

2630 maal 'Dutch'

De Oxford English Dictionary op cd-rom

door H. Brandt Corstius

En ik was zo blij toen ik vijf jaar geleden de twintig delen van de Tweede Druk van de *Oxford English Dictionary* in huis kreeg! Er ligt altijd een van die delen op mijn tafel, omdat ik daar een woord in wou opzoeken, en in dat deel blijf ik dan lezen tot er een ander deel op tafel moet komen.

Ondertussen is er een eenentwintigste deel aan het ontstaan: een boekje, waarin ik op elke bladzij een speciale afwijking bijhoudt. De 150 bladzijden van dit aantekenboekje blijken in drie delen verdeeld te kunnen worden:

1. Naar taal. Van Afrikaans, Baskisch, Belgisch tot zigeuners, Zweeds.
2. Naar de vorm van het woord: *-acism*, acroniem, consonantloos, *-dom*, tot *-ski*, werkwoord and werkwoord, *x-x* (waarmee ik doel op woorden die uit twee gelijke delen bestaan, soms gescheiden door een streepje).
3. Andere curiosa, van Belgisch, betekenisloos, drank, drug, eponiem tot werkwoord in Latijn, ziektes.

Toen gebeurde wat al bij het verschijnen voorspeld was: de CD-versie van de Tweede Druk van de *OED* kwam op de markt. Anthony Burgess schreef er een enthousiast artikel over, omdat hij er zo vaak in geciteerd werd.

Toen onlangs de Apple-Macintosh-versie verscheen, kwam daarover weer een tweede reeks perspublicaties. Wonderlijk is dat altijd wordt gedaan alsof dit de eerste keer is dat een woordenboek machine-leesbaar wordt gemaakt. In Nederland kennen we Lexitron op de eendelige *Van Dale*, maar er zijn er natuurlijk veel meer. Voor zover ik ze ken, is de *OED* de beste, niet alleen omdat de *OED* zelf zo goed is, maar ook omdat het opzoekmechanisme goed is en de resultaten prachtig (met de citaten in blauwe kleur) op je scherm komen.

Bij het gouden plaatje krijg je een keurig instructieboekje van 120 bladzijden, dat alles uitlegt.

Het is handig als je Windows kent, maar nodig is het niet.

Bij het zoeken moet je aangeven wat je zoekt en waar je het wil zoeken.

Je kunt een stuk tekst van méér dan een woord zoeken, dus bijvoorbeeld *black and red*, maar je kunt niet een andere volgorde van die woorden, zoals *red and black* in een zelfde zoekopdracht vinden. Door een vraagteken te zetten of een asterisk, kun je een enkele letter of een reeks van letters willekeurig maken. Zo vind je alle woorden die op *heid* eindigen door *asteriskheid* in te voeren. Heel vervelend is dat je niet kunt eisen dat zulke joker-letters onthouden worden en in meer dan een plaats met dezelfde waarde moeten optreden. Wat zou het heerlijk zijn om alle gevallen waarin drie keer dezelfde letter achter elkaar voorkomt met de zoekopdracht *???* te krijgen. Nu moet je die vraag 26 keer, voor de 26 letters stellen, om uit te vinden dat de drieslag *iii* het vaakst voorkomt.

De plaats waar je wilt zoeken kun je beperken tot alleen het trefwoord (dan vind je het ook zonder de neenvormen), tot de definitie of ook de etymologie. In je zucht niets over te slaan zoek je vaak de hele tekst van de twintig delen af.

Na enkele maanden op de ps/2 Multimedia, bleken mijn zoektochten in vier groepen uiteen te vallen.

1. Specifieke combinaties

Komt *red and black* meer voor dan *black and red*, en is er een verschil tussen die twee volgorden te bepalen? Hoe vaak staat *and* tussen twee woorden die met dezelfde letter beginnen? Eigenlijk gebruik je hier de *OED* als een Engels corpus: veel aandacht gaat uit naar de informatie die in de citaten zit. Hoe goed de *OED* als corpus is, weet ik niet. Het bevat in ieder geval 'alle' Engelse woorden en het bevat teksten uit alle eeuwen. [8]

2. Nederlands in de OED

Al bij de eerste lezing valt op hoeveel Nederlands er in de *OED* voorkomt, en nog foutloos ook.

Het eenvoudigste is om in de tekst het woord *Dutch* te laten zoeken. Maar het heeft geen zin alle 2630 voorkomens daarvan na te gaan. Alleen al het aantal verwijzingen naar Motleys *Geschiedenis van de Nederlandse Republiek* is kolossaal. Beperken we ons tot het steekwoord, dan krijgen we 374 gevallen waarin *Dutch* voorkomt, van *advocaat* tot *Zeeman*. Daar zijn eigennamen bij als *Rembrandt* en *Lorentz*, Nederlandse woorden als *predikant* en *provo*, en verengelste woorden als *korfball* en *pensionary*.

Twee belangrijke bronnen van Nederlandse invloed zijn die via Zuid-Afrika en uit de tijd van drie eeuwen geleden toen veel Nederlandse woorden in Engelse teksten figureerden.

In het veld etymologie komen naast de 6286 gevallen waar het Nederlands, vaak ter illustratie van de woordvorm in een verwante taal, voorkomt, ook nog 563 Vlaamse etymologieën, 30 Kaapse en 272 Zuid-Afrikaanse afleidingen voor. Ook zijn er drie Belgische etymologieën.

Met behulp van een rij hoogfrequente Nederlandse woorden (niet *in* en *is*, die ook in het Engels hoogfrequent zijn), vond ik Nederlandse tekstresten als: *Niet rooken, het huys is van leem gemaakt, kriecken van de dag, mijns weten niet, kruysssen op de zee*.

De achtervoegsels *-heid* en *-je* leveren een honderdtal gevallen op. Bij *-heid* is vaak een verengelsing met *-head* te vinden: *roekeloosheid recklesshead*. Bij *-je* moeten vondsten als *dirai-je* en *equipaje* even verwijderd worden.

3. Engelse woordcuriosa

Ik dacht dat, na het Duitse en Nederlandse palindroomwoord, ook het Engelse palindroomwoord voldoende bekend was. Ik wou alleen een laatste check uitvoeren. Verrassing! Ik vond vele keerwoorden die ik nooit eerder in de Angelsaksische literatuur ontmoet had. Mijn oogst is boven de 200 en ik heb ze nog niet allemaal.

Voor elk tweetalletters *x* en *y* moest ik de zoekopdracht *xy*yx* geven, om daar de palindroomwoorden die met *xy* beginnen uit te ziften. Zo werden juweelties ontdekt als *alula, ebibe, finnif, Nauruan, Ogopogo, Qazaq, rotavator, sbobs, tatarattat, wowwow* en *zerez*. Maar in de *OED* staan niet alle verbuigingen en vervoegingen. Zo moet je naast de wel geregistreerde gevallen als *deifeid* en *dewed* op zoek naar andere de-werkwoorden die misschien een *-ed-vorm* hebben die omkeerbaar blijkt. Zo vind ik *deffed* en *dessed*. Het woord *dew-awed* blijft onvindbaar, hoewel het uit de Opperengelse literatuur bekend is.

Haast elk Engels woord kan van een extra *-s* voorzien worden, soms met een apostrof ervoor. Dus moest voor 25 letters *x* de combinatie *sx*x* onderzocht worden. Dit leverde onder meer op: *Snellen's, shahs* (door Nabokov al in *Pale Fire* voorzien), *stemmet's, Sawbwa's, semi-times*. Een totaalijst hoop ik in het Amerikaanse tijdschrift *Word Ways* te publiceren.

In Lexitron is het mogelijk te vragen naar Nederlandse woorden die een bepaalde verzameling letters in ieder geval bevatten. In *OED* moet je altijd de volgorde aangeven. Zo vindt het zoekwoord *a*e*i*o*u** het prachtige *abstemiously*. Bij *caesious, taedious* en *marvellous* moet je zelf bedenken of je er *-ly* achter mag plaatsen.

Woorden met dezelfde letter vaak gebruikt, waren bijvoorbeeld: *khandalavala, evergreenness, Mmmmmmm, non-intention, chrononhotonthologos, pip-pip-pip, stresslessness, untumultuous* en weer *tatarattat*.

Denkt u dat de reeks *abcde* in een Engels woord voorkomt? Zelfs de reeks *abcdef* komt voor, in 'The rectangle *abcdef*...'

4. Bijzondere woordvorming

Ik was geïnteresseerd in woordvormingen als *portmanteau, rhyming slang, acroniem, onomatopee, university slang*. Maar hoe kan de *OED* zulke gevallen aanduiden; zullen deze aanduidingen in het veld definitie of in het veld etymologie staan? Nu kwamen de inconsistenties in de *OED* duidelijk naar voren.

University slang wordt 2 maal in het etymologie en 23 maal in het definitieveld gegeven. Voor onomatopee lukte het me niet enig systeem te vinden.

Hoe vaak wordt in de *OED* naar het welbekende 'Well-known' in de definitie van een betekenis uitgeweken? Je zoekt dan naar definities waar het woord *well-known* in staat. Ik vond er 134 gevallen van. maar niet altijd zo passend als bij *asparagus* [9] ('plant [...] cultivated for the sake of its vernal shoots, which form a well-known delicacy of the table'). Soms staat het woord in de frase 'less well-known'. Vaak slaat het woord *well-known* op een ander deel van de definitie dan op het gedefinieerde

woord zelf.

Portmanteau (vlecht- of mengwoord) stond 11 maal in het etymologieveld, 26 keer in de definitie. En ik ben allerminst overtuigd dat ik alle *portmantea*-woorden op deze wijze heb gevonden.

Conclusie

Pas bij elektronisch gebruik van een woordenboek komen de inconsistenties aan de oppervlakte. Hoe scheid je precies etymologie en definitie? Hoe dwing je de lexicograaf tot uniforme aanduidingen, zodat de gebruiker zich niet het hoofd hoeft te breken over synoniemen die gebruikt kunnen zijn?

De zoekstrategie van de *OED* is onvergelijkelijk veel mooier dan die *Van Dale* in Lexitron bouwde. Het is dan ook prettig dat de CD-versie van het *WNT* gemaakt zal worden door *AND*, die ook de software voor de *OED* leverde. Ik smeed ze hierbij om, anders dan bij de *OED* gebeurde, van de jokertekens vraagteken en asterisk echte variabelen te maken waarvan de waarde wordt vastgehouden, zoals dat in alle programmeertalen gebeurt.

Een woordenboek dat niet langer op papier verschijnt, is een heel ander woordenboek dan een papieren, gedrukt, woordenboek. Aan de ene kant is er geen belemmering voor een zeer grote omvang en is de alfabetische wurggreep verdwenen. Maar daar staat tegenover dat de maker niet kan rekenen op het verstand achter het oog dat zijn tekst leest - dat oog is nu slechts een computer-oog. Aan de consistentie van termen en aan de indeling van het woordenboekartikel worden veel hogere eisen gesteld. Zo zal het op CD verkrijgbaar stellen van het *WNT* de roep om een nieuwe Tweede Druk van dat eerbiedwaardig woordenboek alleen maar versnellen en luider maken.

(Dit artikel is eerder verschenen in *Trefwoord*, 8, 1994, pp. 7-9. De paginanummering van het origineel staat tussen [] vermeld.)