

Thesis bachelor Kunstmatige Intelligentie

Is het opvolgen van aanwijzingen uit T-Scan een goede manier om teksten te verbeteren in de context van woningcorporaties en is dit een bruikbare tool is om te implementeren bij woningcorporaties zelf?

Auteur

Michiel Schouten
m.schouten5@student.uu.nl

Eerste lezer

M.P. Schraagen
m.p.schraagen@uu.nl

Tweede lezer

F.W. Adriaans
f.w.adriaans@uu.nl

Kunstmatige Intelligentie
7,5 ECTS



**Utrecht
University**

Department Geesteswetenschappen
Universiteit Utrecht
Nederland
11-02-2024

Samenvatting

Hier komt de abstract van de scriptie.

Inhoudsopgave

1	Introductie	3
1.1	Onderzoeksvraag	3
1.2	Relevantie aan AI	4
1.3	Maatschappelijke relevantie	4
1.4	Theoretisch kader	4
1.4.1	Inbedding huidig wetenschappelijk onderzoek . .	4
1.4.2	Wat houdt leesbaarheid in?	6
1.4.3	Hoe kan je een tekst aanpassen ter vereenvoudi- ging?	7
1.4.4	Wat houdt T-Scan/LiNT in?	8
1.4.5	Schatting van taalniveau in woningcorporaties? .	9
2	Methode	10
2.1	Brieven	10
2.2	Leesbaarheidsinstrument	11
2.3	Schrijfprotocol	11
2.4	Herschrijven van brieven	12
2.5	Enquête	12
2.5.1	Gebruikte teksten	12
2.5.2	Opzet	13
2.5.3	Proefpersonen	14
2.6	Verwerking enquête	14
3	Resultaten	16
3.1	Vraag 1	16
4	Conclusie	19
5	Discussie	20
	Appendices	21
A	Gebruikte brieven overzicht	21
B	Overzicht variabelen	23
C	Synoniemen en vaktermen	23
D	Gebruikte alinea's overzicht	23
D.1	Tekst 1	23
D.2	Tekst 2	23

1 Introductie

De meeste teksten die overheden communiceren naar haar burgers zijn notoir lastig te lezen. Zo bleek uit onderzoek dat dat 10% van de brieven aangaande schuldhulpverlening lastig te lezen was voor hoog-opgeleiden. 50% van de teksten was moeilijk te begrijpen voor laag-opgeleide lezers volgens onderzoek van Pander Maat en van der Geest (2021).

Hierdoor begrijpen burgers vaak niet wat de overheid van hen vraagt of wat ze tegen hen wil zeggen. Burgers die (financieel) sterk afhankelijk zijn van de overheid zijn ook vaker laaggeletterd. Dit kan snel problemen opleveren. [bron]. Het aantal laaggeletterden in Nederland kan conservatief geschat worden op 15% (Maat, Kleijn & Frissen, 2023, p. 18).

De laatste jaren wordt er steeds meer bewustzijn gecreëerd over de leesbaarheid van teksten die de overheid communiceert naar haar burgers. Onder andere door de campagne Direct Duidelijk (tegenwoordig Gebruiker Centraal). Semipublieke instellingen als het Openbaar Vervoersbedrijven en woningcorporaties zijn ook bezig om al hun teksten begrijpelijker te maken voor hun klanten.

Deze these zal de focus leggen op de begrijpelijkheid van de teksten die woningcorporaties versturen naar hun klanten. Laaggeletterden hebben over het algemeen een lage socio-economische positie. Op hun beurt zijn mensen in zo'n positie oververtegenwoordigd in het klantenbestand van woningcorporaties. Ook zogenaamde nieuwkomers zijn de Nederlandse taal vaak niet machtig maar zijn wel aangewezen op sociale huurwoningen. Hierdoor is het extra belangrijk dat die teksten leesbaar en begrijpelijk zijn. Op dit moment zijn er teksten die nog niet op het gewenste niveau zijn, wat burgers in de problemen kan brengen.

1.1 Onderzoeksvraag

Er bestaan al jaren automatische tekstanalyse tools die de leesbaarheid van een tekst kunnen analyseren en die kunnen aangeven welke teksteigenschappen ervoor zorgen dat een tekst moeilijk leesbaar is. Een zo'n tool is de T-Scan, ontwikkeld door Pander Maat et al. (2014). Uit de teksteigenschappen die de T-scan teruggeeft kunnen tekstverbeteringen worden gedestilleerd. De belangrijkste teksteigenschappen uit de T-Scan die bepalend zijn voor de tekstcomplexiteit volgt uit onderzoek van Kleijn (2018) en Maat et al. (2023). Een tool hieromheen gebouwd is het Leesbaarheidsinstrument voor Nederlandse Teksten (LiNT). De vraag is of het opvolgen van aanwijzingen uit T-Scan een goede manier is om teksten te verbeteren in de context van woningcorporaties en of dit een bruikbare tool is om te implementeren bij woningcorporaties

zelf. Deze vraag zal worden beantwoord door antwoord te verkrijgen op drie deelvragen.

1. Krijgt de oorspronkelijke tekst een moeilijkere score?
2. Geven mensen dezelfde suggesties, of andere dan de T-Scan?
3. Hebben mensen voor de makkelijkere versie nog suggesties die de T-Scan niet heeft?

1.2 Relevantie aan AI

De T-scan en de toegepaste versie LiNT zijn leesbaarheidstools. Ze rekenen verschillende

1. Doel is dus om de tool T-Scan te testen om daadwerkelijk leesbaardere teksten te maken in dit domein.

1.3 Maatschappelijke relevantie

In zekere zin is deze thesis een voortgang op het onderzoek Monitor Begrijpelijke Overheidsteksten van Pander Maat en van der Geest (2021). Hier werd onderzocht hoe leesbaar teksten van regionale overheden zijn met betrekking tot schuldhulpverlening. Uit dit onderzoek bleek dat “De meeste geanalyseerde overeenkomsten zijn onbegrijpelijk voor lezers van alle opleidingsniveaus, terwijl ondertekening van die overeenkomsten grote consequenties heeft voor het doen en laten van de burger met schulden.” (Pander Maat & van der Geest, 2021, p. 20) Het gaat hier dus over overeenkomsten in de zin van het contractrecht. In dit onderzoek zal worden gekeken hoe dat is bij woningcorporaties gezien deze ook een zeer belangrijke rol hebben in de maatschappij.

De thesis wordt uitgevoerd bij Zig Websoftware. Dit softwarebedrijf levert software voor woningcorporaties die 40% van de sociale huurwoningen beheren. Een groot deel van de Nederlanders die in sociale huurwoningen zitten maken gebruik van de software van Zig. Dit maakt dat deze scriptie een grote maatschappelijke impact kan maken. Indien T-Scan bruikbaar is voor het verhogen van de leesbaarheid, dan kan dit ervoor zorgen dat 2 miljoen Nederlanders begrijpelijker brieven ontvangen.

1.4 Theoretisch kader

1.4.1 Inbedding huidig wetenschappelijk onderzoek

Het onderzoek en onderzochte vragen staan niet op zichzelf. Veel onderzoek in dit veld in de context van de Nederlandse taal is gedaan door Dr. H.L.W. Pander Maat geaffilieerd onderzoeker bij de Universiteit Utrecht. Vele studies van Pander Maat liggen in lijn met het het

onderzoek in deze thesis, hierom zal hij veelvuldig worden aangehaald. Hij was zelf echter niet betrokken bij dit onderzoek.

Er wordt al vele jaren onderzoek gedaan naar de leesbaarheid van teksten en welke factoren daar aan bijdragen zie bijvoorbeeld Freebody en Anderson (1983). Freebody en Anderson deden onderzoek naar de samenhang van het vocabulaire van een tekst en het tekstbegrip van de lezer. Hier bleek dat de (on)bekendheid van woorden in een tekst een grote voorspellende factor is voor het begrijpen van een tekst. Hoewel het niet het hoofddoel is van dit onderzoek zal dit effect nogmaals bestudeerd worden.

Recentere onderzoek naar automatische tekstanalyse is onder andere gedaan voor de Portugese taal Aluisio, Specia, Gasperin en Scarton (2010). Hier werd naar 59 teksteigenschappen gekeken om een tekst te kwalificeren op de non-lineaire schaal Rudimentair, Basis of Geavanceerd. Rudimentair houdt in dat iemand expliciete informatie kan vinden in een korte tekst. Basis houdt in dat iemand teksten met een gemiddelde lengte kan lezen en hier informatie uit kan destilleren. Geavanceerd betekent dat iemand een tekst volledig begrijpt, meningen van feiten kan onderscheiden, de tekst kan samenvatten en er iets over de inhoud kunnen zeggen. Als onderdeel van dit onderzoek hebben ze op basis van de 59 teksteigenschappen een Machine Learning model getraind die met redelijke zekerheid kon zeggen of een tekst was bedoeld voor een Rudimentaire, Basis of Geavanceerde lezer. In combinatie met dit model is er ook een andere tool ontwikkeld: SIMPLIFICA Scarton, Oliveira, Candido Jr, Gasperin en Aluísio (2010). Dit is een programma die automatisch tekst kan versimpelen op basis van vooraf bepaalde regels. Een regel kan zijn om een eenvoudiger synoniem te gebruiken voor een moeilijk woord. De tool doet op zinsniveau lexicale en syntactische suggesties. Het is aan de gebruiker om de wijzigingen over te nemen. Over deze tool is helaas weinig Engelstalige literatuur te vinden, in hoeverre dit programma bijdraagt aan de leesbaarheid van teksten is voor mij onbekend.

Het automatisch herschrijven van teksten met behulp van Large Language Models (LLMs) zoals ChatGPT¹ is sinds een aantal jaar ook mogelijk. Zo is onderzocht of tekst daadwerkelijk leesbaarder en begrijpelijker wordt als een LLM de opdracht krijgt om dit te doen van Raaij, Kolkman en Podoyntsyna (2024). van Raaij et al. kwamen tot de conclusie dat de beste manier om een tekst zo te versimpelen is door ChatGPT de volgende prompts in te laten voeren. Eerst moet een tekst samengevat worden in bulletpoints. Het volgende prompt is om die bulletpoints eenvoudig uit te schrijven tot een lopende tekst. Met deze methode werd de meeste informatie behouden in de versimpelde

¹<https://chatgpt.com>

versie en werd het tekstbegrip verhoogt. Het tekstbegrip werd gemeten door participanten open en gesloten vragen te stellen over de tekst. Dit waren vragen als "Wat houdt een persoonsgebonden budget in?". Daarnaast werd ook gevraagd wat participanten vonden van de toon van een tekst. In deze thesis wordt hier ook naar gevraagd. Deze thesis maakt echter geen gebruik van LLM's om teksten te versimpelen.

Zoals gezegd wordt in deze these ook onderzoek gedaan naar de toon van een versimpelde brief. Er wordt onderzocht of mensen een eenvoudigere tekst betuttelender of neerbuigender vinden dan eenzelfde tekst met een hoger leesniveau. Hier is in het verleden al uitgebreid onderzoek gedaan. Uit de literatuur blijkt dat tekstvereenvoudiging onder, ook hoogopgeleide, lezers beter bevalt Maat en Gravekamp (2022) dan complexere teksten. De aanname is dat als iemand een complex onderwerp eenvoudig kan uitleggen deze persoon meer kennis over dit onderwerp heeft dan iemand die alleen met complexe tekst een complex onderwerp kan uitleggen. Of een versimpelde tekst betuttelend is wordt gevraagd in de enquête.

Er is ook al onderzoek gedaan naar welke tekstkenmerken verschillen tussen eenvoudige en complexe teksten die hetzelfde verhaal vertellen Pander Maat en Ditewig (2017). Uit dit onderzoek werden de verschillen wel onderzocht maar bleef het onduidelijk welk tekstkenmerk precies bijdroeg aan verbeterd tekstbegrip onder lezers.

Deze thesis gaat meer in op de voorspellende/voorschrijvende aard van T-Scan dan de beschrijvende aard zoals bij bovengenoemd onderzoek.

Verder leesbaarheidsonderzoek is gedaan om te onderzoeken welke teksten ten opzichte van elkaar eenvoudiger dan wel complexer zijn D Clerq et al. (2014). Met behulp van zowel twee taalexperts als gemiddelde mensen uit de samenleving werd de leesbaarheid getoetst door teksten onderling te vergelijken. Zowel de rangschikking van de experts als de resultante van de inbreng van alle deelnemers lag zeer op één lijn. Er werd in dit onderzoek niet gevraagd naar specifieke teksteigenschappen die de tekst eenvoudiger of moeilijker maakten ten opzichte van een andere tekst. De teksten die vergeleken werden verschilden qua onderwerp van elkaar, het ging dus niet om een moeilijke en vereenvoudigde versie van eenzelfde tekst.

1.4.2 Wat houdt leesbaarheid in?

Leesbaarheid is op verschillende manieren te meten. Het is zeer lastig om leesbaarheid te meten van een tekst omdat uiteindelijk mensen de tekst moeten begrijpen en interpreteren. Er blijft dus altijd ruimte voor onnauwkeurigheden, gezien leesbaarheid sterk afhangt van de mens zelf.

Leesbaarheid kan zowel worden ingevuld als het gemak waarmee

een tekst wordt gelezen als de begripelijkheid ervan. Dit laatste wordt ook wel tekstbegrip genoemd. Tekstbegrip kan zowel gemeten worden door inhoudelijke vragen te beantwoorden over de gelezen tekst. De vragen kunnen gaan over de hoofdlijnen zoals gedaan in Pander Maat en Ditewig (2017); van Raaij et al. (2024). Er wordt dan onderzocht of lezers meer of minder goede antwoorden geven op deze begripsvragen bij de ene versie ten opzichte van de andere versie. Crossley, Skalicky en Dascalu (2019) bepaalde het tekstbegrip door simpele ja-nee vragen te stellen over twee teksten, welke sneller gelezen was en welke begripelijker was.

Een andere manier om tekstbegrip te meten is door een zogenaamde Cloze-test uit te voeren. In een Cloze-test worden bepaalde woorden weggelaten uit een doorlopende tekst. Het is aan een participant om het juiste woord in te voegen. Een juist woord wordt meestal gekwalificeerd als een woord die dezelfde semantiek bezit. Deze manier wordt gezien als een goede indicator om het tekstbegrip te onderzoeken Kleijn (2018).

Anders dan de eerder genoemde manier om de leesbaarheid van een tekst te bepalen en om te kijken of een versimpelde tekst daadwerkelijk leesbaarder is, wordt in deze thesis de leesbaarheid bepaald doordat gebruikers zelf aangeven in welke mate ze een tekst leesbaar vinden of niet.

”Readability labeling requires a different approach. The ease with which a given reader can correctly identify the message conveyed in a text is, among other things, inextricably related to the reader’s background knowledge of the subject at hand (McNamara et al. 1993).”

1.4.3 Hoe kan je een tekst aanpassen ter vereenvoudiging?

>Lexicaal en syntactisch en vergroten cohesie Een brief is op verschillende manieren aan te passen om de boodschap beter over te brengen. Ten eerste kan het helpen als er infographics in een brief staan die de boodschap van de brief ondersteunt. Daarnaast is het mogelijk om de tekstopbouw aan te passen. Zo kan het doel en de gevolgen voor de lezer eerst helder uitgelegd worden om vervolgens met een onderbouwing daarvoor te komen. Dit raakt al aan het inhoudelijk aanpassen van een tekst. Er kan worden gekozen om bepaalde details weg te laten opdat de lezer het grote geheel beter begrijpt. Daarnaast zijn er op een abstract niveau drie tekstenmerken die gewijzigd kunnen worden die in hoge mate invloed hebben op het tekstniveau zoals beschreven door Pander Maat en Ditewig in 2017. Dit zijn (1) lexicale aanpassingen, denk aan woordgebruik, (2) syntactische aanpassingen dit gaat om zinsopbouw en (3) cohesie-ingrepen, dit gaat om tekstverbanden en signaalwoorden. Deze aanpassingen blijken vaak samen te gaan (Pan-

der Maat & Ditewig, 2017, p. 254). Om een lange zin op te knippen in meerdere kleinere zinnen zijn er verwijswaarden nodig naar die andere zinnen. Ook worden werkwoordsvormen dan aangepast. Uit onderzoek van Freebody en Anderson (1983) blijkt ook al dat woordgebruik een hoge voorspeller is van tekstbegrip.

1.4.4 Wat houdt T-Scan/LiNT in?

In veel van de onderzoeken hierboven wordt T-Scan en of LiNT gebruikt. T-Scan is een automatische tekstanalyse ontwikkeld in 2014 door Pander Maat et al. en Maat et al. (2017). Deze tool meet of berekent 472 teksteigenschappen. Dit gebeurt op basis van vooraf gestelde regels, anders dan de tool van Aluisio et al. (2010) is het geen Machine Learning (ML) model. Het Leesbaarheidsinstrument voor Nederlandse Teksten (LiNT) wordt dubbelzinnig gebruikt. Enerzijds bestaat er de LiNT-formule, deze formule berekent op basis van de woordfrequentie, concrete naamwoorden, de afhankelijkheidslengte en de deelzinslengte van een tekst de leesbaarheid van een tekst Maat et al. (2023). Deze formule is duidelijk, het is geen zogeheten blackbox-model waar een tekst ingaat en een getal uitrolt. Dit maakt het voor tekstschrijvers eenvoudig om te achterhalen wat er nog verbeterd moet worden aan hun tekst. De LiNT-formule geeft een score van 0 tot 100 terug. 0 betekent dat het heel leesbaar is, 100 betekent dat een tekst heel moeilijk is om te begrijpen. Deze schaal is onderverdeeld in 4 niveaus. De niveaus zijn:

1. Het streefniveau, LiNT-score 0-34. Dit niveau is alleen voor de 15% laaggeletterden in Nederland is dit niveau te moeilijk om te begrijpen.
2. LiNT-score 34-46. 30% van de volwassen Nederlanders kan dit niveau niet goed aan.
3. LiNT-score 46-60. 55% van de volwassen Nederlanders kan dit niveau niet goed aan.
4. LiNT-score 60-100. 82% van de volwassen Nederlanders kan dit niveau niet goed aan.

(Maat et al., 2023, p. 19)

Daarnaast is er het leesbaarheidsinstrument. Dit instrument is in feite een grafische invulling van de LiNT-score en T-Scan. De maker geeft zelf aan dat dit instrument onder andere gebruikt kan worden voor het bepalen van de complexiteit van een tekst, probleemopsporing binnen een tekst en originele en herschreven teksten vergelijken (Maat et al., 2023, p. 27). In dit onderzoek wordt naar al deze drie aspecten gekeken.

Een verdere uitleg van T-Scan is al gedaan, evenals onderzoek naar de effectiviteit van de LiNT is al gedaan Maat en Gravekamp (2022). Dit laatste onderzoek heeft ook beargumenteerd waarom T-Scan het beste instrument is voor tekstanalyse.

1.4.5 Schatting van taalniveau in woningcorporaties?

Om te onderzoeken hoe groot het probleem van moeilijke brieven in de sociale huursector is, moet ook duidelijk zijn in welke mate de huurders hiervan geletterd zijn. 15% van de Nederlanders is dan wel laaggeletterd, maar of deze groep onder- of oververtegenwoordigd is onder de klanten van woningcorporaties maakt uit voor de grootte van het probleem. Hier is geen duidelijk onderzoek naar gedaan. Wel zijn er een aantal onderzoeken gedaan in de afgelopen 2 decennia.

Een factsheet van de gemeente Amsterdam uit 2008 stelt dat slechts 20% van sociale huurders hoger onderwijs heeft gevolgd.² Uit welke bron dit getal komt is niet duidelijk. Wat hoogopgeleid precies inhoudt en of dit betekent dat 80% laagopgeleid is, is niet duidelijk.

Deze thesis draagt bij aan het wetenschappelijk discours door te onderzoeken of T-Scan gebruikt kan worden om teksten handmatig te verbeteren en door te onderzoeken of de aanpassingen daadwerkelijk verbeteringen zijn.

²<https://www.fairhuurvoorverhuurders.nl/files/literatuur/de-particuliere-woonsector-en-zijn-bewoners.pdf>

2 Methode

2.1 Brieven

Om te onderzoeken of brieven van woningcorporaties daadwerkelijk leesbaarder kunnen worden gemaakt door gebruik te maken van automatische tekstanalyse zijn er daadwerkelijk brieven nodig. In totaal zijn er voor dit onderzoek in 173 brieven onderzocht van acht verschillende woningcorporaties uit Nederland. Dit zijn brieven die zij op dagelijkse basis versturen vanuit hun systemen. De brieven worden gepersonaliseerd verstuurd naar klanten. Hierom bevatten de brieven variabelen, die door de computer worden ingevuld bij het versturen van de brief. Deze variabelen zijn voor dit onderzoek handmatig ingevuld zie Appendix B.

Daarnaast zijn de brieven verder bewerkt volgens de regels in de T-Scanhandleiding (Maat et al., 2017, p. 7-12). Inhoudelijk zijn er een aantal zaken aan de brieven aangepast. Zo moet T-Scan de aanhef, koptitels en afsluitende groet overslaan. T-Scan ziet de titels van vaak maar enkele woorden als een zin. Hoewel dit feitelijk kan kloppen met een titel als "Heeft u nog vragen?", brengt het de gemiddelde zinslengte sterk naar beneden. Zeker omdat in de gebruikte teksten relatief weinig zinnen staan, hebben de aanhef en bijvoorbeeld de afsluitende groet hier veel invloed op. Inhoudelijk geven ze weinig informatie mee en dragen ze niet bij aan het meer of minder leesbaar maken van een tekst. Zo vond Pander Maat en Ditewig (2017) geen aanwijzing dat kopjes tekstbegrip bevorderen, wel de reproduceerbaarheid van een tekst.

Uiteindelijk blijven er 14 teksten over die herschreven worden, deze hadden de slechtste LiNT-scores. Deze teksten scoorden onder andere slecht op de onbekendheid van woorden. De inhoudswoorden van deze teksten, dat wil zeggen de zelfstandige naamwoorden, bijvoeglijke naamwoorden, werkwoorden en bijwoorden, komen slechts 5000 tot 50000 keer op 1 miljard woorden voor. Ter vergelijking, in teksten in gemiddeld roddelnieuwsblad komen de inhoudswoorden bijna 80000 keer voor op 1 miljard woorden. Ook op de abstractie van de zelfstandignaamwoorden wordt slecht gescoord ($M=68,3$, $SD=19,6$). Deze laatste tekst bestaat wel uit slechts 1 zin. Daarnaast werden er redelijk veel nieuwe woorden gebruikt. Dit wordt uitgedrukt in het percentage inhoudswoorden dat niet voorkwam in de 50 woorden daarvoor ($M=91,5$, $SD=4,2$). De teksten scoorden daarentegen wel goed op het voorkomen van mensen in een tekst. Het gaat hier om verwijzingen naar personen, het aantal persoonlijk voornaamwoorden, het aantal aanspreking en in mindere mate het percentage werkwoorden in de actieve vorm.

»>Al deze teksteigenschappen in een tabel gooien om een overzicht

van de brieven te creëren?«<

Uitleggen elke term.

TODO: Kort benoemen waar deze teksten voornamelijk al goed op scoren en waar notoir slecht.

2.2 Leesbaarheidsinstrument

De tekstkwaliteit van de brieven zal worden onderzocht aan de hand van T-Scan, zoals ook gebruikt in het onderzoek van Pander Maat en van der Geest (2021). De T-Scan registreert welke eigenschappen van de tekst moeilijk te begrijpen zijn. Hieruit kunnen suggesties volgen. Een meer inzichtelijke toepassing van de T-Scan is LiNT zoals eerder beschreven in hoofdstuk 1.4.4.

2.3 Schrijfprotocol

Eerst zal er een protocol worden geschreven over hoe om te gaan met tekstsuggesties. Het protocol wordt geschreven aan de hand van de eerste ervaring met het omschrijven van een brief. Uit deze eerste ervaring zullen de meeste dilemma's opkomen die in bij het herschrijven van volgende brieven relevant zijn. Het protocol geeft een richtlijn aan het herschrijven. Dit protocol zorgt ervoor dat de studie reproduceerbaar is.

Anders dan eerder onderzoek Pander Maat en Ditewig (2017), worden teksten hier herschreven met de handen op de rug. Alleen specifieke tekst wordt aangepast met zo min mogelijk inhoudelijke wijzigingen. Ook wordt de volgorde van zinnen niet aangepast. Daarnaast wordt er geen extra uitleg van processen toegevoegd, ook al kan dat intuïtief wel nodig voelen. Wel kan een vakterm worden uitgelegd indien er geen adequaat synoniem voor bestaat.

Het protocol is zegge opgesteld tijdens het schrijven van een eerste brief. Zo werd duidelijk welke tekstkenmerken worden aangepast en hoe die reproduceerbaar gedaan kan worden. Na een extra verfijning zijn aan de hand van het volgende protocol de 14 brieven met de laagste LiNT-score herschreven.

1. Analyseren van de resultaten van LiNT
 - (a) Moeilijke woorden
 - (b) Moeilijke zinsconstructies
2. Via ishetb1.nl een mogelijke synoniem zoeken die eenvoudiger is
 - (a) De zin zo omschrijven dat het synoniem er in past.
 - (b) Lijst bijhouden met alle vervangingen.
3. Bij speciale (vak)termen een duidelijke korte uitleg toevoegen.
 - (a) Lijst bijhouden welke dit zijn.

4. Lange zinnen opknippen in kleinere zinnen.
5. Als er bijzinnen zijn, dan moeten deze zoveel mogelijk worden herschreven tot aparte zinnen.
6. Zinnen omschrijven naar een actieve werkwoordsvorm
7. Zinnen zoveel mogelijk omschrijven met verwijzingen naar mensen.
8. Reflectie opschrijven over gemaakt keuzes en dilemma's

De lijsten met gebruikte synoniemen en de uitgelegde vaktermen zijn te vinden in Appendix C. Zeker omdat het

2.4 Herschrijven van brieven

Het resultaat van de herschrijvingen zijn te vinden in Tabel ?? . De brieven die 10 punten of meer verbeteren op de LiNT-score zullen worden geselecteerd om vragen over te beantwoorden. Uit elke tekst wordt uiteindelijk een alinea geselecteerd voor in de enquête.

Net als blijkt in het onderzoek (Pander Maat & Ditewig, 2017, p. 254), hebben de herschreven hebben de teksten meer verwijzingswoorden nodig. Opgeknippte zinnen moeten nog wel naar elkaar verwijzen. Voornamelijk bijzinnen moeten verwijzen naar de oorspronkelijk hoofdzin.

Na het herschrijven zijn de LiNT-scores flink verbeterd, zie hiervoor de laatste twee kolommen van Tabel 5

»>Hier dezelfde informatie invoegen over de herschreven teksten als van de originele teksten in sectie Methode.Brieven?«<

Delen van de herschreven brieven zijn te vinden in Appendix D.

Bij het herschrijven zijn voor de uiteindelijk de volgende wijzigingen gedaan.

»>Tabel invoegen welk type wijziging hoe vaak voorkomt«<

2.5 Enquête

2.5.1 Gebruikte teksten

Vervolgens is er een enquête uitgestuurd met daarin 4 vragen over elk van de brieven. Om te voorkomen dat de enquête te lang zou duren is er gekozen om vragen te stellen over slechts 8 van de 14 brieven die oorspronkelijk de slechtste LiNT-score hadden. De 8 brieven hadden de grootste LiNT-score verbetering met minstens 10 punten (M=16.13, SD=8.89) verschil ten opzichte van de originele teksten. Het grote contrast zou het eenvoudiger moeten maken voor respondenten om verschillen te zien tussen tekstversies en om deze te kunnen benoemen.

Om enquête vermoeidheid te voorkomen was het doel om de enquête niet langer dan 30 minuten te laten duren [bron?]. Hierdoor was

het niet mogelijk om alle 8 brieven volledig te tonen. Hierom is er een selectie gemaakt uit deze brieven. Na een korte empirische studie bleek dat de volledige enquête te maken zou zijn binnen 30 minuten als er over maximaal 9 alinea's vragen zouden worden gesteld. Om de diversiteit tussen de teksten te behouden is ervoor gekozen om uit alle 8 brieven 1 alinea te selecteren. De brief genaamd ThuisInLimburg heeft 2 alinea's geleverd aangezien dit een lange brief is die verschillende onderwerpen behandelt. Net als in eerder onderzoek D Clerq et al. (2014) zijn de teksten groter zo'n 100 tot 200 woorden. De teksten zijn op zichzelf te begrijpen. De specifieke alinea's zijn te vinden in D.

2.5.2 Opzet

De enquête is gemaakt in de Qualtrics Experience Management Platform³. De enquête start met een scherm waarop de respondent welkom wordt geheten, waarop er uitleg komt over de inhoud van de enquête alsook informatie over het gebruik van de data. Om door te gaan met de enquête moet de respondent aangeven of deze vrijwillig meedoet, 18 jaar of ouder is en diens data wil delen voor het onderzoek.

Hierna volgt een vraag over het opleidingsniveau van de respondent. Dit kan voor verdere analyse interessant zijn.

Daarna volgen 9 vragenblokken. Deze 9 vragenblokken worden per respondent in een willekeurige volgorde laten zien. Elk blok gaat over 1 alinea tekst. Eerst wordt per blok bepaald of de vragen over de originele tekst of over de bewerkte tekst, de kans hierop is 50%. De enquête-software is zo ingesteld dat het streeft naar een gelijk verdeling vragen over beide versies. De respondent weet niet welke versie van de brief die te lezen krijgt.

De respondent krijgt vervolgens de bepaalde alinea te lezen. Hierna volgen de volgende 3 vragen:

1. In hoeverre begrijpt u de brief?
2. Vindt u de brief betuttelend?
3. Wat zou u aanpassen aan de tekst om hem leesbaarder te maken?
Herschrijf dit gedeelte van de tekst.

De eerste vraag wordt beantwoord door middel van een Likert-schaal Likert (1932). Deze gaat van 'erg goed leesbaar', met waarde 1 naar 'zeer moeilijk leesbaar' met waarde 5. De tweede vraag kan worden beantwoord met 'Ja', 'Nee' en 'Weet ik niet'

Na het invullen van deze eerste 3 vragen volgt er een vierde vraag, de antwoorden op de eerste drie vragen kunnen dan niet meer gewijzigd worden. Deze vraag is als volgt:

³<https://survey.uu.nl>

4. Hier ziet u een aangepaste zin uit de bovenstaande tekst, welke wijzigingen zou u nog toevoegen of er juist uitlaten?

Bovenaan ziet de respondent eerst de versie van de alinea waar die de voorgaande drie vragen over heeft beantwoord. De titel bij deze alinea is "Originele versie". Vervolgens staat onder deze alinea de titel "Bewerkte versie" met daarbij de tegenovergestelde versie van de tekst waar de voorgaande vragen over zijn gegaan. Indien de eerste drie vragen over de daadwerkelijke originele tekst gingen, dan gaat deze vraag over de herschreven versie en vice versa.

Om de respondent te helpen bij het gerichter antwoord geven op deze vraag zijn de verschillende wijzigingen voorzien van verschillende achtergrondkleuren. De kleuren van de wijzigingen in de ene versie corresponderen met de wijzigingen in de bewerkte versie. De alinea's met de kleuren zijn terug te vinden in D.

De antwoorden op vragen één, twee, drie en vier worden vergeleken met de resultaten vanuit de T-Scan.

2.5.3 Proefpersonen

De enquête zal voornamelijk worden beantwoord door hoger opgeleiden, waarvan kan worden aangenomen dat zij in hoge mate geletterd zijn. Dit heeft zowel voordelen als nadelen. Een groot voordeel is dat deze groep zowel een zin kan begrijpen als kan aangeven wat er wel of niet moeilijk of makkelijk is aan deze zin. Daarbij komt het voor dat laag geletterden niet durven aan te geven dat ze een tekst niet begrijpen om zo een moment van schaamte te voorkomen. Het nadeel hiervan is dat de mensen die daadwerkelijk een brief moeten begrijpen niet gevraagd worden naar wat zij vinden. Het zou wel meer representatieve resultaten opleveren. Echter is de tijd voor dit onderzoek maar zeer beperkt. Dit maakt het praktisch onmogelijk om zowel langs de ethische commissie van Universiteit Utrecht te gaan voor goedkeuring van het onderzoek als het aanschrijven van woningcorporaties of hun huurders benaderd mogen worden. Vervolgonderzoek zou zich vooral hierop kunnen toespitsen. Verder is het nog interessant om te onderzoeken of de gebruikte woordfrequenties door T-Scan overeenkomen met de woordenschat voor mensen van wie het Nederlands de tweede taal is.

1. Toevoegen wat het opleidingsniveau is

2.6 Verwerking enquête

1. Hoeveel antwoorden op elke vraagtype
2. Uitleg hoe open vragen zijn geturfd.
3. Uitleg dat elke kleur voor 1 type wijziging staat

4. Uitleg dat negatief antwoord op de bewerkte versie een positief antwoord op de originele versie betekent en vice versa.
- 5.

3 Resultaten

Om te bepalen of er een significant verschil zit tussen de opgegeven leesbaarheidsscores van de originele en bewerkte versie is er een Mann-Whitney U test Mann en Whitney (1947) uitgevoerd in Python⁴. Er is een significant verschil tussen de leesbaarheidsscores van beide versies ($p = 0.026$). Deze teksten worden dus als makkelijker leesbaar ervaren. Deze test is ook uitgevoerd op de vraag of de tekst betuttelend is. Ook dit resultaat is significant ($p = 0.001$). Er kan dus worden aangenomen dat de teksten wel als betuttelender worden ervaren in de bewerkte versie.

”De tekst is qua vorm denk ik eigenlijk best prima. Lekker direct en beknopt. De woordkeuze is alleen fucked up [...]” – Participant

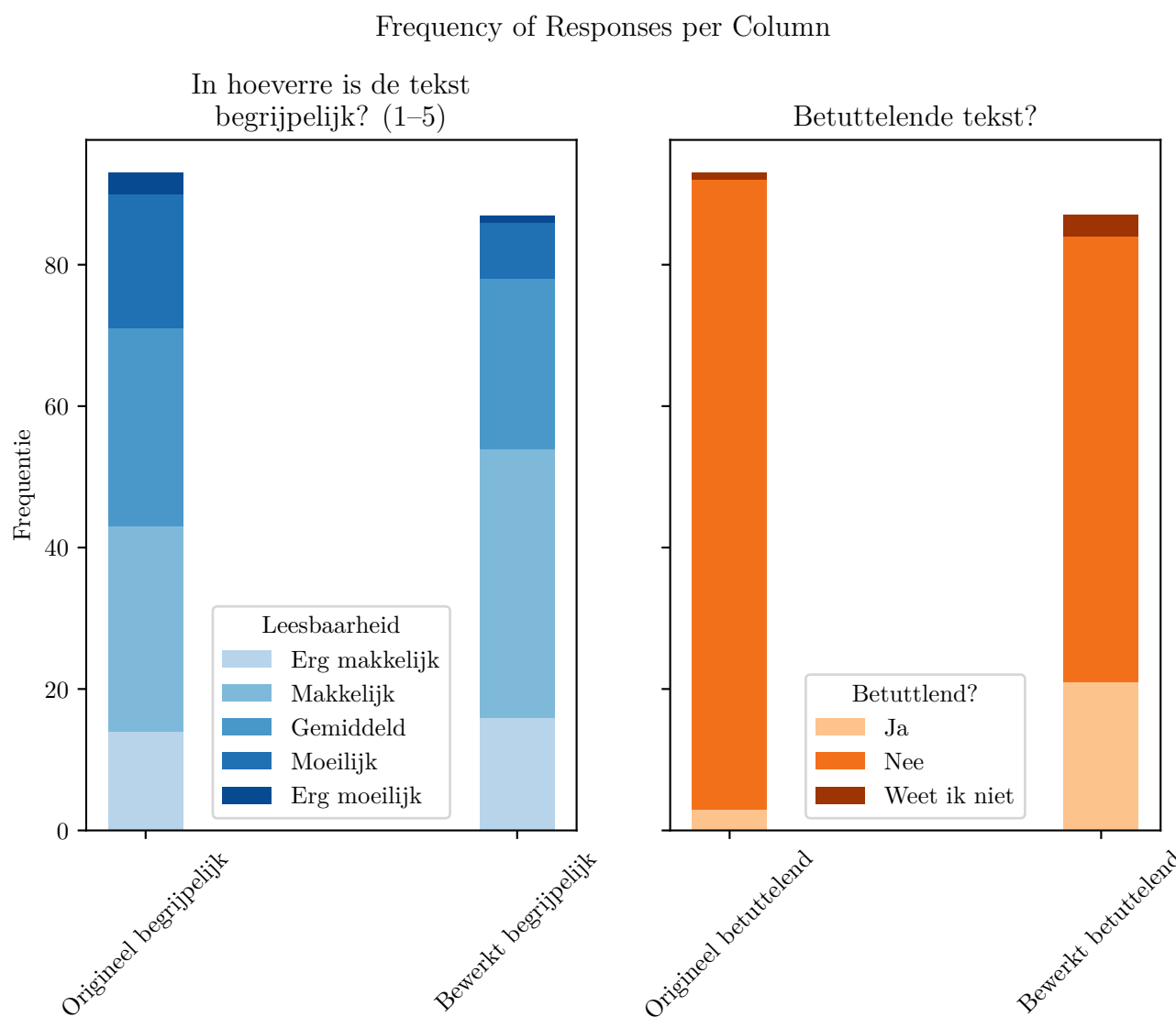
1. Mean + var + SD op elke vraag
2. Cronbachs alfa berekenen?
3. Barcharts moeilijkheid
4. Barcharts betuttelend
5. Resultante elke vraag laten zien.
 - (a) Toevoegen de spreiding van negatief, egal en positieve reacties.
6. Opmerkelijkheden per tekst benoemen.
7. ALS TIJD: labelen van open vraag waar respondenten zelf de tekst moesten herschrijven.

Over het algemeen zijn de positieve scores te vinden bij `w_onbekend`, `w_abstract`, `z_bijzin`, `z_len` en `z_bijzin`. Deze zijn over het algemeen als positieve aanpassingen ervaren.

Omdat er weinig datapunten zijn is het moeilijk om de scores te vergelijken. Zo is `z` afhankelijk als positief ervaren in 65% van de gevallen, maar ging hier ook maar 1 tekst over.

3.1 Vraag 1

⁴<https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.mannwhitneyu.html>



Figuur 1: Leesbaarheid en betuttelendheid per tekstversie vergeleken.

Tabel 1: capt

Aanpassing	Totaal	Weighted	Maximale score	Procentueel positief
w_onbekend	153	5,275862069	580	0,263793103
w_abstract	33	4,714285714	140	0,235714286
w_nieuw	0	0	0	0
z_dzlen	0	0	0	0
z_afhankelijk	13	13	20	0,65
z_bijvoeglijk	0	0	0	0
z_opsomming	1	1	20	0,05
z_bijzin	98	6,533333333	300	0,326666667
z_len	32	6,4	100	0,32
p_men	0	0	0	0
p_pvn	0	0	0	0
p_aspr	-13	-2,166666667	160	-0,08125
p_actief	15	7,5	80	0,1875
o_overig	0	0	0	0

Tabel 2: Bewerkt

Kleur	Totaal	Weglaten	Egal	Behouden	Type
Groen	4	2	3	6	p_actief
Paars	1	4	2	5	p_aspr
Oranje	7	1	2	8	p_actief
Rood	6	2	1	8	p_aspr
Toevoeging	5	0	6	5	w_onbekend

Tabel 3: Origineel

Kleur	Totaal	Weglaten	Egal	Behouden	Type
Groen	5	2	0	7	p_actief
Paars	5	2	0	7	p_aspr
Oranje	3	3	0	6	p_actief
Rood	5	2	0	7	p_aspr
Toevoeging	0	2	5	2	w_onbekend

Tabel 4: Resultante bewerkt totaal

Kleur	Totaal	Weglaten	Egal	Behouden	Type
Groen	-1	0	3	-1	p_actief
Paars	-4	2	2	-2	p_aspr
Oranje	4	-2	2	2	p_actief
Rood	1	0	1	1	p_aspr
Toevoeging	5	-2	1	3	w_onbekend

4 Conclusie

1. Antwoorden op de deelvragen.
2. Antwoorden op de onderzoeksvraag.

5 Discussie

1. De populatie proefpersonen is zeer klein en niet representatief voor mensen in woningcorporaties. In vervolgonderzoek moet de enquête worden uitgezet onder mensen die daadwerkelijk gebruiken van woningcorporaties.
2. Met een grotere samplesize moet de enquête ook veel korter. Reacties van participanten waren dat het lang duurde om in te vullen.
3. De laatste vraag, waar de kleuren worden vergeleken moet een gesloten vraag worden. Participanten moeten aanklikken welke kleuren ze willen behouden en welke ze willen wijzigen. Dit maakt de reacties eenduidiger interpreteerbaar. Daarbij was de vraag tweeledig. Wat wil je weglaten en wat wil je behouden? Dit zorgt voor verwarring en men kan de ene helft van een vraag vergeten. Dit moet in vervolgonderzoek eenduidiger.
4. Verder interessant om contracten te onderzoeken. Vanuit het recht geldt immers: Je hebt pas een geldige rechtshandeling als de wil en verklaring overeenkomen. Het is maar de vraag of mensen bij het ondertekenen van hun contract weten wat ze verklaren.
5. Er moet onderzoek gedaan worden naar in hoeverre mensen in sociale huurwoningen het jargon van woningcorporaties kennen. De corpus van LiNT kende weinig jargon, daar zijn de teksten op aangepast maar wie zegt dat men hier ook geen weet van heeft?
6. Onderzoek doen naar hoe het in de praktijk wordt ervaren. Merken woningcorporaties dat er problemen zijn met huurders door dat hun brieven niet begrepen worden?
7. Tool implementeren bij software woningcorporaties zelf?
8. Vragen meer in lijn met Cox en Cox (2008).
9. Het was achteraf niet altijd duidelijk voor participanten of ze de vragen moesten beantwoorden als zichzelf of dat ze zich moesten inbeelden als ware ze laaggeletterd waren (zucht... hogeropgeleiden)
10. Introductietekst toevoegen bij elke vraag zodat daar geen leesbaarheidsvragen over gesteld kunnen worden.
11. Onderzoeken of LLM's dit werk ook kunnen doen en of dit beter gaat. Obv bijvoorbeeld de data uit LiNT.
12. Het was een goed idee om de teksten niet inhoudelijk te wijzigen. Hierdoor voorkomen dat er over andere onderdelen iets gezegd zou worden.
13. "A Mann-Whitney U test is used to compare the differences between two samples when the sample distributions are not normally distributed and the sample sizes are small ($n < 30$)."

Appendices

A Gebruikte brieven overzicht

Dit zijn de gebruikte brieven:

Id	Tekst	complexity_level	complexity_level_100
0	O_Woonzorg_x00	3	47
1	N_Woonzorg_x00	2	39
2	O_Acantus_x12	3	52
3	N_Acantus_x12	2	38
4	O_ThuisInLimburg_x34	3	51
5	N_ThuisInLimburg_x34	2	43
6	O_FrieslandHuurt_x04	3	51
7	N_FrieslandHuurt_x04	2	38
8	O_Woonzorg_x01	4	70
9	N_Woonzorg_x01	1	31
10	O_ThuisInLimburg_x26	3	50
11	N_ThuisInLimburg_x26	2	37
12	O_DeWoningZoeker_x11	3	49
13	N_DeWoningZoeker_x11	2	39
14	O_ThuisInLimburg_x03	3	50
15	N_ThuisInLimburg_x03	2	37
16	O_ThuisInLimburg_x22	3	49
17	N_ThuisInLimburg_x22	1	32
18	O_DeWoningZoeker_x02	3	49
19	N_DeWoningZoeker_x02	2	39
20	O_ThuisInLimburg_x38	3	46
21	N_ThuisInLimbrug_x38	2	37
22	O_DeWoningZoeker_x10	3	48
23	N_DeWoningZoeker_x10	2	43
24	O_FrieslandHuurt_x08	3	46
25	N_FrieslandHuurt_x08	2	40
26	O_DeWoningZoeker_x23	3	46
27	N_DeWoningZoeker_x23	2	42

Tabel 5: Proces van omschrijven

Name_id	pre complexity level	pre complexity level 100	act complexity level	act complexity level 100	post complexity level	post complexity level 100
Acantus_x00	3	46	2	41		
Acantus_x05	3	47	2	43		
Acantus_x07	3	46	1	29		
Acantus_x12	3	52	3	52	2	38
Acantus_x18	3	48	2	45		
DeWoningZoeker_x02	3	49	3	49	2	39
DeWoningZoeker_x10	3	47	3	48	2	43
DeWoningZoeker_x11	3	50	3	49	2	39
DeWoningZoeker_x23	3	46	3	46	2	42
FrieslandHuurt_x04	3	52	3	51	2	38
FrieslandHuurt_x08	3	46	3	46	2	40
ThuisInLimburg_x03	3	53	3	49		
ThuisInLimburg_x04	3	49	2	50	2	37
ThuisInLimburg_x22	3	49	3	49	1	32
ThuisInLimburg_x26	3	49	3	50	2	37
ThuisInLimburg_x34	3	51	3	51	2	43
ThuisInLimburg_x38	3	48	3	47		
ThuisKompas_x05	3	46	2	46	2	37
ThuisKompas_x08	3	47	2	42		
Valburg_x03	3	46	2	45		
Woonzorg_x00	3	47	3	47	2	39
Woonzorg_x01	4	70	4	70	1	31

B Overzicht variabelen

>

C Synoniemen en vaktermen

>

D Gebruikte alinea's overzicht

>Beter om plaatjes in te voegen. Markeringen met pdftex werken slecht. Scheelt bovendien hoop tijd.

D.1 Tekst 1

Origineel

Beste,

Hierbij herinneren wij u aan de documenten die moeten worden geüpload om de uiteinde

Met vriendelijke groet,
Dirk

Herschreven

Beste,

U heeft een aanbieding ontvangen. U kan de aanbieding pas accepteren als wij een aant
herinneren om deze te uploaden op de website.

Met vriendelijke groet,
Dirk

D.2 Tekst 2

Origineel

Waarom dit bericht?

CORPORATIE is voor de verhuur van woningen aangesloten bij Stichting WONINGCORPORATIES. **om bepaalde personen onder de aandacht te houden, kan deze vanuit de woningcorporatie worden** geplaatst is op deze lijst, wordt de woningcorporatie vanuit **WONINGCORPORATIES** **inschrijven als woningzoekende.** **Plaatsing op de signaleringslijst is voor maximaal 5 jaar.**

Meer informatie over de signaleringslijst **of mogelijkheden tot bezwaar** kunt u opnemen met Dirk.

Herschreven

Waarom dit bericht?

WONINGCORPORATIE is aangesloten bij de Stichting WONINGCORPORATIES. **Om te zorgen dat** **Mensen die dit niet doen, komen op de signaleringslijst.** **De signaleringslijst delen we met andere** **woningcorporatie van Stichting WONINGCORPORATIES weten zij dus van uw problemen als huurder.** **Op de website** kunt u meer informatie vinden over de signaleringslijst. **Wilt u** **kijken.** **U kunt hiervoor** telefonisch contact opnemen met Dirk.

Referenties

- Aluisio, S., Specia, L., Gasperin, C. & Scarton, C. (2010). Readability assessment for text simplification. In *Proceedings of the naacl hlt 2010 fifth workshop on innovative use of nlp for building educational applications* (pp. 1–9).
- Cox, J. & Cox, K. B. (2008). *Your opinion, please!: How to build the best questionnaires in the field of education*. Corwin Press.
- Crossley, S. A., Skalicky, S. & Dascalu, M. (2019). Moving beyond classic readability formulas: new methods and new models. *Journal of Research in Reading*, 42(3-4), 541-561. Verkregen van <https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-9817.12283>
doi: <https://doi.org/10.1111/1467-9817.12283>
- D Clerq, O., HOSTE, V., DESMET, B., VAN OOSTEN, P., DE COCK, M. & MACKEN, L. (2014). Using the crowd for readability prediction. *Natural Language Engineering*, 20(3), 293–325. doi: 10.1017/S1351324912000344
- Freebody, P. & Anderson, R. C. (1983). Effects on text comprehension of differing proportions and locations of difficult vocabulary. *Journal of Reading Behavior*, 15(3), 19–39.

- Kleijn, S. (2018). Clozing in on readability. *How linguistic features affect*.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of psychology*.
- Maat, H. P. & Gravekamp, J. (2022). Kan een tekst te simpel zijn? hoe lager en hoger opgeleiden oordelen over eenvoudige taal [Journal Article]. *Tijdschrift voor Taalbeheersing*, 44(2), 62–90. Verkregen van <https://www.aup-online.com/content/journals/10.5117/TVT2022.2.002.MAAT> doi: <https://doi.org/10.5117/TVT2022.2.002.MAAT>
- Maat, H. P., Kleijn, S. & Frissen, S. (2023). Lint: een leesbaarheidsformule en een leesbaarheidsinstrument. *Tijdschrift voor Taalbeheersing*, 45(1), 2–39.
- Maat, H. P., Kraf, R., Dekker, N., van der Sloot, K., van der Klis, M., van den Bosch, A., ... Kleijn, S. (2017). *Handleiding t-scan*. Utrecht, Nijmegen, Tilburg.
- Mann, H. B. & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, 50–60.
- Pander Maat, H. & Ditewig, S. (2017). Hoe worden onderwijsteksten vereenvoudigd, en helpt dat? *Tijdschrift voor taalbeheersing*, 39(2), 245–263.
- Pander Maat, H., Kraf, R., Van den Bosch, A., Dekker, N., Gompel, M. v., Kleijn, S. d., ... Sloot, K. v. d. (2014). T-scan: a new tool for analyzing dutch text.
- Pander Maat, H. & van der Geest, T. (2021). Monitor begrijpelijke overheidsteksten.
- Scarton, C., Oliveira, M., Candido Jr, A., Gasperin, C. & Aluísio, S. (2010). Simplifica: a tool for authoring simplified texts in brazilian portuguese guided by readability assessments. In *Proceedings of the naacl hlt 2010 demonstration session* (pp. 41–44).
- van Raaij, N. B., Kolkman, D. & Podoyntsyna, K. (2024). Clearer governmental communication: Text simplification with chatgpt evaluated by quantitative and qualitative research. In *Proceedings of the workshop on determit! evaluating text difficulty in a multilingual context@ lrec-coling 2024* (pp. 152–178).