

Tilburg University

Hoe makkelijk is een niet moeilijke tekst?

Kamoen, N.; Holleman, B.; van den Bergh, H.

Published in:
Tijdschrift voor Taalbeheersing

Publication date:
2007

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):
Kamoen, N., Holleman, B., & van den Bergh, H. (2007). Hoe makkelijk is een niet moeilijke tekst? Een meta-analyse naar het effect van vraagformulering in tekstevaluatieonderzoek. *Tijdschrift voor Taalbeheersing*, 29(4), 314-332.

General rights

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

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

Take down policy

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

Hoe makkelijk is een niet moeilijke tekst?

Een meta-analyse naar het effect van vraagformulering in tekstevaluatieonderzoek

1. Inleiding^{1,2}

“Do you think the government should *allow public speeches against democracy?*” Deze vraag stelde Donald Rugg (1941) aan ongeveer 1300 Amerikanen. Maar liefst 75% van de respondenten antwoordde ‘no’. Rugg stelde 1300 andere mensen de vraag: “Do you think the government should *forbid public speeches against democracy?*” Logischerwijs had ongeveer driekwart van de ondervraagden die vraag met ‘yes’ moeten beantwoorden. ‘To forbid’ en ‘to allow’, ofwel ‘verbieden’ en ‘toelaten’, zijn immers elkaars tegengestelden. Deze logica bleek echter niet op te gaan. Slechts 54% van de ondervraagden antwoordde ‘yes’ op de vraag met ‘verbieden’. Rugg concludeerde dat formulering van de vraag effect heeft op de antwoorden die respondenten geven.

De door Rugg getrokken conclusie heeft tot op de dag van vandaag grote methodologische implicaties. Als het anders formuleren van een vraag tot andere antwoorden leidt, hoe valide is de vraag dan? Wat zijn we eigenlijk aan het meten? Inzicht in validiteitsbedreigende factoren en de oorzaken daarvan is van cruciaal belang voor elk terrein waar vragenlijsten als meetinstrument gebruikt worden. In

Samenvatting

Voor het meten van tekstwaardering worden vaak vragenlijsten gebruikt met schaalvragen. Vindt de respondent de tekst bijvoorbeeld makkelijk, of juist moeilijk? Dergelijke vragen kunnen op drie manieren gesteld worden: op een negatieve likertschaal (Ik vind de tekst moeilijk. Eens-Oneens), een positieve likertschaal (Ik vind de tekst makkelijk. Eens-Oneens) of een semantische differentiaal (Ik vind de tekst...Moeilijk-Makkelijk). Deze alternatieve formuleringen worden als dezelfde vragen beschouwd. Masterstudenten Communicatiestudies hebben deze aanname getest. In dit artikel wordt een meta-analyse over deze onderzoeken beschreven, met als doel vast te stellen of een effect van formulering generaliseerbaar is over studies en over woordparen. De keuze van de schaal blijkt cruciaal bij het meten van tekstwaardering. Voor ieder getest woordpaar treedt een formuleringseffect op. Bij tien van de dertien woordparen is dit effect groot. Bij vier woordparen wordt zelfs meer variantie verklaard door de vraagformulering dan door het onderwerp van de tekst. Het formuleren van de vraag op de negatieve likertschaal leidt, in vergelijking met beide andere schalen, vaak tot een positievere weergave van de mening van de respondent.

Hoe makkelijk is een niet moeilijke tekst?

wetenschappelijk onderzoek worden vragenlijsten onder andere gebruikt om tekstwaardering te meten. In vergelijking met andere methoden om tekstwaardering te meten, zoals de plus-en-minmethode, zijn vragenlijsten makkelijk bij een grote groep respondenten af te nemen en kunnen de verkregen data snel verwerkt worden.

Ook voor onderzoek naar webkwaliteit worden vragenlijsten gebruikt. Verschillende onderzoekers hebben recent aandacht besteed aan de manier waarop de kwaliteit van webteksten op een betrouwbare manier met vragenlijsten gemeten kan worden (bijvoorbeeld Elling, Lentz en de Jong, 2007; Muylle, Moeneart en Despontin, 2004). In deze onderzoeken worden de verschillende constructen door meerdere vragen gemeten en wordt steeds ongeveer eenderde deel van de items negatief geformuleerd. Maar maakt het eigenlijk uit of we een respondent vragen hoe makkelijk of juist hoe moeilijk hij de tekst of de website vindt? Geven mensen nog wel dezelfde antwoorden op tekstwaarderingsvragen als ze op zo'n subtiele manier verschillen? In dit artikel wordt een onderzoek beschreven waarin deze kwesties centraal staan. In het resterende deel van deze paragraaf zal worden ingegaan op bestaande literatuur over vraagformulerings-effecten. In de daarop volgende paragrafen wordt besproken hoe die formulerings-effecten worden onderzocht in ruim tien onderzoeken naar tekstwaardering die volgens globaal hetzelfde stramien zijn opgezet, maar onder andere verschillen in de tekst die werd geëvalueerd en in de precieze vragenlijst die werd gebruikt. Vervolgens worden de resultaten besproken en worden conclusies getrokken ten aanzien van formulerings-effecten én ten aanzien van onderzoek naar tekstwaardering.

Na Rugg (1941) is er meer onderzoek gedaan naar formulerings-effecten. Allerlei subtiele variaties in de bewoording van een vraag blijken van belang te zijn, zoals het al dan niet toevoegen van een 'geen mening'-optie en de vraagvolgorde. Eén tak van onderzoek naar formulerings-effecten heeft zich gericht op contradictoire woordparen (bijvoorbeeld: Schuman en Presser, 1981; Hippler en Schwarz, 1986; Holleman, 2000). Volgens de principes van de relationele semantiek zijn contradictoire woordparen, zoals verbieden/toelaten en ja/nee, elkaars tegengestelden en worden ze dichotoom gerepresenteerd (Chaffin en Herrmann, 1984). Als je daarom vóór het verbieden van roken op de werkplek bent, impliceert dit dat je tégen het toelaten bent van datzelfde.

In de literatuur worden nogal wat tegenstrijdige resultaten gerapporteerd over het optreden van formulerings-effecten bij contradictoires. Exemplarisch is Bishop, Hippler, Schwarz en Strack (1988). Zij stelden aan één groep respondenten de vraag of roken in openbare gelegenheden moet worden toegelaten (ja/nee) en een andere groep of roken in openbare gelegenheden moet worden verboden (ja/nee). Bishop et al. voerden dit experiment zowel in Duitsland in het Duits als in de Verenigde Staten in het Engels uit en in beide landen is de vraag zowel schriftelijk als mondeling afgenomen. In de twee mondelinge condities werd geen significant formulerings-effect vastgesteld, maar in de beide schriftelijke condities werd wel een effect geconstateerd. In Duitsland werd, net als bij Rugg (1941), vaker 'nee' geantwoord op de vraag met 'verbieden' dan er 'ja' werd geantwoord op dezelfde vraag geformuleerd met 'toelaten'. In Amerika werd echter het omgekeerde effect vastgesteld. Daar werd vaker 'ja' geantwoord op de vraag met 'verbieden' dan dat er 'nee' geantwoord werd op dezelfde vraag geformuleerd met 'toelaten'.

Op grond van deze resultaten kunnen niet of nauwelijks conclusies getrokken worden over het al dan niet optreden van een formulerings-effect, laat staan over de richting ervan. De 'tegenstrijdige' resultaten kunnen echter verklaard worden vanuit het gekozen onder-

zoeksontwerp: het split-ballot onderzoek. In een split-ballot onderzoek wordt een random getrokken groep respondenten aselekt in twee groepen verdeeld. De ene helft krijgt versie A van een vraag (bijvoorbeeld een verbieden-vraag), de andere helft versie B (bijvoorbeeld een toelaten-vraag). Verschillende factoren van de communicatieve setting (zoals de positie van de vraag in de vragenlijst en mondelinge versus schriftelijke afname) worden gerandomiseerd of gelijk gehouden tussen condities. Daardoor kan een eventueel verschil in antwoorden alleen aan de gemanipuleerde formulering worden toegeschreven. Een consequentie van het onderzoeksontwerp is echter dat getrokken conclusies niet gegeneraliseerd kunnen worden naar andere communicatieve settings (Molenaar, 1989). Als bijvoorbeeld in een mondeling afgenomen enquête een formuleringseffect wordt vastgesteld, mag het effect niet gegeneraliseerd worden naar schriftelijke enquêtes.

Hoewel conclusies op basis van één split-ballot experiment dus moeilijk te generaliseren zijn, hebben verschillende onderzoekers zich al snel bezig gehouden met het verklaren van het tot ‘verbieden/toelaten asymmetrie’ gedoopte formuleringseffect. Schuman en Presser (1981) hypothetiseerden bijvoorbeeld dat de abstractheid van het onderwerp de asymmetrie veroorzaakt. Om dit te testen hebben zij in twee split-ballot experimenten verbieden/toelaten-vragen over concretere onderwerpen gesteld (het tonen van pornofilms en het tonen van sigarettenreclames op de televisie). Bij pornofilms werd een minimaal formuleringseffect geconstateerd ($p < 0.1$). Net als bij Rugg werd er meer ‘nee’ geantwoord op de vraag met ‘verbieden’ dan ‘ja’ op de vraag met ‘toelaten’. Bij de vraag over sigarettenreclames werd geen effect geconstateerd. Schuman en Presser concludeerden daaruit dat abstracte onderwerpen ervoor zorgen dat de keuze voor ‘verbieden’ of ‘toelaten’ meer invloed heeft.

In tegenstelling tot Schuman en Presser, stelden Hippler en Schwarz (1986) wel een significant formuleringseffecten bij een verbieden/toelaten-vraag over een concreet onderwerp (het strooien van zout bij slecht weer) vast. Deze resultaten lijken die van Schuman en Presser tegen te spreken. Als we echter in ogenschouw nemen dat deze experimenten op allerlei aspecten van de communicatieve setting van elkaar verschillen, zijn de resultaten niet zo tegenstrijdig. De geteste ‘verklaring’ van Schuman en Presser is slechts getest in één specifieke communicatieve setting en om die reden is de geteste verklaring niet generaliseerbaar naar andere communicatieve settings. Het split-ballot onderzoeksontwerp is daarom niet geschikt om verklaringen voor formuleringseffecten te testen. Bovendien is het van belang eerst vast te stellen of er, gegeneraliseerd over communicatieve settings, wel een formuleringseffect bestaat en of zoeken naar een verklaring voor het effect dus überhaupt nodig is.

Met behulp van een meta-analyse over meerdere onderzoeken kan vastgesteld worden of een formuleringseffect generaliseerbaar is over communicatieve settings. Holleman (1999a) heeft een meta-analyse uitgevoerd over tweeënvijftig verbieden/toelaten-vragen. Haar dataset bevatte een grote diversiteit aan communicatieve settings en kan gezien worden als een representatieve steekproef uit het totaal aan mogelijke communicatieve settings. Over het algemeen genomen bleken respondenten 14.2% vaker ‘nee’ te antwoorden op een vraag die geformuleerd werd met ‘verbieden’ dan dat ze ‘ja’ antwoorden op dezelfde vraag geformuleerd met ‘toelaten’. Pas nu deze eerste noodzakelijke stap van generaliseren gezet was, is er voldoende basis om te zoeken naar verklaringen voor het formuleringseffect en op basis daarvan advies te geven voor de praktijk (zie Holleman, 2000).

Hoewel er allerlei onderzoek is gedaan naar contradictoire woordparen, is er weinig bekend over formuleringseffecten bij contraire woordparen. Contrairen, zoals makkelijk/moeilijk,

Hoe makkelijk is een niet moeilijke tekst?

zijn elkaars tegenhanger en worden gerepresenteerd op een schaal. Contraire woordparen worden vaak gebruikt om tekstwaardering te meten (bijvoorbeeld: Maes, Ummelen en Hoeken, 1996). In tegenstelling tot dichotoom gerepresenteerde contradictoires, is het bij contraires mogelijk om gradaties aan te geven (Chaffin en Herrmann, 1984). Vaak worden vragen over contraires daarom op een schaal gemeten. Dat kan op drie manieren:

- Een positief geformuleerde likertschaal: *Ik vind de tekst makkelijk eens 0 0 0 0 0 oneens*
- Een negatief geformuleerde likertschaal: *Ik vind de tekst moeilijk eens 0 0 0 0 0 oneens*
- Of een semantische differentiaal: *Ik vind de tekst makkelijk 0 0 0 0 0 moeilijk*

Als de drie vragen hierboven worden geanalyseerd, lijkt de conclusie gerechtvaardigd dat de drie formuleringen hetzelfde zouden moeten meten. Een respondent wordt verondersteld maar één mening te hebben op het denkbeeldige continuüm van een contrair woordpaar. Het eens zijn met de stelling 'ik vind de tekst makkelijk' impliceert het oneens zijn met de stelling 'ik vind de tekst moeilijk' en impliceert de keuze voor de 'makkelijk'-kant van de semantische differentiaal.

Ook in de adviesliteratuur wordt voorbijgegaan aan eventuele formuleringseffecten die op vraagniveau op zouden kunnen treden. Advies over welke vraagstelling de voorkeur verdient, wordt gebaseerd op het voorkomen van antwoordtendenties op vragenlijstniveau:

'Sommige proefpersonen hebben geen zin om de vragenlijst serieus in te vullen. Daarom vullen ze overal hetzelfde in, bijvoorbeeld '5'. Wanneer nu alle Likert-items positief geformuleerd zijn ('de tekst is overzichtelijk, bondig, aantrekkelijk, informatief...') of als alle semantische differentialen het positieve woord aan de zelfde kant hebben ('moeilijk-makkelijk; onlogisch opgebouwd-logisch opgebouwd; onoverzichtelijk-overzichtelijk'), dan zou je concluderen dat deze proefpersoon een licht positief oordeel heeft over de begrijpelijkheid van de tekst. In werkelijkheid heeft de proefpersoon echter niet eens nagedacht over de begrijpelijkheid van de tekst.

Om het effect van onbetrouwbare proefpersonen te minimaliseren kun je gebruik maken van de *balanced-scale techniek*. We gaan ervan uit dat een evaluatief aspect wordt gemeten door een schaal bestaande uit meerdere items. Bij de balanced-scale techniek moet je óf de helft van de Likert-items negatief formuleren óf bij de helft van de semantische differentialen de positie van het positieve woord afwisselen. Dat zorgt ervoor dat in de helft van de items een hoger cijfer een positief oordeel aangeeft, terwijl in de andere helft een lager cijfer een positief oordeel aangeeft.' (Maes, Ummelen en Hoeken, 1996, p. 207-208)

Maar klopt de (impliciete) aanname wel dat de drie formuleringen op vraagniveau hetzelfde meten, of treden er net als bij contradictoires formuleringseffecten op? Friborg, Martinussen en Rosenvinge (2005) zijn een van de weinigen die middels een split-ballot-experiment een vragenlijst met positieve likertschalen hebben vergeleken met een vragenlijst met semantische differentialen. De vragenlijst was bedoeld om het construct 'resilience' (veerkracht) te meten. De data die zijn verkregen met de vragenlijst met semantische differentialen bleken

het beste te passen binnen een vooraf opgesteld model om dit psychologische construct te meten. De drie onderzoekers concluderen dan ook dat de semantische differentiaal beter is voor het meten van psychologische constructen dan de positieve likertschaal. In dit onderzoek is echter niet getest of één van beide schalen op vraagniveau tot een andere weergave van de mening van de respondent heeft geleid.

Twee andere onderzoekers die experimenteel onderzoek naar schaalvragen hebben gedaan, zijn Falzhik en Jolson (1974). Zij hebben, in de context van een marktonderzoek, positieve likertschalen met negatieve likertschalen vergeleken. De negatieve likertschalen bestonden uit de positief geformuleerde stellingen voorafgegaan door het woordje ‘niet’ (antwoorden op de vraag ‘I fully understand unit pricing labels in stores’ werden dus bijvoorbeeld vergeleken met antwoorden op ‘I do not fully understand unit pricing labels in stores’). Uit het onderzoek van Falzhik en Jolson is gebleken dat de positief gestelde vragen tot een positievere weergave van de mening van de respondent leiden in vergelijking met de negatieve likertvragen. De onderzoekers doen daarom de aanbeveling semantische differentialen te gebruiken. Dit advies is echter voorbarig. De semantische differentiaal is immers niet in het onderzoek meegenomen, waardoor niet bekend is of deze schaal wat anders of hetzelfde meet als één van de andere schalen.

Bovenstaande onderzoeken hebben hetzelfde generaliseerbaarheidsprobleem als de genoemde split-ballot onderzoeken naar formuleringseffecten bij het woordpaar verbieden/toelaten. Desondanks wordt op basis van deze onderzoeken advies gegeven voor de praktijk. Het doel van het hier gerapporteerde onderzoek is de noodzakelijke eerste stap te zetten, namelijk vaststellen of er generaliseerbare uitspraken gedaan kunnen worden over formulerings-effecten bij contraire woordparen. Alleen als dit bekend is, weten we of verder onderzoek naar formulerings-effecten bij contrairen van belang is. Vijftien groepen masterstudenten Communicatiestudies van de Universiteit Utrecht hebben onderzoek gedaan naar formulerings-effecten bij contraire woordparen (de beschrijving van deze onderzoeken volgt in paragraaf 2). In dit artikel wordt een statistische meta-analyse over deze studies beschreven. De onderzoeksvragen die in dit artikel centraal staan zijn:

1. *In hoeverre is een effect van vraagformulering bij contraire woordparen generaliseerbaar over studies?*
2. *In hoeverre is een effect van vraagformulering bij contraire woordparen generaliseerbaar over woordparen?*

Een nevenvraag van beide hoofdvragen is die naar de richting van het effect. Falzhik en Jolson (1974) stelden in één enkel split-ballot onderzoek vast dat het stellen van de vraag op de positieve likertschaal tot een positievere weergave van de mening van de respondent leidt in vergelijking met de negatieve likertschaal. Holleman (1999a) vond echter in een meta-analyse van verschillende contradictoire woordparen dat, gegeneraliseerd over communicatieve settings, de negatieve formulering (verbieden) tot een positievere weergave van de mening van de respondent leidt in vergelijking met de positieve formulering (toelaten). In de afzonderlijke onderzoeken die gebruikt werden voor die meta-analyse werd variatie in de richting van het effect gevonden. Om deze redenen wordt verwacht dat bij contraire woordparen de formulering op de negatieve likertschaal tot een positievere weergave van de mening van de respondent leidt.

2. Methode van onderzoek

In de door de masterstudenten Communicatiestudies uitgevoerde onderzoeken werden zowel de positieve likertschaal, als de negatieve likertschaal als de semantische differentiaal op vraagniveau met elkaar vergeleken. Deze in totaal vijftien onderzoeken hadden grotendeels eenzelfde opzet. Proefpersonen kregen eerst een Nederlandstalige tekst van ongeveer één A4 lang te lezen. Daarna moesten zij een vragenlijst invullen, die onder andere bestond uit schaalvragen over tekstwaardering. De proefpersonen verkeerden in de veronderstelling dat zij meededen aan een onderzoek waarin de kwaliteit van teksten centraal stond. Zij wisten niet dat er drie versies van de vragenlijst waren. Elke proefpersoon kreeg slechts één versie onder ogen. Vragen die in de ene versie op een positieve likertschaal gesteld zijn, werden in de andere versie op een negatieve likertschaal gesteld en in de derde versie op een semantische differentiaal.

In elk van de onderzoeken waren de gemanipuleerde schaalvragen gebaseerd op de door Maes, Ummelen en Hoeken (1996) voorgestelde woordparen om het ‘imago van de producent’ en ‘tekstwaardering’ te meten. Deze drie auteurs splitsen het laatste begrip ‘tekstwaardering’ op in twee onderliggende constructen, namelijk: ‘begrijpelijkheid van de tekst’ en ‘aantrekkelijkheid van de tekst’. Elk van de constructen wordt vertaald in een zestal woordparen (zie Tabel 1). Maes, Ummelen en Hoeken zien de woorden die een paar vormen als tegenstellingen. De aanname dat deze woordparen contrair zijn is hierop gebaseerd.

Tabel 1: *Woordparen om tekstwaardering (opgesplitst in begripelijkheid en aantrekkelijkheid) en het imago van de producent te meten (Maes, Ummelen en Hoeken, 1996).*

| Begrijpelijkheid De tekst is... | Aantrekkelijkheid De tekst is... | Imago van de zender De producent lijkt mij... |
|--|---|--|
| Makkelijk/moeilijk | Interessant/oninteressant | Betrouwbaar/onbetrouwbaar |
| Eenvoudig/ingewikkeld | Aansprekend/afstandelijk | Deskundig/ondeskundig |
| Duidelijk/onduidelijk | Uitnodigend/afhoudend | Efficiënt/inefficiënt |
| Overzichtelijk/onoverzichtelijk | Boeiend/saai | Oprecht/onoprecht |
| Logisch/onlogisch opgebouwd | Persoonlijk/onpersoonlijk | Vernieuwend/ouderwets |
| Bondig/omslachtig | Afwisselend/eentonig | Accuraat/inaccuraat |

Naast de gemanipuleerde vragen waren er in elke vragenlijst buffervragen opgenomen, die niet verschilden per versie. Dit waren bijvoorbeeld vragen over demografische gegevens van de respondent.

Hoewel de onderzoeken qua opzet zeer vergelijkbaar zijn, wijken zij op verschillende punten van de communicatieve setting van elkaar af. Zo verschillen de studies op allerlei formuleringsskenmerken. In drie studies zijn de vragen bijvoorbeeld objectief geformuleerd (De tekst is...), in acht studies zijn de stellingen subjectief geformuleerd (Ik vind de tekst...) en in de overige studies zijn subjectieve en objectieve formuleringen afgewisseld. Ook het aantal schaalpunten varieert per studie. In twee studies zijn vijfpuntsschalen gebruikt, in de overige studies zevenpuntsschalen. Daarnaast hebben de gemanipuleerde vragen in verschillende onderzoeken verschillende posities in de vragenlijst. Deze kenmerken hebben volgens Molenaar (1982) invloed op de antwoorden.

Ook verschillen de studies met betrekking tot het type proefpersonen. In één onderzoek zijn ouderen als proefpersonen gebruikt, in twee onderzoeken waren dat middelbare scholieren en in vijf onderzoeken zijn universitaire studenten en HBO-ers gebruikt. In de overige onderzoeken is een heterogene groep volwassenen als proefpersoon gebruikt. In alle gevallen behoorden de respondenten tot de natuurlijke doelgroep van de tekst.

De studies verschillen bovendien op het punt van het gelezen teksttype. In vier studies zijn informatieve teksten gebruikt, in de overige studies zijn persuasieve teksten aan de respondenten voorgelegd. Deze variatie verhoogt de representativiteit van de steekproef uit het totaal aan mogelijke communicatieve settings. Voor meer informatie over de verdeling van deze en andere kenmerken over de studies, zie Bijlage 1.

Doordat het volledige onderzoeksmateriaal van de onderzoeken beschikbaar was, is het mogelijk geweest kwaliteitseisen aan de studies en woordparen te stellen. Woordparen en studies werden alleen meegenomen in de analyse als ze voldeden aan de volgende criteria:

- De manipulatie moet zuiver zijn. Dit betekent dat de verschillende versies van de vragenlijst, op de gemanipuleerde formulering na, identiek moeten zijn. In één van de studies was niet alleen de formulering, maar ook de positie van ‘eens’ en ‘oneens’ op de schaal veranderd. In de ene versie van de vragenlijst stond ‘eens’ links, terwijl in de andere versie ‘oneens’ links stond. Deze dubbele manipulatie geeft ruimte voor een alternatieve verklaring bij een eventueel formuleringseffect. De studie met deze dubbele manipulatie is daarom niet meegenomen in de meta-analyse.
- De onderzoeksgroepen moeten random zijn. Dit is gecontroleerd door te kijken of de onderzoeksgroepen verschillen op basis van hun demografische kenmerken en de scores op de andere buffervragen. Als onderzoeksgroepen binnen een onderzoek verschillen op deze vragen zijn ze niet meegenomen in de meta-analyse, omdat dan niet meer zeker is of verschillen in de antwoorden tussen versies puur aan de formuleringvariantie zijn toe te schrijven. Op basis van dit criterium zijn twee studies niet meegenomen in de meta-analyse.
- De woordparen moeten exact zo geformuleerd zijn als Maes, Ummelen en Hoeken (1996) dat voorstellen. Als in studies een andere formulering is gebruikt voor een woordpaar (bijvoorbeeld boeiend/onboeiend in plaats van boeiend/saai), is het woordpaar niet meegenomen in de analyse.
- De woordparen moeten in minstens zeven studies getest zijn. Na de selectie die gemaakt is op basis van de hierboven beschreven criteria en het feit dat sommige studies toch al niet alle woordparen hadden gebruikt, kwamen de woordparen betrouwbaar/onbetrouwbaar, efficiënt/inefficiënt, oprecht/onoprecht, vernieuwend/ouderwets en accuraat/inaccuraat in minder dan zeven studies voor. Deze woordparen zijn niet verder geanalyseerd, omdat bij minder dan zeven studies het ‘meta’ karakter van de analyse wat ver te zoeken is. De vragen met de dertien overgebleven woordparen zijn door minimaal 897 en maximaal 1619 respondenten beantwoord (zie Bijlage 2).

3. Analyse

Om een antwoord op de twee onderzoeksvragen te krijgen, is per woordpaar een lineaire regressieanalyse uitgevoerd. Hierbij zijn de verschillende studies, antwoordschalen en de interacties tussen studies en antwoordschalen als verklarende variabelen gebruikt om de score

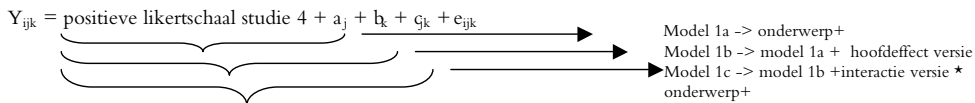
Hoe makkelijk is een niet moeilijke tekst?

op de afhankelijke variabele (het door de respondent aangekruiste schaalpunt op de vraag) te voorspellen. De onafhankelijke variabelen zullen hieronder toegelicht worden.

Allereerst is te verwachten dat de score op een vraag varieert van studie tot studie. De variatie in aspecten van de communicatieve setting (verschillende typen respondenten, verschillende afnamecondities, enzovoorts) speelt hierbij een rol. Ook de feitelijke kwaliteit van de tekst en het onderwerp van de tekst zijn cruciaal. Het onderwerp van de ene tekst is immers begrijpelijker of aantrekkelijker dan het onderwerp van de andere tekst. De dummyvariabele die variantie verklaart op basis van alle niet gedefinieerde verschillen tussen studies is daarom ‘onderwerp+’ genoemd.

Bovendien is het van belang te achterhalen welk deel van de variantie in scores op de vraag toegeschreven kan worden aan de antwoordschalen. Er zijn twee effecten van vraagversie denkbaar. Op de eerste plaats kan er een hoofdeffect zijn. Dit betekent dat voor iedere studie het effect van versie hetzelfde is en dezelfde richting heeft. Deze dummyvariabele is ‘vraagversie’ genoemd. Op de tweede plaats kan er een interactieeffect optreden tussen versie en onderwerp+. In dat geval verschilt het effect van versie van studie tot studie. Deze dummyvariabele is ‘onderwerp+ * vraagversie’ genoemd.

Met behulp van deze drie onafhankelijke variabelen en een arbitrair gekozen uitgangspunt (de positieve likertschaal van studie 4) is het door de respondent aangekruiste schaalpunt voorspeld. Dit gebeurt additief³. Het regressiemodel wordt als het ware steeds uitgebreid met een nieuwe onafhankelijke variabele, waardoor er drie submodellen ontstaan die de score op de vraag steeds specifiek schatten. Stel Y_{ijk} is de score van respondent i ($i = 1, 2, \dots, N$) in studie j ($j = 1, 2, \dots, 15$) op versie k ($k = 1, 2, 3$) dan kan het model als in vergelijking 1 geschreven worden. De term e_{ijk} staat voor de random error.



Vergelijking 1: De regressievergelijking gebruikt voor de meta-analyse.

Eerst wordt de score op het uitgangspunt (de positieve likertschaal van studie 4) berekend. Vervolgens wordt per woordpaar voor iedere studie j berekend hoeveel de score op de positieve likertschaal van die studie afwijkt van dit uitgangspunt. Dit is het effect van onderwerp+ (a_j in de vergelijking). Met Pearson’s correlatiecoëfficiënt r kan de samenhang tussen de op basis van model 1a geschatte scores en de werkelijke scores berekend worden. Door deze waarde te kwadrateren (r^2) krijgen we het percentage van de variantie dat door onderwerp+ wordt verklaard.

Daarna wordt het verschil in het aangekruiste schaalpunt van de respondent voor de andere vraagversies k berekend. Dit wordt eerst gedaan voor studie 4. Als er een hoofdeffect optreedt van versie, zouden de verschillen die optreden bij studie 4 als gevolg van het anders formuleren van de vraag, ook voor de andere studies moeten gelden. De verandering in score voor studie 4 wordt daarom ook toegepast bij de andere studies (b_k in de vergelijking). De geschatte scores op basis van model 1b worden gecorreleerd met de werkelijke scores en opnieuw wordt r^2 berekend. Door het percentage verklaarde variantie van onderwerp+ van het door model 1b berekende percentage verklaarde variantie af te trekken, weten we welk deel van de variantie door het hoofdeffect van versie wordt verklaard.

Vervolgens wordt op basis van model 1c de verandering in score als gevolg van de interactie onderwerp+ * versie berekend (c_{jk} in de vergelijking). Het effect van versie wordt nu voor iedere afzonderlijke studie berekend. Door r^2 uit te rekenen weten we hoeveel van de variantie door de interactie verklaard wordt.

Om een maat te hebben voor de grootte van het effect van vraagformulering, berekenen we per woordpaar de verhouding tussen het percentage verklaarde variantie door beide effecten van versie samen (het hoofdeffect + het interactieeffect) ten opzichte van het totale model. Met behulp van de criteria van Cohen (1977) kan de zo verkregen effectgrootte geïnterpreteerd worden. Zo kan per woordpaar worden vastgesteld of het effect van vraagformulering generaliseerbaar is over studies.

Als is vastgesteld bij hoeveel woordparen het formuleringseffect generaliseerbaar is over studies, kan bepaald worden of er ook over meerdere contraire woordparen gegeneraliseerd kan worden. Vervolgens kunnen we de regressiegewichten uit het model interpreteren om uitspraak te doen over de richting en consistentie van het effect. Hiertoe zijn per woordpaar de significante regressiegewichten in een grafiek getekend en is deze grafiek geïnterpreteerd⁴.

4. Resultaten

Voor ieder woordpaar is de effectgrootte berekend. In Tabel 2 staan de resultaten weergegeven.

Tabel 2: Voor alle woordparen de percentages verklaarde variantie van onderwerp+, de beide effecten van versie, het totale regressiemodel en de op basis hiervan berekende effectgrootte.

| | % variantie onder- werp+ | % variantie versie en onderwerp+ * | % variantie totaal model | Effect- grootte |
|--|-----------------------------|---------------------------------------|-----------------------------|--------------------|
| <i>Interessant/Oninteressant</i> | 5.7 | 8.9 | 14.6 | 0.61 |
| <i>Overzichtelijk/Onoverzichtelijk</i> | 8.3 | 11.8 | 20.1 | 0.59 |
| <i>Afwisselend/Eentonig</i> | 4.6 | 5.8 | 10.5 | 0.55 |
| <i>Duidelijk/Onduidelijk</i> | 13.2 | 14.4 | 27.6 | 0.52 |
| <i>Boeiend/Saai</i> | 8.2 | 5.9 | 14.1 | 0.42 |
| <i>Uitnodigend/Afhoudend</i> | 16.1 | 8.8 | 24.9 | 0.35 |
| <i>Eenvoudig/ Ingewikkeld</i> | 25.1 | 12.6 | 37.7 | 0.33 |
| <i>Aansprekend/Afstandelijk</i> | 15.1 | 7.3 | 22.4 | 0.33 |
| <i>Logisch/Onlogisch opgebouwd</i> | 15.5 | 7.8 | 23.3 | 0.33 |
| <i>Bondig/Omslachtig</i> | 21.5 | 10 | 31.5 | 0.32 |
| <i>Deskundig/Ondeskundig</i> | 12.7 | 4.6 | 17.3 | 0.27 |
| <i>Makkelijk/Moeilijk</i> | 29.2 | 9.7 | 38.9 | 0.25 |
| <i>Persoonlijk/Onpersoonlijk</i> | 21.3 | 4.7 | 26.0 | 0.18 |

Hoe makkelijk is een niet moeilijke tekst?

Als de effectgrootte boven de 0.30 ligt, kan volgens Cohen (1977) geconcludeerd worden dat een effect groot is. Bij tien van de dertien geteste woordparen is het formuleringseffect dus groot. Bij vier woordparen (interessant/oninteressant, overzichtelijk/onoverzichtelijk, afwisselend/eentonig en duidelijk/onduidelijk) verklaart de formulering zelfs meer dan de helft van de variantie binnen het model. Dit betekent dat bij deze woordparen de formulering van de vraag zelfs belangrijker is dan de communicatieve context van het onderzoek, het onderwerp en de kwaliteit van de tekst.

Als de verhouding tussen de beide effecten van versie en het totale model tussen de 0.20 en 0.30 ligt, is er volgens Cohen sprake van een middelgroot effect. Dit is het geval bij twee woordparen (makkelijk/moeilijk en deskundig/ondeskundig). Bij een effectgrootte tussen de 0.10 en 0.20 wordt gesproken van een klein effect. Dit is het geval bij één woordpaar (persoonlijk/onpersoonlijk).

Op basis van deze resultaten kunnen we vaststellen dat de formulering van schaalvragen van cruciaal belang is. Het effect van formulering is immers te generaliseren over studies en over contraire woordparen, omdat er over alle gebruikte studies samengenomen voor ieder woordpaar afzonderlijk een effect bestaat.

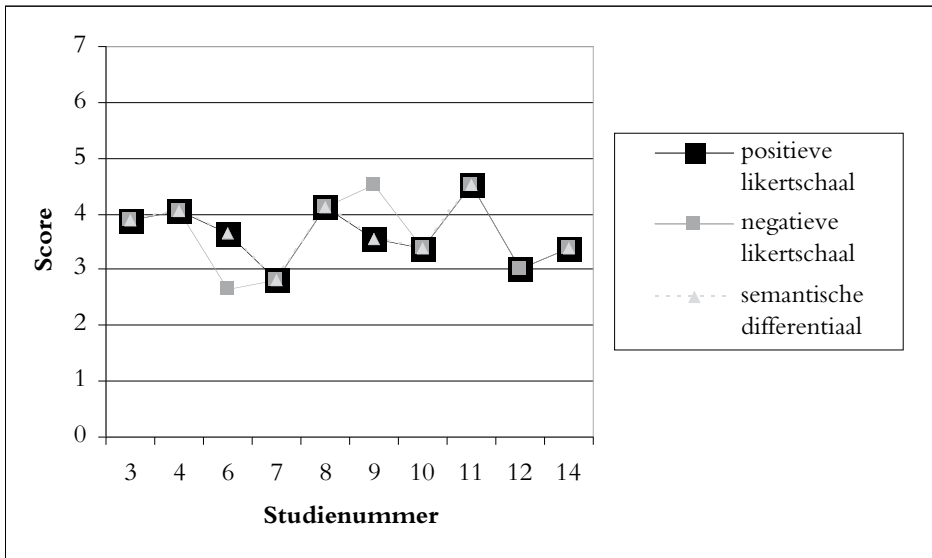
Als naar de vierde kolom uit Tabel 2 wordt gekeken, valt op dat gemiddeld ongeveer 25% van de variantie wordt verklaard door het totale model. Dit lijkt op het eerste gezicht wat weinig. Dat slechts een kwart van de totale variantie wordt verklaard is echter niet vreemd als in ogenschouw wordt genomen dat er geen persoonskenmerken in het model zijn opgenomen. Persoonskenmerken (zoals persoonlijke interesse of leeftijd) zullen logischerwijs een groot deel van de totale variantie in scores verklaren. De grootte van het formuleringseffect is nu vastgesteld ten opzichte van het onderwerp+ van de tekst, zonder rekening te houden met persoonskenmerken.

Wat verder opvalt voor de grootste effecten (kolom 5), is dat het absolute percentage verklaarde variantie door beide effecten van versie (kolom 3) vaak klein is. Dit is niet vreemd. De effectgrootte is een maat die de verhouding van het percentage verklaarde variantie door beide effecten van versie én het percentage verklaarde variantie van onderwerp+ aangeeft. Als het onderwerp+ van de tekst weinig variantie verklaart, hoeft minder variantie toe te schrijven te zijn aan de beide effecten van versie om toch van een groot effect te kunnen spreken.

De effectgrootte is bovendien geen maat voor de consistentie van het formuleringseffect. Het kan voorkomen dat er voor één woordpaar verschillende interactieeffecten significant zijn voor de verschillende studies. Zo kan de negatieve likertschaal voor een aantal studies tot een positievere weergave van de mening van respondenten leiden, terwijl voor een aantal andere studies de negatieve likertschaal tot een negatievere weergave van de mening van de respondent leidt en voor weer andere studies kan er geen verschil zijn tussen de schalen. De score op de afhankelijke variabele wordt dan door de vele significante interacties preciezer voorspeld dan bij één enkel significant hoofdeffect, maar het effect is verre van consistent.

Nu duidelijk is dat een effect van formulering generaliseerbaar is over studies en over contraire woordparen, is het zinvol naar de richting van het formuleringseffect te kijken om zo een uitspraak te doen over de consistentie van het effect. Hiertoe zullen we nu de (verwachte) gemiddelden per conditie grafisch illustreren. Bij het interpreteren van de significante regressiegewichten waren drie tendensen te ontdekken. Van elk van deze drie tendensen zal een voorbeeld besproken worden.

Tendens 1: voor één of enkele studies zijn er verschillen tussen de negatieve likertschaal en beide andere schalen. Als voorbeeld van deze tendens zal het woordpaar afwisselend/eentonig besproken worden (zie Figuur 1):



Legenda:

Score = De op basis van model 1c voorspelde score op de vraag. Een score van 1 staat voor een negatieve evaluatie van het attitudeobject, een score van 7 voor een positieve evaluatie.
 Studienummer = Het nummer van de studie (zie Bijlage 1).

Figuur 1: Weergave van de significante regressiegewichten van het woordpaar afwisselend/eentonig.

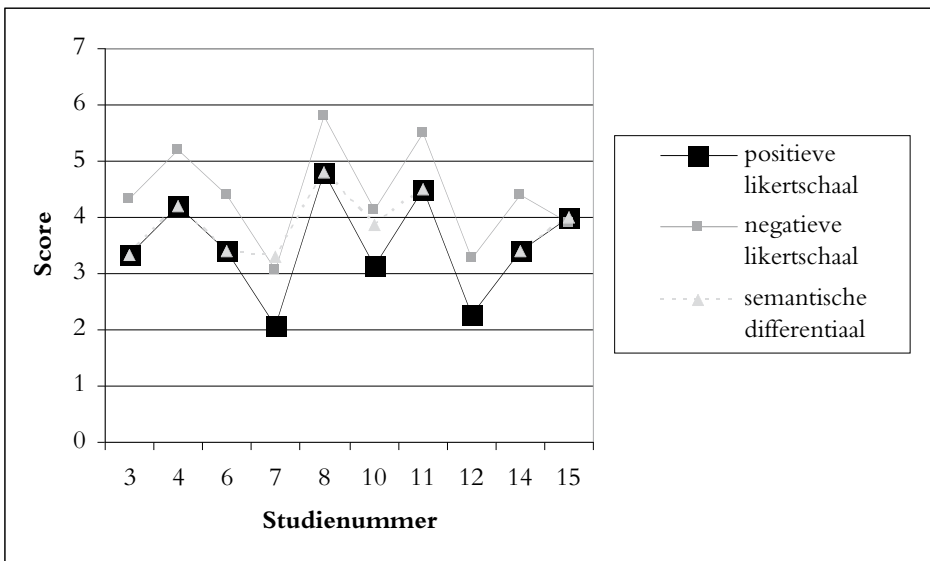
Voor de meeste studies wordt geen significant verschil tussen de schalen vastgesteld. De mening van de respondent wordt op de negatieve likertschaal 0.46 schaalpunten positiever weergegeven in vergelijking met de positieve likertschaal, maar dit verschil is niet significant ($b_2 = 0.46$ met $t = 1.69$ en $p = 0.09$). Op de semantische differentiaal wordt de mening van de respondent 0.24 schaalpunt negatiever weergegeven in vergelijking met de positieve likertschaal, maar ook dit is niet significant ($b_3 = -0.24$ met $t = -0.85$ en $p = 0.39$).

Bij twee studies wordt afgeweken van dit patroon. Bij één studie (studie 6) leidt het stellen van de vraag op een negatieve likertschaal tot een negatievere weergave van de mening van de respondent in vergelijking met de beide andere schalen ($c_{6,2} = -1.00$ met $t = -2.53$ en $p = 0.01$). Voor een andere studie (studie 9) is het tegenovergestelde patroon waar te nemen. Daar leidt het stellen van de vraag op een negatieve likertschaal juist tot een positievere weergave van de mening van de respondent in vergelijking met de beide andere schalen ($c_{9,2} = 0.98$ met $t = 2.42$ en $p = 0.02$). In Figuur 1 lijkt bij studie 12 de negatieve likertschaal af te wijken van de beide andere schalen. Dit is echter niet het geval, in deze studie zijn alleen de positieve en de negatieve likertschaal met elkaar vergeleken en daarom ontbreekt in de figuur de lijn voor de semantische differentiaal.

Hoe makkelijk is een niet moeilijke tekst?

Het patroon waarbij nauwelijks significante verschillen worden geconstateerd, treedt op bij de woordparen: duidelijk/onduidelijk, interessant/oninteressant, boeiend/saai, makkelijk/moeilijk, logisch/onlogisch opgebouwd en eenvoudig/ingewikkeld. Als bij deze woordparen voor één of enkele studies een significant verschil tussen versies wordt geconstateerd, is dat altijd een verschil tussen de positieve en de negatieve likertschaal. In verreweg de meeste gevallen leidt het stellen van de vraag op de negatieve likertschaal tot een positievere weergave van de mening van de respondent in vergelijking met de beide andere schalen. Bij één van de studies (studie 9) wordt dit effect bij alle geteste woordparen vastgesteld.

Tendens 2: voor de meeste studies leidt het stellen van de vraag op de negatieve likertschaal tot een positievere weergave van de mening van de respondent, voor enkele studies zijn er ook verschillen tussen de positieve likertschaal en de semantische differentiaal. Als voorbeeld zal het woordpaar uitnodigend/afhoudend besproken worden:



Legenda:

Score = De op basis van model 1c voorspelde score op de vraag. Een score van 1 staat voor een negatieve evaluatie van het attitudeobject, een score van 7 voor een positieve evaluatie.
Studienummer = Het nummer van de studie (zie Bijlage 1).

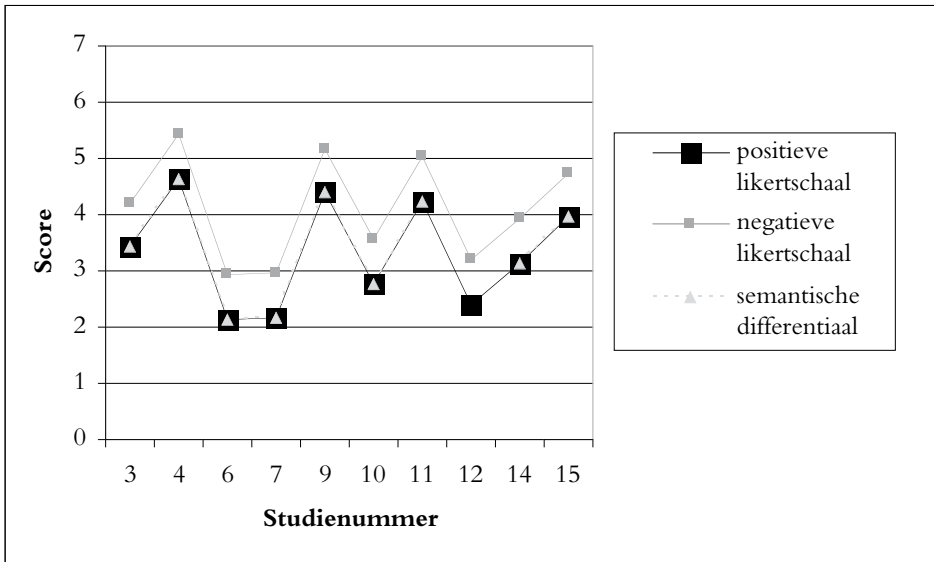
Figuur 2: Weergave van de significante regressiegewichten van het woordpaar uitnodigend/afhoudend.

Bij dit woordpaar leidt het formuleren van de vraag op de negatieve likertschaal voor de meeste studies tot een positievere weergave van de mening van de respondent in vergelijking met de beide andere schalen. Het hoofdeffect van versie is significant, de mening van de respondent wordt 1 schaalpunt positiever weergegeven op de negatieve likertschaal in vergelijking met de beide andere schalen ($b_2 = 1$ met $t = 3.93$ en $p < 0.01$).

Voor één studie (studie 15) wordt dit hoofdeffect van versie teniet gedaan door de interactie tussen de negatieve likertschaal en onderwerp+, waardoor de drie schalen weer vrijwel gelijk komen te liggen ($c_{15,2} = -1.11$ met $t = -2.35$ en $p = 0.02$). Voor twee studies (studie 7 en 10) is tevens een significant verschil tussen de positieve likertschaal en de semantische differentiaal geconstateerd. Voor studie 7 leidt de formulering op de semantische differentiaal tot een weergave van de mening van de respondent die 1.22 schaalpunt positiever is in vergelijking met de positieve likertschaal ($c_{7,3} = 1.22$ met $t = 3.19$ en $p < 0.01$). Voor studie 10 is de weergave van de mening op de semantische differentiaal 0.74 schaalpunt positiever in vergelijking met de positieve likertschaal ($c_{10,3} = 0.74$ met $t = 2.03$ en $p = 0.04$).

Bij de woordparen aansprekend/afstandelijk, overzichtelijk/onoverzichtelijk en bondig/omslachtig is eenzelfde patroon vastgesteld. Ook daar leidt het stellen van de vraag op de negatieve likertschaal bij een aantal of de meeste studies tot een positievere weergave van de mening van de respondent en treedt er bovendien voor één of enkele studies een verschil op tussen de positieve likertschaal en de semantische differentiaal. De richting van dit effect verschilt: de ene keer leidt het stellen van de vraag op de positieve likertschaal tot een positievere weergave van de mening van de respondent, de andere keer tot een negatievere.

Tendens 3: voor alle studies leidt het stellen van de vraag op de negatieve likertschaal tot een positievere weergave van de mening van de respondent in vergelijking met de andere twee schalen. Als voorbeeld van deze tendens wordt het woordpaar persoonlijk/onpersoonlijk besproken (zie Figuur 3):



Legenda:

Score = De op basis van model 1c voorspelde score op de vraag. Een score van 1 staat voor een negatieve evaluatie van het attitudeobject, een score van 7 voor een positieve evaluatie.
 Studienummer = Het nummer van de studie (zie Bijlage 1).

Figuur 3: Weergave van de significante regressiegewichten van het woordpaar persoonlijk/onpersoonlijk.

Hoe makkelijk is een niet moeilijke tekst?

Bij dit woordpaar is het hoofdeffect van de negatieve likertschaal het enige significante effect. De negatieve likertschaal leidt tot een 0.79 schaalpunt positievere weergave van de mening van de respondent in vergelijking met de beide andere schalen ($b_2 = 0.79$ met $t = 2.84$ en $p < 0.01$). Dit patroon komt ook voor bij het woordpaar deskundig/ondeskundig.

Als deze resultaten aan de berekende effectgroottes (zie Tabel 2) worden verbonden, valt op dat de effectgrootte klein is bij de meest consistente effecten. Dit is niet vreemd. Een significant hoofdeffect betekent dat bij alle studies hetzelfde patroon wordt vastgesteld. Er valt dus bij alleen een hoofdeffect minder variantie te verklaren dan bij een interactie. In dat laatste geval is het effect van formulering tegengesteld in verschillende studies en is er daarom altijd meer variantie die verklaard kan worden.

5. Conclusie en discussie

In dit artikel is een meta-analyse beschreven naar formuleringseffecten bij contrairen in vragenlijsten over tekstkwaliteit. Onderzocht is of het uitmaakt of een respondent zijn mening geeft op een positieve likertschaal, een negatieve likertschaal of een semantische differentiaal. Het doel was vast te stellen of een effect van formulering te generaliseren is over studies en over contraire woordparen. Middels een regressieanalyse is daartoe per woordpaar berekend hoeveel procent van de variantie in de score op de vraag kan worden toegeschreven aan de communicatieve context van het onderzoek, het onderwerp en de kwaliteit van de tekst, en hoeveel aan de vraagversie.

Voor elk getest woordpaar bleek de invloed van vraagversie, over de verschillende studies samengenomen, betekenisvol te zijn. Het effect van formulering is daarom generaliseerbaar over studies en over woordparen. De grootte van het effect is geclassificeerd aan de hand van de criteria van Cohen (1977). Bij tien van de dertien geteste woordparen is het effect van formulering als groot te classificeren. Binnen de grote effecten is voor vier woordparen het percentage verklaarde variantie dat toe te schrijven is aan de formulering van de vraag zelfs groter dan het percentage variantie dat verklaard wordt door het onderwerp van de tekst en de communicatieve context van het onderzoek. Dit is tamelijk alarmerend: het maakt in deze gevallen dus meer uit hoe de vraag gesteld wordt dan welke tekst er is gelezen... Naast de tien grote effecten, zijn twee middelgrote effecten en één klein effect geconstateerd.

Als er een significant verschil tussen versies is geconstateerd, leidt het stellen van de vraag op de negatieve likertschaal vaak tot een positievere weergave van de mening van de respondent in vergelijking met de beide andere schalen. Falthzik en Jolson (1974) constateerden het tegengestelde effect. In hun onderzoek leidde een positieve formulering juist tot een positievere weergave van de mening van de respondent. Een verschil tussen het onderzoek van Falthzik en Jolson en het hier gerapporteerde onderzoek is dat de negatieve likertschaal bij eerst genoemden bestond uit een expliciete negatie door middel van het woordje 'niet', terwijl binnen deze meta-analyse de negatieve likertschaal bestond uit een woord met een geïncorporeerde negatie (makkelijk/moeilijk) of een morfologische ontkenning (duidelijk/onduidelijk). De bevindingen van het hier gerapporteerde onderzoek komen wel overeen met het onderzoek van Holleman (1999a). De negatieve formulering met 'verbieden' leidde ook hier tot een positievere weergave van de mening van de respondent.

Hoewel er overduidelijk een tendens in de richting van een formuleringseffect is geconstateerd, verschilt het per studie en per woordpaar hoe vaak deze tendens aangetoond is.

Ook treden er sporadisch andere significante effecten op. Soms leidt het stellen van de vraag op de positieve likertschaal tot een negatievere weergave van de mening van de respondent in vergelijking met de beide andere schalen. Tevens is er een aantal keer een verschil tussen de positieve likertschaal en de semantische differentiaal geconstateerd. Deze verschillen lijken op het eerste gezicht wat vreemd. Ook bij het woordpaar verbieden/toelaten kwamen verschillen tussen studies voor en werden er soms tegengestelde effecten vastgesteld (Holleman, 1999a). Deze verschillen kunnen verklaard worden door de variatie tussen studies in aspecten van de communicatieve setting, tekstkenmerken en persoonskenmerken. Door deze meta-analyse is vastgesteld dat het effect van formulering onafhankelijk van deze factoren bestaat.

Als de bevindingen van deze meta-analyse in een bredere context worden geplaatst, luidt de conclusie dat men de vraag negatief moet stellen als men een positief beeld over iets naar voren wil laten komen. Als bijvoorbeeld de nieuwe gratis krant "De Pers" een positief beeld over de kwaliteit van haar artikelen naar voren wil brengen, kan het beste gevraagd worden hoe slecht respondenten de artikelen vinden. Welke formulering het beste is, het beste benadert wat respondenten echt vinden, is echter in dit type onderzoek niet vast te stellen. Daarvoor zouden alle respondenten alledrie de formuleringen voorgelegd moeten krijgen, zodat inzicht verkregen kan worden in de samenhang tussen antwoorden als deze van meetfouten worden ontdaan (vergelijk Holleman, 1999b).

Hoewel deze meta-analyse veel bruikbare informatie heeft opgeleverd, zitten er ook een aantal haken en ogen aan het gebruikte materiaal en de gehanteerde onderzoeksmethode. Sommige voor deze meta-analyse gebruikte onderzoeken hadden een onderzoeksopzet die een aantal validiteitsbedreigende factoren kent. In deze studies werd een geheel negatieve versie van de vragenlijst vergeleken met een geheel positieve versie en een geheel semantisch differentiale versie. Respondenten zouden bij een geheel negatieve of positieve vragenlijst anders kunnen gaan antwoorden door de algehele negatieve of positieve tendens in de vragenlijst. Bij een andere opzet met twee likertversies met allebei de helft van de items positief en de helft van de items negatief geformuleerd heeft dit geen rol gespeeld.

Een tweede kritiekpunt betreft de gebruikte woordparen. De veronderstelling is gemaakt dat de woordparen allemaal symmetrische contrairen zijn. Hiervoor is een autoriteitsargument gebruikt: Maes, Ummelen en Hoeken (1996) zeggen dat de woordparen tegenstellingen zijn. Buiten symmetrische contrairen worden door Chaffin en Herrmann (1984) een aantal aanverwante tegenstellingen onderscheiden. Het was beter geweest eerst te testen of de gebruikte woordparen door respondenten wel echt als zuivere contrairen gezien worden, zodat er meer zekerheid zou zijn geweest over de constructvaliditeit. Bovendien moet opgemerkt worden dat gebruikte woordparen ook op andere manieren ingedeeld kunnen worden (bijvoorbeeld op basis van Kennedy en McNally, 2005).

Een derde nadeel van deze meta-analyse is dat er niet gegeneraliseerd kan worden over aspecten van de communicatieve setting die niet gevarieerd zijn tussen de gebruikte studies. Deze betreffen bijvoorbeeld de taal (er zijn alleen Nederlandse onderzoeken opgenomen), de afnameconditie (er is alleen gebruik gemaakt van schriftelijke afname) en de context (de stellingen hadden allemaal betrekking op het meten van tekstwaardering). Nader onderzoek zal uitsluitel moeten geven of er ook over deze aspecten gegeneraliseerd kan worden.

De genoemde zwakke punten halen de conclusie niet onderuit. Dat de evaluatieve richting van de vraagvorm van cruciaal belang is, is overtuigend vastgesteld. Vanuit communica-

Hoe makkelijk is een niet moeilijke tekst?

tief oogpunt vragen de resultaten nu om een verklaring. Waar in het vraag-antwoordproces ontstaan verschillen in antwoorden bij vragen die op het eerste gezicht hetzelfde lijken te meten? Zorgt het anders stellen van de vraag ervoor dat de respondent een andere mening ophaalt uit zijn langetermijn-geheugen? Of wordt dezelfde mening opgehaald, maar wordt deze anders vertaald naar de schaal? Holleman (1999b) gebruikte een experimenteel-correlationeel design om dezelfde vragen te beantwoorden voor het woordpaar verbieden/toelaten. Met deze methode kan ook vastgesteld worden welke formulering het meest betrouwbaar is. Resultaten van een dergelijk onderzoek vormen daarom de directe basis voor praktijkadvies. Het formuleringseffect kan ook nader verklaard worden door te specificeren onder welke omstandigheden de verschillen tussen vraagversies het grootst zijn. Hierop kan advies gebaseerd worden in welke situaties de keuze van de vraagformulering (vooral) belangrijk is. Voorlopig is duidelijk dat mensen aangeven een tekst makkelijker te vinden als we ze vragen hoe moeilijk die is.

Noten

- 1 Met dank aan Margriet Brus en de masterstudenten van de cursus 'Vragenlijst en Vraagformulering' uit collegejaar 2005–2006: I. Amez, R. Beerkens, L van den Bosch L., E. Bruens, E. Croes, T. Dankers, A. Dekker, L. Dijkgraaf, C. ter Doest, K. Hoogenboezem, L. in 't Hout, F. Jansen, J. de Jong, M. Joosen, D. Kaland, B. van Kippersluis, M. Kruimelaar, L. Lauterslager, S. de Leeuw, D. Lodewijks, F. Loonstra, J. Luijten, L. Odijk, M. Pelt, D. van Rijswijk, M. Rozenboom, M. van Ruijven, J. Scholten, L. Schoenmaeckers, M. Slieker, G. van Stempvoort, A. Stolk, C. Timmermans, K. Veerman, R. Vlek, L. Victorian, R. Walraven, M. Wegman en L. Zoon.
- 2 Voor correspondentie over dit artikel: Naomi.Kamoen@let.uu.nl, Bregje.Holleman@let.uu.nl
- 3 Elk van de drie schalen van elk van de twaalf gebruikte studies had als uitgangspunt gekozen kunnen worden. Een ander uitgangspunt had weliswaar tot andere getallen, maar tot dezelfde interpretatie geleid. Een eenvoudigd voorbeeld ter illustratie. Stel we doen een regressieanalyse voor het woordpaar interessant/oninteressant. De data zijn zo gecodeerd dat hoe hoger de gemiddelde score is, hoe interessanter proefpersonen de tekst beoordelen. Stel de positieve likertschaal heeft een gemiddelde score van 2 en de negatieve likertschaal van 4. Als de positieve likertschaal als uitgangspunt wordt genomen, is het regressiegewicht van de negatieve likertschaal 2. Wordt de negatieve likertschaal als uitgangspunt genomen, dan is het regressiegewicht van de positieve likertschaal -2. De interpretatie blijft echter hetzelfde. In dit hypothetische geval is de interpretatie dat de negatieve likertschaal tot een positieve weergave van de mening van respondenten leidt.
- 4 Er is gekozen alleen de significante afwijkingen te tekenen, omdat deze het grootste aandeel hebben in het verklaren van variantie.

Bibliografie

- Bishop, G., Hippler, H.J., Schwarz, N., & Strack, F. (1988).** A comparison of response effects in self-administered and telephone surveys. In R.M. Groves, P. Biemer, L. Lyberg, J. Massey, W. Nicholls, & J. Waksberg (eds.) *Telephone survey methodology* (pp. 273–282). New York: Wiley.
- Chaffin, R., & Herrmann, D.J. (1984).** The similarity and diversity of semantic relations. *Memory and Cognition*, 12, 134–141.
- Cohen, J. (1977).** *Statistical power analysis for the behavioural sciences* (revised ed.). New York: Academic Press.

- Elling, S., Lentz, L., & de Jong, M. (2007).** Website Evaluation Questionnaire: Development of a research-based tool for evaluating informational websites. In M.A. Wimmer, J. Scholl, & Å. Grönlund (eds.) *EGOV 2007, LNCS 4656* (pp. 293-304). Berlin, Heidelberg: Springer-Verlag.
- Falzhik, A.M., & Jolson, M.A. (1974).** Statement polarity in attitude studies. *Journal of Marketing Research*, 11, 102-105.
- Friborg, O., Martinussen, M., & Rosenvinge, J.H. (2005).** Likert-based vs. semantic differential-based scorings of positive psychological constructs: A psychometric comparison of two versions of a scale measuring resilience. *Personality and individual differences*, 40, 873-884.
- Hippler, H.J., & Schwarz, N. (1986).** Not forbidding isn't allowing: the cognitive basis of the forbid/ allow asymmetry. *Public Opinion Quarterly*, 50, 87-96.
- Holleman, B.C. (1999a).** Wording effects in survey research. Using meta-analysis to explain the forbid/allow asymmetry. *Journal of Quantitative Linguistics*, 6, 29-40.
- Holleman, B.C. (1999b).** The nature of the forbid/ allow asymmetry. Two correlational studies. *Sociological Methods & Research*, 28, 2, 209-244.
- Holleman, B.C. (2000).** *The Forbid/Allow Asymmetry. On the cognitive mechanisms underlying wording effects in surveys.* (Proefschrift Universiteit Utrecht). Amsterdam/Atlanta: Rodopi Publishers.
- Kennedy, C., & McNally, L. (2005).** Scale structure, degree modification and the semantics of gradable predicates. *Language*, 81, 2, 345-381.
- Maes, F., Ummelen, N., & Hoeken, H. (1996).** *Instructieve Teksten. Analyse, Ontwerp en Evaluatie.* Bussum: Uitgeverij Coutinho.
- Molenaar, N.J. (1982).** Response-effects of "formal" characteristics of questions. In W. Dijkstra, & J. van der Zouwen (Red.) *Response behaviour in the survey-interview* (pp. 49-89). London: Academic Press.
- Molenaar, N.J. (1989).** Onderzoek naar formuleringseffecten in survey-interviews. In J. van der Zouwen, & W. Dijkstra (Red.) *Sociaal-wetenschappelijk onderzoek met vragenlijsten. Methoden, knelpunten, oplossingen* (pp. 53-74). Amsterdam: VU Uitgeverij.
- Muyllé, S., Moeneart, R., & Despontin, M. (2004).** The conceptualization and empirical validation of web site user satisfaction. *Information & Management*, 41, 5, 543-560.
- Rugg, D. (1941).** Experiments in wording questions II. *Public Opinion Quarterly*, 5, 91-92.
- Schuman, H., & Presser, S. (1981).** *Questions and answers in attitude surveys. Experiments on question form, wording and context.* London: Academic Press.

Hoe makkelijk is een niet moeilijke tekst?

Bijlage 1: Overzicht van kenmerken van de uiteindelijk gebruikte studies

| Studie | N | Doel tekst + Onderwerp | Zender Tekst | Doelgroep onderzoek | Omgeving Afname | Formule-ring | N schaal-punten | Soort schaalpunten |
|---------------|-------------|---|---------------------------|----------------------|-----------------|-------------------------|-----------------|--------------------|
| 3 | 151 | Overhalen tot stoppen met roken | Europese Unie | Heterogeen | Druk | Subjectief (Ik vind) | 7 | Cijfers |
| 4 | 155 | Overhalen tot lid worden astmafonds | Astmafonds | Studenten/ HBO-ers | Druk | Subjectief (Ik vind) | 7 | Hokjes |
| 5 | 120 | Overhalen tot aanvraag creditcard | Postbank | Studenten/ HBO-ers | Druk | Subjectief (Ik vind) | 7 | Cijfers + hokjes |
| 6 | 145 | Informeren over lesmethoden vak CKV2 | Overheid | Havo/ VWO-ers | Rustig | Subjectief (Ik vind) | 7 | Cijfers |
| 7 | 127 | Informeren over alternatieve geneeswijzen | Overheid | Havo/ VWO-ers | Rustig | Subjectief (Ik vind) | 7 | Cijfers |
| 8 | 110 | Overhalen tot boeken vakantie | Djoser | Studenten/ HBO-ers | Rustig | Subjectief (Ik vind) | 7 | Hokjes |
| 9 | 130 | Overhalen tot nemen donorcodicil | Overheid | Heterogeen | Druk | Objectief (De tekst is) | 7 | Cijfers |
| 10 | 150 | Overhalen tot standpunt voor sollicitanten 65+ers | Overheid | Studenten/ HBO-ers | Druk | Subjectief (Ik vind) | 7 | Cijfers + hokjes |
| 11 | 130 | Overhalen tot kopen borst vergrotende pillen | Breast gro | Heterogeen (vrouwen) | Druk | Objectief (De tekst is) | 7 | Cijfers + hokjes |
| 12 | 80 | Overhalen tot doneren geld aan armen | Hospitaal broeders | Heterogeen | Druk | Objectief + Subjectief | 5 | Cijfers |
| 14 | 210 | Informeren over nieuwe zorgstelsel | Overheid | Studenten/ HBO-ers | Druk | Subjectief (Ik vind) | 7 | Hokjes |
| 15 | 111 | Leden informeren over epilepsie & epilepsiefonds | Nationaal Epilepsie-fonds | Ouderen | Rustig | Objectief (De tekst is) | 5 | Hokjes |
| Totaal | 1619 | | | | | | | |

Bijlage 2: Overzicht van het aantal studies en proefpersonen per woordpaar

| Woordpaar | Studies nummers | Proefpersonen |
|---------------------------------|---|----------------------|
| Logisch/onlogisch opgebouwd | 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15 | 1619 |
| Makkelijk/moeilijk | 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15 | 1619 |
| Duidelijk/onduidelijk | 3, 4, 5, 6, 7, 8, 9, 10, 11, 14, 15 | 1539 |
| Overzichtelijk/onoverzichtelijk | 3, 4, 5, 6, 7, 8, 9, 10, 11, 14, 15 | 1539 |
| Boeiend/saai | 3, 4, 6, 7, 8, 9, 10, 11, 12, 14, 15 | 1499 |
| Interessant/oninteressant | 3, 4, 6, 7, 8, 9, 10, 11, 12, 14, 15 | 1499 |
| Bondig/omslachtig | 3, 4, 5, 6, 7, 8, 9, 11, 12, 14, 15 | 1469 |
| Persoonlijk/onpersoonlijk | 3, 4, 6, 7, 9, 10, 11, 12, 14, 15 | 1389 |
| Afwisselend/eentonig | 3, 4, 6, 7, 8, 9, 10, 11, 12, 14 | 1388 |
| Uitnodigend/afhoudend | 3, 4, 6, 7, 8, 10, 11, 12, 14, 15 | 1369 |
| Eenvoudig/ingewikkeld | 4, 5, 6, 7, 8, 9, 10, 11, 12, 14 | 1357 |
| Aansprekend/afstandelijk | 4, 6, 7, 11, 12, 14, 15 | 958 |
| Deskundig/ondeskundig | 3, 4, 5, 10, 11, 12, 15 | 897 |