

Kunnen wij leesbaarheid voorspellen?

Een experimenteel onderzoek naar de relatie tussen leeservaring en voorspelkracht van syntactisch gemanipuleerde teksten

Lucas v. H. S. Pompe (L.vanhoeijschilthouwerpompe@students.uu.nl)

4167546

Universiteit Utrecht

27 Juni 2016

Abstract

Dit paper beschrijft een onderzoek naar het verschil tussen de waargenomen leesbaarheid van teksten en de werkelijke leesbaarheid en de relatie daarvan met de leeservaring van de lezer. In een experimentele opzet met 226 proefpersonen bevestigen wij een van de twee hypothesen. Het verschil tussen de voorspelde en werkelijke leesbaarheid van een tekst blijkt negatief samen te hangen met de woordenschat van de lezer. Dit onderzoek kent enkele generalisatieproblemen, met name op het gebied van opleidingsniveau.

Inleiding

De popularisatie van het internet zorgde voor een extreme toename in de hoeveelheid beschikbare informatie. Reclame teksten, maar ook verzekeringspolissen en bijsluiters zijn tegenwoordig makkelijk te vinden en te lezen op websites van de relevante organisaties. Vooral organisaties met sterk gespecialiseerde functies proberen manieren te vinden om hun ingewikkelde boodschappen toch aan een zo groot mogelijk publiek over te dragen. Veel van deze organisaties moeten afwegingen maken over het opofferen van specialistische informatie om de leesbaarheid van hun boodschappen te verbeteren. Om de leesbaarheid van dergelijke boodschappen te toetsen zijn in het verleden verschillende leesbaarheidsformules ontwikkeld. De formules bekijken enkele variabelen in teksten, en bepalen zo een leesbaarheidsscore. Een voorbeeld hiervan is de *Flesch-Kincaid* formule. Deze formule definiëert de leesbaarheidsscore als een functie van het aantal woorden in de tekst, het aantal zinnen in de tekst en het aantal lettergrepen in de tekst, maar de validiteit van deze formule verschilt sterk per tekst (Benjamin, 2011). Naast dit soort traditionele formules zijn met de toename van de rekenkracht van computers ook statistische afhankelijkheidsmethodes zoals *Hidden Markov modellen* (HMM) in populariteit toegenomen. Deze methodes bestaan in allerlei vormen, maar kunnen bijvoorbeeld de complexiteit van een zin berekenen door middel van de zeldzaamheid van woordcombinaties.

Hiernaast kiezen de eerder genoemde organisaties vaak voor een meer menselijke aanpak zoals focusgroepen of simpelweg experimenteel onderzoek. In dit type onderzoek zien we vaak proefpersonen die moeten invullen hoe leesbaar ze denken dat teksten zijn. Deze focusgroepen zijn vaak een te kleine steekproef om uitspraken te doen over de werkelijkheid. Bovendien zullen niet alle lezers even veel ervaring hebben met lezen. Deze weinig ervaren lezers kunnen meer moeite hebben met het onderscheiden van teksten met verschillende moeilijkheidsgraden. Wellicht zullen zij dus ook

meer moeite hebben met het voorspellen van de leesbaarheid van teksten.

Waar eerder twijfels waren over computationele methodes om leesbaarheid te voorspellen, zullen wij in dit paper kijken naar de nuances van het voorspellen van leesbaarheid door mensen.

Metabegrip

Om uitspraken te doen over de leesbaarheid van een tekst zullen lezers hun eigen kennis moeten kunnen beoordelen. De leesbaarheid van een tekst zal namelijk per persoon verschillen, afhankelijk van de ervaring als lezer. Het begrip van een lezer over zijn eigen begrip noemen we *metacomprehension* (metabegrip). Een zeer ervaren lezer met een hoge mate van metabegrip zal zo bijvoorbeeld kunnen inzien dat een tekst met lage begrijpelijkheid voor hem waarschijnlijk begrijpelijker is dan voor de rest van de populatie.

Begg et al. (1989) lieten een groep proefpersonen over een verzameling woorden voorspellen hoe goed zij dachten dat ze de woorden konden onthouden. Een andere groep kreeg de opdracht de verzameling woorden uit het hoofd te leren, en vervolgens de woorden te beoordelen op hoe goed zij dachten de woorden te hebben onthouden. Beide groepen beoordeelden hoog frequente woorden als makkelijker te onthouden dan laag frequente woorden. Na het toetsen van de geleerde woorden bleek dat de proefpersonen hoog frequente woorden minder goed hadden onthouden dan laag frequente woorden. Zij concluderen dat het gemak waarmee men woorden verwerkt een basis vormt voor de voorspellingen die zij maken over hoe goed zij die woorden onthouden. Dit is de *ease of processing* hypothese.

Rawson & Dunlosky (2002) breidden deze hypothese vervolgens uit door deze niet slechts te toetsen op lijsten met woorden, maar op paren van coherente zinnen. Rawson & Dunlosky (2002, p. 69) voerden twee experimenten uit om deze hypothese op volledige zinnen te toetsen, gevolgd door twee experimenten met hele stukken tekst. In deze experimenten zijn de stukken tekst gemanipuleerd op causale coherentie. Een lagere causale coherentie zou de *ease of processing* moeten verminderen vanwege moeizame integratie in het *situatiemodel* (Rawson & Dunlosky, 2002, p. 70). Het *situatiemodel* is een mentale representatie van de situatie die een tekst schetst (Zwaan & Radvansky, 1998). Als de tekst nieuwe informatie presenteert zal de lezer deze informatie in zijn huidige situatiemodel proberen te integreren. Als de lezer nieuwe informatie probeert te integreren in het situatiemodel,

maar deze informatie is weinig coherent met het bestaande model, zal de integratie moeizamer verlopen. Dit moeizame verloop zal voor de lezer worden ervaren als een lage *ease of processing*. Aan de hand van de gelijknamige hypothese stellen Rawson & Dunlosky (2002, p. 70) dat een lagere causale coherentie dus zal zorgen voor een lagere beoordeling op hoe makkelijk stukken tekst te onthouden zijn. In de eerste twee experimenten manipuleerden Rawson & Dunlosky (2002) causale coherentie.

De resultaten van beide experimenten ondersteunden de *ease of processing* hypothese. Proefpersonen rapporteerden dat zij de hoog coherente paren van zinnen beter dachten te onthouden. De volgende twee experimenten richtten zich op stukken tekst in plaats van zinnen. Ook in deze paragrafen vonden Rawson & Dunlosky resultaten die de *ease of processing* hypothese ondersteunen. Proefpersonen lazen paragrafen met gemanipuleerde causale samenhang. Een lagere samenhang resulteerde in proefpersonen die de tekst als hoger beoordeelden op voorspelde moeilijkheid van het onthouden van de tekst.

De resultaten van Begg et al. (1989) en Rawson & Dunlosky (2002) wijzen er op dat de *ease of processing* een sterke invloed heeft op voorspellingen die men maakt over de prestaties in het onthouden van teksten, maar zij keken niet naar de ervaring die lezers hadden met lezen. Maki et al. (2005) bouwden voort op de bevindingen van Rawson & Dunlosky (2002) door verschillen te onderzoeken tussen lezers van verschillende ervaringsniveaus. Als uitgangspunt gebruikten zij het onderzoek van Kruger & Dunning (1999): "Unskilled and unaware of it". Dit paper vormt de basis voor wat nu het *Kruger-Dunning* effect heet, de illusie van onwetendheid. Dit houdt in dat onwetendheid eerder resulteert in zelfvertrouwen dan dat de aanwezigheid van kennis dit doet.

Maki et al. (2005) verwachtten dan ook een positieve relatie tussen de nauwkeurigheid van de voorspellingen - de voorspelkracht - en het verbale uitdrukkingsvermogen. De resultaten van het bijbehorende experiment ondersteunen deze verwachtingen. Maki et al. (2005) verdeelden 159 studenten in drie groepen, afhankelijk van hun prestaties op het verbale gedeelte van de scholastic aptitude test (SAT). Deze drie groepen zijn onderling ook weer in drie groepen verdeeld, voor lastige, middellastige, en makkelijkere teksten. De lastigheid van de teksten is gemanipuleerd door laagfrequente woorden te vervangen door hoogfrequente synoniemen. De proefpersonen gaven aan hoe goed ze dachten de tekst te hebben onthouden, en beantwoordden daarna vragen die dit toetsten. Het resultaat was een interactie effect van de SAT-prestatie en voorspelling op de lastige teksten. bovendien waren minder ervaren studenten typisch overmoedig waar dit voor ervaren lezers andersom was.

Hacker et al. (2000) deden een vergelijkbaar onderzoek. Studenten werden verdeeld door middel van prestaties bij een vak dat ze volgden. Over een langere periode werden studenten 5 maal getoetst. Voor en na iedere toets maakten studenten voorspellingen over hun resultaten. Hieruit

bleek dat studenten met hogere prestaties in het vak betere voorspellingen maakten dan studenten met lagere prestaties. Bovendien vond ook het paper van Hacker et al. (2000) dat proefpersonen met hogere prestaties zich hadden onderschat, terwijl proefpersonen met lagere prestaties zich hadden overschat. In dit paper zullen wij proberen deze resultaten uit te breiden naar een bredere groep proefpersonen. Zo komen we op onze eerste hypothese:

H_1 : Het verschil tussen de voorspelde en de werkelijke leesbaarheid van een tekst hangt negatief samen met geletterdheid.

Dit verschil is in werkelijkheid de inverse van de voorspelkracht zoals Maki et al. (2005) die noemden. Dat wil zeggen dat mensen met hogere geletterdheid een voorspelling zullen maken die dichter bij de werkelijke leesbaarheid van de tekst ligt dan voorspellingen van lezers met lagere geletterdheid. Of andersom: de voorspellingen van lager geletterde lezers zullen minder dicht bij de werkelijke leesbaarheid van de tekst liggen dan voorspellingen van hoger geletterde lezers.

Ervaring

Ervaring heeft volgens Rawson & Dunlosky (2002) een sterke samenhang met de mate van *metabegrip*. Als men beter is in *metabegrip* verwachten we dus dat men een meer accurate voorspelling kan maken over de begrijpelijkheid van een tekst. Bovendien verwachten we dat het verschil tussen de voorspelling en de werkelijke leesbaarheid een overschatting zal zijn, en dat deze overschatting groter zal zijn bij minder geletterde proefpersonen dan bij meer geletterde proefpersonen (Maki et al., 2005; Hacker et al., 2000). De geletterdheid van proefpersonen is in dit onderzoek gemeten door middel van een woordenschattoets. Hoger geletterden zullen namelijk waarschijnlijk meer bekend zijn met laag frequente woorden dan lager geletterden.

Syntactische complexiteit

Het bepalen van de moeilijkheid van een tekst kan op verschillende manieren. In Rawson & Dunlosky (2002) gebeurt dit door het manipuleren van de causale samenhang. Omdat de groei in gebruik en kwaliteit van computationele modellen een belangrijke aanleiding vormt voor dit paper zoeken wij echter naar een meer algoritmische manier. Een mogelijke manier is zoals Maki et al. (2005) de aanwezigheid van hoog- en laagfrequente woorden manipuleren, maar niet alle woorden hebben noodzakelijk een leidende rol in betekenisgeving binnen een zin. In het geval van sommige woorden zal de lezer dus de lagere *ease of processing* van een laag frequent woord in zijn geheel kunnen vermijden door de betekenis uit de rest van de zin te halen. Een andere mogelijke manier voor begrijpelijkheidsmanipulatie is het verhogen van de *integratiekosten* in een zin (Temperly, 2007, p. 2). De *integratiekosten* van een zin komen voort uit de *dependency locality theory*. Deze theorie zegt dat de syntactische complexiteit van een zin - en daarmee waarschijnlijk de *ease of*

processing - afhankelijk is van twee factoren: *opslagkosten* en *integratiekosten* (Gibson, 1998; 2000). *Opslagkosten* zijn afhankelijk van de hoeveelheid andere zinsdelen die woorden met zich mee moeten brengen om een zin correct te laten verlopen. In het Nederlands verwachten we bijvoorbeeld meestal dat een onderwerp een werkwoord met zich meebrengt. Hoe meer verwachtingen we over een bepaald woord hebben, hoe hoger de *opslagkosten*. *Integratiekosten* nemen toe naarmate de afstand tussen twee verbonden woorden toeneemt (Gibson, 2000).

1. *Hoge integratiekosten*: Jan - die weinig boeken las - ging naar school.
2. *Lage integratiekosten*: Jan ging naar school. Jan las weinig boeken.

Dit brengt ons op onze volgende hypothese:

H_2 : *Hoge integratiekosten* resulteren in lagere voorspelde leesbaarheid

Temperly (2007) en Gibson (1998, 2000) stellen dat de complexiteit van een zin toeneemt met hogere integratiekosten, en dat daarmee de leesbaarheid afneemt. Als we deze samen nemen met onze eerste hypothese verwachten we in een experiment rond leesbaarheid van teksten dus een lagere voorspelde leesbaarheid te vinden in teksten waar de *integratiekosten* zijn verhoogd door de afstand tussen afhankelijke tekstdelen te vergroten. Bovendien verwachten we dat de overschatting - het verschil tussen voorspelde begripelijkheid en werkelijke begripelijkheid - groter zal zijn bij minder geletterde proefpersonen.

Methoden

Materiaal

De teksten voor ons experiment zijn overgenomen uit het onderzoek van Kleijn (2014). Dat onderzoek gebruikte 10 unieke teksten over 5 verschillende onderwerpen: geschiedenis, economie, aardrijkskunde, ongelukken & ziekten en sociale wetenschappen. Deze teksten zijn gemanipuleerd op afhankelijkheidslengtes. Zo ontstonden voor elke tekst twee versies, een makkelijke (kleine afhankelijkheidslengte) en een moeilijke (grote afhankelijkheidslengte).

In het experiment van Kleijn (2014) hebben proefpersonen een *cloze*-test afgelegd met deze teksten. In een *cloze*-test is (meestal) in elke zin een woord weggelaten op een vaste positie in de zin. De proefpersoon moet vervolgens de weggelaten woorden invullen. De *cloze*-test is gebruikt om te meten in hoeverre de proefpersonen de teksten begrepen. Een kritiekpunt op traditionele *cloze*-tests is dat deze tests bijvoorbeeld ook lidwoorden kunnen weglaten. Het toetsen van kennis over lidwoorden is geen goede methode om begrip te meten (Gellert & Elbro, 2012). Om dit soort zwaktepunten van de *cloze*-test te vermijden is in het experiment van Kleijn (2014) gekozen om een semiautomatische variant van de *cloze*-test

te gebruiken. Hierbij is in eerste instantie elk woord op een bepaalde positie in een zin weggelaten, tenzij het woord een lidwoord of voorzetsel was. Bij lidwoorden of voorzetsels is een ander woord handmatig weggelaten. In dit paper zijn de resultaten van deze tests gebruikt als indicatie van het begrip van de teksten.

Voor dit experiment zijn dezelfde teksten gebruikt als in het onderzoek van Kleijn (2014). De manipulatie van de afhankelijkheidslengte in deze teksten is intact gehouden. Er zijn dus 10×2 teksten gebruikt. Iedere tekst ging gepaard met een vragenlijst. Naast demografische gegevens als leeftijd en opleidingsniveau bevatte de vragenlijst vragen over het begrip van de tekst. In deze vragen werd proefpersonen gevraagd om op een 7-punts Likert-schaal te bepalen hoe leesbaar de gelezen tekst was. Aan het einde van de vragenlijst kregen proefpersonen een multiple-choice woordenschattoets bestaande uit 25 vragen. Deze toets komt grotendeels overeen met de toets van Kleijn (2014), behalve 9 vragen waar sommige van de afleidende antwoorden geen goede afleiders bleken te zijn. Bij deze 9 vragen zijn die afleiders herzien.

Procedure

De teksten en bijbehorende vragenlijsten zijn online verspreid. Proefpersonen lazen online 4 verschillende teksten. Binnen elke tekst kregen proefpersonen na iedere alinea een vraag over de leesbaarheid op een 7-punts likert schaal. Na elke tekst kregen proefpersonen vragen over interesse en expertise, maar deze zijn niet meegenomen in de analyse. Zo lazen de proefpersonen 4 verschillende teksten. Proefpersonen lazen telkens 2 moeilijke en 2 makkelijke teksten. De volgorde waarin de proefpersonen de verschillende moeilijkheidsversies van de teksten lazen was een om-en-om volgorde, waarbij de helft van de proefpersonen begon met een makkelijke tekst, en de andere helft met een moeilijke. De combinatie van teksten die proefpersonen te lezen kreeg was identiek aan een van de combinaties uit Kleijn (2014).

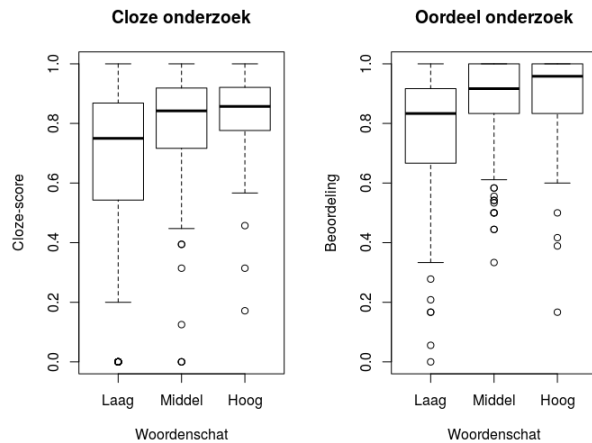
Proefpersonen

Een gedeelte van onze respondenten is overgenomen uit de dataset van Kleijn (2014). Hiernaast zijn door middel van de bovenstaande procedure 112 proefpersonen verzameld. De resultaten zijn maar gedeeltelijk relevant voor dit onderzoek omdat naast syntactische manipulatie sommige teksten ook op connectieven zijn gemanipuleerd. Deze zijn niet relevant voor dit paper, en zijn dus - samen met de bijbehorende proefpersonen - verder buiten beschouwing gelaten. Dit resulteerde in een totaal van 226 proefpersonen. De verdeling van de proefpersonen per opleidingsniveau is te vinden in tabel 1. Tussen de proefpersonen van Kleijn (2014) en de proefpersonen van ons onderzoek bleek een significant verschil in opleidingsniveau. $\chi^2(5, N = 226) = 10.87, p < .01$. De implicaties hiervan zullen wij verder bespreken in de discussie.

Proefpersonen waren gemiddeld 36 (sd = 15.97) jaar oud, en waren tussen de 18 en 80 jaar oud. Leeftijden tussen de twee onderzoeken verschilden ook significant. $t(216.01) = 5.70, p < .001$. In het *cloze* onderzoek waren proefpersonen

Tabel 1: Verdeling van proefpersonen over opleidingsniveaus

Opleiding	Oordelen (N = 126)	Cloze (N = 100)	Overall (N = 226)
Laag	11	22	33
Midden	65	42	107
Hoog	50	36	86



Figuur 1: Cloze-scores en beoordelingen per woordenschat tertiel. De boxplots weergeven in hun uiteinden minima en maxima van de kwartielen. De randen binnen de boxplot weergeven de kwartielen, waar de tweede (dikgedrukte) grens dus de mediaan is.

gemiddeld 43 jaar oud, terwijl het oordeelonderzoek een gemiddelde leeftijd kende van 32. Dit vormt enkele problemen voor generaliseerbaarheid van dit onderzoek. Hier zullen wij in de discussie wederom op terugkomen. Woordenschatsscores verschillen niet tussen de twee onderzoeken $t(43.45) = 1.01, p > .31$.

Resultaten

Tekstoordelen van alinea's bleken hoog samenhangend ($\alpha > 0.96$). Deze zijn derhalve samengenomen. De beoordelingen en scores van het Cloze-onderzoek zijn gesorteerd op hun bijbehorende woordenschatsscores en verdeeld in tertielen. Vervolgens zijn de Cloze-scores en beoordelingen naar hetzelfde interval $[0,1]$ geschaald. Om te bepalen of proefpersonen binnen de tertielen dezelfde geletterdheid hadden zijn voor ieder tertiel T-toetsen uitgevoerd tussen de woordenschatsscore. Er was een significant verschil in woordenschatsscore tussen de Cloze-groep en de oordeelgroep in het laagst geletterde tertiel. ($t(147) = -5.87, p < .01$) Dit verschil was er niet bij het tweede en derde tertiel. (respectievelijk: $p > 0.11$ en $p > 0.27$) Dit vormt weer een betrouwbaarheidsprobleem.

De geschaalde beoordelingen en Cloze-scores per woordenschat-tertiel staan afgebeeld in figuur 1. De figuur vertoont enige samenhang tussen Cloze-scores en beoor-

Tabel 2: Verschillen en significanties tussen beoordelingen en Cloze-scores

Woordenschat Tertiel	Beoordeling	Cloze-score	Verskil en sig.
Laag	0.77	0.64	0.13**
Midden	0.87	0.78	0.09**
Hoog	0.89	0.83	0.06**

** voor $p < 0.01$, Welch two-sample.

delingen, maar vooral de proefpersonen die laag scoorden op de woordenschattoets vertonen ook grote variantie in Cloze-scores. Om onze hypothese H_1 te bevestigen of te ontkrachten zijn we vooral geïntereseerd in het verschil tussen beoordelingen en Cloze-scores, dat wil zeggen: het verschil tussen de twee figuren binnen figuur 1. Deze verschillen - en significanties van T-toetsen - staan gerapporteerd in tabel 2. Deze T-toetsen zijn gebruikt om binnen elk tertiel de significantie tussen de beoordeling en de cloze-score te bepalen. De richting van het verschil vertelt ons of het om een over-, of een onderschatting gaat. Dat wil zeggen dat proefpersonen in alle tertielen de leesbaarheid van de tekst hebben overschat.

Om het effect van de mogelijke verhoogde integratiekosten te bekijken - zoals we in hypothese H_2 hebben besproken - is een gekeken naar een mogelijk verschil tussen de beoordelingen van de moeilijkere en makkelijkere teksten van het oordeelonderzoek. Een Welch two-sample T-test onthulde geen significant effect van de moeilijkheid van de tekst. $t(501.11) = -1.48, p > 0.13$. Dit komt niet overeen met onze hypothese.

Discussie

De verschillen tussen de beoordelingen en Cloze-scores ondersteunen hypothese H_1 , maar komen niet geheel overeen met de bevindingen van Hacker et al. (2000) en Maki et al. (2005). Onze hypothese stelde dat het verschil tussen de werkelijke en de voorspelde leesbaarheid van de tekst negatief samenhangt met de ervaring van de lezer. Het verschil nam in ons geval inderdaad af van 0.13 naar 0.06, maar in het geval van de proefpersonen in het hoge woordenschat tertiel zien we nog steeds een overschatting. Terwijl de papers van Hacker et al. (2000) en Maki et al. (2005) een significante onderschatting rapporteren in de hoogst geletterde groep. Hier kunnen enkele redenen voor zijn. Waarschijnlijk de belangrijkste is het verschil in type proefpersonen. Hacker et al. (2000) en Maki et al. (2005) gebruikten uitsluitend studenten. Dat heeft als gevolg dat proefpersonen in dit onderzoek gemiddeld wellicht minder leeservaring hadden dan de onderzoeken van Hacker et al. (2000) en Maki et al (2005). Als wij uitsluitend studenten hadden gebruikt hadden wij wellicht eenzelfde resultaat gevonden. Als men de trend van onze resultaten zou doortrekken, zou een onderschatting als die van Maki et al. (2004) niet ondenkbaar zijn. Bovendien is in het laagste tertiel een verschil van woordenschat gevonden. Dat

zorgt ervoor dat deze significantie niet direct is toe te wijzen aan het verwachte effect.

De afwezigheid van een significant effect tussen beoordelingen in de moeilijke en makkelijke condities van de beoordeling is niet direct toe te wijzen aan een variabele die ongelijk is tussen de twee condities. Woordenschat, leeftijd en opleidingsniveau waren namelijk gelijk tussen de twee condities.

Dit onderzoek en de vergelijking met het onderzoek van Kleijn (2014) kennen enkele generaliseerbaarheidsproblemen. Zoals in de proefpersonensectie al is besproken, verschilden de leeftijd en het opleidingsniveau tussen de twee onderzoeken. Dit zou als gevolg kunnen hebben dat het significante effect dat wij eerder aan hypothese H_1 koppelden in werkelijkheid wordt veroorzaakt door een -of beide- van deze factoren. Hiernaast is het onderzoek dat voor dit paper is uitgevoerd digitaal afgenomen, terwijl de gegevens van Kleijn (2014) analoog zijn verkregen. Het invullen van vragenlijsten verschilt fundamenteel tussen de twee media. Digitale invulling is bijvoorbeeld wellicht vaker gepaard gegaan met afleidende factoren rond de proefpersoon. Het contrast tussen onze twee hypothesen creëert een ongemakkelijke situatie waar de hypothese H_1 wel met een significant effect gepaard gaat, maar lijdt aan verschillende betrouwbaarheidsproblemen terwijl dit niet voor hypothese H_2 - met meer constante variabelen - geldt. Hoewel de twee hypothesen niet simpelweg met elkaar te vergelijken zijn, wekt dit twijfel op over de betrouwbaarheid van het onderzoek. Vervolgonderzoek is dus ten sterkste aan te raden.

Aanbevelingen

De suggesties voor vervolgonderzoek zijn vooral gecentreerd rond het verhelpen van enkele van de betrouwbaarheidsproblemen van het onderzoek dat in dit paper is beschreven. Naast een herhalingsonderzoek met beter gecontroleerde proefpersonen zou een vervolgonderzoek sterk gebaat zijn bij een binnen-proefpersonen ontwerp. Met een dergelijk ontwerp kan beter bepaald worden in hoeverre proefpersonen zichzelf over-, of onderschatten. Een onderzoek zou bijvoorbeeld - hoewel intensief - proefpersonen kunnen vragen een Cloze-test, voorspelling en een woordenschattoets te maken.

Verdiepingsonderzoek rondom onze eerste hypothese H_1 zou dieper kunnen ingaan op de kenmerken van de tekst. In ons geval zijn alleen persoonskenmerken bekeken -namelijk woordenschat-, maar zoals Rawson & Dunlosky (2002) beschrijven zullen teksteigenschappen ongetwijfeld ook een rol spelen in de beoordeling die men maakt over een tekst. Moderne analytische programma's als T-scan kunnen bijvoorbeeld worden gebruikt om te komen tot model om voorspellingen te maken over de mate waarin proefpersonen zichzelf overschatten.

Referenties

Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language*, 28(5), 610-632.

doi:10.1016/0749-596x(89)90016-8

Benjamin, R. G. (2011). Reconstructing Readability: Recent Developments and Recommendations in the Analysis of Text Difficulty. *Educational Psychology Review*, 24(1), 63-88. doi:10.1007/s10648-011-9181-8

Gellert, A. S., & Elbro, C. (2012). Cloze Tests May be Quick, But Are They Dirty? Development and Preliminary Validation of a Cloze Test of Reading Comprehension. *Journal of Psychoeducational Assessment*, 31(1), 16-28. doi:10.1177/0734282912451971

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, 68, 176

Gibson, E. (2000). The dependency locality theory: a distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, language, brain* (pp. 95-126). Cambridge, MA: MIT Press.

Hacker, D. J., Bol, L., Horgan, D. D., & Rakow, E. A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology*, 92(1), 160-170. doi:10.1037/0022-0663.92.1.160

Kleijn, S. (2014). The LIN cloze-procedure as a measure of text comprehension *ongepubliceerd*

Kraf, R., Lentz, L., & Pander Maat, H. (2011). Drie Nederlandse instrumenten voor het automatisch voorspellen van begrijpelijkheid - Een klein consumentenonderzoek. *tijdschrift voor taalbeheersing*, 33(3), 249-265. doi:10.5117/tvt2011.3.drie411

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77, 1121-1134.

Maki, R. H., Shields, M., Wheeler, A. E., & Zacchilli, T. L. (2005). Individual Differences in Absolute and Relative Metacomprehension Accuracy. *Journal of Educational Psychology*, 97(4), 723-731. doi:10.1037/0022-0663.97.4.723

Rawson, K. A., & Dunlosky, J. (2002). Are performance predictions for text based on ease of processing? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(1), 69-80. doi:10.1037/0278-7393.28.1.69

Temperley, D. (2007). Minimization of dependency length in written English. *Cognition*, 105(2), 300-333. doi:10.1016/j.cognition.2006.09.011

Zwaan, R. A., & Radvansky, G. A. (1998). Situation mo-

dels in language comprehension and memory. *Psychological Bulletin*, 123, 162-185.