



GÖTEBORGS UNIVERSITET

Institutionen för språk och litteraturer

Nederländska

Het woord na de moord

De invloed van de moord op Theo van Gogh op latente waarderingen
van *moslims* en *islam* in Nederlandse krantentaal

Björn Sjödén

Kandidatuppsats

HT-11

Handledare: Evie Coussé

Examinator: Mona Arfs

Het woord na de moord: de invloed van de moord op Theo van Gogh op latente waarderingen van *moslims* en *islam* in Nederlandse krantentaal

Samenvatting.

De algemene vraag van deze scriptie is hoe maatschappelijke gebeurtenissen het dagelijkse taalgebruik beïnvloeden. Als specifiek voorbeeld werd bestudeerd hoe de moord op Theo van Gogh in 2004 de waarderingen van moslims en de islam in dagelijkse krantentaal heeft beïnvloed. Twee hoofdvragen werden nader onderzocht: (1) Heeft de moord op Van Gogh op korte en lange termijn geleid tot een meer of minder positief taalgebruik over *moslims* en *islam* in Nederlandse dagbladen? (2) Welke woorden en onderwerpen karakteriseren de berichtgeving over *moslims* en *islam* vóór en ná de moord? Om het zeer grote materiaal van tienduizenden krantenartikelen te analyseren werd de gecomputeriseerde methode van Latente Semantische Analyse (LSA) toegepast. Waarderingen van *moslims* en *islam* werden beoordeeld op basis van hun nabijheid van positieve en negatieve geladen woorden in de krantenteksten. Het resultaat liet zien dat *moslims* en *islam* (1) in significant meer positieve contexten voorkomen ná de moord dan daarvoor en (2) in veranderde thematische contexten optreden die een overgang aanduidt van associaties met het buitenland naar godsdienst en binnenlandse politiek. Een tentatieve conclusie is dat de gemeenschappelijke talige referentiekaders rond moslims en de islam veranderd zijn als gevolg van de moord. De studie is een eerste poging om LSA te gebruiken voor deze vragen; ten laatste wordt een aantal mogelijke opvolgingsstudies voorgesteld.

Ordet efter mordet: hur mordet på Theo van Gogh påverkat latenta värderingar kring *muslimer* och *islam* i nederländskt tidningspråk

Sammanfattning.

Den övergripande frågeställningen i denna uppsats är hur samhällseliga händelser påverkar det dagliga språkbruket. Som specifikt exempel undersöktes hur mordet på Theo van Gogh 2004 påverkat värderingarna kring *muslimer* och *islam* i dagligt [nederländskt] tidningspråk. Två huvudfrågor undersöktes närmare: (1) Har mordet på van Gogh på kort och lång sikt lett till ett mer eller mindre positivt språkbruk om muslimer och islam i nederländska dagstidningar? (2) Vilka ord och ämnen karaktäriserar nyhetsrapporteringen kring muslimer och islam före och efter mordet? För att analysera det mycket stora materialet omfattande tiotusentals tidningsartiklar tillämpades den datoriserade metoden Latent Semantisk Analys (LSA). Värderingar kring muslimer och islam skattades på basis av närheten till positivt och negativt laddade ord i tidningstexterna. Resultatet visade att muslimer och islam (1) förekommer i signifikant mer positiva kontexter efter mordet än dessförinnan, och (2) uppträder i förändrade tematiska kontexter som antyder en övergång från associationer med utlandet, till religion och inrikespolitik. En tentativ slutsats är att de gemensamma språkliga referensramarna runt muslimer och islam förändrats som följd av mordet. Studien är ett första försök att använda LSA för dessa frågor; avslutningsvis föreslås ett antal möjliga uppföljningsstudier.

Inhoudsopgave

1	Inleiding.....	4
1.1	Motief – een persoonlijke ervaring.....	4
1.2	Probleemgebied en doelstelling.....	4
1.3	Structuur van de scriptie.....	5
2	Taalwetenschappelijke achtergrond.....	6
2.1	Het verband tussen taal en maatschappij.....	6
2.2	Cognitivismen als integrerend perspectief op talige en sociale ontwikkeling.....	7
2.3	Latente Semantische Analyse (LSA): het kwantificeren van betekenissen.....	8
2.3.1	LSA om betekenissen te meten.....	8
2.3.2	LSA om waarderingen te meten.....	10
2.4	Conclusies van de theoretische achtergrond.....	11
3	Onderzoeksvragen.....	12
4	Methode.....	12
4.1	Materiaal.....	12
4.1.1	Nederlandse taalruimte.....	12
4.1.2	Woorden met positieve/negatieve valentie.....	12
4.1.3	Corpora.....	13
4.2	Procedure.....	14
4.3	Statistische tests.....	14
5	Resultaten.....	15
5.1	Valentiewaarden van krantenartikelen over <i>moslims</i> en <i>islam</i> vóór en ná de moord.....	15
5.2	Kenmerkende woorden van krantenartikelen over <i>moslims</i> en <i>islam</i> vóór en ná de moord.....	16
5.3	Een poging tot kwalitatieve validatie.....	17
6	Discussie.....	19
7	Referenties.....	22
7.1	Internetbronnen.....	22
7.2	Boeken en artikelen.....	22
8	Appendix: lijst van woorden met positieve/negatieve lading.....	24

1 Inleiding

1.1 Motief – een persoonlijke ervaring

Op 2 november 2004 werd de Nederlandse schrijver en filmmaker Theo van Gogh op straat doodgeschoten, een paar blokken van mijn toenmalige woonplaats in Amsterdam. Het motief van de dader, een zelfbenoemde fundamentalistisch islamist, was dat Van Gogh door zijn heftige uitspraken – in het bijzonder zijn en Ayaan-Hirsi Alis kortfilm *Submission*¹ – de islam en daarmee alle moslims had beledigd. Zijn woorden waren, zonder enige ironie, moordend. Voor mij wekte de gewelddadige gebeurtenis vragen over de soms overweldigende emotionele kracht van het woord.

Uit een taalkundig perspectief rijst de vraag hoe woorden hun emotionele lading en sociale functie krijgen om saamhorigheid of vijandschap, goedkeuring of afkeuring, uit te drukken. Woorden werken op verschillende niveaus, van neutrale beschrijvingen tot iets positiefs of negatiefs. Omdat de moord op Theo van Gogh zo veel en uitgebreide negatieve reacties opleverde, lijkt het logisch dat de context rond de moord in negatieve termen wordt beschreven. Maar heeft de negativiteit van de moord ook gevolgen gehad buiten de specifieke context, als wij over moslims en de islam in andere dagelijkse contexten spreken?

Sinds de moord op Theo van Gogh is er een hevig debat ontstaan rond islamisme en de positie van moslims in Nederland. In de media lijkt een polarisatie tussen islamitische voorstanders en hun critici te ontstaan. Het taalgebruik is soms direct vijandig, bijvoorbeeld als moslims met “terroristen” worden geassocieerd en er aangemaand wordt “dat ze kleur bekennen” (Van Walsum in De Volkskrant, 5 januari 2011). De controversiële politicus en islam-criticus Geert Wilders komt zelfs voor de rechter wegens “het aanzetten tot haat en discriminatie” tegen moslims en hun geloof (geraadpleegd op www.rechtspraak.nl).

Volgens recente debatboeken (Buruma, 2006; Zuiderveld, 2010) kan de moord op Van Gogh in 2004, en eerder op de politicus Pim Fortuyn in 2002, zelfs gezien worden als een indicatie van een grotere maatschappelijke en politieke ontwikkeling in Nederland waar een algemeen beeld van de Nederlandse cultuur in twijfel is getrokken. De beschrijving van die ontwikkeling wordt gekarakteriseerd door begrippen zoals *tolerantie*, *immigratie*, *assimilatie*, *fundamentalisme* en *vrijheid*. De positieve of negatieve lading van deze begrippen zijn niet vanzelfsprekend maar zijn afhankelijk van de context en de taalgebruiker.

Duidelijk is dat zowel Theo van Gogh als Geert Wilders pejoratieve termen in hun beschrijvingen van moslims hebben gebruikt. De kwestie is weer actueel geworden: wat precies maakt een uitdrukking pejoratief of zelfs een expressie van haat? Hoe komt het dat een woord zoals *tolerantie* in de algemene Nederlandse context een positieve lading heeft en *islamisme* een negatieve lading heeft gekregen? Is het zelfs maar mogelijk dat soort vragen objectief en zonder eigen, persoonlijke bias te analyseren?

¹ In verband met de moord op Van Gogh was ook Hirsi-Ali met de dood bedreigd, maar niet vermoord. Bovendien liet zich Van Gogh negatief uit over de islam in veel andere contexten, b.v. in krantencolumns, tv-programmas, zijn eigen website en boeken. Zie verder b.v. Van Gogh (2003) en Buruma (2006).

1.2 Probleemgebied en doelstelling

Door de boven gestelde vragen probeer ik een groter taalkundig probleemgebied te beschrijven: hoe sociale factoren en gebeurtenissen in de maatschappij (zoals de moord op Van Gogh en de redenen ervoor) samenhangen met talige ontwikkeling (zoals de ontwikkeling van betekenissen, waarderingen en connotaties van een begrip). Een algemeen doel van mijn onderzoek is deze soorten verbanden te bestuderen met de moord op Theo van Gogh als concreet voorbeeld. Centraal staat de invloed van die gebeurtenis op (latente) waarderingen van moslims en islam in het Nederlandse massamediale debat². Het probleem hoe dit wetenschappelijk onderzocht kan worden, beschouw ik net zo belangrijk als de uitkomst van het onderzoek. Mijn hypothese is dat de moord op Theo van Gogh, als ongeëvenaarde gebeurtenis met enorme belichting in de media, de algemene waarderingen van moslims en islam heeft beïnvloed. In het bijzonder wil ik de toepassing van een relatief nieuwe maar reeds geëtablerde techniek voor deze doeleinden, *Latent Semantic Analysis* (LSA), onderzoeken.

Het basismateriaal van mijn onderzoek is een groot corpus Nederlandse krantenteksten van vóór en ná de moord op Theo van Gogh dat met gecomputeriseerde methodes (LSA) statistisch geanalyseerd kan worden. Door veel – tienduizenden of meer – verschillende contexten en miljoenen woorden te analyseren, kunnen aparte woorden een kwantitatieve waarde³ toebedeeld worden die aanduidt of het woord vaker in negatieve of positieve contexten voorkomt. Daardoor kunnen latente waarderingen van een begrip of onderwerp ontmaskerd worden met een minimum aan subjectieve interpretaties. Voor zover ik weet, zijn er geen eerdere studies over de specifieke invloed van zo'n enkele, maatschappelijke gebeurtenis op het Nederlandse taalgebruik. Mijn doel is ook een methodologische discussie te voeren die hopelijk kan bijdragen aan toekomstige tekstgebaseerde analyses van latente waarderingen.

1.3 Structuur van de scriptie

Hoofdstuk 2 bevat een brede taalwetenschappelijke achtergrond om het onderzoeksgebied af te bakenen en relevante theoretische perspectieven en terminologie te introduceren. Ik wil hier ook de toegepaste techniek, LSA, op conceptueel niveau beschrijven. Dit leidt tot een precisering van mijn onderzoeksvragen in hoofdstuk 3. Hoofdstuk 4 gaat nader in op de methode en de selectie van het materiaal (tekstcorpora) voor de huidige studie. In hoofdstuk 5 presenteer ik de resultaten. Ten slotte discussieer ik in hoofdstuk 6 de resultaten en probeer enige conclusies te trekken.

² Met "latente waarderingen" bedoel ik waarderingen die niet expliciet uitgedrukt worden, maar door de omgevende context te begrijpen zijn. Dit wordt verder uitgelegd in paragraaf 2.2–2.3.

³ De "kwantitatieve waarde" benoem ik de *valentie* van een woord die gespecificeerd kan worden met een *valentiewaarde* (b.v. "het woord 'slecht' heeft een negatieve valentie en een valentiewaarde van -1"). In dagelijkse taal en met kwalitatieve termen zouden wij eerder daarover spreken als de positieve/negatieve *lading* of *emotionele lading* van het woord dat de *waardering* in een bepaalde context uitdrukt (b.v. "slecht" is een negatief geladen woord; "dat was slecht gedrag" is een negatieve waardering over het gedrag). Zie ook 2.3.2.

2 Taalwetenschappelijke achtergrond

In redelijke taal weerklinkt wat in werkelijkheid gebeurt.
G. J. P. J. Bolland (1854-1922), Nederlands taalkundige en filosoof
(Bolland, 1909, p. 888)

2.1 Het verband tussen taal en maatschappij

Er zijn tegenwoordig meerdere disciplines van de linguïstiek waarin bestudeerd wordt hoe talige uitdrukkingen verband houden met sociale of maatschappelijke factoren. Op algemeen niveau lijkt dit een kwestie voor de sociolinguïstiek (b.v. Appel, Hubers & Meijer, 1981) en daarmee gerelateerde disciplines zoals taalsociologie (Van der Plank, 1985) en tekstsociologie (Nijhof, 2003). Enige typische onderwerpen binnen deze disciplines zijn de invloed van het Frans op de Nederlandse woordenschat na de Franse revolutie, hoe nieuwe taalconventies zijn ontstaan als gevolg van nieuwe vormen van communicatie, zoals slang, straattaal en sms-taal, en hoe nieuwe werkwoorden rond internet en computers zijn ontstaan uit “genederlandiseerd” Engels (*surfen, googlen, chatten, downloaden*, enz.; zie b.v. Boon, 2007).

Deze voorbeelden maken een aspect van het inleidende citaat van Bolland duidelijk: dat verschijnselen in de werkelijkheid – zoals maatschappelijke gebeurtenissen en technische ontwikkelingen – in het taalgebruik weerklinken. Men kan ook een ander aspect op het verband tussen maatschappelijke en talige ontwikkeling benadrukken, namelijk, dat ontwikkelingen in de taal beïnvloeden hoe wij de werkelijkheid waarnemen. Bijvoorbeeld: Heeft de groeiende technische woordenschat in dagelijkse taal (rond computers, mobiele telefoons, internet etc.) geleid tot een meer technische visie op de wereld? Gaat de grote invloed van het Engels op het Nederlands ook gepaard met een culturele invloed die onze begrippen en interpretaties vormt? Het idee dat de taal als een soort cultureel raster de waarneming zelf beïnvloedt en daarmee in bepaalde mate leidt of beperkt wat wij (kunnen) denken, is bekend als de Sapir-Whorf-hypothese (Dirven & Verspoor, 2001, p. 176f). Een vaak genoemd voorbeeld daarvan is de veel verschillende woorden die de Eskimo's gebruiken voor *sneeuw*.

Samenvattend is het zowel mogelijk dat de maatschappij de taal beïnvloedt als dat de taal de maatschappij beïnvloedt. Met andere woorden, “de woorden weerspiegelen de maatschappij”.⁴ Maar zulke verbanden zijn normaal niet direct observeerbaar en ontwikkelen zich over lange tijdsperiodes. Als theoretisch deelgebied worden beide aspecten omvat door de interculturele en transculturele semantiek, bijvoorbeeld in het onderzoek van communicatiestijlen en betekenisontwikkeling van woorden binnen een taalgemeenschap (Dirven & Verspoor, 2001, hs. 6).

Opvallend is echter dat veel onderzoek gericht is op verschillende communicatiegedragingen en betekenisverschillen, terwijl veel minder bekend is over het proces waardoor waarderingen rond woorden veranderen. Veel woorden worden in een context positief en in een andere context negatief gewaardeerd. Maar waarderingen kunnen ook in de tijd en door historische redenen veranderen – een duidelijk voorbeeld is de wisselende populariteit van eigennamen (zoals de naam *Adolf* vóór en ná de Tweede Wereldoorlog). Mode en trends veranderen

⁴ Dit is ook de titel van een Zweeds boek over het onderwerp, “Orden speglar samhället” (Allén, Gellerstam & Malmgren, 1989).

continu wat als "cool" of "sullig" beleefd wordt, en als een woord vaak in hetzelfde soort waarderende contexten voorkomt, zal dit ook positieve of negatieve associaties wekken als het woord opnieuw gebruikt wordt. Een voorbeeld daarvan zijn de hoger of lager gewaardeerde eigenschappen die door de geschiedenis geassocieerd zijn met mannen en vrouwen, die een positieve of negatieve lading geven aan wat (stereotiep als) *mannelijk* of *vrouwelijk* beleefd wordt (zie b.v. Barreto, Ryan & Schmitt, 2009). Dus, de betekenisontwikkeling van *mannelijk* en *vrouwelijk* volgt niet noodzakelijk, maar verloopt wel parallel met, de ontwikkeling van hoe mannelijk en vrouwelijk gedrag gewaardeerd wordt. Om zulke verbanden te verklaren, moeten wij buiten de talige context kijken en ook de sociale context van de taalgebruikers bekijken.

2.2 Cognitivismen als integrerend perspectief op talige en sociale ontwikkeling

Hoe kan men het zeer complexe proces van betekenis- en waarderingsontwikkeling – dat bovendien steeds elkaar beïnvloedt en continu onder verandering staat – in één begrijpelijk beeld uitleggen? Hoewel de vraag van belang is voor het huidige onderzoek, probeer ik hem hier niet meer dan kort te schetsen. De analyse (met name de methode van LSA) rust echter op een aantal gerelateerde theoretische veronderstellingen die binnen het perspectief van de cognitieve taalkunde samenvallen.

Een eerste veronderstelling is het tekstsociologische idee (Nijhof, 2003) dat de sociale werkelijkheid alleen in de interpretaties van mensen bestaat. Teksten vermelden, schriftelijk of mondeling, beelden van de werkelijkheid die mensen reïnterpreteren of "construeren". Men bekijkt dan de taal vanuit een cognitieve basis, dat wil zeggen dat de taal deel uitmaakt van het globale cognitieve systeem van de mens, met betrekking tot de begrippen en categorisaties die gebruikt worden om de wereld waar te nemen (Dirven & Verspoor, 2007). Met andere woorden: ook al bevat de werkelijkheid een onoverzichtelijk complexiteit, toch kan hij begrepen worden in termen van onze conceptualisatie van die werkelijkheid. Tekstanalyse kan een ingang bieden om na te gaan hoe deze conceptualisatie in woorden wordt uitgedrukt. Centraal voor het huidige onderzoek is hoe ook latente, dus niet expliciete, betekenissen door de tekstuele samenhang afgelezen kunnen worden (zie verder paragraaf 2.3).

Een tweede veronderstelling volgt het idee van de constructiegrammatica (b.v. Goldberg, 2003; Langacker, 1987/1991; Verhagen, 2005) dat talige uitingen uit vaste relaties tussen vorm en betekenis bestaan die overeenkomen met onze cognitieve categorieën. Ook al is het niet mijn bedoeling om zulke zogenaamde constructieschema's specifiek te onderzoeken, gaat de analyse ervan uit dat de betekenis van een tekst op het abstracte, algemene niveau van geformaliseerde patronen geconstrueerd wordt. Zulke patronen kunnen schematisch en wiskundig gerepresenteerd worden. Tevens is kennis over het pragmatische taalgebruik nodig, dat behalve een tekstuele ook een sociale context bevat. De focus ligt dus op welke betekenissen woorden krijgen in verhouding tot andere woorden, eerder dan de traditionele syntactische categorieën of lexicaal definities. De kwestie of een woord hier "juist" of "onjuist" gebruikt wordt, is niet relevant.

Een derde veronderstelling betreft de interactie tussen talige en sociale ontwikkeling, want "alles gebeurt in het hoofd": naast een vaste betekenis is er ook een dynamisch proces waarbij die betekenis wordt geactualiseerd in een bredere context *en* een cognitieve realiteit. Kortom, de betekenis wordt gemodificeerd in de context door ons waarnemingsvermogen op individueel niveau en de context schept de betekenis op sociaal niveau. Tegelijkertijd bevatten

teksten hun eigen, interne context die onafhankelijk van de omgeving bestaat. Een nuttig begrip onderscheid wordt gemaakt tussen de *co-text* en de *context* (Bergs & Diewald, 2009, met verwijzing naar Catford, 1965). De *co-text* verwijst naar “intra-linguïstische” factoren binnen de tekst zelf, zoals syntactische eigenschappen, grammaticale regels en constructieschema’s. Door de *co-text* krijgt de tekst zijn interne samenhang en karakter, en schept soms bepaalde verwachtingen voor de inhoud, bijvoorbeeld door idiomatische uitdrukkingen. Aan de andere hand verwijzen de *context* of “extra-linguïstische” factoren naar omstandigheden buiten de tekst zelf, zoals tijd, ruimte, sprekers en communicatiegedrag. Mijn analyse is op de *co-texten* gebaseerd, maar de resultaten worden in een grotere context besproken.

Ten laatste, ook al is de dynamiek tussen betekenis, context, *co-text* en conceptualisatie vaak opvallend, toch zijn deze factoren moeilijk theoretisch te vatten. Een opmerkelijk aspect is dat de *co-text* blijft zoals hij is, terwijl de *context* waarin de interpretatie gebeurt, verandert. Het verschil bepaalt welke conclusies men kan trekken over de *context* op basis van het *co-textuele* materiaal (zoals over de moord op Theo van Gogh op basis van krantenartikelen). Om de analyse zin te geven en een al te subjectieve bias van de interpretatie te vermijden is zowel maatschappelijke als historische en culturele kennis nodig.

2.3 Latente Semantische Analyse (LSA): het kwantificeren van betekenissen

2.3.1 LSA om betekenissen te meten

Een probleem met de traditionele Chomskyaanse taalkunde (Chomsky, 1957) is hoe te verklaren dat vrijwel alle mensen een taal spreken en vooral hoe wij zoveel woorden van onze moedertaal binnen zo’n korte periode als kinderen leren. Als filosofische kwestie maakt dit deel uit van “Plato’s probleem” ofwel het probleem van inductie: Hoe kan een mens zo veel van *alles* leren op basis van de relatief weinig informatie uit directe waarneming? *Latent Semantic Analysis* (Latente Semantische Analyse, LSA) is een poging om deze vraag te beantwoorden met een theoretisch model van hoe de betekenissen van woorden gemaakt worden in hun talige context (of beter gezegd, de *co-text*, volgens de terminologie in 2.2).

Toen de psycholinguïsten Landauer en Dumais (1997) de theorie van LSA oorspronkelijk presenteerden, was het als een cognitief model van kennisrepresentatie. In tegenstelling tot Chomsky’s veronderstelling dat de brein “hard-wired” voor het leren van taal moet zijn, lieten Landauer en Dumais zien dat de *co-text* voldoende is om betekenissen van onbekende woorden af te leiden. Ze gebruikten daarvoor een computerprogramma met een leeralgoritme dat de frequentie en afstand tussen woorden in duizenden teksten berekende. Daarop kon de gelijkheid van woorden en tekststukken wiskundig gerepresenteerd worden vanuit hun *co-textuele* relatie met andere woorden. Het model werd getest door het algoritme toe te passen op een synoniemmeerkeuzetoets. Het bleek dat het computerprogramma synoniemen kon identificeren (en dus door het algoritme nieuwe woorden “leren”) op hetzelfde niveau als dat van schoolkinderen (die een aanzienlijke capaciteit vertonen om nieuwe woorden te leren; zie b.v. Anglin, 1993).

Het experiment leek dus te laten zien dat de betekenis van teksten “gekwantificeerd” konden worden enkel op basis van associaties in schriftelijke documenten. Daarvoor was geen menselijke kennis of interpretatie buiten het toegepaste leeralgoritme nodig. De studie wordt in detail beschreven in Landauer en Dumais (1997, p. 221ff.).

De basis van de LSA is de theoretische abstractie van semantische ruimten. Een “ruimte” is in dit geval een maat van hoe dicht een woord bij andere woorden in een tekst staat. Iets vereenvoudigd, een bepaald woord dat vaak samen met een ander woord voorkomt, creëert een kleine semantische ruimte tussen de twee woorden (wat dan een lage waarde krijgt), terwijl twee woorden die niet zo vaak samen voorkomen een grotere ruimte creëren (die een hogere waarde krijgt). Elke vergelijking tussen woorden krijgt zodanig een gelijkeniswaarde of “similarity estimate” (Landauer, Foltz & Laham, 1998). Bijvoorbeeld, door het tellen van woorden in duizenden teksten zal het Nederlandse woord *hond* semantisch dichtbij het woord *pup* gevonden worden maar ver van het woord *pop*. Het woord *hond* zou wiskundig beschreven worden ten opzichte van zijn gelijkenis of “semantische afstand” (als dimensies in een vector) van een vooraf gekozen aantal woorden in een tekstbestand. Door de onderlinge relaties tussen woorden in kaart te brengen en de gelijkeniswaarden binnen zinnen en tussen stukken tekst te rangschikken, kan op die manier elk woord een maat van “hondheid” worden toegeschreven.

Wellicht komt uit de zeer complexe, statistische basis van LSA voor de niet-wiskundige een misleidende maat van exactheid naar voren. Het is belangrijk er rekening mee te houden dat LSA gewoon een beschrijving van de co-textuele betekenis van een woord oplevert. LSA kan “latente” betekenissen duidelijker maken maar geeft geen precieze definities. Het resultaat van de LSA-analyse kan alternatief worden begrepen als de “gemiddelde betekenis” van een woord vanuit alle geanalyseerde tekstpassages waarin het staat. Behalve van het corpus waarop het LSA-algoritme wordt gebruikt, hangt de juistheid en kwaliteit van de beschrijving af van het aantal dimensies van de semantische ruimte (normaal 100-300 dimensies; zie Landauer & Dumais, 1997; Landauer, McNamara, Dennis & Kintsch, 2007). Tegelijkertijd moet LSA niet verkeerd begrepen worden als gewoon “het tellen van woorden”, als Landauer, Foltz en Laham (1998) benadrukken:

It is important to note from the start that the similarity estimates derived by LSA are not simple contiguity frequencies, co-occurrence counts, or correlations in usage, but depend on a powerful mathematical analysis that is capable of correctly inferring much deeper relations (thus the phrase “Latent Semantic”), and as a consequence are often much better predictors of human meaning-based judgments and performance than are the surface level contingencies that have long been rejected /... / by linguists as the basis of language phenomena. (Landauer, Foltz & Laham, 1998, p. 4)

De wiskundige basis van LSA-algoritmes wordt in detail besproken door Martin en Berry (2007).

Uit een toegepast perspectief zal opgemerkt worden dat de validiteit van de LSA-analyse, dat wil zeggen de kwaliteit van het algoritme om semantische associaties te vinden, geëvalueerd kan worden door middel van empirische methoden. Zoals eerder genoemd, kan het resultaat van de hoogst gerangschikte associaties van een woord vergeleken worden met de woorden in een synoniemwoordenboek (thesaurus). Dat zou ook een indicatie geven in welke mate de betekenissen uit de bestaande co-texten representatief zijn voor het conventionele taalgebruik. Woorden in een thesaurus zijn echter contextloos. De validiteit van de LSA-analyse wordt beter gemeten door co-textuele relaties, bijvoorbeeld hoe consistent negatief of positief het gebruik van bepaalde woorden zijn in verschillende passages van een tekst of tussen teksten van dezelfde aard (zie verder 2.3.2).

Het feit dat LSA exclusief gebaseerd is op de aanwezigheid van enkele woorden lijkt een ander potentieel probleem met zich mee te brengen, bijvoorbeeld met negaties en conditionele uitdrukkingen. Als, bijvoorbeeld, *lekker* op dezelfde manier als *niet lekker* wordt

geclassificeerd en conditionele zinnen zoals *dat zou lekker zijn* ook als een uitdrukking van *dit is lekker* geïnterpreteerd wordt, lijkt de analyse misleidend. Volgens het citaat boven wordt dit probleem echter minder door de complexiteit van LSA-algoritmes, vooral als zij op een groter tekstbestand uitgeprobeerd worden. Een reden daarvoor is simpel statistisch: op grote schaal (tienduizenden tot miljoenen woorden) zijn er met betrekking tot “slecht eten” behalve uitdrukkingen zoals *het eten was niet lekker* zeer waarschijnlijk ook vergelijkbare co-texten met andere woordkeuzen, zoals *het eten was slecht / vreselijk / smaakloos* enz. Als ik dan iets over de mate van “lekkerheid van het eten” wil weten, zou de door LSA op grote basis berekende “gemiddelde betekenis” wel negatief worden.

De grootschalige, statistische benadering van LSA laat ook een sterkte van de methode zien: zeer ongewone formuleringen of constructies, net als extreme uitdrukkingen, krijgen over het geheel geen merkbare invloed op het resultaat. Tegelijkertijd kan meer informatie door LSA geanalyseerd worden dan een persoon redelijkerwijs kan lezen. Een LSA-studie door Sikström en Schmer-Galunder (2007) bevatte alle Reuters nieuwsteksten van een jaar, een bestand van meer dan 800.000 artikelen. Dit gezegd zijnde, specifieke taal- en cultuurkennis blijft nodig om de opgeleverde data in een zinvolle samenhang te brengen.

2.3.2 LSA om waarderingen te meten

Na het analyseren van betekenissen, eist LSA nog een stap om de waardering van een woord te schatten. De schatting neemt aan dat men enige voorkennis heeft van de emotionele lading, ofwel valentie, van de associaties binnen de semantische ruimte van een geanalyseerd woord. Soms blijkt de emotionele lading van een woord heel duidelijk, zelfs per definitie, zoals van het woord *lief*. Als het dan blijkt uit een tekstcorpus dat het woord *hond* vaak samen (dus, in dergelijke co-texten) met het woord *lief* voorkomt, kan ook aan *hond* een positieve waarde toegeschreven worden. Een mogelijke conclusie is dan dat *hond* over het algemeen als een lief dier opgevat wordt. Dit wordt in feite een soort kwalitatieve analyse van hoe honden in het tekstmateriaal worden opgevat, maar zonder de betrokkenheid van een of meerdere menselijke lezers. De enige premisse is de voorkennis van welke woorden in het algemene taalgebruik positief en negatief zijn.

Men kan een LSA-gebaseerde waarderingsanalyse ook uitvoeren met het waarderende woord als uitgangspunt. In een onderzoek over de semantische associaties van het duidelijk positieve woord *geluk* (Zweeds: *lycka*) werd door Garcia en Sikström (in progress), op basis van 1,5 miljoen woorden in Zweedse dagbladen, gevonden dat *geluk* positief geassocieerd werd met persoonsnamen en inclusief voornaamwoorden (zoals *wij, ik en jij*) en negatief geassocieerd met bedrijfsnamen en exclusief voornaamwoorden (zoals *zij*)⁵. Dus, de functie van de LSA-analyse was in dit geval de semantische associaties (de voornaamwoorden) van een al bekend positief begrip zichtbaar te maken, in plaats van de waarderende associaties van een neutraal woord (zoals *hond*) te vinden. Het resultaat leent zich daarna voor diepere psychologische interpretatie (zoals dat mensen over het algemeen gelukkiger worden samen met anderen dan door hun materiële eigendom).

Ter conclusie, een voorwaarde voor een LSA-gebaseerde waarderingsanalyse is dat men tenminste één, positief of negatief, referentiewoord heeft waarmee andere woorden (associaties en zoektermen) gerelateerd kunnen worden. Om subjectieve waarderingen zo beperkt mogelijk te houden, moeten de referentiewoorden een duidelijke emotionele lading

⁵ Dit onderzoek werd populair nieuws voor de avondbladen:
<http://www.expressen.se/nyheter/1.2302114/zlatan-gor-dig-glad-har-ar-de-andra-orden>

hebben in de algemene (dus niet contextspecifieke) taal voor een representatief aandeel van de taalgebruikers. Positief-waarderende woorden krijgen dan een positieve waarde (+1) toegeschreven en negatief-waarderende woorden een negatieve waarde (-1). De cijfers representeren de gemeten emotionele valentie. Daarna kan de valentie van gezochte woorden berekend worden op basis van de semantische nabijheid van plus- en min-woorden binnen de dimensies van de semantische ruimte.

2.4 Conclusies van de theoretische achtergrond

Samenvattend heb ik in de secties 2.1–2.3 de achtergrond van mijn onderzoek willen schetsen. Een uitgangspunt is dat de taal weerspiegelt hoe mensen in een maatschappij of cultuur denken en (psychologisch) werken. Kennis over ons cognitief construeren van de werkelijkheid door de taal kan toegang geven tot dit waarnemingsproces. Het idee dat mensen hun wereld actief “construeren” raakt een taalwetenschappelijke grensgebied aan, tussen de taalsociologie en takken van de semantiek. Ik pleit voor de hypothese dat constructies expliciet gemaakt kunnen worden op statistische basis door middel van uitgebreide tekstanalyse, met name door LSA, die de relaties tussen woorden in termen van semantische ruimten berekent. Door gedetailleerde analyse van de co-texten kan ook de latente waarderingen rond een begrip of thema transparant worden. Latente waarderingen worden expliciet uitgedrukt met valentiewaarden op het niveau van aparte woorden en/of documenten.

Ik kies voor LSA als analysemethode niet als vervanging van een traditionele kwalitatieve analyse maar omdat LSA het structureren en categoriseren van veel grotere bestanden menselijke kennis toelaat dan anders overzichtelijk zou zijn. Daarmee krijgen de resultaten ook een graad van objectiviteit, in de zin van algemeenheid en kwantificeerbaarheid, die boven de persoonlijke bias van individuele analisten staat. Het resultaat van de LSA-analyse kan ook beschouwd worden als een strikt geformaliseerde beschrijving van de co-text, waarna verdere kwalitatieve analyse op andere (sociale, maatschappelijke) aspecten van de context gefocust kan worden.

Behalve als schokkende misdaad onderscheidt zich de moord op Theo van Gogh als maatschappelijke gebeurtenis met invloed op een aparte semantische domein, namelijk het taalgebruik rond moslims en de islam. Ook al zijn de woorden in hun (lexicale) definitie niet veranderd, toch laat het massamediale debat na de moord duidelijk zien dat men de rol van moslims en de islam in een grotere sociale context probeert te positioneren (wat in detail besproken wordt door Buruma, 2006). Dat is een debat vol van waarderingen. De taalwetenschappelijke achtergrond helpt vooral bij het formuleren van specifieke onderzoeksvragen om na te gaan of en hoe moslims en de islam op een nieuwe manier gewaardeerd zijn als gevolg van de moord.

3 Onderzoeksvragen

De hoofdvragen van het huidige onderzoek zijn:

1. Heeft de moord op Theo van Gogh op 2 november 2004 beïnvloedt hoe *moslims* en *islam* gewaardeerd worden in Nederlandse krantentaal, specifiek in termen van positieve/negatieve valentie?
 - a) Zo ja, in welke richting (positief of negatief) blijkt deze verandering?
 - b) Onderscheiden de resultaten zich op korte en lange termijn ná de moord, of lijkt het effect van de moord op de algemene waarderingen van *moslims* en *islam* relatief constant het jaar direct ná de moord (2004) in vergelijking met vijf jaren daarna (2009)?
2. Welke woorden en onderwerpen karakteriseren de Nederlandse berichtgeving met betrekking tot *moslims* en *islam* vóór en ná de moord?

Bovendien rijst een consequentievraag: hoe relateren de resultaten uit 1-2 tot de algemene afbeelding van moslims en de islam in de kranten? Dat wil zeggen, hoeveel kan men uit de resultaten van een LSA-onderzoek aflezen? Daarmee wil ik een startpunt voor een kwalitatieve discussie van de resultaten zetten, wat de focus voor het laatste hoofdstuk (hs. 6) wordt. In het volgende hoofdstuk beschrijf ik mijn methode om deze vragen te onderzoeken en de specifieke maten die daarvoor gebruikt zijn.

4 Methode

4.1 Materiaal

4.1.1 Nederlandse taalruimte

Om het LSA-algoritme aan de corpusrelevante taal (in dit geval krantenteksten) te leren moest eerst een taalruimte van het Nederlands gecreëerd worden, waarop het algoritme “getraind” wordt om Nederlandse co-texten te analyseren⁶. Daarvoor werd het Twentse NieuwsCorpus (TwNC) gebruikt, bestaande uit alle krantenartikelen van negen verschillende dagbladen van 1999 (exclusief de ondertiteling van nieuwsprogramma’s).⁷ Dit vormde een basis van 410 Mb tekst waarop een taalruimte van (als default) 150 dimensies berekend werd. De huidige analyse baseert zich op het programma *Semantic*⁸, ontwikkeld door professor Sverker Sikström aan Lunds universiteit, Zweden, onder het grotere wiskundig programma *Matlab*.

4.1.2 Woorden met positieve/negatieve valentie

Om algemene woorden met positieve of negatieve lading te verzamelen – voortaan de *bronwoorden* benoemd – werden tien personen met Nederlands als moedertaal geconsulteerd (7 vrouwen en 3 mannen) volgens de instructies in het Appendix. Niemand die met woorden bijdroeg werd geïnformeerd over het ware doel van het onderzoek. Een lijst met 35 positieve en 16 negatieve bronwoorden werd samengesteld (zie Appendix).

⁶ Het “trainen” van een algoritme betekent dat het algoritme verbeterd wordt om patronen in de taal te vinden door herhaalde toepassing (iteraties) op authentieke taaldata (teksten). Zie verder Martin en Berry (2007).

⁷ Dank aan Roeland Ordelman aan de Universiteit Twente voor de hulp met toegang tot het TwNC.

⁸ *Semantic* is gratis toegankelijk, na contact met de programmeur, via <http://www.lu.se/sverker.sikstrom/LSALAB.html>

4.1.3 Corpora

Krantenartikelen werden gehaald uit het online bestand van de Lexis Nexis Krantendatabase, toegankelijk voor geregistreerde leden via de Nederlandse Koninklijke Bibliotheek (<http://academic.lexisnexis.nl.access.authkb.kb.nl/kb/>). Krantenartikelen werden geselecteerd op basis van de volgende zoekcriteria:

- Alleen dagelijkse kranten (uitgaven tenminste elke werkdag)
- Alleen nationale kranten (geen regionale of lokale kranten)
- Geen Vlaamse of Friese kranten

De analyse werd op drie deelcorpora gericht: een tekstbestand bestaande uit krantenartikelen over moslims en/of islam exact één jaar vóór de moord op Theo van Gogh (Deelcorpus 1, 03-11-2003–02-11-2004), een tekstbestand bestaande uit krantenartikelen exact één jaar ná de moord (Deelcorpus 2, 03-11-2004–02-11-2005) en een tekstbestand bestaande uit krantenartikelen op lange termijn, ongeveer vijf jaar, ná de moord (Deelcorpus 3, 01-01-2009–31-12-2009)⁹.

De zoekcriteria leverden acht krantentitels op: *Trouw*, *De Volkskrant*, *NRC Handelsblad*, *Het Parool*, *Algemeen Dagblad*, *De Telegraaf*, *Metro* en *Spits!*. (De gratis kranten *Metro* en *Spits!* waren alleen verkrijgbaar voor Deelcorpus 3 omdat de Krantendatabase geen teksten van deze dagbladen uit de jaren van de eerste twee deelcorpora bevat.) Het onderzoeksmateriaal werd verder beperkt door de genoemde zoekstrings *moslims* en *islam* (met Booleaanse zoektermen getypt *moslims* OR *islam*). Het aantal krantenartikelen in elk deelcorpus wordt in Tabel 1 weergegeven.

Opmerkingen aangaande de selectie van zoekstrings en het bronmateriaal zijn:

- geen Vlaamse dagbladen te includeren omdat de gevolgen van de moord op Van Gogh vooral Nederland als (nationaal) politiek en maatschappelijk gebied raken.
- *moslims* in het meervoud gebruikt om expliciet de groep moslims en niet *moslim* op individueel niveau te betekenen.
- het zoekterm *van Gogh* alleen te includeren voor Deelcorpus 1, aangezien de doelstelling van het onderzoek om het algemene taalgebruik over *moslims* en *islam* te bestuderen. Het zou anders een risico zijn, dat de massale berichtgeving ná de moord eigenlijk de waarderingen van de moord zelf (als steeds negatieve gebeurtenis), eerder dan algemene veranderingen in het gebruik van *moslims* en *islam* weerspiegelt.
- door technische redenen kon helaas de volledige artikelbestanden alleen voor Deelcorpus 1 en Deelcorpus 3 geanalyseerd en gedownload worden. Deelcorpus 2 bestaat uit het maximum, op het moment van zoeken, aantal displaybare en door de zoekmachine gelabelde “meest relevante” resultaten, beperkt door een bovengrens van 1000 artikelen.¹⁰

⁹ Het jaar 2009 werd gekozen omdat het, op het moment van schrijven, het meest recente, complete jaar van krantenartikelen in de Lexis Nexis Krantendatabase was. Ook al dit niet exact vijf jaar naar de moord was, maar wel krantenartikelen rond en inclusief de vijfjaardag op 02-11-2009 bevat, verwijs ik ook naar Deelcorpus 3 als “vijf jaar ná de moord”.

¹⁰ Oorspronkelijk bevatte de zoekmachine een functie om alle (>5000) zoekresultaten automatisch uit de database te halen. Bij het downloaden van Deelcorpus 2, enige weken daarna, was een technisch-administratieve beperking van de database in werking getreden, die een maximum downloaden van 10 artikelen per keer toelaat. Honderd aparte downloads werden daarom uitgevoerd om Deelcorpus 2 te samenstellen; meer leek niet praktisch uitvoerbaar.

Tabel 1. Overzicht van het onderzoeksmateriaal uit de Lexis Nexis Krantendatabase.

	Aantal krantenartikelen in deelcorpora		
	Deelcorpus 1 Vanaf 3/11/2003 tot en met 2/11/2004	Deelcorpus 2 Vanaf 3/11/2004 tot en met 2/11/2005	Deelcorpus 3 Vanaf 1/1/2009 tot en met 31/12/2009
Zoekstringen	<i>moslims OR islam</i>	<i>moslims OR islam AND NOT van Gogh</i>	<i>moslims OR islam AND NOT van Gogh</i>
Naam bron			
Trouw	1264	268	1297
De Volkskrant	1089	239	992
NRC Handelsblad	1229	187	943
Het Parool	702	133	577
AD/Algemeen Dagblad	622	93	438
De Telegraaf	420	74	391
Metro (NL) ***	0	0	225
Spits! ***	0	0	293
Som aantal artikelen	5326*	994**	5156*

* totaal aantal documenten uit de periode; ** “meest relevante resultaten” volgens de zoekmachine

*** De Krantendatabase bevat geen teksten van *Metro* of *Spits!* uit de jaren 2003–2005.

4.2 Procedure

De verzameling van data werd in de volgende zes stappen uitgevoerd:

1. De Nederlandse taalruimte creëren om het LSA-algoritme op het Nederlands te trainen: formatering van het TwNC in het programma *Semantic* (volgens 4.1.1).
2. Een lijst met positieve en negatieve woorden van moedertalige Nederlanders samenstellen (volgens 4.1.2 en Appendix).
3. “Eigen” tekstcorpus oprichten door het downloaden van relevante krantenteksten uit de Lexis Nexis Krantendatabase (volgens 4.1.3): drie verschillende tekstbestanden (Deelcorpus 1–3) werden geformateerd voor bewerking in *Semantic*.
4. Toepassing van het LSA-algoritme in *Semantic* op de deelcorpora ten opzichte van de valentiewaarden van stap (2); dit resulteert in een valentiewaarde per deelcorpus.
5. Lijsten van kenmerkende woorden voor elk deelcorpus opmaken. De functie “Semantic keywords” (eigenlijk: semantische sleutelwoorden) in *Semantic* laat zien welke woorden het vaakst in elk geanalyseerd tekstbestand voorkomen, in dit geval de woorden die het vaakst in co-texten met *moslims* en *islam* genoemd worden, in Deelcorpus 1, 2 en 3 respectievelijk, met Deelcorpus 1 als referentiepunt.
6. Statistische testen uitvoeren om de significantie van de verkregen verschillen te evalueren (zie details beneden).

4.3 Statistische tests

De focus van de huidige studie ligt niet op statistiek; wel zijn een paar opmerkingen nodig om de validiteit van de LSA-analyse te controleren. Het toegepaste LSA-algoritme is gevalideerd in meerdere, peer-reviewed studies op het Zweeds en Engels (b.v. Arvidsson, Sikström & Werbart, 2011; Gustafson & Sikström, 2008; Gustafson & Sikström, 2011, Sikström & Garcia, 2010; Willander & Sikström, 2011). Alle tests zijn in persoonlijke communicatie met Sverker Sikström, als ontwikkelaar van het programma, uitgevoerd. Om de huidige studie voor een grotere groep taalonderzoekers toegankelijk te maken, heb ik bewust voor een eenvoudige statistische evaluatie en een nadruk op de conceptuele discussie gekozen.

Een belangrijke voorwaarde voor een zinvol LSA-gebaseerde waarderingsanalyse is dat de positieve/negatieve woorden herkenbaar zijn in het geanalyseerde materiaal. Toegepast, als het LSA-algoritme goed werkt, zullen de positieve bronwoorden in het Appendix ook geïdentificeerd worden als positieve woorden in de krantenteksten (ideaal met een valentiewaarde van +1, bijvoorbeeld, als het woord *baby* in dezelfde context met alle bronwoorden van de “positieve lijst” zou voorkomen). Ook, voor de negatieve bronwoorden, zou op basis van de onderlinge gevonden associaties in de krantenteksten een ideale valentiewaarde van -1 berekend worden (bijvoorbeeld, als het woord *boos* in dezelfde context met alle andere woorden van de “negatieve lijst” zou voorkomen).

In feite werd de correlatie tussen de gehypotiseerde valentiewaarden (b.v. *boos* = -1) en de geprediceerde (volgens het LSA-algoritme berekende) valentiewaarden in *Semantic* berekend tot 0,675. Dit is, op basis van eerdere studies, relatief hoog en indiceert dat het ontwikkelde LSA-algoritme over het algemeen goed werkt om de valentie van woorden in onbekend materiaal te beoordelen (S. Sikström, persoonlijke communicatie, 12 september 2011).

Na verwerking van de data in *Semantic* werden de resultaten geïmporteerd voor statistische analyse in *SPSS Statistics 17.0*. Conventionele *t*-tests werden gekozen om de verschillen in valentiewaarden tussen de deelcorpora te testen, in de eerste plaats tussen Deelcorpus 1 en 3 (de grootste afstand representerend en de grootste tekstbestanden vóór en ná de moord). De gekozen significantiegrens was $p = 0,05$.

5 Resultaten

5.1 Valentiewaarden van krantenartikelen over *moslims* en *islam* vóór en ná de moord

Op basis van de positieve (+1) en negatieve (-1) bronwoorden (in Appendix) genereerde het LSA-algoritme een valentiewaarde voor elk krantenartikel in het geanalyseerde corpus. De valentiewaarden waren tussen -2,32 en +11,12. De gemiddelde valentiewaarde was 6,89 voor Deelcorpus 1, 7,29 voor Deelcorpus 2 en 7,23 voor Deelcorpus 3. Figuur 1 illustreert deze data op een tijdslijn die suggereert dat *moslims* en *islam* in vergelijking met het tijdstip vóór de moord, in hoger gewaardeerde co-texten optreden het jaar direct ná de moord, en later (2009), ongeveer vijf jaar ná de moord, in iets lager gewaardeerde co-texten optreden.

Een eenweg-ANOVA van de drie deelcorpuswaarden liet zien dat er statistisch hoog significante valentiever verschillen waren tussen de deelcorpora; $F(2, 11473) = 54,520$; $p < 0,001$. Een *t*-test liet zien, specifiek, dat zelfs de co-texten op lange termijn ná de moord (Deelcorpus 3) significant meer positief waren dan de co-texten vóór de moord (Deelcorpus 1); $t(1, 10480) = 9,512$; $p < 0,001$. Er was echter geen significant verschil tussen de valentiewaarden van de laatste twee deelcorpora, Deelcorpus 2 en Deelcorpus 3; $t(1,6148) = 1,025$; $p = 0,305$. Dus, de valentiewaarden van co-texten ná de moord waren significant meer positief dan vóór de moord, maar er was geen significant verschil tussen de twee valentiewaarden ná de moord.

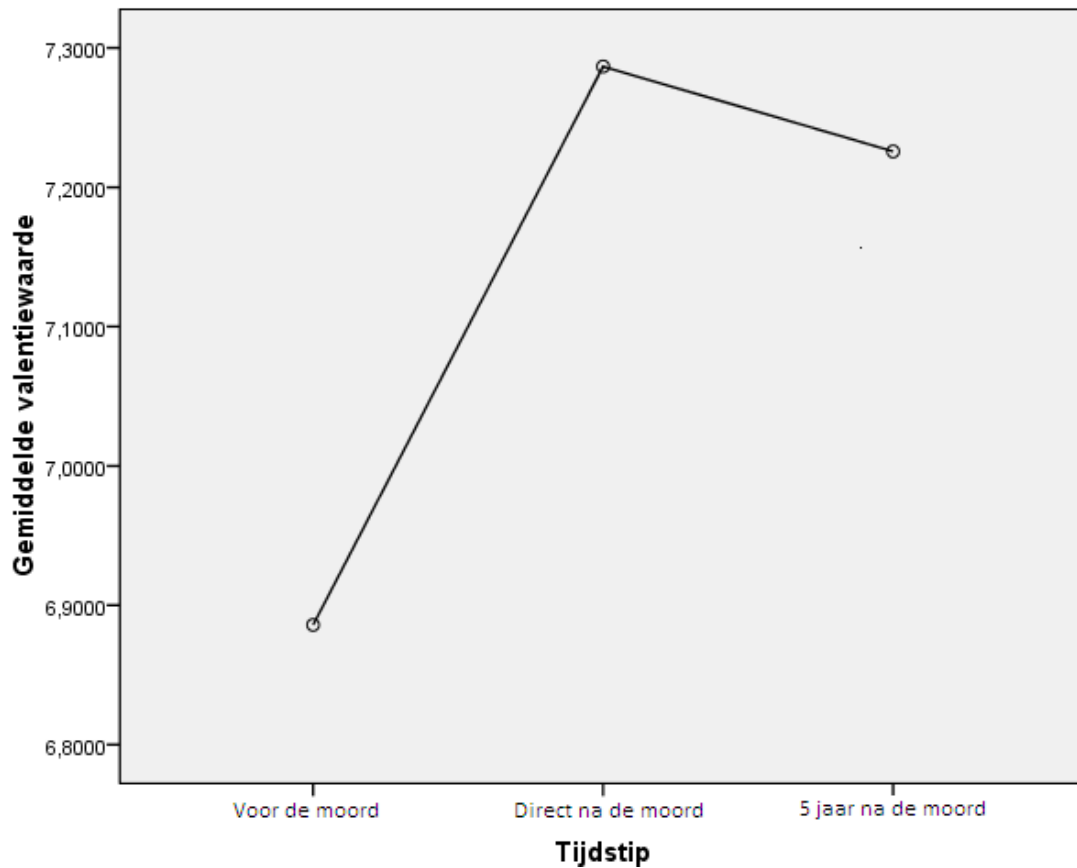





Fig. 1. Valentiewaarden van krantenteksten met woorden islam en moslims uit drie tijdstippen rond de moord op Theo van Gogh.

5.2 Kenmerkende woorden van krantenartikelen over *moslims* en *islam* vóór en ná de moord

De functie “Semantic keywords” (“semantische sleutelwoorden” ofwel kenmerkende woorden, in 4.2 beschreven) in *Semantic* werd toegepast op drie aparte vergelijkingen van deelcorpora. Ten eerste was het doel te zien welke woorden het vaakst voorkwamen in het deelcorpus vóór de moord op Van Gogh in vergelijking met de krantenteksten vijf jaar daarna (d.w.z. Deelcorpus 1 vs Deelcorpus 3). Dus, woorden die alleen in het laatste deelcorpus voorkwamen of in zowel Deelcorpus 1 als Deelcorpus 3 voorkwamen, werden in de eerste vergelijking geëxcludeerd. Daarna werden de vaakst voorkomende woorden in het geanalyseerde materiaal door het programma opgesomd. Om de kenmerkende woorden van Deelcorpus 3 te identificeren werd dezelfde analyse in de omgekeerde orde uitgevoerd (d.w.z. Deelcorpus 3 vs Deelcorpus 1). Tenslotte werd Deelcorpus 2 (dat minder artikelen bevatte) op dezelfde manier met Deelcorpus 1 vergeleken.

Tabel 2 toont de twintig vaakst voorkomende woorden van elk deelcorpus in vergelijking met Deelcorpus 1 als referentiepunt. Eigennamen die samen horen maar op verschillende plaatsen in de frequentielijst verschijnen (b.v. *Geert + Wilders*, *Said + Qutb*), worden samengevoegd en als één woord beschouwd. Deze en andere verduidelijkingen worden tussen haakjes aangegeven.

Tabel 2. De 20 meest kenmerkende woorden voor elk deelcorpus over *moslims* en *islam*.

<p>A. Deelcorpus 1 (vóór de moord) vs. Deelcorpus 3 (vijf jaar ná de moord)</p> 	<p>B. Deelcorpus 2 (direct ná de moord) vs. Deelcorpus 1 (vóór de moord)</p> 	<p>C. Deelcorpus 3 (vijf jaar ná de moord) vs. Deelcorpus 1 (vóór de moord)</p> 
<p>Irak Hirsi [Ali] aanslagen Iraakse Saddam [Hussein] terrorisme Turkije Aivd [Algemene Inlichtingen- en Veiligheidsdienst] Bush terroristen Bin [Laden] Saoedi-Arabië Madrid artsen Irakezen september Amerikanen terreur Cliteur Bagdad</p>	<p>Koran imams islamitische [de] Telegraaf profeet Jason [W] [Said] Qutb Mohammed [de] paus Leeds Londen islamdebatten radicale Nussimbaum Saadawi Hofstadgroep Selim geloof Karimov Leonard [Geluk]</p>	<p>[Geert] Wilders PVV Obama taliban [Ahmed] Marcouch Ramadan partij Somalië [de] paus [Ali] Eddaoudi [Eberhard Van der] Laan Ahmadinejad Oeigoeren PvdA PVV-leider Slotervaart Iraanse Groot-Brittannië financiële België</p>

Noot: Het getinte deel van de cirkels illustreert welke deelverzameling van twee vergeleken deelcorpora de woorden representeren.

5.3 Een poging tot kwalitatieve validatie

Zijn de resultaten uit 5.1-5.2 kwalitatief aanvaardbaar? Met andere woorden, komt de output van de gecomputeriseerde, kwantitatieve analyse overeen met een interpreterende, kwalitatieve analyse van hetzelfde materiaal? Om een eerste indicatie daarvan te krijgen deed ik een poging om de data te relateren tot de inhoud van de krantenteksten, door een beperkt (praktisch hanteerbaar) aantal krantenartikelen met de hoogste en laagste valentiewaarden na te lezen.

De deelverzameling krantenartikelen die door LSA de hoogste valentiewaarden (dus meest positieve waarderingen) kregen, bevatte een artikel over goedkoop leven en dat de islam aalmoezen bevordert (“De kunst van het klaplopen”, AD/Algemeen Dagblad, 21 november 2009), een artikel over de vreedzame aard van het boeddhisme (“Zen geeft zin”, NRC Handelsblad, 13 juni 2009) en een portret over politicus Judith Belinfante die “in de Kamer zat, samen met Joden, moslims, Palestijnen” (“Laveren tussen kunst en Kamer”, Trouw, 7 januari 2004). De krantenartikelen met de laagste valentiewaarden (dus meest negatieve waarderingen) bevatte een korte tekst over geweld in een moskee (“Al Aqsa Israëlische oproerpolitie bestormt heilige moskee”, De Volkskrant, 26 oktober 2009), een artikel over vermoorde moslims in Srebrenica (“Serviërs in Bosnië geven moorden Srebrenica toe”, De

Volkscrant, 12 juni 2004) en een artikel over botsingen tussen christenen en moslims in Nigeria (“Tientallen doden bij onlusten”, AD/Algemeen Dagblad, 4 mei 2004). De onderwerpen van de hoogst en laagst gewaardeerde artikelen (co-texten) lijken dus redelijk te overeenkomen met positieve en negatieve berichten – ook al zou het moeilijk zijn de criteria voor een meer fijnkorrelige categorisatie of onderlinge rangschikking van artikelen in dit opzicht te specificeren.

De semantische sleutelwoorden lijken ook relevant te zijn voor de berichtgeving over moslims en de islam in Nederland uit deze periode. Wel lijken de woorden van elk deelcorpus rond bepaalde thema’s te draaien. Veel woorden zijn namen van plaatsen en personen. Krantenteksten van vóór de moord bevatten meer duidelijk negatieve woorden (*terrorisme, aanslagen, terreur*) dan daarna. Ook zijn er meer associaties met het buitenland (*Irak, Turkije, Bush, Saoedi-Arabië, Madrid*).

Het jaar direct ná de moord (Deelcorpus 2) bevat woorden met religieuze associaties, vooral met betrekking tot de islam (*Koran, imams, islamitische, Said Qutb, Mohammed* maar ook *de paus, geloof*). Een opvallend onderwerp is de terreuraanslag in Londen op 7 juli 2005 (*Londen, Leeds*) maar ook verschijnselen binnen Nederland (de krant *De Telegraaf*, de terroristenorganisatie *Hofstadgroep*, CDA-politicus *Leonard Geluk*) worden vermeld. Deelcorpus 3 wordt vervolgens gekarakteriseerd door woorden met nadere betrekking tot het Nederlandse politieke debat: vooral de islam-kritische *Geert Wilders* (ook *PVV-leider, PVV*) en zijn politieke tegenstanders binnen de *PvdA*, *Eberhard Van der Laan* en *Ahmed Marcouch* (ook bekend als de “Sheriff van Slotervaart”, de Amsterdamse wijk waarvan hij in 2009 stadsdeelvoorzitter was).

Hoe de semantische sleutelwoorden relateren tot de berekende valentie en de grotere context van het taalgebruik over moslims en islam wordt een kwestie van discussie die de verdere interpretatie van de resultaten beneden volgt. Als samenvatting tot dusver, zijn de volgende tentatieve antwoorden op de onderzoeksvragen gevonden:

1. De LSA-analyse geeft zeer significante verschillen betreffende de valentiewaarden van krantenteksten rond *moslims* en *islam* vóór en ná de moord op Theo van Gogh.
 - a) Krantenartikelen ná de moord hebben over het algemeen hogere valentiewaarden, dus *moslims* en *islam* worden daar positiever gewaardeerd, dan vóór de moord.
 - b) De valentiewaarden zijn significant hoger zowel direct ná de moord (2004) als vijf jaar daarna (2009), maar de jaren 2004 en 2009 onderscheiden zich ten opzichte van valentie niet significant van elkaar (ook al wijzen de data naar lagere valentie vijf jaar ná de moord).
2. Tabel 2 vat de woorden samen die kenmerkend zijn voor de krantenartikelen met betrekking tot *moslims* en *islam* vóór en ná de moord. Vóór de moord zijn *moslims* en *islam* vooral geassocieerd met buitenlandse verschijnselen en eigennamen van politici en debaters. Het jaar direct ná de moord zijn er meer associaties met religieuze en politieke leiders en debaters. Vijf jaar ná de moord verwijzen de kenmerkende woorden meer specifiek naar een Nederlandse politieke context.

6 Discussie

De scriptie begon met de algemene vraag hoe maatschappelijke gebeurtenissen het taalgebruik kunnen beïnvloeden. De moord op Theo van Gogh is in veel opzichten een exceptionele gebeurtenis die een nadere kijk op dit verband toelaat. Omdat de moord gepleegd werd door iemand met een (vermeende) islamitisch-fundamentalistische agenda is mijn focus gericht op het gebruik van de woorden *moslims* en *islam*. De massale berichtgeving in Nederlandse dagbladen vóór en ná de moord vormt een uitstekend bron van relevante informatie die – wegens de omvang op zich – een traditionele kwalitatieve analyse moeilijk maakt maar met een gecomputeriseerde methode als LSA wel aan te pakken is. Is dan de berichtgeving over moslims en de islam veranderd als gevolg van de moord?

Het antwoord was (voor mij) onverwacht: de resultaten geven aan dat moslims en islam in *meer positief* gewaardeerde contexten voorkomen ná de moord dan daarvoor. Dit lijkt zo te zijn zowel op korte als op lange termijn na de gebeurtenis. Waarschijnlijk is er zelfs een oorzakelijk verband, want de emotionele lading van de woorden gaat duidelijk omhoog in de berichtgeving van de periode onmiddellijk ná de moord. Duidelijk is ook dat *moslims* en *islam* in veranderde thematische contexten optreden. Maar hoe verregaande conclusies kunnen worden getrokken op basis van deze observaties? Zijn de waarderingen van moslims en de islam echt veranderd in het algemene Nederlandse taalgebruik en, wellicht nog belangrijker, betekent dit dat Nederlanders echt een positiever beeld van moslims hebben gekregen? Zo ja, dan zijn de meningen dat er een sterker wantrouwen tegen moslims in de maatschappij heerst misschien overdreven.

Direct moet gezegd worden dat deze vragen niet definitief kan beantwoorden. Wel wijzen de vragen op de waarde van een scherpere controle op de beperkingen en verdiensten van de onderzoeksmethode. Ten eerste, er werd benadrukt dat LSA de co-texten van islamgerelateerde krantenteksten, eerder dan de volle maatschappelijke context, rond het tijdstip van de moord analyseert. Alles komt natuurlijk niet in druk.

Ten tweede, waarderingen zijn niet binnen hetzelfde soort co-texten geanalyseerd, dus geen conclusies kan getrokken worden of moslims positiever gewaardeerd zijn specifiek met betrekking tot, bijvoorbeeld, berichten over immigratie, over godsdienst, over arbeid of andere specifieke maatschappelijke domeinen.¹¹ Noch zegt het resultaat iets over de emotionele waarde van de woorden *moslims* en *islam* op zich; de valentiewaarde verwijst alleen maar naar de krantenartikelen (dus co-texten) in welke die woorden genoemd worden. Bovendien, potentieel relevante teksten die een indirecte invloed op het beeld van moslims zouden kunnen hebben, zoals over immigratiebeleid, Marokkanen (een gestigmatiseerde groep in deze context) of allochtonen over het algemeen, maar niet *moslims* of *islam* bevatten, zijn niet in de analyse inbegrepen.

Welke zijn dan de voornaamste verdiensten van het onderzoek? Ten eerste, de resultaten komen goed overeen met de theoretische doelstelling van LSA om “latente” waarderingen te ontmaskeren. De basis van enkele emotionele woorden (zie Appendix) lijkt voldoende om een geautomatiseerde, kwalitatief goede waardering te maken van een reeks verschillende onderwerpen in tienduizenden krantenartikelen. Dit wordt geïllustreerd door het feit dat artikelen over aalmoezen en vrede hoge waarden kregen terwijl artikelen over geweld en oorlog lage waarden kregen. De resultaten hebben een hoge graad van soliditeit en objectieve

¹¹ Met andere woorden, de co-text wordt, net als de valentiewaarden, als een afhankelijke variabele behandeld.

controleerbaarheid want iedereen met toegang tot hetzelfde materiaal en programma zou hetzelfde uitslag bereiken. Ook al gaat LSA letterlijk over “semantische analyse”, meen ik dat de methode het allerbest van toepassing komt om latente *waarderingen* aan te tonen. Want waarderingen zijn een complex sociaal fenomeen dat niet gewoon uitgedrukt wordt als “wij zijn goed” en “zij zijn slecht”, maar door meer subtiele contexten verspreid worden. LSA lukt het deze subtiele, ofwel latente, contexten expliciet te maken en kwantitatief te beoordelen.

Ten tweede, waarschijnlijk gaat een verandering in gebruik van een woord – weerspiegeld in de verschillende co-texten en hun semantische sleutelwoorden – een verandering van betekenis vooraf. Dus, ook al raakt de analyse niet de waardering of betekenis van *moslims* en *islam* direct, toch zou het resultaat een eerste indicatie op zo’n talige ontwikkeling kunnen zijn. De ontwikkeling van gebuiksdomeinen hoort ook niet noodzakelijk samen met de ontwikkeling van sociale waarderingen. In feite, als er een trend van afnemende valentiewaarden op langere termijn ná de moord blijkt (wat de data aanduiden maar hier niet statistisch vastgesteld kon worden) zou het een teken van normalisatie kunnen zijn. Dat wil zeggen, het is mogelijk dat de aparte invloed van de moord op lange termijn verdwijnt en dat de co-texten dan weer de oorspronkelijke valentiewaarden aannemen. Aan de andere hand, het zou onwaarschijnlijk zijn dat de gebuiksdomeinen en kenmerkende woorden naar het vorige taalgebruik terugkeren.

Er werd meerdere keren in deze scriptie benadrukt dat goede maatschappelijke, culturele en historische kennis nodig is om de resultaten zin te geven. Een inzicht van het onderzoek is dat hoewel veel objectieve data (cijfers en woordenlijsten) door LSA gegenereerd kunnen worden, de bruikbaarheid daarvan wordt toch beperkt door de kennis en interpretatievermogen van een menselijke vertolker – en daarmee ook een mate van subjectiviteit. Gewoon een bevredigende uitleg over de context van de gehaalde sleutelwoorden zou een overweldigende taak zijn.

Een voorbeeld van dit laatste punt is waarom *De Telegraaf* bij de meest genoemde woorden in het tweede deelcorpus staat. Men zou eerst de algemene achtergrondkennis moeten hebben dat *De Telegraaf* de aard van een sensatieblad heeft met een politieke tendens naar rechts en meer dan enkele keren met uitdagende koppen in islamdebatten meedoet. Specifiek zou men moeten weten dat de krant in 2005 een veelbesproken (valse) dreigmail over een aanstaande islamitische terreuraanslag ontving, wat ook in andere dagbladen publiciteit kreeg. Tevens, de aanslagen in Londen in dezelfde periode zijn wel bekend, maar misschien minder bekend is de toen zware kritiek tegen de islamofobische Oezbeeks dictator *Karimov* vanwege zijn schending van mensenrechten (met name tegen de moslims), wat verklaart waarom zijn naam ook op de lijst staat. Kortom, de grote reeks van co-texten eist grote kennis.

Een iets meer speculatieve conclusie is dat de verandering van valentiewaarden een soort emotionele reactie weerspiegelt, die een verandering van woordgebruik katalyseert en/of specificeert. De resultaten wijzen op een relatief duidelijke overgang van de aanwezigheid van *moslims* en *islam* in buitenlandse contexten vóór de moord, naar meer religieuze contexten direct ná de moord en later naar een meer specifieke binnenlandse, politieke context (vooral rond Geert Wilders, zijn partij en antagonisten). De stijgende valentiewaarden zouden dan kunnen worden verklaard doordat *godsdiens*t en *binnenland* over het algemeen positiever geladen domeinen zijn dan *buitenland*, hoewel noch George W Bush, de Hofstadgroep noch Geert Wilders in een positief licht verschijnen. Indien moslims en de islam ook positiever worden beschouwd vanwege duidelijkere associaties met een Nederlandse “thuis”-context is nog een open vraag. Een interessante gedachte is dat de veranderingen in zowel valentie als

onderwerpen van de co-texten meehelpen in de ontwikkeling van de “common ground” van de taalgebruikers: met de tijd veranderen de referentiekaders door de contextverschillen en uiteindelijk ook de lexicale betekenissen van het woord in gebruik.

Ten laatste, de voornaamste bijdrage van deze studie is te hebben gedemonstreerd hoe LSA op een vruchtbare wijze toegepast kan worden in het Nederlands om latente waarderingen in authentieke, massale tekstcorpora te analyseren. Ik rond af door een aantal punten voor te stellen over hoe men met deze methode in toekomstig onderzoek verder kan gaan:

- Een herhaalde LSA-analyse op basis van zinnen, in plaats van artikelen, als de eenheid van analyse zou een betere indicatie kunnen geven van hoe belangrijk een samenhangende co-text – dus de nabijheid van associaties – is voor het resultaat. Het is ook mogelijk de wiskundige premissen van het LSA-algoritme, zoals het aantal dimensies van de semantische ruimte, in dit opzicht aan te passen.
- Een verdere kwalitatieve validatie zou gedaan kunnen worden door een grotere aantal (willekeurig) geselecteerde krantenartikelen door te lezen en de teksten laten scoren, graag door meerdere personen, ten opzichte van emotionele valentie, zonder kennis van de berekende waarde door LSA. Het resultaat zou een betere indicatie geven van hoe goed de gecomputeriseerde beoordeling overeenkomt met die van een menselijke taalgebruiker.
- Een nadere kwalitatieve analyse (interpretatie van de resultaten in een sociale en culturele context) zou bevorderd kunnen worden door het categoriseren en sub-categoriseren van de kenmerkende woorden, om bijvoorbeeld bepaalde inhoudelijke of psychologische themata te identificeren.
- Betreffende het voorbeeld van exceptionele maatschappelijke (nieuws)gebeurtenissen die het taalgebruik beïnvloeden, zou het zeer leerzaam zijn een conceptuele replicatie (vergelijking) van de huidige studie uit te voeren in andere relevante talen en contexten. Waarderingen van *moslims* en *islam* zijn relevant ook met betrekking tot de ontwikkeling in de Verenigde Staten vóór en ná 9/11 (2001). Een ander exceptionele gebeurtenis op een specifieke datum die een toekomstige analyse mogelijk maakt – ten opzichte van zijn invloed op bepaalde onderwerpen zoals *democratie* en *extremisme* – is de terroristendaad en massamoord door Anders Behring Breivik in Noorwegen op 22 juli 2011.

Er zijn dus meerdere mogelijkheden om de hypothesen en methodologie van de huidige studie te testen en voor andere onderwerpen aan te passen. Een allerlaatste reflectie is wat het zou betekenen het tegenovergestelde uitgangspunt van het onderzoek te kiezen: Zou een bewuste verandering van het taalgebruik ook maatschappelijke waarderingen kunnen veranderen? Als de studie juist is op het punt dat waarderingen cognitief geconstrueerd worden op basis van de contexten die wij in verschillende teksten aflezen – waarom zou dan ook niet een verandering van hoe teksten geschreven worden een verandering van waarderingen met zich mee kunnen brengen? Wetenschappelijk zou zo’n oorzaaksverband moeilijk vast te stellen zijn. Maar als een minder geladen taalgebruik in bepaalde domeinen tot een minder vijandig of gewelddadig maatschappij kan leiden, blijft dit een vraag van belang – zowel voor taalonderzoekers als voor iedere verantwoordelijke taalgebruiker.

7 Referenties

7.1 Internetbronnen

Lexis Nexis Krantendatabase: <http://academic.lexisnexis.nl.access.authkb.kb.nl/kb/>

LSALAB, Sverker Sikström: http://www.lu.se/sverker.sikstrom/LSALAB_intro.html

7.2 Boeken en artikelen

Allén, S., Gellerstam, M., & Malmgren, S-G. (1989). *Orden speglar samhället*. Stockholm, Allmänna förlaget.

Anglin, J. M. (1993). Vocabulary development: A morphological analysis. *Monographs of the Society for Research in Child Development*, 58 (10).

Appel, R., Hubers, G., & Meijer, G. (1981). *Sociolinguïstiek* (4e ed.). Utrecht, Spectrum.

Arvidsson, D, Sikström, S., & Werbart, A. (2011). Changes in self and object representations following psychotherapy measured by a theory-free, computational, semantic space method. *Psychotherapy*, 21, 430-446.

Barreto, M., Ryan, M. K., & Schmitt, M. T. (Eds.). (2009). *The Glass Ceiling in the 21st Century: Understanding Barriers to Gender Equality*. Washington DC, VS, American Psychological Association.

Bergs, A. & Diewald, G. (2009). *Contexts and constructions*. Amsterdam, John Benjamins.

Bolland, G. J. P. J. (1909). *Zuivere rede en hare werkelijkheid: Het boek der spreuken*. Leiden, Adriani.

Boon, T. (2007). *Ik googel, jij googelt: Engelse werkwoorden in het Nederlands*. Nijmegen, BnM.

Boves, T., & Gerritsen, M. (1995). *Inleiding in de sociolinguïstiek*. Utrecht, Spectrum.

Buruma, I. (2006). *Murder in Amsterdam: The death of Theo Van Gogh and the limits of tolerance*. Londen, Atlantic books.

Chomsky, N. (1957). *Syntactic Structures*. Den Haag, Mouton.

Dirven, R., & Verspoor, M. (2001). *Cognitieve inleiding tot taal en taalwetenschap*. Leuven, Uitgeverij Acco.

Garcia, D., & Sikström, S. Quantifying the Semantic Representations of Adolescents' Memories of Positive and Negative Life Events. *In progress*

Goldberg, A. (2003). Constructions: a new theoretical approach to language. *TRENDS in Cognitive Sciences*, 7 (5), 219-224.

- Gustafsson, M. & Sikström, S. (2008). Latent semantic analysis reveals higher valence for ingroups than outgroups. *International Journal Of Psychology*, 43, 621-621.
- Gustafsson, M. & Sikström, S. (2011). Ingroup allocation model - Redistributing resources through language to improve fitness. *Submitted*
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211-240.
- Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to Latent Semantic Analysis. *Discourse Processes*, 25, 259-284.
- Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (2007). *Handbook of latent semantic analysis*. Mahwah, New Jersey, Lawrence Erlbaum.
- Langacker, R. W. (1987/1991). *Foundations of Cognitive Grammar (Vols I & II)*. Stanford, California, Stanford University Press.
- Martin, D., & Berry, M. (2007). Mathematical foundations behind latent semantic analysis. In: Landauer et al. (red.), 35–55.
- Nijhof, G. (2003). *Tekstsociologie: over de talige constructie van de sociale werkelijkheid*. Amsterdam, Aksant.
- Sikström, S., & Garcia, D. (2010). Studying Subjective Well-being by Quantifying the Semantic Content of Hedonic Autobiographical Memories. *The 10th ISQOLS International Conference - Understanding Quality of Life and Building a Happier Tomorrow*, 8-11 december, Bangkok, Thailand.
- Sikström, S., & Schmer-Galunder, S. (2007). Measuring Value Systems in Large Text Corpora. *Poster gepresenteerd bij Psychology and Social Justice Conference*. New York University, New York, April 20, 2007.
- Van Gogh, T. (2003). *Allah weet het beter*. Amsterdam, Uitgeverij Xtra.
- Van der Plank, P. (1985). *Taalsociologie: een inleiding tot de rol van taal in het maatschappelijk verkeer*. Muiderberg, Coutinho.
- Van Walsum, S. (2011). *Een welkom geluid*. De Volkskrant, 05-01-2011; gedownload van http://opinie.volkskrant.nl/artikel/show/id/7536/Een_welkom_geluid
- Verhagen, A. (2005). Constructiegrammatica en 'usage based' taalkunde. *Nederlandse Taalkunde*, 10, 197–222.
- Willander, E., & Sikström, S. (2011). What is Sacred in a Space of Numbers?. In Linda, W. (Red.) *Innovative Methods in the Study of Religion*. Oxford University Press.
- Zuiderveld, R. (2010). *Fallet Wilders: holländarna, islam och den mångkulturella drömmens död*. Stockholm, Timbro.

Appendix: lijst van woorden met positieve/negatieve lading

Tien volwassen moedertalige Nederlanders beantwoordden de volgende instructies:

Schrijf a.j.b. zoveel positieve en negatieve woorden (graag in twee kolommen) die je kan bedenken tijdens 4-5 minuten. Het kan wat voor woord dan ook zijn en je hoeft de woorden in geen bepaalde volgorde te schrijven. Wel is het belangrijk dat je niet te veel nadenkt – noteer gewoon wat spontaan in je opkomt – en dat het woord voor jou een duidelijke positieve óf negatieve lading heeft.

Alleen woorden die meer dan een keer werden genoemd (dus door meer dan een persoon) werden in het onderzoek geïncludeerd. Eigennamen (van personen) werden geëxcludeerd. Sommige woorden kwamen drie, vier of vijf keer voor, maar werden niet onderling gegradeerd. Ieder negatief/positief woord kreeg een waarde van -1/+1 toebedeeld, respectievelijk.

Positieve woorden (+1)	Negatieve woorden (-1)
BABY	BOOS
BINNEN	DOOD
BLIJ	IRRITANT
BLOEMEN	KOUD
ETEN	LELIJK
FANTASTISCH	MOE
FEEST	OORLOG
GELUK	PIJN
GEWELDIG	REGEN
GEZELLIG	ROUW
GOED	SAAI
GRAAG	SLECHT
GROEN	STRESS
HOOP	VERVELEND
HUIS	VIES
KOKEN	ZIEK
LACHEN	
LEKKER	
LEZEN	
LICHT	
LIEFDE	
MOOI	
MUZIEK	
PRIMA	
REIZEN	
RING	
RUST	
SAMEN	
THUIS	
VAKANTIE	
WARM	
VREDE	
VROLIJK	
ZACHT	
ZON	