

# ***Semantic maps voor vogelnamen in Brabantse dialecten***<sup>1</sup>

Joost Zwarts  
Universiteit Utrecht

## **Inleiding**

Eén van de meest in het oog springende aspecten van het onderzoek naar dialecten is de dialectgeografie, die de lexicale, fonologische of syntactische variatie tussen dialecten in kaart brengt. Zo kunnen we bijvoorbeeld in één oogopslag zien hoe de ui in verschillende plaatsen in het Nederlandse taalgebied aangeduid wordt en wat de regionale verspreiding is van het woord *ajuin*. Dialectkaarten en -atlassen zijn daarmee een belangrijk visueel hulpmiddel bij het bestuderen van ruimtelijke patronen en factoren in taalvariatie.

In dit artikel wil ik een ander type visualisering van lexicale dialectvariatie verkennen, de *semantic map*, een uit de taaltypologie afkomstige methode (Haspelmath 2003). Zoals een traditionele woordkaart de spreiding van woorden over de geografische ruimte afbeeldt, zo brengt een semantic map in beeld hoe woorden over een 'conceptuele ruimte' van betekenissen zijn verdeeld. Dat helpt ons om een beter inzicht krijgen in de semantische structuur van variatie. Deze semantische 'kartografie' is al met vrucht toegepast in de bestudering van het gebruik van bijvoorbeeld naamvallen, voorzetsels en voornaamwoorden. In dit artikel zal ik de methode toepassen op de benoeming van vogels in Brabantse dialecten, zoals beschreven in Swanenberg (2000), die zich daarbij baseerde op materiaal uit het *Woordenboek van de Brabantse Dialecten* (WBD).

Na een korte bespreking van de verschillende dimensies van lexicale dialectvariatie in paragraaf 1, introduceer ik de methode van semantic maps in paragraaf 2 en pas die toe op vijf groepen vogels in paragraaf 3. Hoe we de maps kunnen gebruiken om semantische eigenschappen van variatie te visualiseren en exploreren is het onderwerp van paragraaf 4. In de afsluitende paragraaf 5 bespreek ik de uitdagingen waarvoor deze methode zich gesteld ziet.

## **1 Lexicale variatie**

De belangrijkste bron van lexicale variatie in dialecten is *heteronymie*: een bepaalde zaak wordt op verschillende plaatsen op een andere manier benoemd. Er zijn bijvoorbeeld 30

---

<sup>1</sup> Dit artikel is de uitwerking van een voordracht op het symposium 'De systematische woordenboeken van de Zuid-Nederlandse dialecten: lexicale variatie als taaltheoretisch thema' georganiseerd door *Taal en Tongval* op 17 november 2006 in Gent. Ik dank het publiek daar voor hun nuttige vragen en opmerkingen en ook in het bijzonder Jos Swanenberg, Roeland van Hout en Folkert de Vriend voor commentaar naar aanleiding van een eerste versie. Het onderzoek voor dit artikel is mogelijk gemaakt door een subsidie van de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) (subsidie 220-70-003 voor het PIONIER-project 'Case Cross-Linguistically').

verschillende woorden opgetekend voor de veldleeuwerik, zoals *klimleeuwerik* (Turnhout), *tjur* (Hoogerheide), *graspieper* (Gilze-Rijen). Het omgekeerde verschijnsel bestaat ook: met een bepaald woord worden op verschillende plaatsen andere zaken benoemd. Zo wordt met *franse lijster* in Oudenbosch een zanglijster bedoeld, in Wechelderzande een grote lijster en in Aarschot een koperwiek. Ik zal dit verschijnsel hier *heterosemie* noemen. Hoewel niet zo opvallend als heteronymie, speelt dit verschijnsel wel degelijk een rol in de data. Van de 190 woordtypen die Swanenberg voor de vijf mezen opsomt zijn er 48, dus ongeveer een kwart, heteroseem. Dat percentage ligt nog hoger als we in aanmerking nemen dat veel verschillende woordtypen eigenlijk varianten van hetzelfde woord zijn, zoals *blauwkoolmus* en *blauwe koolmus*. Het is dit type van semantische 'spreiding' van woorden dat een belangrijke rol zal spelen bij het maken van semantic maps voor lexicale variatie.

Bij het onderzoek van lexicale variatie spelen in elk geval drie dimensies een rol:

*Lexicale* dimensie: verschillende benoemingen

*Semantische* dimensie: verschillende betekenissen

*Geografische* dimensie: verschillende plaatsen

In het algemeen zijn we erin geïnteresseerd hoe de ene dimensie met de andere samenhangt. Traditioneel gaat het daarbij om de verdeling van woorden over de geografische ruimte; voor dit artikel is belangrijk dat woorden ook de *betekenisruimte* op een bepaalde manier verdelen.

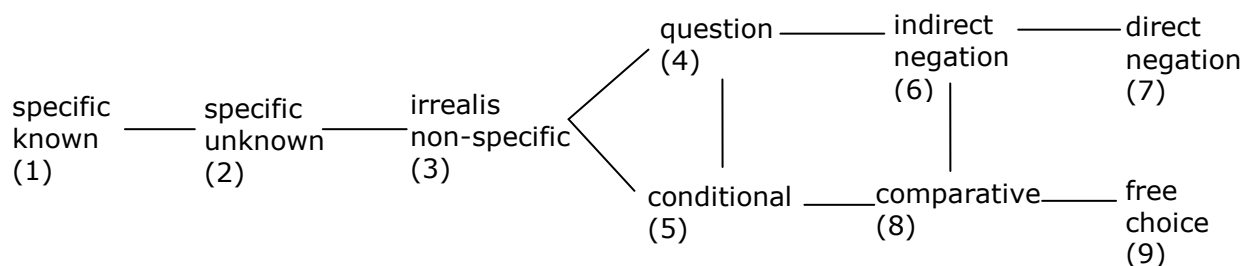
De geografische dimensie heeft een duidelijke structuur, bepaald door de fysieke afstanden tussen de plaatsen. Het is echter ook mogelijk om de afstanden tussen die plaatsen taalkundig te definiëren, op basis van de overeenstemming in woordenschat of andere talige kenmerken (Heeringa en Nerbonne 2006, De Vriend, Van Hout en Boves 2007). We krijgen dan een kaart waarin plaatsen of dialecten dichter bij elkaar staan naarmate ze meer op elkaar lijken. In het eerste geval is de geografische ruimte *a priori* gegeven, in het tweede geval *a posteriori* uit de taaldata afgeleid.

Ook wat betreft de semantische dimensie kunnen we in theorie dezelfde twee wegen bewandelen. Bij de *a priori* benadering proberen we de structuur van een bepaald domein, bijvoorbeeld de lijsters, onafhankelijk van hun benoeming te bepalen. De relatieve afstanden tussen de grote lijster, de merel en de koperwiek worden dan bepaald door de mate waarin ze objectieve of subjectieve kenmerken delen (zoals kleur, zang en dergelijke). Zo'n benadering is voor lijsters en andere natuurlijke soorten niet principieel onmogelijk, maar op dit moment is er nog geen goed gemotiveerde, kant-en-klare conceptuele ruimte voor bepaalde vogelgroepen (zoals die er wel is voor kleuren,

bijvoorbeeld, zie Gärdenfors 2000).<sup>2</sup> Het is daarom op dit moment voor dit domein realistischer om te proberen de structuur van een betekenisruimte af te leiden uit de manier waarop de betekenissen benoemd worden. Dat brengt mij bij de typologische methode van *semantic maps*.

## 2 Semantic maps

Semantic maps zijn een taaltypologische methode om te visualiseren hoe een verzameling woordbetekenissen of grammaticale functies samenhangt (zie Haspelmath 2003 voor een algemeen overzicht en Zwarts 2006 voor een toepassing op de lexicale semantiek van *om* en *rond*). Het bekendste en best uitgewerkte voorbeeld van een semantic map is te vinden in Haspelmath (1997). In zijn studie onderscheidde hij negen verschillende betekenissen of functies van onbepaalde voornaamwoorden. Vervolgens stelde hij van een groot aantal talen in de wereld vast hoe ze die negen betekenis benoemden en op basis van die benoemingen definieerde hij het volgende netwerk (de *conceptuele ruimte* of *conceptuele structuur* van de onbepaalde voornaamwoorden):<sup>3</sup>

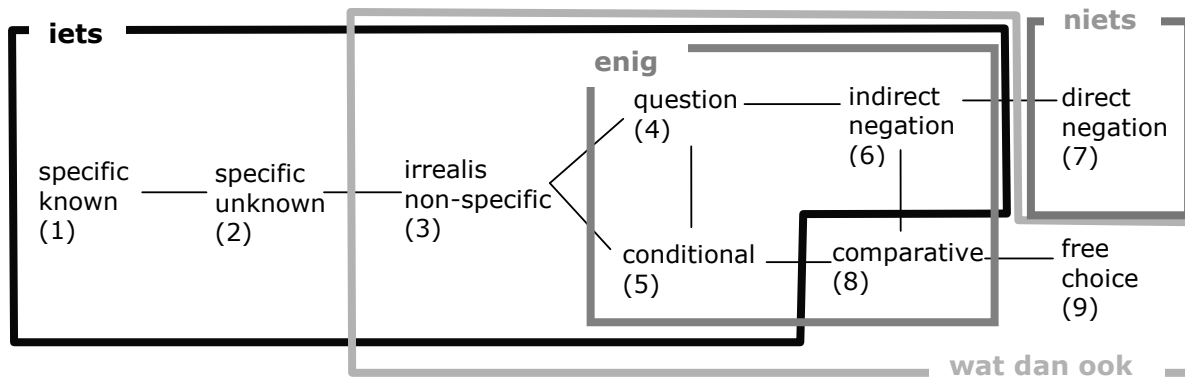


Figuur 1: Semantic map van onbepaalde voornaamwoorden

Dit netwerk is zo opgesteld dat elk voornaamwoord in elke taal uit Haspelmaths database met een samenhangend gebied correspondeert. De situatie in het Nederlands ziet er als volgt uit:

<sup>2</sup> Er is natuurlijk wel een hiërarchische taxonomie van vogelsoorten, maar die biedt vooral inzicht in de opbouw van de vogelklasse als geheel, maar vertelt ons weinig over de structuur van vogelgroepen die zich laag in de taxonomie bevinden, zoals het geslacht van de echte lijsters (waartoe merel, zanglijster, grote lijster, koperwiek en kramsvogel behoren).

<sup>3</sup> Sommige auteurs onderscheiden de *conceptual structure* (de betekenissen met hun netwerk van relaties) en de *semantic map* (die laat zien hoe woorden daarover verspreid zijn). Figuur 1 geeft dan dus een conceptual structure en Figuur 2 een semantic map. Ik gebruik de term semantic map voor beide.



Figuur 2: Onbepaalde voornaamwoorden in het Nederlands

De eigenschap van *samenhang* is cruciaal. Als een woord *w* twee betekenissen *a* en *b* heeft, dan moet er een pad van betekenissen zijn tussen *a* en *b* die ook in het gebied van *w* vallen. Er kan dus geen woord bestaan dat betekenissen (1) en (3) dekt, zonder dat ook (2) erbij zit en er is geen woord met (5) en (6) zonder dat het ook (4) of (8) omvat.

De veronderstelling is nu dat de semantic map in Figuur 1 laat zien welke betekenissen dichter bij elkaar liggen dan andere betekenissen. Twee betekenissen die direct verbonden zijn met een lijn hebben een nauwe verwantschap, terwijl de semantische afstand tussen (1) aan de ene kant en (7) en (9) aan de andere kant juist heel groot is. Tegelijk is elke semantic map daarmee de belichaming van een simpele maar belangrijke hypothese, namelijk dat de verschillende betekenissen van een woord altijd nauw verwant met elkaar zijn.

Hoe maken we een semantic map? Er is geen algoritme bekend uit de literatuur dat ons vanuit een verzameling woorden naar een netwerk brengt, maar het algemene idee is duidelijk. Stel dat we een verzameling *W* van woorden hebben die elk een of meer betekenissen uitdrukken die komen uit een verzameling *B*. Zet de betekenissen van *B* als punten op een vel papier en trek het kleinste aantal lijnen tussen die punten dat nodig is om elk woord uit *W* te laten corresponderen met een samenhangend netwerk.<sup>4</sup> Laten we nu bekijken hoe we dat idee kunnen toepassen op de vogelbenamingen uit Swanenberg (2000).

### 3 Semantic maps van vijf vogelsoorten

In zijn proefschrift bespreekt Swanenberg de benoeming van in totaal 22 vogelsoorten, in vijf groepen:

*Zwaluwen*: boerenzwaluw, huiszwaluw, oeverzwaluw; gierzwaluw, nachtzwaluw

<sup>4</sup> Cysouw (2007) laat zien dat dit uitgangspunt niet altijd tot een eenduidig resultaat leidt, zoals we hieronder ook al bij een aantal vogelkaarten zullen zien.

*Mezen*: koolmees, pimpelmees, kuifmees, zwarte mees, staartmees  
*Lijsters*: zanglijster, grote lijster, koperwiek, kramsvogel, merel  
*Leeuweriken*: veldleeuwerik, kuifleeuwerik, boomleeuwerik  
*Mussen*: huismus, ringmus; grasmus, heggenmus

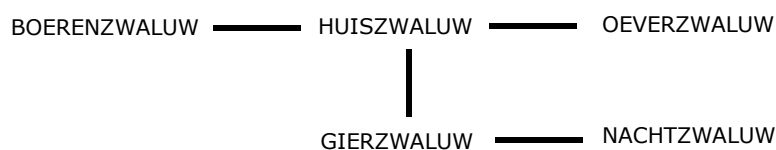
Van elk van deze vijf groepen heb ik een map gemaakt op basis van de data in het proefschrift. Die zal ik nu een voor een bespreken.

### 3.1 De zwaluwen-map

Het domein van de zwaluwen, zoals beschreven in Swanenberg (2000), bestaat uit 5 soorten en 104 verschillende woordtypen. De soorten zijn drie echte zwaluwen (BOERENZWALUW, HUISZWALUW, OEVERZWALUW) en twee leden van andere orden die lijken op zwaluwen (GIERZWALUW, NACHTZWALUW). Veel van de 104 verschillende benamingen zijn uniek voor een bepaalde soort, maar er zijn een aantal heteroseme benamingen. Van belang zijn allereerst de zwaluwwoorden die twee soorten bij elkaar nemen:

BOERENZWALUW + HUISZWALUW: boerenzwalm, huiszwalf/m/uw, nonnetje  
 HUISZWALUW + GIERZWALUW: gier(zwalm/uw), kerkzwalling, steen/torenzwaluw  
 HUISZWALUW + OEVERZWALUW: oeverswalm/uw,  
 GIERZWALUW + NACHTZWALUW: pieper

Op basis van deze benamingen kunnen we namelijk al direct de vijf soorten als volgt verbinden:



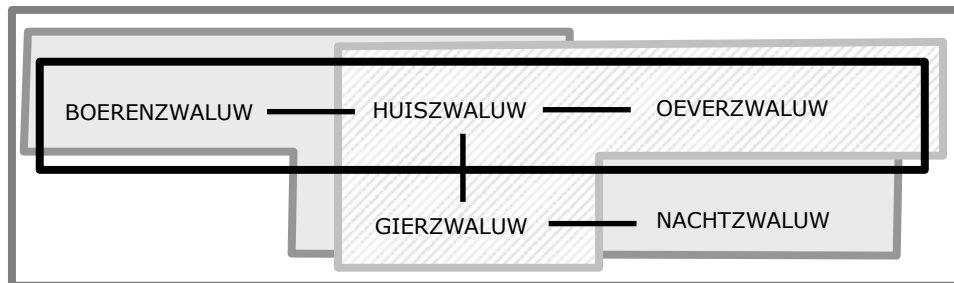
*Figuur 3: Semantic map voor zwaluwen*

De bovenste regel representeert de echte zwaluwen, de onderste regel de twee andere soorten. Als we nu de andere heteroseme benoemingen erbij nemen, dan zien we dat het bovenstaande netwerk voldoende is om elk woord een samenhangend gebied toe te kennen:

BOEREN-, HUIS- en OEVERZWALUW: zwalling  
 HUIS-, OEVER- EN GIERZWALUW: steenzwalm  
 BOEREN-, HUIS-, GIER- en NACHTZWALUW: zwalver

alle soorten: zwalm, zwaluw

Figuur 3 geeft het kleinst mogelijke netwerk weer dat voor al deze woorden samenhang garandeert. In Figuur 4 zijn de gebieden van die woorden aangegeven:



*Figuur 4: Semantic map voor zwaluwen*

### 3.2 De mezen-map

Bij de mezen vinden we ook vijf soorten: KOOLMEES, PIMPELMEES, KUIFMEES, ZWARTE MEES en STAARTMEES, met het volgende patroon van heterosemie:

KOOLMEES + PIMPELMEES: bijbijter, bijenbeestje, bijenmeesje, bijkoolder, bijkuiver, bijmeestertje, bijmewuwtje, bijmook, bijmuisje, bijmus, bijpikker, bijteutje, blauwkopje, blauwmeesje, dubbele bijdief, dubbele bijmeester, enkele bijmeester, keeskeut, keesmees, keesmusje, keespook, keespreut, kezenmeesje, kezenmietje, kezenmus, kooldertje, stijfselkopje

KOOLMEES + ZWARTE MEES: koolmeesje, koolmusje

KOOLMEES + STAARTMEES: ossenhoofdje

KOOLMEES + PIMPELMEES + ZWARTE MEES: bijdiefje, bijenbijter, bijmees, bijteut, keesmietje, keesmus, keut, kezenmees, kezenmusje, zwartkop, zwartkopmees

KOOLMEES + PIMPELMEES + STAARTMEES: meesje

PIMPELMEES + ZWARTE MEES + STAARTMEES: ossenkop

KOOLMEES + PIMPELMEES + ZWARTE MEES + STAARTMEES: bijmeesje

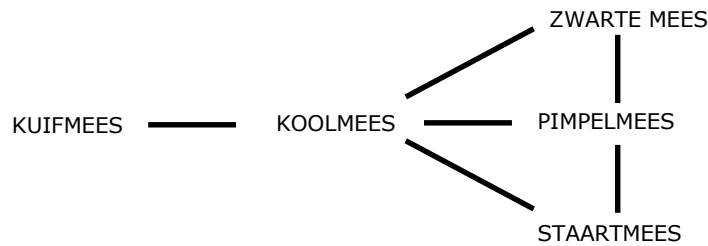
KOOLMEES + KUIFMEES + ZWARTE MEES + STAARTMEES: koolmus

KOOLMEES + PIMPELMEES + KUIFMEES + ZWARTE MEES: mees

KOOLMEES + PIMPELMEES + KUIFMEES + STAARTMEES: ossenkopje

alle mezen: bijdief

Eén van de drie structuren die bij deze data past ziet er als volgt uit:<sup>5</sup>



*Figuur 5: Semantic map voor mezen*

Er loopt alleen een lijn van de kuifmees naar de koolmees omdat elke lexicale combinatie van de kuifmees met een andere soort altijd ook de koolmees omvat, maar niet noodzakelijkerwijs een van de andere drie mezen.

### 3.3 De lijstster-map

Swanenberg geeft voor vijf lijststers (MEREL, ZANGLIJSTER, GROTE LIJSTER, KOPERWIEK, KRAMSVOGEL) in totaal 105 woordtypen, met het volgende patroon van heterosemie:

GROTE LIJSTER + KRAMSVOGEL: bal(k)lijster, dubbele lijstster, dubbellijster, kanekster, sjachel, tjakker, tjaklijster, vlierscheut

GROTE LIJSTER + KOPERWIEK: paplijster

GROTE LIJSTER + ZANGLIJSTER: klijster, schijtlijster, vosse merel

ZANGLIJSTER + KOPERWIEK: kleine lijstster, pitser

ZANGLIJSTER + KRAMSVOGEL: grijze lijstster

KOPERWIEK + KRAMSVOGEL: winterlijster

GROTE LIJSTER + ZANGLIJSTER + KOPERWIEK: franse lijstster

ZANGLIJSTER + KOPERWIEK + KRAMSVOGEL: titser

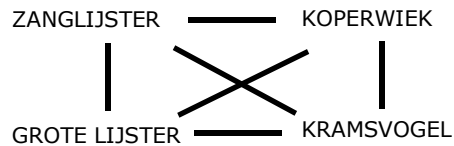
GROTE LIJSTER + ZANGLIJSTER + KOPERWIEK + KRAMSVOGEL: treklijster

GROTE LIJSTER + ZANGLIJSTER + KOPERWIEK + KRAMSVOGEL + MEREL: lijstster

Als we de merel even buiten beschouwing laten, dan vinden we in dit geval een netwerk waarin alle vier punten verbonden zijn:

---

<sup>5</sup> In de andere twee minimale structuren zijn ZWARTE MEES dan wel STAARTMEES gewisseld met PIMPELMEES. De data geven dus geen uitsluitel over de manier waarop deze drie mezen met elkaar verbonden zijn.



*Figuur 5: Semantic map voor lijsters uitgezonderd de merel*

Deze vier soorten bevinden zich dus alle op dezelfde 'afstand' van elkaar.

De positie van de merel is echter bijzonder. Van de 21 woordtypen voor merel is er slechts één heteroseem, namelijk *lijster*, de term die ook voor de andere vier soorten gebruikt kan worden. Dit betekent dat de data ons geen uitsluitsel geven met welk van de vier soorten we merel moeten verbinden om *lijster* samenhangend te maken. Dit is een algemene tekortkoming van de bestaande semantic maps die kan worden opgelost door meer structuur in een map aan te brengen. Ik vat de vier dichterbij elkaar staande lijsters op als één punt dat vervolgens als geheel met de merel wordt verbonden.



*Figuur 6: Semantic map voor lijsters*

De speciale status van de merel komt zo duidelijk in het netwerk tot uitdrukking.

### 3.4 De leeuweriken-map

Bij de leeuweriken (KUIFLEEUEWERIK, VELDLEEUEWERIK, BOOMLEEUEWERIK) lijkt geen enkele mogelijkheid uitgesloten te zijn:

KUIF- EN VELDLEEUEWERIK: akkerleeuwerik

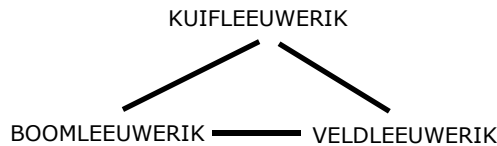
BOOM- EN VELDLEEUEWERIK: grasleeuwerik, graspieper, schrijver, zoet liefje

KUIF- EN BOOMLEEUEWERIK: tureluur, tureluut

KUIF-, BOOM- EN VELDLEEUEWERIK: heileeuwerik, kantleeuwerik, leeuwerik, strontpikker, tjur, wijntapper

De semantic map is daarom een driehoek:





*Figuur 7: Semantic map voor leeuweriken*

Deze structuur drukt uit dat de drie soorten even dicht bij elkaar staan.

### 3.5 De mussen-map

De vijfde en laatste groep vogels die in Swanenberg (2000) behandeld wordt, zijn de mussen. De HUISMUS en RINGMUS zijn beide van de familie van de Passeridae, terwijl de GRASMUS en de HEGGENMUS tot andere families behoren.

HEGGENMUS + GRASMUS: aardmus, blauwpiepertje, koolmus, kwetter

GRASMUS + RINGMUS: grasmus, graspieper

HEGGENMUS + RINGMUS: hegmus, vliermus

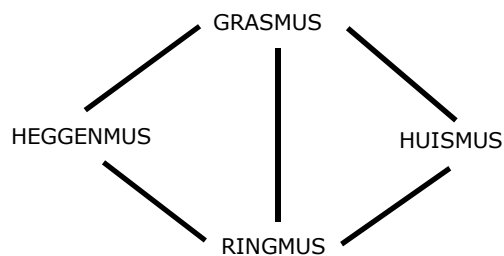
RINGMUS + HUISMUS: korenmus, panmus, potmus, sjierp, sjurk, steenmus, zjefke

HEGGENMUS + GRASMUS + HUISMUS: musje

HEGGENMUS + GRASMUS + RINGMUS: rietmus

alle vier: mus

Een van de twee maps die dit patroon ondersteunt ziet er als volgt uit:



*Figuur 8: Semantic map voor mussen*

In de andere mogelijke map zijn de grasmus en de heggenmus omgewisseld.

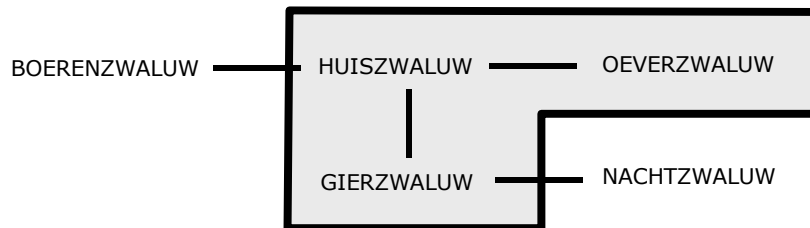
## 4 Patronen op de semantic map

We hebben in paragraaf 3 een aantal 'kale' semantic maps gedefinieerd. Dat biedt ons natuurlijk op zich al een visualisering van de manier waarop een bepaald domein in elkaar zit. Maar de bruikbaarheid van deze visualisatie ligt ook in de manier waarop we

nu vervolgens lexicale data op een map kunnen projecteren, om zo patronen en generalisaties te kunnen ontdekken. In deze sectie zal ik kijken naar de spreiding van individuele woorden, van benoemingsmotieven, en van generieke benoemingen.

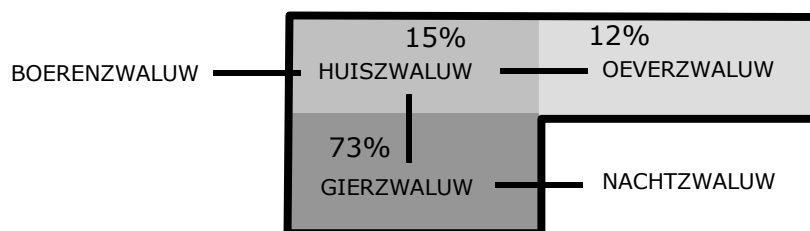
#### 4.1 De spreiding van woorden

We zagen al dat woorden afgebeeld kunnen worden als gebieden op de semantic map, zoals bijvoorbeeld *steenzwalm*:



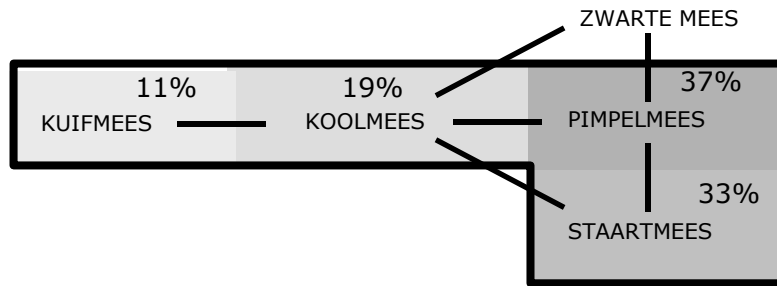
Figuur 9: Spreiding van steenzwalm

Door de manier waarop de map is gestructureerd zal dit per definitie altijd een samenhangend gebied zijn. De map wordt interessanter als we een extra dimensie eraan toevoegen, door aan te geven hoeveel opgaven er voor *steenzwalm* zijn per soort. Er zijn 19 opgaven voor de gierzwalm ( $\approx 73\%$  van het totaal), 4 ( $\approx 15\%$ ) voor de huiszwalm en 3 ( $\approx 12\%$ ) voor de oeverzwalm. Dat is in Figuur 10 weergegeven door middel van grijstinten. Hoe donkerder het gebied dat een soort bedekt, hoe vaker het woord voor die soort is opgegeven:



Figuur 10: Spreiding van steenzwalm met percentages

Zo kunnen we zien hoe dat het woord *steenzwalm* een piek heeft bij GIERZWALUW en afloopt via HUISZWALUW naar OEVERZWALUW. Ook woorden in andere domeinen kunnen we zo weergeven, zoals *ossenkopje* bij de mezen (steeds met afgeronde percentages):



Figuur 11: Spreiding van ossenkopje

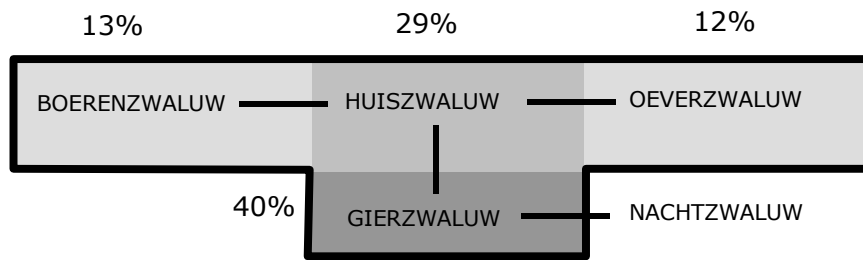
Hier wordt de piek van het woord gevormd door PIMPELMEES en vandaar loopt het af naar andere soorten.

Deze twee voorbeelden suggereren een hypothese over de spreiding van woorden over het netwerk. Die hypothese luidt dat elk woord ten hoogste één piek heeft: er is één betekenis waarvoor het woord het best vertegenwoordigd is en vanuit die betekenis nemen de percentages geleidelijk af. Het kan dus niet zo zijn onder deze hypothese dat een woord als *ossekopje* 37% van de opgaven heeft bij PIMPELMEES, afneemt naar 19% bij KOOLMEES en vervolgens weer toeneemt naar 30% bij KUIFMEES. Woordbetekenissen zijn dus niet alleen samenhangend op het netwerk, maar ze zijn ook op een statistisch normale manier verdeeld over het netwerk. Anders gezegd: het aantal opgaven voor een woord neemt monotoon af naarmate de afstand tot de lexicale piek toeneemt. Alle vogelwoorden uit het gegevensbestand lijken in overeenstemming te zijn met deze hypothese.

Er is een voor de hand liggende hypothese over wat aan dit patroon ten grondslag ligt. Een benoeming als *steenwalm* is wellicht begonnen als aanduiding voor de gierzwaluw en vandaar stapsgewijs uitgebreid naar soorten die er semantisch het dichtst bij staan, eerst de huiszwaluw en dan de oeverzwaluw. Belangrijk is daarbij dat de uitbreiding van een benoeming niet willekeurig verloopt, maar via de paden die op de map gedefinieerd zijn.

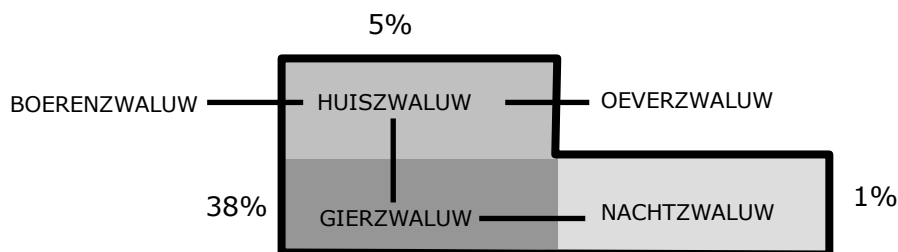
#### 4.2 De spreiding van benoemingsmotieven

Een belangrijk onderdeel van de studie van Swanenberg vormen de benoemingsmotieven voor vogelnamen, dat wil zeggen in hoeverre eigenschappen als kleur, zang, formaat of biotoop een rol spelen bij de benoeming van vogels. Om dat op het spoor te komen heeft Swanenberg ook de relatieve frequenties bepaald van de benoemingsmotieven bij verschillende vogelgroepen. Zulke frequenties kunnen we ook weergeven op een map, zoals te zien is in Figuur 12 voor het benoemingsmotief B1 ('broedt in de buurt van mensen, bouwt het nest tegen de muur onder het dak', Swanenberg 2000:124):



Figuur 12: Spreiding van benoemingsmotief B1

Het gaat hier om benoemingen die de vorm hebben van een samenstelling, met een van de nomina *boer, huis, stal, kerk, toren, venster, muur, steen, stad, rots, steen, kot*, en *pan* als eerste lid en het generieke nomen *zwaluw* (of een variant daarvan) als tweede lid, bijvoorbeeld *steen+zwalw*. De percentages komen uit Swanenberg (2000:126) en zijn (anders dan in Figuur 10 en 11) verkregen door 'het aantal opgaven van de benamingen die bij een bepaald benoemingsmotief zijn ondergebracht te delen door het totaal aantal opgaven voor de betreffende referent.' (Swanenberg 2000:125).<sup>6</sup> Een ander voorbeeld is het benoemingsmotief C1 ('hoge schrille toon'):



Figuur 13: Spreiding van benoemingsmotief C1

De generalisatie die we voor de zwaluwen allereerst kunnen maken is dat een motief steeds correspondeert met een samenhangend gebied op de map. Dit is een belangrijke uitkomst: het laat zien dat een netwerk gebaseerd op de heterosemie van individuele woorden (in paragraaf 3) ook iets zegt over de samenhang in een *groep* van woorden met een gemeenschappelijk lexicaal kenmerk. Het gebied dat gedekt wordt door deze benoemingen is niet alleen samenhangend, maar vertoont ook weer een typerend verloop. Het benoemingsmotief is steeds het sterkst vertegenwoordigd bij één soort en het loopt vandaar monotoon en via de lijnen van het netwerk af naar andere soorten. In deze twee voorbeelden ligt de piek bij de gierzwaluw, maar andere soorten kunnen de kern vormen van andere benoemingsmotieven.

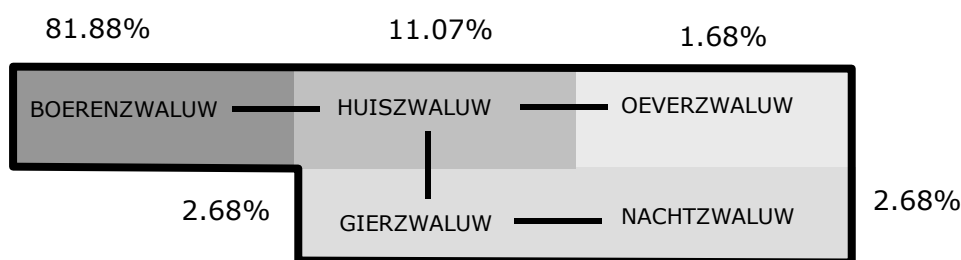
<sup>6</sup> Een anonieme reviewer van *Taal en tongval* merkt op dat het aantal opgaven (*tokens*) van een bepaalde benoeming te zeer van toeval afhankelijk is om methodologisch een goede basis te vormen voor deze maat. Ik ga hier uit van Swanenbergs getallen, maar een maat die gebaseerd is op benoemingstypes zou hier wellicht betrouwbaarder kunnen zijn.

Het is belangrijk om te benadrukken dat de gebieden in Figuur 12 en 13 uitsluitend gebaseerd zijn op het gebruik van *benoemingen* met een bepaald motief voor bepaalde soorten, niet noodzakelijk op het daadwerkelijk aanwezig zijn van de betreffende eigenschap bij die soorten. Swanenberg meldt alleen voor de huiszwaluw, oeverzwaluw en gierzwaluw een schrille toon, het gebied met een stippellijn omkaderd in Figuur 13. De nachtzwaluw wordt dus op één plaats een pieper genoemd terwijl hij niet piept en de oeverzwaluw wordt nergens in de data op basis van zijn schrille toon benoemd. Er is dus geen identiteit van de (linguïstische) benoeming en het (biologische) motief.

### 4.3 De spreiding van generieke benamingen

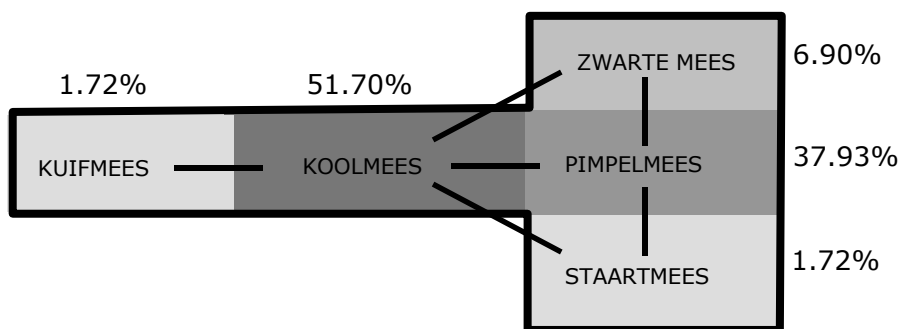
Swanenberg laat zien dat *autohyponymie* een belangrijk fenomeen is in de benoeming van vogels: een benoeming die gebruikt wordt voor een groep als geheel (de *primaire* of *generieke* naam), wordt ook bijna altijd gebruikt voor individuele soorten. Zo duidt *lijster* niet alleen het geslacht van de echte lijsters aan, maar ook specifiek de zanglijster. Nu gebeurt dit voor sommige lijsters meer dan voor andere en dat gegeven gebruikt Swanenberg om in een domein een prototype aan te wijzen. Zo wordt de zanglijster het meest als lijster aangeduid en de kramsvogel het minst: de zanglijster is daarmee de meest prototypische lijster.

Door nu de verdeling van zulke generieke benamingen in kaart te brengen kunnen we visualiseren wat in een bepaald domein de prototypestructuur is. In de volgende semantic map is weergegeven wat de percentages zijn van het gebruik van varianten van *zwaluw* voor de verschillende soorten in verhouding tot het totale aantal opgaven van *zwaluw* (Swanenberg 2000:121):



Figuur 14: Prototypicaliteit van zwaluwen

Bij de boerenzwaluw vinden we hier de prototypische piek en vandaar lijkt de prototypicaliteit van de zwaluwen af te lopen naar de andere uiteinden van de map. Een zelfde soort patroon zien we bij het gebruik van *mees* (en varianten) (Swanenberg 2000:180):



Figuur 15: Prototypicaliteit van mezen

Ook hier biedt de structuur van de semantic map, zoals we die in paragraaf 3 hebben afgeleid op basis van heterosemie, een goed fundament voor het begrijpen van een onafhankelijk semantisch verschijnsel, namelijk autohyponymie en prototypicaliteit.

## 5 Conclusies

Dit artikel biedt een eerste verkenning van het gebruik van een heel ander soort semantische 'kartografie', semantic maps, toegepast op lexicale dialectdata. Ik heb laten zien hoe het verschijnsel heterosemie gebruikt kan worden om netwerken van vogelsoorten te bouwen, waarop we vervolgens de verdeling van individuele woorden, benoemingsmotieven en prototypicaliteit kunnen afbeelden. Deze methode biedt allereerst een visualisering van de semantische structuur van lexicale variatie en het geeft daarmee belangrijke heuristische mogelijkheden. Maar de methode belichaamt tevens een hypothese over de manier waarop lexicale variatie beperkt wordt door de structuur van het onderliggende conceptuele domein. De data over benoeming van vogels suggereert bovendien dat een woord niet alleen samenhangend is op de semantic map, maar dat het bovendien ook één enkele 'heuvel' vormt. Het gaat hier om een eerste verkenning op basis van één domein en die roept natuurlijk allerlei empirische, methodologische en theoretische vragen op.

In hoeverre zijn de semantic maps werkelijk *semantisch*? We zouden de verbanden, zoals die tussen vogelsoorten op basis van de dialectdata worden onthuld, ook onafhankelijk bevestigd willen zien, op basis van niet-talige gegevens die ons iets kunnen vertellen over de conceptuele verwantschap tussen de soorten en hoe die voor verschillende soorten sprekers een rol spelen in hun taalgebruik. En als de structuur niet of niet helemaal semantisch is, willen we nog steeds weten waarom de maps een bepaalde vorm hebben.

Hoe kunnen we de semantic maps *statistisch* verfijnen? De niet-statistische methode die ik hier heb gebruikt is eigenlijk te grof. De leeuweriken-map in Figuur 7 verbindt alle betekenissen met elkaar, waardoor geen enkele beperking of structuur

zichtbaar wordt en alle leeuweriken even dicht bij elkaar staan. In werkelijkheid zijn er veel meer benoemingen voor de veldleeuwerik en boomleeuwerik samen dan voor de veldleeuwerik en kuifleeuwerik samen en we zouden dus de link tussen de eerste twee veel sterker willen laten uitkomen dan tussen de laatste twee. Recentelijk worden bij het maken van semantic maps ook allerlei statistische methoden ingezet om taalvariatie in kaart te brengen (Levinson en Meira 2003, Majid et al. 2004, Croft en Poole 2007, Cysouw 2007). Dat leidt tot een type map waarin betekenissen dicht bij elkaar worden gezet als ze meer woorden gemeenschappelijk hebben. Het zou ook voor lexicale dialectdata de moeite waard zijn om zulke methoden toe te passen. De rijkdom aan variatie die besloten ligt in dialectwoordenboeken kan daarmee met bestaande statistische technieken als *multidimensional scaling* gemakkelijker ontsloten worden.

Hoe kunnen we bestaande methoden van kartografie en visualisatie gebruiken om semantic maps te maken? De semantic maps in dit artikel zijn met de hand gemaakt. Het maken van semantic maps voor dialectdata wordt pas echt interessant als we de data uit dialectwoordenboeken automatisch op een 'semantische landkaart' kunnen laten zetten, zoals dat ook in de dialectgeografie gebeurt, door een lexicale database te koppelen aan de afbeelding. We willen daarbij niet alleen flexibiliteit in het selecteren en projecteren van bepaalde woorden, maar ook meer mogelijkheden in de weergave van allerlei eigenschappen door middel van kleur en reliëf (zie ook Old 2002 en Skupin en Fabrikant 2003).

Wat kunnen semantic maps ons eigenlijk leren over lexicale variatie? Hoewel we dankzij de hier toegepaste technieken inzicht kunnen krijgen in de manier waarop woorden verdeeld zijn over betekenissen, blijven er natuurlijk veel vragen over het hoe en waarom van lexicale variatie in dialecten (De Caluwe en Devos 1993, Van Bree 2000). Waarom is er meer variatie bij de ene soort dan bij de andere? Waarom is het ene woord flexibeler in zijn toepassing dan een ander woord? Hoe wordt de lexicale variatie bepaald door factoren in ruimte en tijd? Semantic maps bieden natuurlijk geen directe antwoorden op zulke vragen, maar ze kunnen wel helpen in het zoeken naar die antwoorden.

## **Referenties**

- Bree, C. van (2000). Taalbewustzijn, taalverandering en regionale taalvariatie: Een benadering. *Taal en Tongval* 52, 22-46.
- Caluwe, J. De en M. Devos (1993). Universele tendenzen in de lexiconstructuur van de Nederlandse dialecten? *Handelingen van de Koninklijke Zuidnederlandse Maatschappij voor Taal- en Letterkunde en Geschiedenis* XLVIII, 25-37.

- Croft, W. en K.T. Poole (2007). Inferring universals from grammatical variation: multidimensional scaling for typological analysis. *Theoretical Linguistics* 34.1, 1-37.
- Cysouw, M. (2007). Building semantic maps: the case of person marking. In M. Miestamo en B. Wälchli (reds.). *New Challenges in Typology*. Berlin: Mouton, 225-248.
- Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought*. Cambridge, MA: MIT Press.
- Haspelmath, M. (1997). *Indefinite pronouns*. Oxford: Oxford University Press.
- Haspelmath, M. (2003). The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. In M. Tomasello red., *The new psychology of language* 2. Mahwah, NJ: Lawrence Erlbaum, 211-242.
- Heeringa, W. en J. Nerbonne (2006). De analyse van taalvariatie in het Nederlandse dialectgebied: methoden en resultaten op basis van lexicon en uitspraak. *Nederlandse Taalkunde* 11.3, 218-257.
- Levinson, S.C. en Meira, S. (2003). 'Natural concepts' in the spatial topological domain. *Language*, 79.3, 485-516.
- Majid, A., M. van Staden, J.S. Boster en M. Bowerman (2004). Event categorization: A cross-linguistic perspective. *Proceedings of the 26th Annual Meeting of the Cognitive Science Society*, 885-890
- Old, L.J. (2002). Information Cartography Applied to the Semantics of Roget's Thesaurus. *Proceedings of the 13th Midwest Artificial Intelligence and Cognitive Science Conference (MAICS'02)*, Chicago, Illinois.
- Skupin, A. en S.I. Fabrikant (2003). Spatialization methods: A cartographic research agenda for non-geographic information visualization. *Cartography and Geographic Information Science* 30.2, 95-115.
- Swanenberg, J. (2000). *Lexicale variatie cognitief-semantisch benaderd: Over het benoemen van vogels in Zuid-Nederlandse dialecten*. Proefschrift, Katholieke Universiteit Nijmegen.
- Vriend, F. de, R. van Hout en L. Boves (2007). Data driven clustering van de gegevens uit het *Woordenboek van de Brabantse Dialecten* op basis van lexicale afstanden. Deze aflevering.
- Zwarts, J. (2006). *Om en rond: een semantische vergelijking*. *Nederlandse Taalkunde* 11.2, 101-123.