



Royal Netherlands Academy of Arts and Sciences (KNAW)

KONINKLIJKE NEDERLANDSE
AKADEMIE VAN WETENSCHAPPEN

Aardbei = rood, lekker, zomer

van der Sijs, N.

published in

NRC Handelsblad

2012

document version

Publisher's PDF, also known as Version of record

[Link to publication in KNAW Research Portal](#)

citation for published version (APA)

van der Sijs, N. (2012). Aardbei = rood, lekker, zomer. *NRC Handelsblad, Wetenschapsbijlage*, 2-2.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the KNAW public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the KNAW public portal.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

pure@knaw.nl

Aardbei = rood, lekker, zomer

Een taal is een regelmatige verzameling van woorden', schreef de taalkundige Adriaan Kluit omstreeks 1759. Met de vraag volgens welke regel die woorden het best ingedeeld kunnen worden, wordt al eeuwenlang geworsteld. In de 16de eeuw, toen de eerste Nederlandse woordenboeken verschenen, koos men zowel voor een alfabetische indeling als voor een thematische. Alle namen voor ziektes, vogels of bomen werden in betekenisvelden bij elkaar gezet.

De alfabetische ordening is objectief en praktisch, maar betekenisloos: ze laat niet zien hoe woorden via betekenissen aan elkaar zijn verbonden. In 1898 keerde de Nederlandse taalkundige J.W. Muller zich dan ook tegen 'de tyrannie der volslagen onwetenschappelijke [...] alphabetische volgorde'. Woorden hebben immers allerlei betekenisrelaties. Zo vallen *herfst*, *lente*, *voorjaar* onder de overkoepelende benaming *jaargetijd*. *Lente* en *voorjaar* zijn synoniemen, *lente* en *herfst* zijn elkaars tegengestelden. *Vader* en *moeder* zijn via *kind* in betekenis aan elkaar gerelateerd.

In navolging van filosofische en biologische classificaties probeerde men in de 19de eeuw de hele woordenschat wetenschappelijk in te delen. Het eerste woordenboek dat geheel naar begrippen was gerang-

schikt, was de *Thesaurus of English words and phrases* van P.M. Roget uit 1852. In 1876 ontwierp M. Dewey voor bibliotheken een vergelijkbaar indelingssysteem, de Dewey Decimale Classificatie.

De omvangrijkste thesaurus van het Nederlands is *Het juiste woord* van L. Brouwers, waarvan in 1931 de eerste editie verscheen. In dit systematisch ingedeelde woordenboek is bijvoorbeeld *lente* opgenomen onder de hoofdrubriek 'De stoffelijke wereld', subrubriek 'dampkring'. Via een alfabetisch register kan men *lente* vinden en alle woorden die hieraan in betekenis zijn gerelateerd, waaronder ook verouderde. Definities van de woorden ontbreken echter, zodat de betekenisrelatie aan de gebruiker wordt overgelaten. Dat gaat niet altijd goed. Zo zei een Russische neerlandicus ooit van iemand: "Hij slacht zijn vader." Hij had uit Brouwers' woordenboek de totaal verouderde betekenis geleerd: 'Hij lijkt op zijn vader.'

Wetenschappelijke indelingen zijn ongeschikt voor dialectwoordenschat, die ontstaan zijn in een agrarische tijd. Zo vallen granen volgens de botanische indeling onder de grassenfamilie. Dat is echter niet de indeling die boeren hanteren: zij maken een scherp onderscheid tussen waardeloze grassen en nuttige graansoorten. De bekende Nijmeegse dialectoloog A. Weijnen heeft dan

ook een speciale thematische indeling van de dialectwoordenschat gemaakt, die is uitgewerkt in regionale woordenboeken van het Brabants, Limburgs en Vlaams.

De menselijke geest legt weer heel andere, eigen verbanden tussen woorden. Dat onze begripsassociaties niet samenvallen met wetenschappelijke classificaties, is geruime tijd bekend. Maar tot voor kort hadden we daarover nauwelijks harde gegevens. Sinds 2003 loopt aan de universiteit Leuven een grootschalig onderzoeksproject waarbij aan duizenden deelnemers is gevraagd welke drie woordassociaties ze bij een bepaald begrip hebben (<http://nrch.nl/kxu>). 'Aardbei' blijken de meeste mensen te associëren met 'rood', 'lekker' en 'zomer'. Deze woordassociaties geven inzichten in ons mentale lexicon: de manier waarop woorden in de hersenen zijn opgeslagen.

Nog weer andere indelingen komen naar voren als we bekijken hoe woorden in grote tekstbestanden in combinatie optreden. Zoals de Engelse taalkundige J.R. Firth al in 1957 zei: 'You shall know a word by the company it keeps'. De computer kan berekenen welke woorden vaak bij elkaar klitten. Op die manier vind je meer of minder vaste verbindingen van woorden, die informatie over de betekenis prijsgeven. Zo spreken we wél van *een school haringen* of *kwallen*,

Granen vallen onder grassen, maar boeren onderscheiden graan van nutteloos gras

maar niet van *een school herten*.

Een superthesaurus waaraan sinds 1985 aan de universiteit van Princeton wordt gewerkt is Wordnet. Begrippen zijn in domeinen verdeeld (volgens de Dewey-classificatie), en daarbinnen zijn woorden en betekenissen, betekenisrelaties, frequentiegegevens en voorbeeldzinnen gerangschikt. Inmiddels bestaan er Wordnetten van honderd talen. Als die in de toekomst onderling worden gelinkt, zal dit veel nieuwe praktische, psychologische en taalkundige inzichten opleveren. Zo blijkt dan dat het Engelse *vegetables*, anders dan zijn Nederlandse tegenhanger *groenten*, ook aardappelen insluit.

De Nederlandse Wordnet is door Piek Vossen van de VU ontwikkeld onder de naam Cornetto. Binnen het NWO-project DutchSemCor, dat eind september is voltooid (nrch.nl/kxv), zijn de betekenissen die in Cornetto worden onderscheiden, gekoppeld



aan woorden in miljoenen tekstbestanden. Hierdoor is het mogelijk niet alleen woorden maar ook de betekenissen in context te onderzoeken.

De genoemde projecten en indelingen staan tot nu toe los van elkaar. Het zou een enorme meerwaarde opleveren als de gegevens aan elkaar worden gekoppeld en voor iedereen zonder technische kennis eenvoudig doorzoekbaar gemaakt. Dan ontstaat een enorm begrippennetwerk waarin men bijvoorbeeld de relatie tussen woordassociaties en frequenties kan onderzoeken. Komen *aardbei* en *fruit* in een tekstbestand veel samen voor? Dit zou meteen ook kunnen resulteren in een digitale, gemoderniseerde editie van Brouwers' *Het juiste woord*, waar veel vraag naar is.

Dringend noodzakelijk is dat het begrippennetwerk ook een historische dimensie krijgt. Dan kunnen we opzoeken hoe vroeger een 'hemellichaam' heette (*ster*, *planeet*, *dwaalder*, *wachter*, *vuurbol*, *satelliet*). Kennis van oudere benamingen is onmisbaar wanneer men gegevens zoekt in historische teksten. Momenteel scharrelt iedere onderzoeker zijn eigen materiaal bij elkaar. Dat kost onnodig tijd, en het is te hopen dat er eindelijk eens wordt begonnen aan de samenstelling van een historische thesaurus van het Nederlands.