed. De papieren versie raadpleegde ik lang niet voor ieder probleem: het opzoeken kostte veel tijd, en bij veel woorden denk je al bij voorbaat dat ze er wel niet in zullen staan, dus je begint niet eens met zoeken. Maar zoeken op de cd-rom kost helemaal geen tijd, en is een woord niet opgenomen als trefwoord: niet getreurd, dan zoek je toch in de rest van het bestand.

Het zou prachtig zijn als het *WNT* na voltooiing niet alleen wordt aangevuld, maar als de redactie ook tijd zou krijgen om het bestaande bestand redactioneel recht te trekken, consequenter te maken, te 'updaten', dateringen bij de bronnen te vermelden in die delen waar dat nog niet gebeurd is, latere aanvullingen in het lemma te integreren, enzovoorts. Zo blijven er altijd wensen en dromen over, maar wat zijn de mogelijkheden nu al groot! Dit is een mooie tijd voor lexicografen.