

Meer teksten, natuurlijk – maar hoeveel en welke?

Hans Hoeken

TT 38 (2): 153–155

DOI: 10.5117/TVT2016.2.HOEK

Abstract

More messages, of course – but how many and what kind of messages?

It is solid advice to employ more messages in one's experiments when conducting research on message variables' effects. The question is how many messages should be used? And what characteristics should these messages have? Compared to the insights and guidelines available for the selection mode and number of participants when aiming to generalize to a population, such insights and guidelines are sorely lacking when aiming to generalize over messages.

Keywords: methodology, external validity, message effects, message design

Het is een goed advies: als je onderzoek doet naar de effecten van tekstkenmerken, kun je niet volstaan met één tekst. De (vaak impliciete) claim van veel onderzoek op ons terrein is dat het gevonden effect van de onderzochte manipulatie, voor deze specifieke tekst en voor deze specifieke proefpersonen, ook zal optreden als we deze manipulatie uitvoeren op een vergelijkbare tekst en die vervolgens laten lezen door een vergelijkbare groep proefpersonen. De p-waarden van de uitgevoerde statistische toetsen geven daarbij een indicatie van de waarschijnlijkheid dat het gevonden effect op toeval berust; naarmate die kans kleiner is, neemt ons vertrouwen toe dat we opnieuw een effect zouden vinden als het onderzoek wordt herhaald met een vergelijkbaar aantal vergelijkbare proefpersonen.

Zoals Meuffels en Van den Bergh (2006) al hebben betoogd, zegt dit echter niets over de kans dat we een vergelijkbaar effect zouden vinden als we dit onderzoek zouden herhalen met eenzelfde manipulatie in een andere tekst, zelfs niet als we die nieuwe tekst zouden voorleggen aan precies dezelfde proefpersonen uit de eerste studie. Om een uitspraak te

doen over de generaliseerbaarheid van het effect van de manipulatie voor andere teksten, zul je die manipulatie in een groot aantal teksten moeten uitvoeren om zicht te krijgen op de generaliseerbaarheid van het gevonden resultaat.

1 Een populatie van teksten?

Het lijkt heel simpel: net zoals je een groep proefpersonen gebruikt in je onderzoek om te kunnen generaliseren naar andere mensen, moet je een groep teksten gebruiken om te kunnen generaliseren naar andere teksten. Het grote verschil is echter dat er bij proefpersonen richtlijnen zijn voor hoe onderzoekers die moeten selecteren om te kunnen generaliseren. Als ze bijvoorbeeld een uitspraak willen doen over de invloed van een tekstkenmerk op het Nederlandse volk, dan houdt dat in dat ze een lijst moeten aanleggen van de gehele bevolking, om vervolgens een steekproef van voldoende omvang te trekken waarbij elk lid van de Nederlandse bevolking precies evenveel kans heeft om in die steekproef opgenomen te worden. Dat we dit in ons onderzoek nooit doen, en dat we ons vaak tevreden stellen met gelegenheidssteekproeven, doet voor mijn betoog niet ter zake. Het gaat mij er om dat er duidelijke richtlijnen zijn voor de onderzoeker die graag uitspraken zou willen doen over de Nederlandse bevolking, want wie een onderzoek wil uitvoeren waarbij hij of zij een uitspraak wil doen over “alle Nederlandse teksten” moet dergelijke richtlijnen ontberen.

Daar zijn verschillende redenen voor. In de eerste plaats ontbreekt een steekproefkader: we hebben geen lijst van alle Nederlandse teksten zodat we (1) geen idee hebben hoe groot onze steekproef zou moeten zijn en (2) niet in staat zijn om een steekproef te trekken waarbij elke tekst een even grote kans heeft om in die steekproef opgenomen te worden. Maar zelfs al zouden we die hebben, dan is er nog een probleem: in taalbeheersings-experimenten is vaak maar één van de experimentele tekstversies (en soms zelfs géén van die versies) een in het echt voorkomende tekst, zoals een gezondheidsbrochure of gemeentelijke website. Van die tekst (of site) wordt een zorgvuldige gemanipuleerde versie afgeleid, eentje die slechts op het te onderzoeken tekstkenmerk verschilt van de originele tekst. Dus het liefst hebben we een tekst over hetzelfde thema, die even lang is en vanuit een lexicaal, syntactisch en discourseperspectief even complex. Dat is vanuit experimenteel oogpunt wenselijk en noodzakelijk, maar leidt wel tot een vergelijking tussen een bestaande en een door de onderzoeker ge-

creëerde tekst. Die onderzoekers creëren daarmee dus een parallel universum van Nederlandse teksten en een te generaliseren effect zou dus voor de helft op niet bestaande teksten zijn gebaseerd (of meer, als beide versies door de onderzoekers zijn ontworpen).

Een mogelijke oplossing zou zijn om eerst een corpusanalyse uit te voeren waarin de kenmerken worden beschreven van de teksten waarin het te onderzoeken kenmerk voorkomt. Dan krijg je een beeld van de variatie in lengte, thema's, doelgroepen en genres waarin het kenmerk voorkomt en daarmee ook een idee van de variatie in het experimenteel materiaal dat je moet nastreven om tot een verantwoorde generalisatie te komen. Wellicht dat 'big data'-achtige benaderingen van grote tekstcorpora een dergelijk overzicht in de toekomst tot stand kunnen brengen. Dergelijk corpora zouden ook de verschuivingen in ontwerpkenmerken in kaart kunnen brengen die in de loop der tijd ongetwijfeld zullen optreden; gemeentelijke websites van nu, zien er heel anders uit dan die van tien jaar geleden. Maar voor het zover is, blijven we met de vraag zitten: meer teksten, natuurlijk, maar hoeveel en welke?

Referentie

Meuffels, B. & Bergh, H. van den (2006). De ene tekst is de andere niet: The language-as-a-fixed-effect fallacy revisited: Statistische implicaties. *Tijdschrift voor Taalbeheersing*, 28 (4), 323-345.

Over de auteur

Hans Hoeken is hoogleraar Communicatie- en Informatiewetenschappen aan de Universiteit Utrecht bij het Departement Talen, Literatuur en Communicatie en bij het Utrechts Instituut voor Linguïstiek OTS.

E-mail: j.a.l.hoeken@uu.nl

