

12 De invloed van pretest-instructies op de bruikbaarheid van de ontlokte oordelen

U. Schuurs, H. van den Bergh en G. Verhoeven

1. Inleiding

In de vakliteratuur wordt vaak gepropageerd om teksten te pretesten, zeker wanneer de betreffende teksten bedoeld zijn voor een groot publiek en/of de teksten een duidelijk persuasieve strekking hebben. In Nederland lijken vooral grotere organisaties en de overheid meer en meer de consequenties te trekken uit dit taalbeheersers-advies. Met name de zgn. plus-en-min-methode lijkt daarbij te worden gebruikt.

Met opzet houden we de nodige slagen om de arm: we spreken van "het lijkt erop dat" en "naar alle waarschijnlijkheid", omdat er over de feitelijke gang van zaken weinig bekend is. Er is geen systematisch onderzoek verricht naar de mate waarin gebruik gemaakt wordt van de bestaande pretest-methoden; evenmin is er iets bekend over de manier waarop teksten het beste aan een pretest onderworpen kunnen worden, over het rendement van de verschillende methoden, en over de mate waarin de gang van zaken bij het pretesten het resultaat van de pretest beïnvloedt.

In deze bijdrage gaan we in op de volgende vraag: in hoeverre wordt het commentaar van de proeflezers beïnvloed door de aard van de instructie? Het belang van deze vraag moge duidelijk zijn: als de resultaten van een pretest systematisch beïnvloed worden door de aard van de instructies, dan krijgt men ongewild een vertekend beeld van positief en negatief gewaardeerde tekstenmerken. Eerst moet dit soort bias onderkend zijn, wil men die in het vervolg kunnen vermijden. Omgekeerd valt te overwegen om een eventuele systematische invloed vanuit de aard van de pretest bewust te gebruiken om juist het soort commentaar te ontlokken waar je, gegeven een tevoren welomschreven doelstelling, met name in geïnteresseerd bent.

Eerst bespreken we enkele soorten pretest-methoden die in Nederland bekend zijn (par. 2), daarna beschrijven we het door ons ondernomen experiment: in par. 3 beschrijven we de onderzoeksofzet, in par. 4 presenteren we de resultaten, en in par. 5 bespreken we die.

2. Soorten pretests

In de wereld van de p.r. en de voorlichting is de pretest-activiteit vooral bekend geraakt door het werk van Van Woerkum (1987, p. 48): die benadrukt in zijn 'werkplan van de massamediale voorlichting' het belang van meermaals pretesten. Met name gaat hij in op de zgn. plus-en-min-methode (p. 278 e.v.).

Het pleidooi van Van Woerkum heeft, althans in de vakliteratuur, navolging gekregen. Zo wordt in Janssen (1989) een hoofdstukje gewijd aan een drietal verschillende pretest-methoden:

- de praktische effectiviteitstoetsing: een concept-tekst wordt voorgelegd aan een aantal leden van de doelgroep; het handelen naar aanleiding van de tekst, of expliciet commentaar (al dan niet in een hardop-denktprotocol verwoord) kan aanwijzingen voor verbetering geven.
- de interview-methode: nadat een concept-tekst is gelezen door een aantal vertegenwoordigers van de doelgroep, volgt met hen een gesprek waarin gevraagd wordt naar de mening van de lezers.
- de plus-en-min-methode: proefpersonen lezen de concepttekst, en plaatsen in de marge plusstekens als zij een passage positief waarderen, en een minteken als zij een passage afkeuren. In een nagesprek wordt ingegaan op de motieven die hebben geleid tot de plus- en de mintekens.

Ook in de Formulierenwijzer (Jansen et al. 1989, p. 171-196) is er expliciete aandacht voor verschillende manieren van pretesten. Al is de bespreking daarvan nadrukkelijk toegesneden op het toetsen van de bruikbaarheid van formulieren, toch geeft die bespreking een aardig beeld van uiteenlopende manieren om zinnig commentaar te ontlokken aan vertegenwoordigers van de doelgroep van het document in kwestie. Er worden drie verschillende benaderingen onderscheiden:

- de checklist: derden beoordeelen het formulier aan de hand van een lijst met aandachtspunten, of aan de hand van gesloten vragen;
- gesprekken met deskundigen: deze gesprekken kunnen plaatsvinden binnen een pretest-afname volgens het stramien van de eerder genoemde plus-en-min-methode; een variant is die waarbij beoordelaars achteraf gezamenlijk praten over hun bevindingen, zodat zij feedback kunnen geven op elkaars commentaar;
- de hardopwerk-aanpak: proefpersonen vullen het formulier in naar aanleiding van een fictieve situatieschets; tijdens het invullen verwoorden de proefpersonen zoveel mogelijk hun gedachten.

Wat valt op aan deze methodische besprekingen? In de eerste plaats zijn er weinig fundamentele verschillen: zo wordt niet duidelijk waarin de interview-methode zich nu precies onderscheidt van bijv. het gesprek met deskundigen.

In de tweede plaats valt er op grond van de methodische verschillen weinig of niets te voorspellen over de aard van het commentaar dat mag worden verwacht: we krijgen geen aanwijzingen over de methode die de voorkeur verdient als we bijv. snel inzicht willen krijgen in de persuasieve kracht van een folder, of als we geïnteresseerd zijn in de vraag of de schrijfstijl geschikt is voor het te bereiken doel.

In de derde plaats zijn niet alleen de verschillen tussen de besproken methoden onderling erg klein, ook dreigt het onderscheid tussen raadpleging van deskundigen en raadpleging van vertegenwoordigers van de doelgroep te vervagen. Zo lijken in de Formulierenwijzer de vertegenwoordigers van de doelgroep op één hoop te worden gegooid met de beoordelaars die inhoudelijk deskundig mogen worden geacht. Nu lijkt dit onderscheid van zichzelf al weinig helder: deze twee partijen hoeven elkaar immers zeker niet altijd uit te sluiten. Maar ook is duidelijk dat ze zelden of nooit volledig zullen samenvallen. Enig onderscheid tussen deskundigen en vertegenwoordigers van de doelgroep lijkt echter zeker wenselijk als we nadenken over de vraag hoeveel waarde de opmerkingen van beide groepen zou moeten worden toegekend.

Tenslotte valt op dat er maar één methode is die in beide overzichten wordt genoemd, namelijk de plus-en-min-methode. Deze methode is door Vroom (1987) uitvoerig besproken en geïllustreerd. In de opleiding Communicatiekunde aan de R.U. Utrecht fungeerde het artikel van Vroom in studiejaar 1989-'90 in maar liefst

drie syllabi als ingrediënt. Hoe werkt deze methode, wat zijn de opbrengsten en wat de nadelen?

Voor de beschrijving van de methode baseren we ons grotendeels op eigen ervaringen en op de beschrijving van Vroom (1987). De methode komt er kort en goed op neer, dat een tekst die bedoeld is voor massale verspreiding door een aantal leden van de doelgroep wordt beoordeeld. Daartoe wordt de tekst in zgn. content-units ingedeeld, en wordt de beoordelaars gevraagd om in de marge een plus te noteren als de passage (of de illustratie) goed vinden, en een minteken als ze de tekst minder geslaagd vinden. De beoordelingstaak heeft een volstrekt ongericht karakter: het is nadrukkelijk aan de beoordelaars om te bepalen aan welke aspecten ze aandacht schenken. Daardoor zijn de plus- en mintekens op zich weinigzeggend; het is dan ook noodzaak om met elke beoordelaar een nagesprek te houden waarin de motieven voor hun oordelen aan de orde komen¹. Naar aanleiding van de bevindingen wordt dan besloten welke content-units voor herschrijving in aanmerking komen². Vroom hanteerde voor de beslissing tot herschrijving de volgende norm: als 50% van de proefpersonen negatief reageerde op een passage, kwam die voor herschrijving in aanmerking, ook al waren andere beoordelaars met de betreffende passage tevreden.

Wat levert een dergelijke werkwijze op? Terecht gaat Vroom in tegen de suggestie van Van Woerkum als zou deze methode een combinatie zijn van effect-meting en beoordeling. Natuurlijk is niet direct aan de plus- en mintekens af te lezen of de boodschap door de doelgroep zal worden geaccepteerd. Ook is aan de reacties niet af te lezen of de beoogde doelgroep spontaan zelfs maar kennis zou willen nemen van de tekst. Wel levert de methode mogelijk aanknopingspunten voor verbetering, en dan met name op punten waarvoor de schrijver van de tekst zelf blind is.

Maar er kleven ook nadelen aan deze methode. We noemen er vijf: - voor het bespreken van de toegekende plussen is nadrukkelijk geen plaats in de methode. Natuurlijk is eenieder vrij om de methode in dezen naar eigen hand te zetten, maar het is geen standaard-optie³. Dat is vreemd, niet omdat de tekstschrijver verlegen zou zitten om pedagogisch verantwoorde schouderklopjes, maar omdat uit positieve reacties wellicht consequenties kunnen worden getrokken voor de inrichting van andere tekstdelen die minder positief zijn ontvangen; - de aspecten waarover een oordeel wordt uitgesproken, zijn zeer divers: nu eens worden -al dan niet vermeende- spelfouten bekritiseerd, dan weer wordt de persuasieve kracht van de tekst ter discussie gesteld of worden twijfels geuit aan het waarheidsgehalte van de tekst. In dit licht bezien is het overigens onthutsend, dat veel proefpersonen in het onderzoek van Vroom een minteken hebben geplaatst bij passages waarover ze heel tevreden waren: de reden voor het minteken was erin gelegen dat ze het in de betreffende passage beschreven gedrag afkeurden...

- een derde nadeel is, dat veel proefpersonen weliswaar fouten en gebreken signaleren, maar dat ze weinig aanwijzingen genereren die bruikbaar zijn in de revisiefase⁴; - op basis van de nagesprekken concludeert Vroom, dat maar liefst dat 44% (!) van de kritiekpunten door de beoordelaars niet wordt vertaald in plussen en minnen; ook onze ervaring is, dat veel proefpersonen lang niet alle kritiek weergeven in een minteken. Vergeleken bij de opmerkingen in de nagesprekken blijken de mintekens een nogal willekeurige steekproef uit de opmerkingen die de proefpersonen zoal hadden kunnen maken. De validiteit van de plus-en-min-methode is, anders gezegd, erg laag;

- tevoren moet de tekst worden ingedeeld in 'content units'. Dit moet gebeuren zonder dat de proefpersonen deze grenzen te zien krijgen, maar ook wel wordt gepretest terwijl de grenzen tussen deze units wel voor de proefpersonen zichtbaar zijn. In dat geval komt het er dan vaak op neer, dat de structuur van de tekst over de grenzen van de content-units heen niet meer in de beschouwing wordt betrokken. Een oordeel over de samenhang tussen units wordt daarmee uitgesloten, terwijl het succes van een folder vaak juist afhangt van de consistentie en de overtuigingskracht van de tekst als geheel...

De vraag is nu, of deze nadelen niet zijn te ondervangen. Wellicht is een meer gerichte instructie aan de beoordelaars al een stap op weg: wanneer voor de beoordelaars duidelijk is waarop met name zou moeten worden gelet, kunnen zij de onderzoeker wellicht meer aanwijzingen geven over de aspecten waarin die met name is geïnteresseerd; de beoordelaars kunnen zich dan op enkele aspecten concentreren, waardoor zij vollediger kunnen zijn in hun kritiek, en wellicht ook beter duidelijk kunnen maken in welke richting een verbetering te vinden is. Zo wordt ook de kans op onterechte kritiek (d.w.z. op kritiek gebaseerd op afkeuring van de inhoud) geminimaliseerd. Tenslotte zou met een gerichtere instructie ook de signaleerde verontachtzaming van de samenhang over content-unitsgrenzen heen kunnen worden tegengegaan.

3. Onderzoeksofzet

Wij wilden onderzoeken of de aard van de instructie bij het pretesten invloed heeft op het soort commentaar dat ermee wordt ontlokt⁵. Om op die vraag gefundeerd antwoord te geven, hebben we een experiment ondernomen waarbij ongeveer 100 proefpersonen twee verschillende, persuasieve teksten moeten beoordelen volgens drie varianten op de plus-en-minmethode. Iedere te beoordelen tekst bevatte opzettelijk aangebrachte fouten op vier dimensies (spelling, zinsbouw, tekststructuur, publiekgerichtheid). Het experiment was opgezet volgens het zgn. completely randomized balanced block design.

De drie beoordelingsmanieren zijn als volgt te beschrijven:

- a. een "globale" methode waarbij de tekst beoordeeld wordt zonder dat die is ingedeeld in content units;
- b. de "analytische" plus-en-min-methode: beoordelaars geven commentaar bij een tekst die tevoren door de proefleider is ingedeeld in content units;
- c. een "specifieke" methode waarbij beoordelaars de instructie hebben gekregen om met name te letten op specifieke tekstvariabelen. In ons onderzoek hebben we de proefpersonen in deze conditie gevraagd met name te letten op tekstfragmenten die afbreuk zouden kunnen doen aan de publiekgerichtheid en de overtuigingskracht van de tekst.

De te beoordelen teksten bestaan uit door ons bewerkte fragmenten uit bestaande folders over resp. de Treintaxi en over het Rookverbod in openbare ruimten. In elk van de teksten zijn vijf spelfouten en vijf grammaticale fouten ingebracht; daarnaast bevat elke tekst vijf zinsfragmenten die vanuit het oogpunt van publiekgerichtheid als zeer onhandig kunnen worden gekenschetst. Bijvoorbeeld begint in de door ons gebruikte versie over het rookverbod een zin met de woorden "Een enkele verstokte

roker zal zich misschien nog afvragen ...", terwijl de proefpersonen eerder was duidelijk gemaakt dat de tekst als doel had bij rokers begrip te kweken voor de maatregel. Tenslotte bevat elke tekst een drietal structureringsfouten: bijvoorbeeld wordt in de tekst over de treintaxi de informatie over de heenrit pas in de laatste alinea gepresenteerd⁶. Figuur 1 bevat, bij wijze van voorbeeld, een overzicht van de in totaal 18 fouten in de tekst over Rookvrije ruimten.

Figuur 1: Overzicht van fouten in de tekst Rookvrije ruimten

Spelfouten:

1. bedoelt ipv bedoeld: "Daarmee worden gebouwen bedoelt die ..."
2. binen ipv binnen
3. vind ipv vindt: "En tot slot vind 75 % van de bevolking dat rokers ..."
4. verkeerde afbreking: benauwdheid ipv benauwdheid
5. betekend ipv betekent: "Gedwongen meeroken betekend voor mensen ..."

grammaticale fouten:

1. andere ipv anderen
2. dat is waar het (ROOKVERBOD) voor bedoeld is
3. Het aantal rokers in ons land fors daalt
4. roken moet <worden> verboden <worden>
5. giftige bestanddelen (...) blijft

vorm/verwijzing
ref. verwijzing
woordvolgorde
weglating
incongruentie

publiekgerichtheid:

in de tweede zin van de toets was door ons nadrukkelijk vermeld: "Misschien hebt u er als roker al een en ander over gehoord". Conclusie: de rokers vormen de primaire doelgroep van de folder. Het ligt voor de hand dat die aanstoot nemen aan frasen als:

1. "Eindelijk een verbod waarbij we allemaal opgelucht adem kunnen halen"
2. "Een enkele verstokte roker zal zich misschien nog afvragen ..."
3. "(deze maatregel) is voor ons een hele opluchting..."
4. "Roken is slecht voor de gezondheid. Dat weet ieder kind."
5. "Daarom is roken a-sociaal."

alinea-/tekststructuur.

1. In de toets luidt de eerste alinea als volgt:

(1) *Eindelijk* een verbod waarbij we allemaal opgelucht adem kunnen halen; (2) het rookverbod in *openbare gebouwen* is per 1 januari wet. (3) Over het hoe en waarom van dit verbod kunt u in deze folder meer lezen. (4) Misschien hebt u er als roker al een en ander over gehoord of gelezen.

Deze alinea is raar opgebouwd. Beter is b v. de volgorde zin 2 - 4 - 3 (en deletie van zin 1 want die ontkracht de event. persuasieve kracht van de tekst in hoge mate), of volgorde 4 - 2 - 3.

2: een alinea die duidelijk als afsluiting is bedoeld (en in het origineel ook als zodanig fungeerde) hebben we naar voren gehaald, zodat drie krachtige argumenten tegen gedwongen laten meeroken nu komen vóór een meer technische uiteenzetting over de schadelijke effecten van roken die niet als afsluitende alinea kan fungeren...

- 3: de toets bevat onder meer het volgende fragment:

"Terecht dat veel niet-rokers zich afvragen of ze risico's lopen wanneer ze *gedwongen* worden de rook van anderen in te ademen. Roken is slecht voor de gezondheid".

Hier suggereert het woord "terecht" dat er voor de eerste zin van de alinea al een verklaring is gegeven; de verklaring volgt er echter op.

Nadrukkelijk hebben we in beide toetsen dezelfde soorten fouten opgenomen, zij het dat de volgorde ervan niet in beide toetsen gelijk gehouden kon worden. Dit illustreren we aan de hand van de grammaticale fouten in beide toetsen (figuur 2):

Figuur 2: De grammaticale fouten in beide tests

grammaticale fouten in toets 1:

1. andere ipv anderen
2. dat is waar het (ROOKVERBOD) voor bedoeld is
3. Het aantal rokers in ons land fors daalt
4. roken moet <worden> verboden <worden>
5. giftige bestanddelen (...) blijft

onterecht enkelvoud
ref. verwijzing
woordvolgorde
weglating
incongruentie

grammaticale fouten in toets 2:

1. deelt u hem >>> deelt u de taxi
2. U moet wel een geldig vervoerbewijs <hebben> gekocht <hebben>
3. Als u in een van de taxi's gestapt bent, wachten die ...
4. Verder worden degene(N) die ...
5. Als u weer wilt terug

vgl. test 1:

2
4
5
1
3

De verwachting was, dat

- de globale methode tot betere resultaten zou leiden waar het de tekststructuur betreft ($a > b = c$).
 - de analytische methode tot betere resultaten zou leiden op microniveau, dus dat bij die instructie de spelfouten en de grammaticale fouten het meest commentariseerd zouden worden ($b > a = c$).
 - de specifieke methode tot betere resultaten zou leiden op macroniveau, dus dat bij de specifieke instructie de fouten op het vlak van de publiekgerichtheid en in de structuur het meest commentariseerd zouden worden ($c > b = a$).
- Deze verwachtingen kunnen als volgt worden gevisualiseerd (een * duidt een significant hoger aantal herkenningen aan in de betreffende conditie):

Figuur 3: Visualisering van gehanteerde hypothesen

	analytisch	specifiek	globaal
spelfouten	*		
grammatica	*		
publiek		*	
structuur			*

Als de gehanteerde hypothesen juist zijn, zou men bewust kunnen kiezen voor het gebruik van een bepaalde methode afhankelijk van de eisen die men aan de uiteindelijke foldertekst wil stellen.

In het onderzoek hebben we gebruik gemaakt van het 'completely randomized balanced block design'. Daarmee konden we de drie levels van de onafhankelijke variabele (resp. de Globale methode, de Analytische methode en de Specifieke methode) combineren met de twee levels van de afhankelijke variabele (resp. de Treintaxi en Rookvrij ruimten): eventuele volgorde-effecten zijn uitgesloten door bij de ene helft van de proefpersonen de volgorde Treintaxi - Rookvrij te gebruiken, en bij de andere helft de volgorde Rookvrij - Treintaxi. Het resultaat is een volledig counter-balanced design.

In het experiment fungeerden als proefpersoon de studenten die in mei-juni 1990 deelnamen aan de doctoraal-cursus "Formuleren" aan de R.U. te Utrecht. In het eerste en het laatste werkcollege werd telkens een half uur uitgetrokken om de studenten een van de twee teksten te laten beoordelen.

4. Resultaten

De toetsen werden begin mei resp. eind juni gemaakt door studenten die deelnamen aan de cursus Formuleren van afdeling Communicatiekunde R.U.U. Na afloop bleken 93 proefpersonen aan zowel voor- als natoets te hebben deelgenomen. Van deze proefpersonen zijn de toetsresultaten in de verdere analyse betrokken. Omdat in dit onderzoek alleen de foutherkenning in verschillende condities centraal stond, hebben we alleen op de toegekende minnetjes gelet; daarvan hebben we alleen die waarden geselecteerd die -blijkens het gevraagde schriftelijke commentaar- daadwerkelijk op de door ons ingebrachte fouten betrekking hadden.

In tabel 1 presenteren we de gemiddelde aantallen foutherkenningen, gesommeerd over teksten en condities:

	SPELLING	GRAMMATICA	PUBLIEK	STRUCTUUR
Overall (n=93)	2.16 (1.6)	2.09 (1.4)	1.61 (1.3)	.39 (.6)

Tabel 1: Aantal foutherkenningen, gemiddeld over teksten en condities

We hebben de resultaten getoetst met behulp van een multivariate variantie-analyse voor een 'within-subjects'-design. Deze analyse laat zien, dat de gemiddelde som-scores over de vier soorten fouten heen niet significant worden beïnvloed door de conditie ($p = .099$), en evenmin door de tekst die voorgelegd is ($p = .612$). Ook blijkt er geen interactie op te treden tussen tekst en conditie ($p = .154$). We kunnen op basis hiervan concluderen dat er geen verschil is tussen de *somscores* per conditie.

Omdat deze eerste analyse betrekking heeft op het totaal aantal herkenningen over foutsoorten heen, hebben we de interactie tussen foutsoort en conditie aan een nadere analyse onderworpen: in de hypothesen (par. 3) gaat het immers ook om de interactie tussen foutsoort en conditie: de analytische methode zou leiden tot meer herkenningen van fouten op microniveau, de specifieke methode tot meer herkenningen van fouten op het gebied van de publiekgerichtheid en de globale methode tot meer herkenningen van gebreken in de structurering.

Ook de interactie tussen foutsoort en tekst hebben we aan een nadere analyse onderworpen. Deze interactie is van belang: we willen natuurlijk dat een eventuele invloed vanuit de gebruikte methode op de oordelen zich systematisch voordeelt, ongeacht de tekst die we de proefpersonen voorleggen (convergente validiteit). De interactie tussen foutsoort en tekst zou afwezig moeten zijn, omdat we anders moeten concluderen dat weinig méér kunnen concluderen dan dat het aantal herkenningen beïnvloed wordt door het tekstonderwerp, en er in feite een heel nieuw onderzoeksterrein wordt blootgelegd.

Uit deze analyse blijkt dat er een interactie bestaat tussen foutsoort en conditie ($p = .039$). De verschillende foutsoorten worden kennelijk in de ene conditie meer opgemerkt dan in de andere. Dit wordt duidelijk in tabel 2, waarin we de gemiddelden per conditie per foutsoort presenteren:

	SPELLING	GRAMMATICA	PUBLIEK	STRUCTUUR
conditie 1 (n=68)	2.13 (1.6)	2.15 (1.5)	1.90 (1.4)	.37 (.6)
conditie 2 (n=68)	2.43 (1.6)	2.28 (1.4)	1.35 (1.2)	.44 (.6)
conditie 3 (n=50)	1.84 (1.6)	1.76 (1.3)	1.56 (1.2)	.36 (.6)

Tabel 2: Gemiddeld aantal foutherkenningen per conditie per foutsoort

Conform de tweede hypothese leidt de "analytische" pretest-methode (met een voor de proefpersonen zichtbare indeling in content units) tot het grootste aantal foutherkenningen op het gebied van de spelling en de grammatica. De eerste hypothese blijkt niet juist: het is niet de globale methode die leidt tot het grootste aantal signaleringen van gebreken in de tekststructuur, maar -verrassend genoeg- de analytische aanpak. Ook de derde hypothese moet worden verworpen: het is niet de specifieke, maar juist de globale methode die leidt tot het grootste aantal herkenningen van gebreken in de publiekgerichtheid; dat is verrassend, omdat in de derde methode de proefpersonen nadrukkelijk werd gevraagd te speuren naar dit soort fouten.

Verder zouden we nagaan of er een convergente validiteit bestaat tussen de twee gebruikte teksten. De analyse van de interactie tussen foutsoort en tekst laat echter ook een significant resultaat zien ($p = .002$). De invloed die de tekst heeft op het aantal herkenningen per foutsoort is af te lezen aan tabel 3:

	SPELLING	GRAMMATICA	PUBLIEK	STRUCTUUR
Treintaxi (n=94)	2.38 (1.6)	1.95 (1.3)	1.40 (1.1)	.27 (.5)
Rookvrij (n=92)	1.93 (1.6)	2.24 (1.6)	1.82 (1.4)	.52 (.6)

Tabel 3: Gemiddeld aantal foutherkenningen per tekst per foutsoort

Bij de tekst over de Treintaxi wordt wat meer gelet op spelfouten, en minder op de andere drie foutsoorten. Een aanvullende variantie-analyse laat zien dat er tussen de twee teksten significante verschillen bestaan waar het betreft de fouten op het gebied van

- de spelling ($p = .03$);
- de publiekgerichtheid ($p = .05$);
- de tekststructuur ($p = .00$).

Waar het de grammaticale fouten betreft doen zich geen significante verschillen voor.

De mate waarin de verschillende soorten fouten worden opgemerkt, is dus in hoge mate afhankelijk van de tekst die de proefpersonen krijgen voorgelegd. Daardoor is het niet mogelijk om *in het algemeen* te stellen dat de aard van de instructie bij het pretesten invloed heeft op het soort commentaar dat ermee wordt ontlokt.

5. Discussie

Onze vraag was of het soort instructie bij het pretesten effect had op de oordelen van de respondenten. Voor drie soorten van instructie hebben we bij wijze van hypothese aangegeven wat het mogelijke effect zou kunnen zijn voor verschillende tekstkenmerken. Gebleken is, dat de instructie conform de hypothesen een wisselende invloed heeft per tekstkenmerk: inderdaad leidt visualisering van de grenzen tussen content units tot meer aandacht voor het micro-niveau (spelling en grammatica). Deze visualisering leidt ook tot iets meer aandacht voor structureringsproblemen.

Problemen die de publiekgerichtheid betreffen, worden het beste opgemerkt wanneer de proefpersonen een "kale" tekst wordt gepresenteerd zonder verdere aanwijzingen. De instructie om met name te letten op de publiekgerichtheid van de tekst heeft geen enkel positief effect. Het is de vraag hoe dat komt. Wellicht was de taakstelling te complex, al zou men juist van doctoraal-studenten Nederlands op dit punt enige vaardigheid mogen verwachten.

Terwijl de verschillende instructies leiden tot uiteenlopende effecten per foutsoort, doen zich ook significante verschillen voor tussen de reacties op de twee gebruikte teksten. Dit resultaat is des te verontrusterend omdat we hebben gestreefd naar zoveel mogelijk gelijkheid tussen de fouten in de twee teksten. Het lijkt erop dat bij een tekst met een wat minder emotioneel beladen onderwerp (de Treintaxi) de aandacht vooral uitgaat naar op zich meer triviale fouten op het gebied van de spelling, terwijl een tekst over een onderwerp dat meer aanspreekt vooral wordt gelet op de publiekgerichtheid en op grammaticale aangelegenheden. Over de oorzaken hiervan kunnen we alleen speculeren. Waar het de publiekgerichtheid betreft, geldt misschien het volgende: moet een tekst worden beoordeeld waarvan het onderwerp intrigeert, dan is het wellicht zo dat beoordelaar fungeert als een geboeid lezer die elk gebrek aan publiekgerichtheid als "van nature" zal opvallen. Waar het de grammatica betreft, fungeren wellicht de gemanipuleerde grammaticale aspecten als een onderdeel van de publiekgerichtheid.

Doordat er geen sprake is van convergente validiteit, zijn er geen generaliserende uitspraken mogelijk over het effect van de instructie op de mate van fourth-kenning in pretest-situaties. Het lijkt erop dat in toekomstig onderzoek de aandacht

mede uit moet gaan naar de mate waarin de lezer geboeid is door het onderwerp van de tekst.

Noten

1. Vroom (1987) heeft niet alleen met de 'klassieke' mondelinge variant gewerkt, maar ook met een schriftelijke variant: ook hier plaatsen de proefpersonen plussen en minnen in de marge, maar zij voorzien die na afloop van schriftelijk commentaar buiten de afwezigheid van de proef-leider.
2. In de praktijk wordt wel gevraagd één oordeel per content-unit te geven; Vroom (1987) heeft zijn proefpersonen gevraagd om allerlei zaken van een plus- of minteken te voorzien, geheel vrij naar het inzicht van de beoordelaars. De indeling in content-units fungeerde daar alleen om de proefleider duidelijkheid te geven omtrent de grenzen waarbinnen eventueel herschrijving zou moeten plaatsvinden.
3. Vergelijk ook Vroom (1987, p. 262), die stelt dat met name van positieve oordelen vaak niet duidelijk is op welk tekstelement ze eigenlijk betrekking hebben.
4. In het algemeen geldt overigens, dat er bijzonder weinig bekend is over de kenmerken die feedback bruikbaar maken in een revisiefase. Verder onderzoek zou ook betrekking moeten hebben op de vraag wat voor soort opdracht zou moeten worden verstrekt om te waarborgen dat de feedback ook de gewenste kenmerken bezit.
5. Bewust hebben we niet gekozen voor onderzoek naar de vraag welke methode tot de beste, d.w.z. tot de meest bruikbare aanwijzingen zou leiden. Zulk normatief onderzoek impliceert immers dat er eenduidige normen zouden bestaan waaraan iedere voorlichtingstekst zou moeten voldoen. Waar het bijv. de publiekgerichtheid of de inhoudskeuze betreft zijn er naar ons idee geen algemene normen te formuleren. Juist bij gebrek aan zulke normen wordt gebruik van enigerlei vorm van pretest zinnig.
6. Overigens was dat ook het geval in de door de NS uitgegeven tekst.

Literatuur

- Jansen, C., M. Steehouder, K. Edens, J. Mulder, H. Pander Maat, P. Slot
1989 *Formulierenwijzer / Handboek Formulieren Redigeren*. 's Gravenhage.
- Janssen, D. (red.)
1989 *Zakelijke communicatie II*. Groningen.
- Vroom, B.
1987 *Publieksonderzoek met behulp van de plus-en-min methode*. In: *Tijdschrift voor Taalbeheersing* 9: 256-271.
- Van Woerkum, C.
1987 *Massamediale Voorlichting / Een werkplan*. Meppel.