

Mededelingen van de
Nijmeegse Centrale
voor Dialect- en
Naamkunde

Themanummer

De Automatisering van
Dialectwoordenboeken

onder redactie van Dr. G. Janssens

XXI / 1987-1988

Inhoudsopgave

- Inleiding: De automatisering van dialectwoordenboeken (G. Janssens) 1
- Automatisering van WBD en WLD (P. Goossens) 7
- Automatische Datenverarbeitung am Westfälischen Wörterbuch (H. Taubken) 27
- Het WVD-programma voor de automatische verwerking van dialectvragenlijsten (J. Van Keymeulen, P. Valck) 33
- De automatisering van een plaatselijk dialectwoordenboek: concrete ervaringen en een algemene beschouwing (S. Kroon, H. Paijmans) 47
- **Automatisering binnen het project "De Brabantse Spreekwoorden" van de Hein Mandosstichting te Waalre (W. Balemans, C. van de Ven) 69**
- Het WDD en de computer (G.H. Kocks) 81
- Automatisering op het P.J. Meertens-Instituut; in het bijzonder de automatische kartografie (H. Hogerheijde) 85
- De automatisering van het Vroegmiddelnederlands Woordenboek (W.J.J. Pijnenburg, T.H. Schoonheim) 99
- Kroniek 109

Automatisering binnen het project "De Brabantse Spreekwoorden" van de Hein Mandosstichting te Waalre

Will Balemans
Caspar van de Ven

0. Inleiding

Op 28 augustus 1985 werd de, in Waalre gevestigde, Hein Mandosstichting opgericht met als doel het archief dat drs. Hein Mandos naliet in stand te houden, te completeren, te bewerken en te doen publiceren, in het bijzonder de door hem verzamelde spreekwoorden en zegswijzen die door Brabanders gebruikt zijn of worden. Kortom, de Hein Mandosstichting stelde zich als eerste doel een spreekwoordenboek samen te stellen op basis van de door Hein Mandos tussen 1926 en 1978 aangelegde collectie Brabantse Spreekwoorden. (De term "spreekwoord" wordt hier als overkoepelende term gebruikt voor: uitdrukking, gezegde, spreekwoordelijke vergelijking enz. De discussie omtrent het formele onderscheid valt buiten het kader van dit artikel). Na zijn dood werd het archief Mandos aangevuld en bewerkt door zijn vrouw Miep Mandos- van de Pol. Zij nam de samenstelling van "De Brabantse Spreekwoorden" (DBS) ter hand op basis van het principe van alfabetisering van het eerste kernwoord. Het corpus staat geheel op fiches. Ten behoeve van de verwerking met de computer werd op basis van voorstellen van Miep Mandos door de technisch adviseur van de Hein Mandosstichting een computerbestand opgezet. Om de toepasbaarheid zo algemeen mogelijk te laten zijn, gaan de auteurs ervan uit dat men met de opzet van een spreekwoordenboek vanuit "fase 0" vertrekt.

Nadat men het principe (alfabetisch en/of thematisch enz.) en de manier van archiveren heeft gekozen zal het spreekwoordenmateriaal moeten worden vastgelegd in een databestand van een computer: een eigentijdse variant op het aloude en beproefde kaartsysteem. Het zal duidelijk zijn

dat terugzoeken en bewerken van spreekwoordenmateriaal met behulp van een computer vele malen sneller gaat dan het handmatig verwerken van allerlei fiches in kaartenbakken. Hoe kan men met een computer een systeem opzetten dat het mogelijk maakt een spreekwoordenboek te componeren? In de volgende paragraaf beschrijven we hoe men een database op kan zetten ten behoeve van dat doel en hoe men kan komen tot een spreekwoordenboek.

1.0. Het archiveren van spreekwoorden met behulp van een computer

1.1. De benodigde hard- en software

Voor het opzetten van een "kaartsysteem" met behulp van een computer is het wenselijk dat men de beschikking heeft over een personal computer met een databaseprogramma. Een computer met een harde schijf is te verkiezen boven het werken met één of meerdere floppy's. Een adequaat programma - dat overal te koop is - is dBase III. Wanneer we het computergeheugen met dit programma geladen hebben, biedt dat ons twee mogelijkheden n.l. de mogelijkheid een aantal velden te definiëren (a.h.w. de kaartenbak) en de mogelijkheid om binnen het programma de velden te manipuleren d.w.z. dat we de in het bestand ingevoerde gegevens b.v. alfabetisch kunnen sorteren, of dat we - wanneer we de voorwaarde omschreven hebben - kunnen selecteren.

1.2. Het definiëren van de velden

Het definiëren van de velden is populair gezegd te vergelijken met het bepalen wat er op het oude, vertrouwde fiche komt te staan. Met andere woorden: je gaat nu al bepalen hoe je lemma eruit komt te zien. Belangrijk is dat je de "computerfiche" straks weer terug kunt vinden. Je kunt b.v. het eerste kernwoord (=trefwoord) gebruiken om de gegevens straks weer terug te zoeken.

Een mogelijke lemma-opbouw wordt dan:

1. tekst in dialect
2. tekst in A.B.
3. toelichting
4. eerste kernwoord
5. overige kernwoorden
6. plaats en/of tijd.....
(bron) (laatst opgetekend gebruik)*
7. spreekwoordelijke vergelijking
8. zeispreuk

* Het laatst opgetekend gebruik geeft aan of het spreekwoord of gezegde nog in de samenleving functioneert.

N.B. De overige kernwoorden worden bij DBS in het databestand opgenomen om verwijzingen in de bewerking van de gegevens mogelijk te maken: ze worden straks niet in het lemma van het woordenboek opgenomen, omdat elke verwijzing een zelfstandig lemma in het woordenboek wordt.

Bijv. bakker (zie brouwer)

Wor den braauwer is, kan den bekker niej zèn.

1.3. Het inbrengen van alle gegevens

Het intoetsen van de gegevens is een energie- en tijdrovend werk: het is het "monnikenwerk" van de twintigste eeuw. In de praktijk van DBS komt het erop neer dat er zo'n 2.000.000 karakters ingebracht moesten worden.

Bij het inbrengen van de gegevens kunnen er allerlei problemen ontstaan. Zo hebben de velden een vaste lengte van maximaal 254 karakters. Die lengte van een veld moet gedefinieerd worden om het langste gegeven erin op te kunnen slaan. Elk lemma heeft in principe die volle ruimte nodig. Dus een gezegde als: "Zit er 'n vliegske in den kraaiap-

pel, dan hebben we een korte winter, want dan vliegen we erdoor; zit er 'n wurmke in, dan hebben we een lange winter, want dan kruipen we erdoor" bepaalt de te reserveren ruimte voor een gezegde als "Dirk slaan". In het laatste geval houden we dus een aantal oningevulde karakters over. Omdat je dat grote aantal karakters moet reserveren kun je in de problemen komen met je beschikbare opslagcapaciteit.

Om dit probleem op te lossen kan men twee bestanden naast elkaar maken. Een bestand waarvan de velden groot genoeg zijn om het langste spreekwoord in op te slaan en een bestand waarvan de velden aanmerkelijk (bijv. factor 3) kleiner zijn. Het blijkt dat ongeveer 65% van de spreekwoorden in het kleinere bestand "past". Op deze manier wordt een aanzienlijke besparing van de opslagcapaciteit verkregen. Overtreft een gegeven de maximale veldlengte van 254 karakters dan bestaat de mogelijkheid om het gegeven over meerdere velden te verdelen.

Voor het veld "Toelichting" is een lengte van 240 karakters gedefinieerd. Het programma zorgt ervoor dat een toelichting die korter is dan 120 karakters weggeschreven wordt naar het kleine bestand (vgl. kaartenbak). Is de toelichting meer dan 240 karakters, dan wordt deze weggeschreven naar een ander bestand. Dit gebeurt automatisch; de gebruiker merkt er niets van.

Voor het veld "Overige kernwoorden" zijn 59 karakters gereserveerd. Indien het totaal aantal karakters van het aantal overige kernwoorden groter is dan 59, dan wordt een kernwoord minder opgenomen. Uitbreiding van het aantal kernwoorden blijft in principe mogelijk door het programma aan te passen. De veldlengte kan in een later stadium naar believen vergroot of verkleind worden. Ook kunnen nog velden worden toegevoegd of worden weggelaten.

Voor de velden "Spreekw. verg." en "Zeispreuk" is één karakter gereserveerd: een J of een N van J(a) of N(ee).

Bij het programma waarmee de Hein Mandosstichting werkt, haalt de computer een gedeelte van het gegevensbestand van een harde schijf (het achtergrondgeheugen) en schrijft dit naar het werkgeheugen. We voeren in dat werkgeheugen bewerkingen uit en het resultaat wordt later weer teruggeschreven naar de harde schijf. Een harde schijf met een opslagcapaciteit van 20 Mb heeft voldoende ruimte om naast de benodigde programmatuur te manipuleren met een gegevensbestand dat 6000 of meer lemma's bevat. Om te voorkomen dat, door technische problemen met de harde schijf, de ingebrachte gegevens verloren zouden gaan moeten de gegevens van de harde schijf regelmatig worden gekopieerd naar floppy disks (de zg. "backups"). Aangezien de informatie op floppy disks niet langer dan ongeveer 7 jaar betrouwbaar bewaard blijft (omdat de floppy zijn magnetische eigenschappen verliest), zal men genoodzaakt zijn om voor het verstrijken van die termijn (bijv. om de 5 jaar) nieuwe kopieën van de floppy disks te maken.

1.4. Voorbeeld van een menugestuurd computerprogramma

Voor het intypen van alle gegevens is binnen het dBase III pakket een speciaal programma geschreven. Een van de randvoorwaarden was, dat iemand met een minimale kennis van computers dit programma moest kunnen gebruiken. Daarom is voor een zg. "menugestuurd" programma gekozen. Bij het inschakelen van de computer verschijnt het hoofdmenu op het beeldscherm (zie figuur 1). De gebruiker kan nu uit een aantal mogelijkheden kiezen. Wordt gekozen voor "1..... Nieuwe gegevens toevoegen", dan verschijnt figuur 2, echter met lege velden. Is zoals in dit concrete voorbeeld het veld "Toelichting" te klein, dan verschijnt na positieve beantwoording van de vraag "Wilt u meer toelichting geven" figuur 3 op het scherm. Indien we bij het afdrukken van de gegevens op papier een bepaald deel van de tekst willen cursiveren, dan wordt dit deel tussen " " -tekens geplaatst. Het programma zal dan zorgen voor een cursieve tekst.

Wordt bij het hoofdmenu gekozen voor "2..... Gegevens wijzigen", dan verschijnt figuur 4 en wordt gevraagd naar het eerste kernwoord

van het spreekwoord dat gewijzigd moet worden. Hierna verschijnen de gegevens van het gevraagde kernwoord (figuur 5) waarbij het mogelijk is de gewenste wijzigingen aan te brengen. Indien meerdere spreekwoorden opgenomen zijn met hetzelfde eerste kernwoord, dan zullen deze één voor één op het scherm verschijnen. Op analoge wijze kunnen vanuit het hoofdmenu gegevens verwijderd worden.

Voor het maken van overzichten zijn drie mogelijkheden aanwezig nl. op het beeldscherm, in een bestand op de schijf en op de printer. Indien gekozen wordt voor "4..... Overzichten op beeldscherm", verschijnt een nieuw menu (zie figuur 6). Vanuit dit menu zijn een aantal selecties mogelijk. Bij de keuze "1..... Selecteren op kernwoord" is een verdere specificatie mogelijk (zie figuur 7). Na beantwoording van enkele vragen (zie figuur 8) wordt het gewenste resultaat op het beeldscherm verkregen. Door de keuze "2..... Selecteren op een gedeelte van de tekst" verschijnt een nieuw menu (zie figuur 9). Dit menu opent de mogelijkheid om te selecteren op een geheel woord of op een woorddeel. Dit geeft bijv. de mogelijkheid om alle spreekwoorden te selecteren waarin het woord "bakker" voorkomt. Op analoge wijze kan een selectie afgedrukt worden op de printer, of worden weggeschreven in een bestand op schijf ten behoeve van de drukker.

De in het hoofdmenu vermelde mogelijkheid van het maken van "backups" wordt gebruikt om de kans dat reeds ingevoerde gegevens verloren gaan tot een minimum te beperken. Herindexeren kan nodig zijn indien bijv. door het foutief afbreken van het programma, een interne fout ontstaat. Het hoofdmenu geeft de gebruiker de mogelijkheid om het programma op legale wijze te beëindigen.

In de twee jaar dat dit programma nu in gebruik is geweest, zijn geen noemenswaardige problemen opgetreden. Wel zijn enkele verbeteringen aangebracht om de responstijden te verbeteren. Dit programma blijkt een goed hulpmiddel te zijn om op efficiënte manier een grote hoeveelheid gegevens in de computer in te voeren en om deze gegevens te manipuleren. Het is zo flexibel opgebouwd, dat nieuwe wensen gemakkelijk in het programma kunnen worden opgenomen.

HEIN HANDOSSTICHTING Waalre	
GEGEVENSBESTAND VOOR BRABANTSE SPREEKWOORDEN EN GEZEGDEN HOOFDMENU TOTAAL 5824	
[1]...	Nieuwe gegevens toevoegen
[2]...	Gegevens wijzigen
[3]...	Gegevens verwijderen
[4]...	Overzichten op beeldscherm
[5]...	Overzichten dumpen in file
[6]...	Overzichten op printer
[7]...	Backups maken
[8]...	Herindexeren
[0]...	Einde programma

Figuur 1. Hoofdmenu

Gegevens toevoegen		Totaal 5825 records
Tekst in dialect :		
De slak hei net zo gauw Neijaor as den haos.		
Tekst in AB :		
De slak heeft even vlug Nieuwjaar als de haas.		
Toelichting :		
Haasten heeft vaak geen zin. Varianten: @'n Slak en 'nen haos zèn even gauw on de meij@ (Oost-Brab. '84); @'n Slak komt 'r net zo goed als 'n haos@ (Breda 1892); ... @mar nie zeu vlug@ (Kempenland '58); @'n Slek en 'n haos hebben even tijd		
Eerste kernwoord :	slak	
Ov. kernwoorden :	afstand, voet, tijd, haas, mel, Nieuwjaar	
Plaats en/of tijd:	Dinther (Bi) '40	Spreekw. verg. : <input type="checkbox"/> N Zeispreuk : <input type="checkbox"/> N
Zijn bovenstaande gegevens correct (J/N) (Q = stoppen) :	<input type="checkbox"/> J	
Wilt U meer toelichting geven :	<input type="checkbox"/> J	

Figuur 2. Gegevens toevoegen

EXTRA TOELICHTING
<p>Haasten heeft vaak geen zin. Varianten: @'n Slak en 'nen haos zèn even gaauw on de mei]@ (Oost-Brab. '84); @'n Slak komt 'r net zo goed als 'n haos@ (Breda 1892); ... @mar nie zeu vlug@ (Kempenland '58); @'n Slek en 'n haos hebben even tijd</p>
<p>Nieuwjaar@ (Waalre '74); @Een slek kumt er ook@ (Oost-Brab. '84); @De afstand i s even ver te voet als met de tram@ (Eindhoven '84). (Zie ook: @rap, gauw@.)</p>
<p>Is deze extra toelichting correct (J/N) : <input type="checkbox"/> J</p>

Figuur 3. Extra toelichting

Gegevens wijzigen	Totaal 5825 records
<p>Voer het verlangde kernwoord in : <input type="text" value="slak"/></p>	

Figuur 4. Kernwoord t.b.v. wijzigen gegevens

Gegevens wijzigen		Totaal 5825 records
Tekst in dialect :		
De slak hei net zo gauw Meijaor as den haos.		
Tekst in AB :		
De slak heeft even vlug Nieuwjaar als de haas.		
Toelichting :		
Haasten heeft vaak geen zin. Varianten: @'n Slak en 'nen haos zen even gaauw on de meij@ (Oost-Brab. '84); @'n Slak koat 'r net zo goed als 'n haos@ (Breda 1892); ... @mar nie zeu vlug@ (Kempenland '58); @'n Slek en 'n haos hebben even tijd		
Eerste kernwoord :	slak	
Ov. kernwoorden :	afstand, voet, tijd, haas, mel, Nieuwjaar	
Plaats en/of tijd:	Dinther (B1) '40	Spreekw. verg. : <input type="checkbox"/> N <input type="checkbox"/> Zelspreuk : <input type="checkbox"/> N <input type="checkbox"/>
Zijn bovenstaande gegevens correct (J/N) (Q = stoppen) :	<input type="checkbox"/> J <input type="checkbox"/> N	
Wilt U meer toelichting geven :	<input type="checkbox"/> J <input type="checkbox"/> N	

Figuur 5. Gegevens wijzigen

Overzicht op beeldscherm
[1]... Selecteren op kernwoord
[2]... Selecteren op een gedeelte van de tekst
[3]... Selecteren op plaats en/of tijd
[4]... Selecteren op spreekwoordelijke vergelijkingen
[5]... Selecteren op zelspreuken
[0]... Einde overzicht op beeldscherm

Figuur 6. Overzicht op beeldscherm

Overzicht op beeldscherm geselecteerd op kernwoord

{1}...Selecteren op eerste kernwoord

{2}...Selecteren op overig kernwoord

{3}...Selecteren op alle kernwoorden

{0}...Einde selecteren op kernwoord

Figuur 7. Overzicht geselecteerd op kernwoord

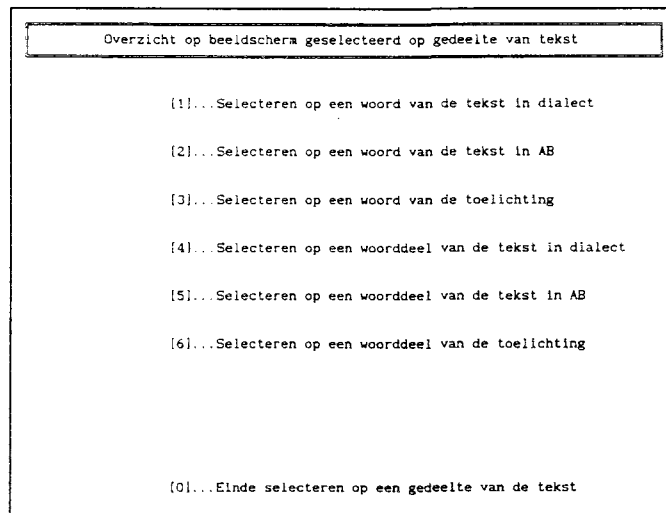
Overzicht op beeldscherm geselecteerd op eerste kernwoord

Wilt U een pauze na ieder kernwoord (J/N) (Q = stoppen) :

Overzicht start met kernwoord :

Overzicht eindigt met kernwoord :

Figuur 8. Kernwoorden t.b.v. overzicht



Figuur 9. Overzicht geselecteerd op tekst