

# **Scholieren met dyslexie van de derde graad middelbaar onderwijs ondersteunen bij het lezen van wetenschappelijke papers via tekstvereenvoudiging.**

## **Vergelijkende studie en de opbouw van een prototype.**

---

**Dylan Cluyse.**

Scriptie voorgedragen tot het bekomen van de graad van  
Professionele bachelor in de toegepaste informatica

**Promotor:** Mevr. L. De Mol

**Co-promotor:** J. Decorte; J. Van Damme; M. Dhondt

**Academiejaar:** 2022–2023

**Eerste examenperiode**

**Departement IT en Digitale Innovatie .**

**HO  
GENT**



# Woord vooraf

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

# Samenvatting

Ingewikkelde woordenschat en zinsbouw hinderen scholieren met dyslexie in het derde graad middelbaar onderwijs bij het lezen van wetenschappelijke artikelen. Adaptieve tekstvereenvoudiging helpt deze scholieren bij hun lees- en verwerkingsnelheid. Daarnaast kan artificiële intelligentie (AI) dit proces automatiseren om de werkdruk bij leraren en scholieren te verminderen. Dit onderzoek achterhaalt met welke technologische en logopedische aspecten AI-ontwikkelaars rekening moeten houden bij de ontwikkeling van een AI-toepassing voor adaptieve en geautomatiseerde tekstvereenvoudiging. Hiervoor is de volgende onderzoeksvraag opgesteld: "Hoe kan een wetenschappelijk artikel automatisch worden vereenvoudigd, gericht op de unieke noden van scholieren met dyslexie in het derde graad middelbaar onderwijs?". Een vergelijkende studie beantwoordt deze onderzoeksvraag en is uitgevoerd met bestaande toepassingen en een prototype voor adaptieve en geautomatiseerde tekstvereenvoudiging. Uit de vergelijkende studie blijkt dat toepassingen om wetenschappelijke artikelen te vereenvoudigen, gemaakt zijn voor een centrale doelgroep en geen rekening houden met de unieke noden van een scholier met dyslexie in het derde graad middelbaar onderwijs. Adaptieve software voor geautomatiseerde tekstvereenvoudiging is mogelijk, maar ontwikkelaars moeten meer inzetten op de unieke noden van deze scholieren.

# Inhoudsopgave

<b>Lijst van figuren</b>	<b>viii</b>
<b>1 Inleiding</b>	<b>1</b>
1.1 Probleemstelling	2
1.2 Onderzoeksvraag	3
1.3 Onderzoeksdoelstelling	3
1.4 Opzet van deze bachelorproef	3
<b>2 Stand van zaken</b>	<b>5</b>
2.1 Onderzoeken rond dyslexie	5
2.1.1 Centraal zicht op dyslexie	5
2.1.2 Centraal zicht op de doelgroep	6
2.1.3 Diagnosecriteria	6
2.1.4 Mogelijke drempels voor mensen met fonologische dyslexie.	7
2.1.5 Bewezen effecten van tekstvereenvoudiging -en aanpassing bij scholieren met dyslexie.	10
2.1.6 Samenvattend schema over aanpakken	13
2.1.7 Conclusie	13
2.2 Wetenschappelijke artikelen	13
2.2.1 Wetenschappelijke geletterdheid in Vlaanderen	14
2.2.2 Trends rond wetenschappelijke artikelen	15
2.2.3 Woordenschat en vakjargon	18
2.2.4 Aanpak voor het lezen van wetenschappelijke artikelen	18
2.2.5 Conclusie	19
2.3 Tekstvereenvoudiging	19
2.3.1 Manuele tekstvereenvoudiging	20
2.3.2 Natural Language Processing	22
2.4 De verschillende soorten tekstvereenvoudiging	27
2.4.1 Lexicale vereenvoudiging	27
2.4.2 Syntactische vereenvoudiging	28
2.4.3 Conceptuele of semantische vereenvoudiging	28
2.4.4 Overige vormen van vereenvoudiging	29
2.4.5 Tekstvereenvoudiging automatiseren	29

2.5	Samenvatten . . . . .	29
2.5.1	Extraherend samenvatten . . . . .	30
2.5.2	Abstraherend samenvatten . . . . .	33
2.5.3	Hybride samenvatten . . . . .	34
2.5.4	Evaluatie . . . . .	34
2.5.5	Tekstvereenvoudigingstechnieken voor scholieren met dyslexie. 36	
2.5.6	Conclusie . . . . .	37
2.6	Valkuilen en uitdagingen voor AI-ontwikkelaars bij tekstvereenvoudi- ging met AI . . . . .	38
2.6.1	Uitdagingen voor softwarebedrijven . . . . .	38
2.6.2	Ambigüiteit, synoniemen en homoniemen . . . . .	38
2.6.3	Paternalisme en ethische overwegingen . . . . .	39
2.6.4	Valkuilen bij prompt engineering . . . . .	40
2.6.5	Evaluatie en interpretatie . . . . .	40
2.7	Beschikbare software voor tekstvereenvoudiging . . . . .	41
2.7.1	Momenteel ingezet in het onderwijs . . . . .	41
2.7.2	Proof-of-concepts en online webapplicaties . . . . .	42
2.7.3	GPT-3 . . . . .	42
2.7.4	Bing AI . . . . .	48
2.7.5	Meta LLaMa . . . . .	49
2.7.6	Samenvattend schema AI NLP-modellen . . . . .	50
2.7.7	Conclusie . . . . .	50
2.8	Conclusie . . . . .	50
<b>3</b>	<b>Methodologie</b>	<b>51</b>
3.1	Requirementsanalyse . . . . .	51
3.1.1	Software in het onderwijs . . . . .	51
3.1.2	Tekstanalyse . . . . .	51
3.1.3	Lexicale vereenvoudiging . . . . .	51
3.1.4	Conclusie . . . . .	52
3.2	Ontwikkelen van een prototype . . . . .	52
3.2.1	Schetsing met Python notebooks . . . . .	52
3.2.2	Integreren naar Flask . . . . .	52
<b>4</b>	<b>Conclusie</b>	<b>53</b>
<b>A</b>	<b>Onderzoeksvoorstel</b>	<b>55</b>
A.1	Introductie . . . . .	55
A.2	State-of-the-art . . . . .	57
A.2.1	Tekstvereenvoudiging . . . . .	57
A.2.2	Noden van scholieren met dyslexie . . . . .	57
A.2.3	Huidige toepassingen . . . . .	59

A.2.4 Ontwikkelen met AI . . . . .	59
A.3 Methodologie . . . . .	60
A.4 Verwacht resultaat, conclusie . . . . .	62
<b>Bibliografie</b>	<b>63</b>

# Lijst van figuren

2.1	Afbeelding uit Bezem en Lugthart (2016). Deze afbeelding bootst fixatie disparatie na. . . . .	9
2.2	Afbeelding van DuBay (2004) . . . . .	10
2.3	Afbeelding van Rello, Baeza-Yates, Dempere-Marco e.a. (2013). Volgens de richting van de pijl wordt de ideale situatie benaderd, gekenmerkt door doelwaarden. Deze waarden worden bereikt door mensen zonder dyslexie onder optimale omstandigheden. Het gebruik van vaak voorkomende woorden vermindert de decodeertijd en verbetert de leesbaarheid voor mensen met dyslexie. . . . .	11
2.4	Afbeelding uit Rello en Baeza-Yates (2013). Verticaal wordt de gemiddelde mening van de bevroegden weergegeven. Horizontaal worden de lettertypes gerangschikt op gemiddelde leestijd van alle bevroegden. Dit onderzoek wijst uit dat Arial, CMU, Helvetica en Times de populaire keuzes zijn. Arial en CMU behoren hierbij tot de drie best scorende lettertypes rond gemiddeld leestempo. . . . .	12
2.5	Afbeelding uit Rello en Baeza-Yates (2013). . . . .	12
2.6	Figuur van De Meyer e.a. (2019). Op alle PISA-domeinen scoren de Vlaamse vijftienjarigen in ASO, BSO en TSO significant slechter dan de eerste metingen. De Meyer e.a. (2019) noemen dit een achteruitgang in alle onderwijsvormen. . . . .	14
2.7	Figuur van De Meyer e.a. (2019). Het leesplezier van Vlaamse 15-jarigen. Zij uitten zich uiterst negatief op stellingen over leesplezier. Volgens de enquête vond de helft van de scholieren begrijpend lezen enkel tijdsverlies en slechts 17% gaf aan dat lezen één van hun favoriete hobby's is. Er is wel een significant verschil tussen de mening van jongens en meisjes, waar jongens negatiever antwoorden op lezen. . . . .	15
2.8	Figuur van De Meyer e.a. (2019). De wetenschappelijke geletterdheid bij vijftienjarigen op internationaal niveau. Vlaanderen scoort significant slechter dan acht deelnemende landen. . . . .	15



2.9 Afbeelding uit Plavén-Sigray e.a. (2017). Links wordt de evolutie per FRE-score getoond. Hoe hoger de score, hoe hoger de gemiddelde complexiteit van een tekst. Rechts wordt de evolutie volgens de NDC-score getoond. Hoe hoger de score, hoe lager de gemiddelde complexiteit van een tekst. Het onderzoek schat dat nu een kwart van alle wetenschappelijke artikelen gebruik maken van Engels op het niveau van een masterstudent, ofwel een FRE onder nul. . . . .	16
2.10 Afbeelding uit Plavén-Sigray e.a. (2017). Horizontaal worden het aantal auteurs per wetenschappelijk artikel aangeduidt. Verticaal wordt de gemiddelde NDC-score weergegeven. HOe hoger de NDC-score, hoe hoger de vereiste leesgraad om de tekst te kunnen lezen. . . . .	17
2.11 Afbeelding van Murdos e.a. (2014) Volgens deze grafiek scoren de wetenschappelijke artikels rond fysica gemiddeld het best op de FRE-score. Al scoren de wetenschappelijke artikels rond microbiologie gemiddeld het zwakst op de FRE-score, ze scoren gemiddeld beter op de FOG-score. . . . .	18
2.12 Voorbeeld van PoS-labeling op de Engelstalige zin "She sells seashells on the seashore". Afbeelding van Bilici (2021) . . . . .	24
2.14 Afbeelding uit McFarland (2023). Een illustratie over de werking en begeleiding van prompt engineering bij een taalmodel. . . . .	25
2.13 Voorbeeld van sequence labeling op de Engelstalige zin "She sells seashells on the seashore". Afbeelding van Bilici (2021) . . . . .	25
2.15 Voorbeeld van tekstvereenvoudiging. Oorspronkelijke tekst uit Historia 5 bron toe te voegen . . . . .	27
2.16 Afbeelding van Althunayyan en Azmi (2021). Deze pipeline wordt in meerdere onderzoeken rond lexicale vereenvoudiging toegepast, zoals Bingel e.a. (2018), Bulté e.a. (2018) en Paetzold en Specia (2016) . . .	28
2.17 Afbeelding van Chauhan (2018). De manier waarop teksten automatisch samengevat kunnen worden, is afhankelijk van drie verschillende domeinen. . . . .	30
2.18 Afbeelding uit Bulté e.a. (2018). Deze pipeline omvat de stappen die de toepassing aflegt. . . . .	37
2.19 Afbeelding uit Bulté e.a. (2018). TODO . . . . .	37
2.20 Schermafbeelding van SciSpace. . . . .	42
2.21 Afbeelding van Gooding 2022. De invoertekst is een paragraaf uit een niet-vermeld boek van de Russische schrijver Dostoevsky. Het resultaat van de meegegeven prompt is een transformatie dat iedere vorm van vooraf aangehaalde vereenvoudiging weergeeft. Lexicale, conceptuele en syntactische vereenvoudiging worden op de invoertekst toegepast. . . . .	43

2.22 Afbeelding van Gooding 2022. Gooding haalt verder aan dat modellen zoals ChatGPT op twee vlakken de leesbaarheid van een tekst kan bevorderen. Allereerst door het verlenen van verschillende mogelijke versies van een vereenvoudigingstaak. . . . .	43
2.23 Afbeelding van Gooding 2022. . . . .	44
2.24 Afbeelding van Binz en Schulz (2023). Dit toont de <i>mean regret</i> aan tussen de vier engines en de menselijke antwoorden. . . . .	45
2.25 Afbeelding van Simon (2021). De evolutie van pre-trained taalmodellen wordt hier weergegeven tot eind 2022. De performantie van de modellen ten opzichte van de grootte volgt een lineaire functie. . . . .	46
2.26 Afbeelding van Ribas (2023). . . . .	48
2.27 In deze afbeelding wordt er een online wetenschappelijk artikel meegegeven. Er wordt geen titel of onderwerp meegegeven, maar de Bing AI chatbot is in staat om een abstraherende samenvatting te maken van het artikel. Daarna geeft de chatbot verder uitleg over een bepaald onderwerp en geeft het extra referenties mee. . . . .	49
A.1 (Readable, 2021) . . . . .	61

# 1

## Inleiding

Het middelbaar onderwijs staat op springen. Dagelijks sneuvelen leerkrachten en scholieren van het middelbaar onderwijs onder de te harde werkdruk (Minnen e.a., 2018). Lezen is ingebed in ons dagelijks leven in de vorm van Nederlandstalige nieuwsartikelen tot de ondertiteling van televisieseries. Mensen van eender welke leeftijdsgroep kunnen lezen niet ontsnappen en dit moet van jongs af aan geprikkeld worden (Daoud, 2023).

### Vakmiddelen -en didactiek in het onderwijs

Lerarenopleidingen benadrukken nu het gebruik van verschillende bronnen in lessen. De leesgraad van deze bronnen verandert echter niet, want de noodzaak aan bronnen met diverse leesgraden is bedoeld om scholieren uit te kunnen dagen (Surma e.a., 2019). Het Amerikaanse onderwijs stampte C.R.E.A.T.E.<sup>1</sup> uit de grond. Dit initiatief zet scholieren tussen 12 en 18 jaar aan om wetenschappelijke artikelen te lezen in plaats van enkel boeken. Scholieren komen zo in direct contact met wetenschappelijk onderzoek. Ze begrijpen hoe wetenschappers experimenten uitvoeren, plannen en resultaten analyseren en interpreteren. Vlaamse STEM-leerkrachten in de derde graad middelbaar onderwijs moeten volgens het M-decreet en de leerplannen van zowel het katholiek<sup>2</sup> als het gemeenschapsonderwijs<sup>3</sup> hun theorielessen op een toegankelijke manier aanbieden, zodat alle scholieren worden meegenomen in het verhaal.

### Kunstmatige intelligentie in Vlaanderen

Met een jaarlijks budget van 32 miljoen is België een pionier in het vakgebied artificiële intelligentie (AI) op de werkvloer (Crevits, 2022). Zo stampte de Vlaamse overheid verschillende AI-projecten uit de grond, om Vlaamse AI-ontwikkelingen te on-

---

<sup>1</sup><https://teachcreate.org/>

<sup>2</sup><https://pro.katholiekonderwijs.vlaanderen/basisoptie-stem/ondersteunend-materiaal>

<sup>3</sup><https://g-o.be/stem/>

dersteunen en om AI-softwarebedrijven te inspireren. Het amai!-project<sup>4</sup> brengt AI-softwarebedrijven uit verschillende domeinen samen. Dit project leidt tot het ontstaan van AI-toepassingen die processen automatiseren om de werkdruk te verminderen, zoals *real-time* ondertiteling in de klas en een taalassistent voor leerkrachten in meertalige klasgroepen.

## 1.1. Probleemstelling

Ghesquière (2018) schat dat 5 tot 9 % van de Nederlandstalige bevolking<sup>5</sup> nu de diagnose van dyslexie heeft. Lissens e.a. (2020) benadrukt dat de impact van leerstoornissen niet stopt na het middelbaar onderwijs. Scholieren met dyslexie in het middelbaar onderwijs kampen met unieke uitdagingen. Ze worden niet aan hun lot overgelaten en kunnen rekenen op ondersteuning van coaches en beschikbare hulpmiddelen om hun leesachterstand te beperken. Zoals eerder aangehaald, stimuleert het leerplan voor STEM-vakken het gebruik van wetenschappelijke artikelen, maar houdt niet altijd rekening met de moeilijkheidsgraad ervan. De complexe woordenschat en zinsopbouw in deze artikelen vormen een barrière voor de begripelijkheid van een tekst, waardoor scholieren met dyslexie de kerninhoud moeilijk kunnen doorgronden. Een oplossing hiervoor is om de tekst te vereenvoudigen, zonder de kern- en bijzaken te verliezen.

Leesvaardigheid is een belangrijke vaardigheid om te slagen op school en op de arbeidsmarkt. Mensen met een lagere leesvaardigheid kunnen echter negatieve gevolgen ondervinden op hun psychosociaal welbevinden, waaronder een lagere zelfwaardering. Hoewel werkgevers zich voor inclusiviteit en diversiteit openstellen volgens Lissens e.a. (2020), kampen mensen met een leesstoornis nog steeds met problemen bij de overgang naar de arbeidsmarkt. Moeizame en stroeve automatisering bij het lezen en spellen kunnen dit proces verstoren. Ondersteuningsmaatregelen -en hulpmiddelen kunnen wel helpen bij het bereiken van succes, maar het is belangrijk om te erkennen dat de noden van deze groep niet altijd even goed begrepen worden naarmate ze ouder worden (Lissens e.a., 2020).

Wetenschappelijke artikelen vereenvoudigen vraagt tijd en energie van docenten in de derde graad middelbaar onderwijs. Het Vlaamse middelbaar onderwijs staat nu onder druk en docenten hebben moeite om met deze werkdruk boven water te blijven. Daarom is er nood aan software die wetenschappelijke artikelen automatisch kan vereenvoudigen, specifiek gericht op de unieke noden van scholieren met dyslexie. Een dergelijke toepassing ondersteunt het scholieren met dyslexie in de derde graad middelbare onderwijs bij het lezen van een wetenschappelijk artikel. Daarnaast vermindert deze toepassing het routinematige werk van STEM-docenten.

---

<sup>4</sup><https://amai.vlaanderen/>

<sup>5</sup>Deze schatting is gebaseerd op de Vlaamse en Nederlandse bevolking.

## **1.2. Onderzoeksvraag**

De volgende onderzoeksvraag is opgesteld: "Hoe kan een wetenschappelijke artikel automatisch vereenvoudigd worden, gericht op de unieke noden van scholieren met dyslexie in de derde graad middelbaar onderwijs?". Daarnaast worden de volgende deelvragen beantwoord.

- Welke aanpakken zijn er voor geautomatiseerde tekstvereenvoudiging? Aansluitende vraag: "Hoe worden teksten handmatig vereenvoudigd voor scholieren met dyslexie?"
- Welke specifieke noden hebben scholieren met dyslexie van de derde graad middelbaar onderwijs bij het begrijpen van complexere teksten?
- Wat zijn de specifieke kenmerken van wetenschappelijke artikelen?
- Met welke valkuilen bij taalverwerking met AI moeten ontwikkelaars rekening houden?
- Welke toepassingen, tools en modellen zijn er beschikbaar om Nederlandse geautomatiseerde tekstvereenvoudiging met AI mogelijk te maken?
- Welke functies ontbreken AI-toepassingen om geautomatiseerde tekstvereenvoudiging mogelijk te maken voor scholieren met dyslexie in de derde graad middelbaar onderwijs? Aansluitende vraag: "Welke manuele methoden voor tekstvereenvoudiging komen niet in deze tools voor?"

## **1.3. Onderzoeksdoelstelling**

Het doel van dit onderzoek is om te achterhalen met welke technologische en logopedische aspecten AI-ontwikkelaars rekening moeten houden bij de ontwikkeling van een adaptieve AI-toepassing voor geautomatiseerde tekstvereenvoudiging. Het resultaat van dit onderzoek is een prototype voor een toepassing die de tekstinhoud van een wetenschappelijke paper zal vereenvoudigen, naargelang de specifieke noden van een scholier met dyslexie in de derde graad middelbaar onderwijs. Het prototype houdt rekening met de transformatie van het bronbestand, bijvoorbeeld een PDF of een afbeelding, naar de tekstinhoud. Hiervoor bestaan er kant-en-klare pakketten dat de transformaties al voor de ontwikkelaar doen. De invoer van dit prototype is een wetenschappelijk artikel van minstens 500 woorden lang.

## **1.4. Opzet van deze bachelorproef**

De rest van deze bachelorproef is als volgt opgebouwd:

In Hoofdstuk 2 wordt een overzicht gegeven van de stand van zaken binnen het onderzoeksdomein, op basis van een literatuurstudie.

In Hoofdstuk 3 wordt de methodologie toegelicht en worden de gebruikte onderzoekstechnieken besproken om een antwoord te kunnen formuleren op de onderzoeksvragen.

- Welke aanpakken zijn er voor geautomatiseerde tekstvereenvoudiging? Aansluitende vraag: "Hoe worden teksten handmatig vereenvoudigd voor scholieren met dyslexie?"
- Welke specifieke noden hebben scholieren met dyslexie van de derde graad middelbaar onderwijs bij het begrijpen van complexere teksten?
- Wat zijn de specifieke kenmerken van wetenschappelijke artikelen?
- Met welke valkuilen bij taalverwerking met AI moeten ontwikkelaars rekening houden?
- Welke toepassingen, tools en modellen zijn er beschikbaar om Nederlandstalige geautomatiseerde tekstvereenvoudiging met AI mogelijk te maken?
- Welke functies ontbreken AI-toepassingen om geautomatiseerde én adaptieve tekstvereenvoudiging mogelijk te maken voor scholieren met dyslexie in de derde graad middelbaar onderwijs? Aansluitende vraag: "Welke manuele methoden voor tekstvereenvoudiging komen niet in deze tools voor?"

In Hoofdstuk 4, tenslotte, wordt de conclusie gegeven en een antwoord geformuleerd op de onderzoeksvragen. Daarbij wordt ook een aanzet gegeven voor toekomstig onderzoek binnen dit domein.

# 2

## Stand van zaken

### 2.1. Onderzoeken rond dyslexie

Lezen is een essentieel onderdeel van ons dagelijks leven en speelt een belangrijke rol in onze communicatie en begrip. Dyslexie kan het functioneren in het dagelijks leven belemmeren. Het begrijpen van de noden en hindernissen voor een scholier met dyslexie is van belang om deze doelgroep te ondersteunen en hun kwaliteit van lezen te verbeteren. Deze sectie zal ingaan op de unieke noden en bespreken hoe mensen met dyslexie kunnen worden geholpen bij het lezen. De volgende onderzoeksvraag wordt in deze sectie beantwoordt: "Welke specifieke noden hebben scholieren met dyslexie van de derde graad middelbaar onderwijs bij het begrijpen van complexere teksten?"

#### 2.1.1. Centraal zicht op dyslexie

Lezen is onnatuurlijk en volgens de geschiedenis van de mens een recent begrip. Pas 5000 jaar geleden werd de geschreven taal bedacht. Mensen worden niet met leesvaardigheid geboren, maar leren dit zelf aan en daarvoor moet het brein heringericht worden (Bonte, 2020; van der Meer, 2022). De visuele cortex is een hersengebied dat instaat voor de aanschouwelijke waarneming en herkenning van objecten, zoals meubels of letters. Lezen traint dit gebied bij het herkennen van letters, maar dit is niet evident en vordering bij letterherkenning gebeurt pas na meerdere pogingen.

Dyslexie betekent letterlijk 'beperkt lezen'. Het voorlezen verloopt radend, letter-voor-letter en langzaam. Lezen gaat niet soepel en leesproblemen kunnen beperkend zijn. Elke zin verloopt langzamer. Dyslexie is genetisch en erfelijkheid speelt een rol. Goede woordenschat ontwikkeling of vaak voorlezen is een beschermende factor tegen dyslexie. Ten slotte noemen Bonte (2020) en Vellutino e.a. (2004) dyslexie een verborgen stoornis, want de diagnose kan niet gesteld worden met her-

senscans en vereist daarmee een nauwe opvolging bij de diagnose. Onderzoeken halen drie verschillende types van dyslexie aan, namelijk fonologische dyslexie, *surface dyslexia* en *deep dyslexia*. Dezelfde onderzoeken wijzen erop dat een overlap van kenmerken over de drie types heen mogelijk is (Rello, Kanvinde e.a., 2012; Vellutino e.a., 2004).

### 2.1.2. Centraal zicht op de doelgroep

#### Statistieken rond prevalentie en comorbiditeit

Vlaamse en Nederlandse onderzoeken van Desoete (2017) en Wentink e.a. (2008) wijzen uit dat gemiddeld 4% van de Nederlandstalige bevolking de diagnose van dyslexie heeft. De transparantie van de taal beïnvloedt volgens Association (2013) de prevalentie van dyslexie in een taalgebied. Spaans, Italiaans en Chinees zijn transparant en hebben een sterkere grafeem-foneem en foneem-grafeem koppeling. Zo is bijvoorbeeld slechts 1% van de Chinese sprekers dyslectisch. Engels is een minder transparante taal waar klanken op verschillende manieren geschreven kunnen worden, bijvoorbeeld *eight* en *late*. De prevalentie van dyslexie ligt met 20% dan ook hoger in Engelstalige landen. Het aantal scholieren met dyslexie in het lager en middelbaar onderwijs wereldwijd loopt op tot 15% ingeschat (Bonte, 2020; van der Meer, 2022).

#### Scholieren met dyslexie in de derde graad middelbaar onderwijs

Mensen met dyslexie ondervinden een last bij het technisch lezen van een tekst. Al zijn er onderzoeken naar adolescenten en volwassenen met dyslexie, onderzoeken rond dyslexie richten zich vaker op kinderen in het kleuter- en basisonderwijs. Volgens Bonte (2020) is dit om voldoende onderzoek te hebben voor de meest geschikte diagnose en ondersteuning op jonge leeftijd. Deze diagnose zo vroeg mogelijk vastleggen vereist voldoende onderzoek. Bonte haalt een sneeuwbal-effect aan wanneer kinderen met dyslexie een zwakke ondersteuning krijgen. Deze last wordt megedragen naar volgende levensfasen. Volgens Lissens e.a. (2020) zijn jongvolwassenen en ouderen een leeftijdsgroep die bij onderzoeken rond dyslexie over het hoofd word gezien. Het onderzoek van Lissens e.a. (2020) benadrukt deze vaak over het hoofd geziene doelgroep. Mensen met dyslexie ervaren uitdagingen bij het lezen en schrijven. Op sociaal vlak worden deze mensen niet begrepen door anderen, omdat ze woorden vaak verkeerd uitspreken of verkeerd spellen. Het tempo van anderen bijhouden kan stroef verlopen en het verwerken van teksten vereist meer energie. Ondanks het vele oefenen kan snel en veel lezen moeilijk blijven.

### 2.1.3. Diagnosecriteria

Dyslexie is geen lijst van 'kwalen' waaraan een scholier moet voldoen om dit te bezitten. Zo halen Ghesquière (2018) en Kleijnen e.a. (2008) samen drie beschrijvende cri-



teria aan waarmee de diagnose van dyslexie wordt vastgesteld. Het achterstands-criterium wijst aan dat een persoon ten opzichte van andere leeftijdsgenoten niet even hoog scoort op het vlak van lezen. Vervolgens houdt het hardnekkigheids-criterium in dat de lees- of spellingsachterstand het gevolg is van een moeizame automatisering van het lees- en spellingsproces. Ten slotte wijst het exclusiviteits-criterium volgens Ghesquière (2018) erop dat een persoon enkel de diagnose van dyslexie heeft, zonder bijhorende lees- of spellingsstoornissen.

Verder halen Ghesquière (2018) en Van Vreckem en Desoete (2015) de volgende kenmerken aan die kunnen verschillen per individu.

- De ernst of uitgebreidheid van een stoornis.
- De gevolgen van een stoornis, zoals faalangst.
- De mate waarin iemand al dan niet kan compenseren.
- De secundaire kenmerken zoals problemen met werkhouden en structuur.

Het onderzoek van Van Vreckem en Desoete (2015) achterhaalt of het groepsprofiel overeenkomt met de individuele profielen bij begrijpend lezen en spelling. De onderzoeksresultaten bij een experiment met zeventien kinderen wijst uit dat er een geïndividualiseerde analyse en effectieve behandeling op maat nodig is bij begrijpend lezen. Onvoldoende beheerste leerstof en leesstrategieën per kind moet achterhaald worden om zo specifieke begeleiding te kunnen bieden.

#### **2.1.4. Mogelijke drempels voor mensen met fonologische dyslexie.**

Spellingsregels toepassen hindert mensen met dyslexie. Dit kan leiden tot een onzeker gevoel en angst om fouten te maken, frustratie en stress, waardoor het voor mensen met dyslexie moeilijker wordt om hun vaardigheden te verbeteren. Vooroordelen aan dyslexie zijn prevalent volgens Diels (2022). Mensen met dyslexie kunnen in het bedrijfsleven en binnen de schoolmuren als minder intelligent of traag worden afgestempeld. Sommige mensen met dyslexie proberen deze leesstoornis te maskeren door bijvoorbeeld moeilijke woorden in afkortingen te schrijven. Toch motiveren onderzoeken zoals Bonte (2020), Ghesquière (2018) en Lissens e.a. (2020) deze doelgroep door hun sterk doorzettingsvermogen, geduldig en luisterend karakter en probleemoplossend denkvermogen te benadrukken. Deze motivatie en inspiratie helpt deze doelgroep om te blijven oefenen en hun vaardigheden te ontwikkelen, ondanks de uitdagingen. Onderzoeken rond digitale toepassingen voor kinderen en scholieren met dyslexie reiken prevalentie moeilijkheden en struikelblokken voor deze doelgroep aan. De onderzoeken beklemtonen welke unieke noden scholieren met dyslexie kunnen hebben.

- Langzame woordbenoeming

- Hardnekkig letter-voor-letter lezen
- Woordherkenning en -herinnering
- Visuele disfunctie
- Letter- en klankvorming
- Homofonische of pseudo-homofonische woordenschat
- Begripsproblemen

### Langzame woordbenoeming

Het correct spellen van pseudowoorden en regelmatig gespelde woorden is mogelijk met beheerste letterklankkoppelingen. Echter verloopt het automatiseren van moeilijke en nieuwe woorden stroef, met een trage woordbenoeming tot gevolg. Lezers kunnen met dit leesprobleem veel woorden niet als één geheel herkennen. Filipiak (2020) raadt aan om pseudowoorden en het herkennen te oefenen als mogelijke hulp. De meeste schrijffouten komen voor in onregelmatig gespelde woorden waardoor een fonologische route die wel tact is, leidt tot het schrijven van *gedaan* als *guhdaan*.

### Begripsproblemen

Typerende symptomen de verstoring van leesbegrip en het spreken in qua betekenis onbedoelde klanken, woorden en woordgroepen. Bij fonologische dyslexie kan bijvoorbeeld gelezen worden als boom. Begripsproblemen bij het lezen kunnen goed visueel, met steun van film en afbeeldingen ondersteund worden, beter dan alleen via gedrukte woorden. Daarbij moet de lezer die het geschreven woord wil ontcijferen, gebruik kunnen maken van bronnen van kennis op een hoger niveau: een grote woordenschat en een goed redeneervermogen. Schriftelijke expressie is uit den boze.

### Hardnekkig letter-voor-letter lezen

Bonte (2020) haalt een minder optimale informatieverwerking in visuele gebieden aan bij scholieren met dyslexie, wat het gevolg is van een minder optimale leesontwikkeling. Deze verwerking is belangrijk voor letter- en woordherkenning. Lezers zijn niet in staat woorden goed te lezen, zelfs niet met een langzame en spellende letter-klankroute. Lange woorden worden moeizaam gelezen en scholieren hebben de neiging om visueel gedesoriënteerde te raken. Er is verwarring over de richting van de letters.

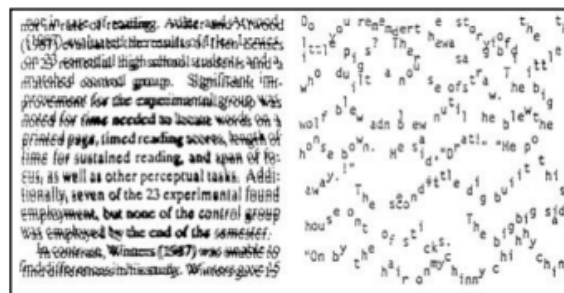
### Woordherkenning

Oogbewegingsonderzoeken zoals Rivero-Contreras e.a. (2021) en M. Zhang e.a. (2021) laten zien dat geoefende lezers pauzeren bij 50-80% van de woorden om zich erop

te fixeren tijdens een saccade (een snelle oogbeweging). Kinderen kunnen de letters correct herkennen, maar last hebben van verspringende letters tussen woorden in zinnen, wat minder voorkomt bij het lezen van woorden in lijsten en opsommingen dan bij doorlopende tekst en volzinnen.

### Visuele disfunctie

Het boek van Bezem en Lugthart (2016) maakt ouders en leerkrachten bewust rond de principes van visuele disfunctie. Mensen met visuele disfunctie observeren objecten anders dan de meeste mensen. Fixatie disparatie is een vorm van visuele disfunctie dat een onjuiste samenwerking van de ogen veroorzaakt. Ogen doen hun werk vanuit een verschillende positie en nemen zaken afzonderlijk waar. Deze afzonderlijke observaties moeten samengenomen worden. Als deze fixatie niet correct verloopt, dan wordt er van disparatie gesproken. Het gevolg is dat mensen met drie letters tegelijk zien en daardoor een woordbeeld missen of een zwakke spellingstijl gebruiken.



**Figuur (2.1)**

Afbeelding uit Bezem en Lugthart (2016). Deze afbeelding bootst fixatie disparatie na.

In het onderwijs worden er initiatieven ingezet. Zo krijgen scholieren in het lager en middelbaar onderwijs een vaste plaats. Deze scholieren worden niet geplaatst bij het raam, want het invallende licht in de ogen heeft een negatieve invloed op fixatie disparatie, waardoor het lastiger wordt om de ogen te laten samenwerken (Bezem & Lugthart, 2016). Het onderzoek van Rivero-Contreras e.a. (2021) wijst uit dat teksten aanpassen, zoals lexicaal vereenvoudigen en samenvatten, een significant effect heeft op de leessnelheid en woordherkenning van een kind met visuele dysfunctie.

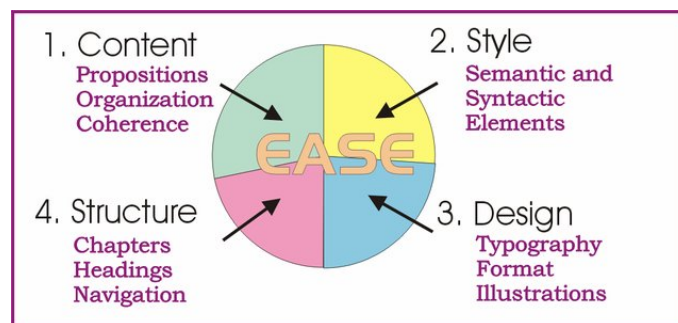
### Letter- en klankverwarring

Wanneer letterverwarring en letterklankproblemen samenkomen, kan dit leiden tot diverse leesfouten zoals klank- en letterverwisselingen, weglaten van letters en moeite met auditieve analyse en synthese. Om deze problemen effectief aan te pakken, is het van groot belang om ze snel te constateren en gebruik te maken van

methoden voor technisch lezen. Er zijn verschillende digitale en tekstuele tools beschikbaar om leesproblemen aan te pakken. Enkele voorbeelden zijn educatieve apps en software, e-books en luisterboeken, woordspelletjes en puzzels, en tekst-naar-spraak technologie. Hoewel deze hulpmiddelen nuttig kunnen zijn, is het belangrijk om te benadrukken dat ze niet de plaats innemen van persoonlijke begeleiding en ondersteuning door een professional.

### 2.1.5. Bewezen effecten van tekstvereenvoudiging -en aanpassing bij scholieren met dyslexie

Dyslexie kan zich op verschillende manieren uiten bij elke leeftijdsgroep. Een ondersteunende toepassing moet worden ontworpen met een individuele analyse van de specifieke behoeften en uitdagingen van elke leerling in gedachten (Gooding, 2022). Instructies moeten op een begrijpelijke en geïndividualiseerde manier worden gepresenteerd om de leerlingen te helpen bij het begrijpen en toepassen van de informatie. Het is belangrijk om te erkennen dat dyslexie zich bij verschillende kinderen op verschillende manieren kan uiten. Een bijkomende stoornis heeft bijvoorbeeld geen impact op de spellingprestaties van een kind. Het is daarom belangrijk om een toepassing te ontwerpen met de diversiteit van dyslexie in het achterhoofd.



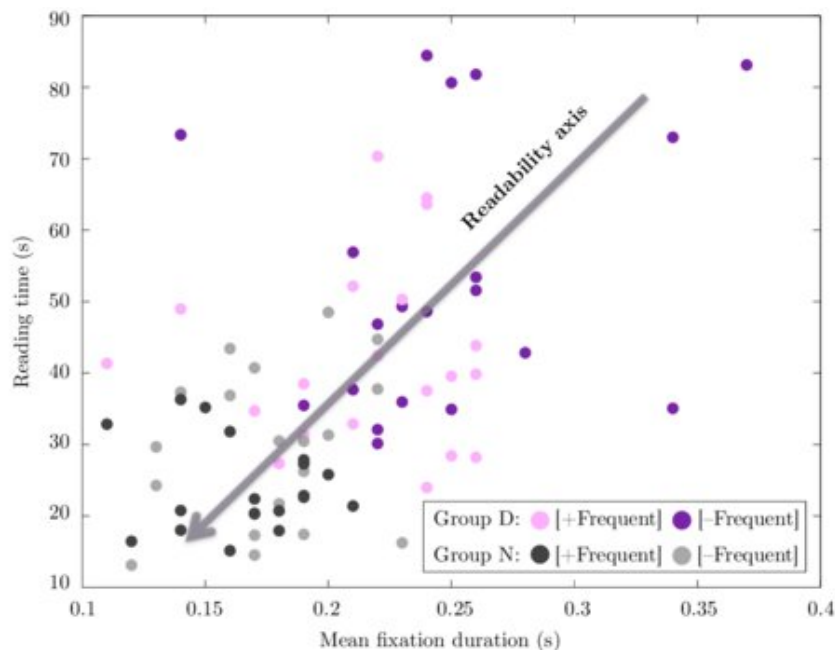
**Figuur (2.2)**

Afbeelding van DuBay (2004)

### Effecten op wijzigingen op lexicaal niveau

Het experiment van Rello, Baeza-Yates, Dempere-Marco e.a. (2013) wijst uit dat de decodeertijd of ontcijfertijd vermindert significant bij mensen met dyslexie wanneer de tekst een frequent woordgebruik toepast. De decodeertijden bij deelnemers met dyslexie waren significant korter bij het gebruik van frequente woordenschat, terwijl de decodeertijd bij minder frequente woordenschat hoger lag. Het verschil bij deelnemers zonder dyslexie was amper merkbaar.

Bevraagden met dyslexie maken minder leesfouten bij een tekst met een gereduceerde lexicale complexiteit volgens Gala en Ziegler (2016). De leessnelheid bij de kinderen lag hoger zonder een invloed op de begrijpelijkheid van de tekst. Het experiment benadrukt dat de bevraagden moeite hadden bij het lezen van woor-



**Figuur (2.3)**

Afbeelding van Rello, Baeza-Yates, Dempere-Marco e.a. (2013). Volgens de richting van de pijl wordt de ideale situatie benaderd, gekenmerkt door doelwaarden. Deze waarden worden bereikt door mensen zonder dyslexie onder optimale omstandigheden. Het gebruik van vaak voorkomende woorden vermindert de decodeertijd en verbetert de leesbaarheid voor mensen met dyslexie.

den met meer dan zeven karakters. Daarnaast zorgden onregelmatige en infrequente lettergreepcombinaties voor moeilijkheden bij de bevroagde kinderen met dyslexie.

### Effecten bij grammatische en syntactische wijzigingen

Onderzoek rond de effecten op syntactische vereenvoudiging bij kinderen en scholieren met dyslexie zijn in schaarse hoeveelheid. Het aanpassen causale structuren bij kinderen en jongeren met een lage leesgraad had een significant effect op het leestempo en de foutenmarge van de bevroagden uit het experiment van Lindholm e.a. (2000). Bij de revisies werden coherentieonderbrekingen werden hersteld door extra uitleg te voorzien, alsook door tekstgebeurtenissen in een temporele of tijdsafhankelijke volgorde te plaatsen. Zowel vaardige als minder vaardige lezers hadden baat bij de revisies. Verbale parafrasering heeft geen significant effect op lezers met dyslexie volgens Rello, Baeza-Yates en Saggion (2013). De bevroagden waren 13 tot en met 37 jaar oud met een gemiddelde leeftijd van 21 jaar. Het tekstformaat bleef ongewijzigd, maar lettertypes werden wel aangepast.

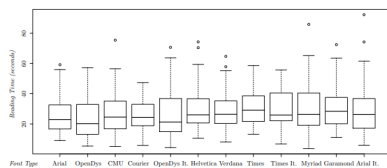
### Effecten bij wijzigingen op formaat en structuur

Geassisteerd samenvatten bevoordeelt de leesbaarheid van een scholier met dyslexie volgens het experiment van Nandhini en Balasundaram (2013). De geassisteerde samenvatting is gebaseerd op onaangepaste zinnen afkomstig uit de oor-

spronkelijke tekst. Het ontwerp bij dit experiment haalt de belangrijkste zinnen on-aangepast uit de oorspronkelijke tekst, herorganiseerd deze volgens de structuur van de oorspronkelijke tekst en presenteert deze aan de lezer. Al werd de logische structuur van de gepresenteerde zinnen in vraag gesteld, de leesbaarheid van de bevroagden was significant beter dan bij de oorspronkelijke tekst zonder een nadelig effect op de verstaanbaarheid van de bevroagden.

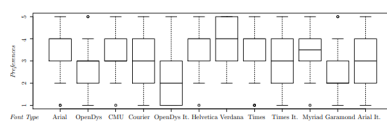
### Tekstweergave voor scholieren met dyslexie

Onderzoeken wijzen uit hoe ontwikkelaars rekening kunnen houden met de visuele weergave van een applicatie. De leessnelheid en decodeertijd van scholieren met dyslexie vergroot wanneer het lettertype groter is dan 14pt bij digitale en uitgeprinte documenten. De decodeertijd bij 18 tot 26pt was significant sneller dan teksten met 14pt bij bevroagden met en zonder dyslexie. Daarnaast wordt een sans-serif, monospaced of roman lettertype zoals Arial, Helvetica of Verdana aangeraden. Lettertypes zoals OpenDys die ontworpen zijn voor scholieren met dyslexie hebben een neutraal effect op lezers met of zonder dyslexie. Cursieve lettertypes hebben een nadelig effect op het leestempo van zowel gewone scholieren als scholieren met dyslexie en wordt afgeraden (Rello & Baeza-Yates, 2013; Rello & Baeza-Yates, 2015).



**Figuur (2.4)**

Afbeelding uit Rello en Baeza-Yates (2013). Verticaal wordt de gemiddelde mening van de bevroagden weergegeven. Horizontaal worden de lettertypes gerangschikt op gemiddelde leestijd van alle bevroagden. Dit onderzoek wijst uit dat Arial, CMU, Helvetica en Times de populaire keuzes zijn. Arial en CMU behoren hierbij tot de drie best scorende lettertypes rond gemiddeld leestempo.



**Figuur (2.5)**

Afbeelding uit Rello en Baeza-Yates (2013).

Op basis van experimenten op oogfixaties gaven Rello en Baeza-Yates (2015) de voor- en nadelen aan van verschillende visuele parameters. De bevroagden met dyslexie hadden bij de oorspronkelijke weergave een langere decodeertijd en tragere leessnelheid vergeleken met bevroagden zonder dyslexie. De bevroagden met dyslexie waren meer gevoelig aan de veranderingen van tekstweergave bij lettertype, character spacing, tekst en achtergrondkleur en tekst grijsschaal. Er zijn

geen duidelijke aanbevelingen voor een grijze achtergrond of het gebruik van grijs-schalen bij de bevroagden met dyslexie, maar licht grijs wordt voorgesteld zonder verschil in objectieve leesbaarheid. De bevroagden in het experiment van Rello en Baeza-Yates (2015) kozen voor een kleurencombinatie van een zwart lettertype over een gele achtergrondkleur.

De achtergrondkleur aanpassen naar zachtgeel, -groen of lichtblauw heeft een verbeterd effect op scholieren met dyslexie in het lager en middelbaar onderwijs (Bezemer & Lugthart, 2016; Rello & Bigham, 2017). Anthony (2020) haalt vijf ontwerpmethoden aan die voordelig zijn voor mensen met dyslexie. Een minimalistisch ontwerp met duidelijke pictogrammen en statische inhoud helpen mensen met dyslexie bij het concentreren op de belangrijke informatie. Toegankelijke kleurschema's zijn noodzakelijk voor een aangename leeservaring. De eindgebruiker moet in staat zijn om zelf een kleurschema te kiezen.

Het bevoorordelend effect van aanschouwelijkheid via afbeeldingen of schema's wordt benadrukt in het onderzoek van Rello, Saggion e.a. (2012). De visualisaties maken de tekstinhoud toegankelijk en hebben een significant effect op de leesbaarheid van een tekst. Bij de bevroagden was de fixatietijd lager en de leessnelheid sneller wanneer een tekst visuele schema's gebruikt. De impact op de begrijpelijkheid van een tekst verandert weinig tot niet.

### **2.1.6. Samenvattend schema over aanpakken**

### **2.1.7. Conclusie**

Scholieren met dyslexie in de derde graad van het middelbaar onderwijs kunnen moeite hebben bij het lezen van teksten. De ondersteuning kan op zowel syntactisch, lexicaal als visueel niveau gebeuren. Frequent woordgebruik, woordenschat korter dan zeven karakters, lettergrootte tussen 18 en 26 en een gepast lettertype -en stijl hebben een positief effect op deze doelgroep bij het lezen van teksten. Verbale parafrasering en het gebruik van grijstinten hebben een neutraal effect. Woordenschat langer dan zeven karakters, kleine lettergroottes en een cursief lettertype hebben een nadelig effect bij het lezen van teksten bij scholieren met dyslexie in de derde graad van het middelbaar onderwijs. Een adaptieve en gepersonaliseerde aanpak voor deze doelgroep is nodig.

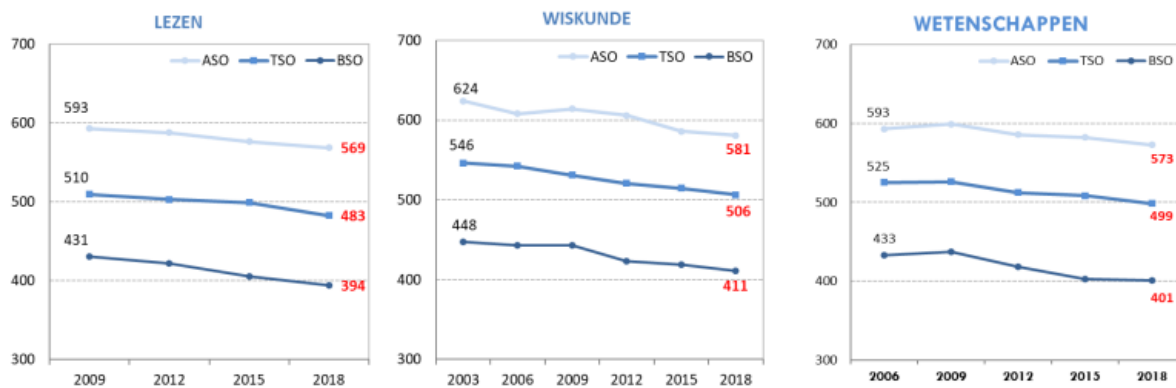
## **2.2. Wetenschappelijke artikelen**

Wetenschappelijke artikelen zijn een communicatiemiddel voor en door onderzoekers en wetenschappers. Deze artikelen worden echter ingezet als leermiddel voor jongeren in de laatste graad van het middelbaar onderwijs, alsook het hoger onderwijs en verdere studies. Wetenschappelijke artikelen kennen een uniform formaat qua structuur, genaamd IMRAD. De gebruikte syntax en lexicale woordenschat verschilt echter. Deze sectie verduidelijkt de verschillende uitdagingen dat het lezen van een wetenschappelijk artikel met zich meebrengt, alsook hoe docenten dit kun-

nen aanpakken in de derde graad van het middelbaar onderwijs. In deze sectie wordt de volgende onderzoeksvraag beantwoordt: "Wat zijn de specifieke kenmerken van wetenschappelijke artikelen?"

### 2.2.1. Wetenschappelijke geletterdheid in Vlaanderen

De *Programme for International Student Assessment* of PISA-test<sup>1</sup> van OESO is een driejaarlijkse test bij vijftienjarigen. Deze test bestudeert de wiskundige en wetenschappelijke geletterdheid<sup>2</sup> van 15-jarigen in geïndustrialiseerde landen, wat op ongeveer 79 landen komt. 4822 Vlaamse scholieren van vijftien jaar namen deel aan deze test. Dit onderzoek baseert op de cijfers van 2018, aangezien de testen van 2022 pas eind 2023 worden gepresenteerd. Deze testen houden echter geen rekening met leer- en leesstoornissen, waaronder dyslexie en dyscalculie. Het is echter nodig om deze cijfers mee te geven, om een idee te geven waar de doelgroep staat voor de start van de derde graad middelbaar onderwijs.



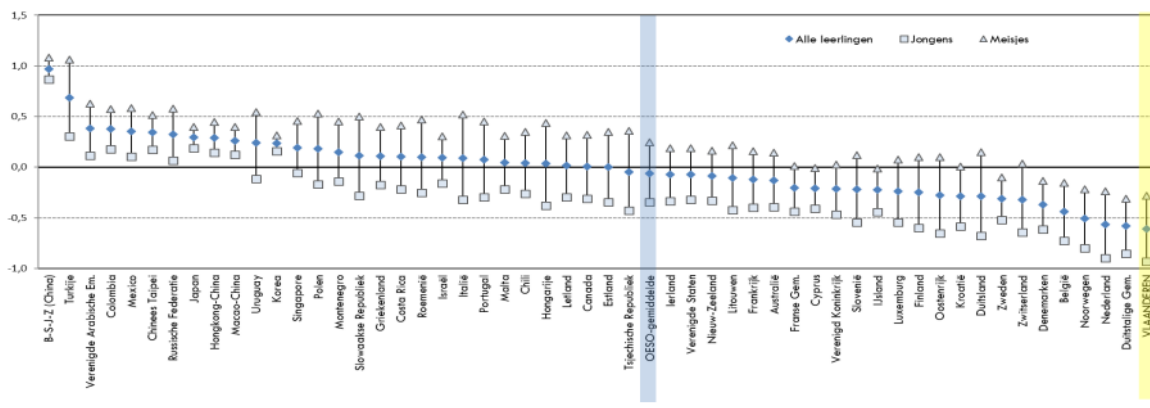
**Figuur (2.6)**

Figuur van De Meyer e.a. (2019). Op alle PISA-domeinen scoren de Vlaamse vijftienjarigen in ASO, BSO en TSO significant slechter dan de eerste metingen. De Meyer e.a. (2019) noemen dit een achteruitgang in alle onderwijsvormen.

<sup>1</sup><https://www.pisa.ugent.be/resultaten/pisa-2022>

<sup>2</sup>"Het beheersen van vaardigheden om als kritische burger om te gaan met wetenschappelijke onderwerpen en ideeën." volgens De Meyer e.a. (2019)





**Figuur (2.7)**

Figuur van De Meyer e.a. (2019). Het leesplezier van Vlaamse 15-jarigen. Zij uitten zich uiterst negatief op stellingen over leesplezier. Volgens de enquête vond de helft van de scholieren begrijpend lezen enkel tijdsverlies en slechts 17% gaf aan dat lezen één van hun favoriete hobby's is. Er is wel een significant verschil tussen de mening van jongens en meisjes, waar jongens negatiever antwoorden op lezen.

Land	Gem.	SE
B-S-J-Z (China)	590	2,7
Singapore	551	1,5
Macao (China)	544	1,5
Estland	530	1,9
Japan	529	2,6
Finland	522	2,5
Korea	519	2,8
Canada	518	2,2
Hongkong (China)	517	2,5
Chinees Taipei	516	2,9
Polen	511	2,6
<b>VLAANDEREN</b>	<b>510</b>	<b>3,3</b>
Nieuw-Zeeland	508	2,1
Slovenië	507	1,3
Verenigd Koninkrijk	505	2,6
Nederland	503	2,8
Duitsland	503	2,9
Australië	503	1,8
Verenigde Staten	502	3,3
Zweden	499	3,1
België	499	2,2
Tsjechische Republiek	497	2,5
Ierland	496	2,2
Zwitserland	495	3,0
Frankrijk	493	2,2
Denemarken	493	1,9
Portugal	492	2,8
Noorwegen	490	2,3
Oostenrijk	490	2,8
OESO-gemiddelde	489	0,4
Litland	487	1,8
Frans Gemeenschap	485	2,8
Duitstalige Gem.	483	7,4

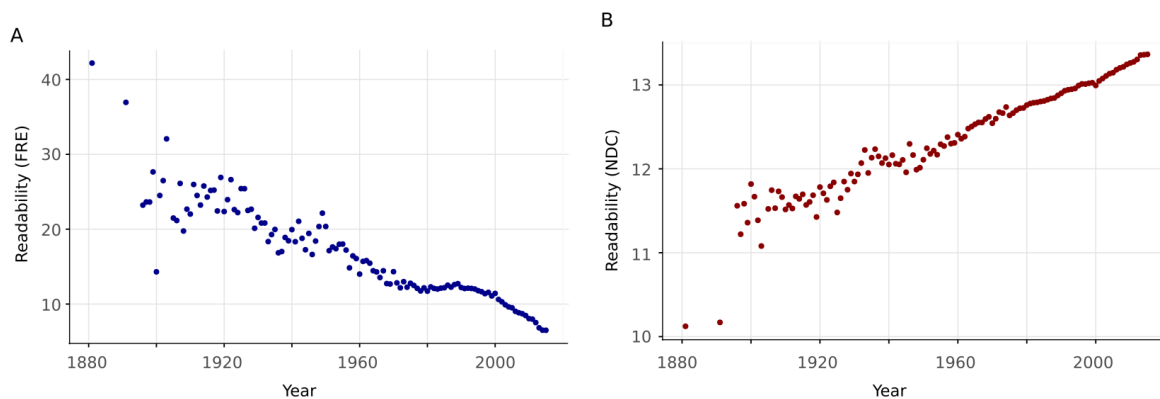
**Figuur (2.8)**

Figuur van De Meyer e.a. (2019). De wetenschappelijke geletterdheid bij vijftienjarigen op internationaal niveau. Vlaanderen scoort significant slechter dan acht deelnemende landen.

### 2.2.2. Trends rond wetenschappelijke artikelen

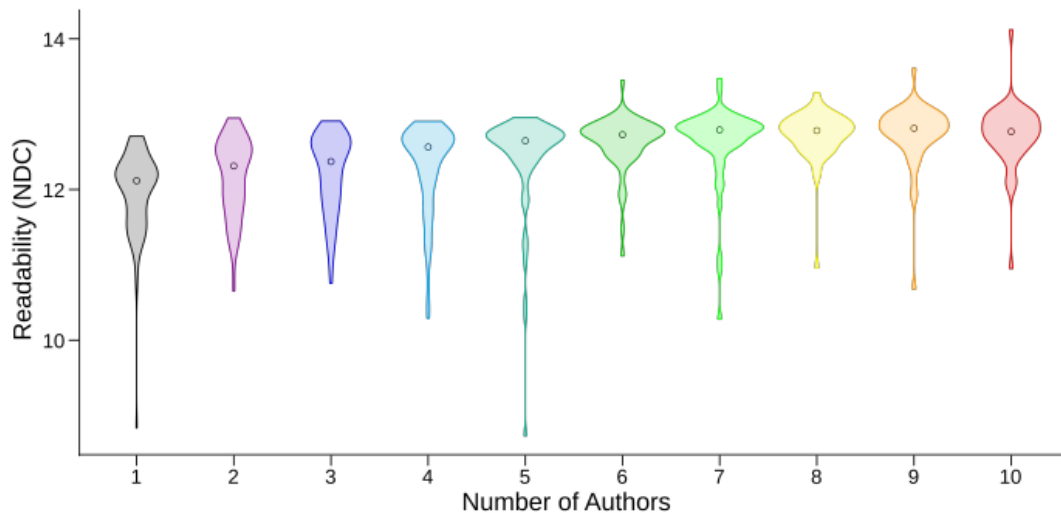
De leesgraad van wetenschappelijke teksten volgt al sinds de tweede helft van de twintigste eeuw een stijgende trend (Hayes, 1992). Meerdere onderzoeken in de voorbije tien jaar besluiten dat de complexe woordenschat en zinsbouw deze wetenschappelijke artikelen ontoegankelijk maakt voor doelgroepen naast onderzo-

kers (Ball, 2017; Jones e.a., 2019; Plavén-Sigraay e.a., 2017). Plavén-Sigraay e.a. (2017) onderzoekt de verschillende trends waarom wetenschappelijke artikelen alsmaar moeilijker te lezen worden. De relatie tussen de leesbaarheid van een abstract werd vergeleken met het jaar waarin het wetenschappelijk artikel werd gepubliceerd. De *Flesch-Reading-Ease* of FRE score werd gebruikt om de leesgraad van een wetenschappelijk artikel te beoordelen. Om te bevestigen dat de relatie tussen de complexiteit van een abstract overeenstemt met die van de volledige tekstinhoud, werden er vergelijkingen gemaakt met zes verschillende wetenschappelijke journalen. De overeenkomst tussen de leesgraad van het abstract en de overige tekstinhoud in een wetenschappelijk artikel werd eerder bevestigd door Dronberger en Kowitz (1975). Dat onderzoek benadrukte dat een abstract complexer werd geschreven, vergeleken met de rest van een wetenschappelijk artikel.



**Figuur (2.9)**

Afbeelding uit Plavén-Sigraay e.a. (2017). Links wordt de evolutie per FRE-score getoond. Hoe hoger de score, hoe hoger de gemiddelde complexiteit van een tekst. Rechts wordt de evolutie volgens de NDC-score getoond. Hoe hoger de score, hoe lager de gemiddelde complexiteit van een tekst. Het onderzoek schat dat nu een kwart van alle wetenschappelijke artikelen gebruik maken van Engels op het niveau van een masterstudent, ofwel een FRE onder nul.

**Figuur (2.10)**

Afbeelding uit Plavén-Sigraay e.a. (2017). Horizontaal worden het aantal auteurs per wetenschappelijk artikel aangeduidt. Verticaal wordt de gemiddelde NDC-score weergegeven. HOe hoger de NDC-score, hoe hoger de vereiste leesgraad om de tekst te kunnen lezen.

De hoge leesgraad van wetenschappelijke artikelen beperkt volgens Plavén-Sigraay e.a. (2017) twee aspecten: de toegankelijkheid en de herproduceerbaarheid.

### Toegankelijkheid

Bronnen worden minder toegankelijk tot het algemene publiek. Wetenschappelijke artikels worden enkel toegankelijk tot mensen die wetenschappelijk geletterd zijn of een leesgraad daarboven hebben. F en Ennals (2010) zegt dat wetenschap ons de nauwkeurige kennis moet geven, omdat mensen zich zorgen maken dat moderne samenlevingen minder streng worden met feitelijke waarheden en deze vervangen door *post-facts* die waar lijken te klinken. Wetenschappelijke inhoud moet volgens hem zo toegankelijk mogelijk worden gemaakt, zodat een zo breed mogelijk publiek de kern begrijpt.

### Reproduceerbaarheid

Onbegrijpelijke en ontoegankelijke zinsstructuren hinderen ook vakexperten. Het herschrijven van abstracten vergroot de begrijpbaarheid bij academici volgens Hartley (1999) en Snow (2010). De wetenschap bouwt voort op betrouwbare ontdekkingen en het reproduceren van experimenten is een belangrijke manier voor wetenschappers om vertrouwen te krijgen in hun besluiten. De inhoud van het wetenschappelijke artikel moet gecontroleerd kunnen worden. Voor de reproduceerbaarheid van onderzoeken is het volgens McNutt (2014) belangrijk dat de methodologie en resultaten begrijpelijk zijn. Een lage leesgraad en duidelijke zinsbouw beperkt het aantal misopvattingen en verwarringen bij onderzoekers. Experimenten uit Hubbard en Dunbar (2017) wijzen erop dat de bevraagde onderzoekers zo-

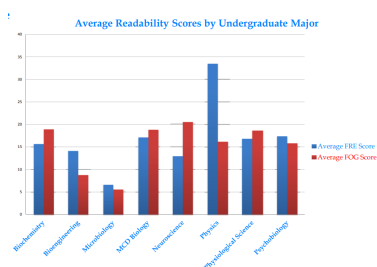
wel de methodologie als de resultaten de twee componenten vonden die een hoge leesgraad vergden.

### 2.2.3. Woordenschat en vakjargon

Complexe processen, methoden en ideeën worden in wetenschappelijke artikelen verwoord met gebruik van grammatische embeddings, doordachte en abstracte woordenschat en naamwoordstijlen. De kenmerken van academische taal variëren afhankelijk van de discipline, het onderwerp en de vorm, maar er zijn gemeenschappelijke kenmerken die wetenschappelijke taal onderscheiden van taal van een lagere leesgraad. (F & Ennals, 2010; Snow, 2010)

Wetenschappelijke artikelen dienen volgens Plavén-Sigraay e.a. (2017) in eerste instantie als uitwisseling van kennis tussen vakexperten. Daarnaast moet er rekening worden gehouden met de lengte wat een nadelig effect heeft op de beschikbare uitleg voor deze terminologie.

Snow (2010) beklemtoont dat deze zaken in het onderwijs moeten betrokken worden. STEM-vakken of vakken waar deze wetenschappelijke artikelen aan bod komen, moeten stil staan bij voldoende uitleg over de toegepaste grammatica en woordenschat voorzien tijdens de lessen.



**Figuur (2.11)**

Afbeelding van Murdos e.a. (2014) Volgens deze grafiek scoren de wetenschappelijke artikels rond fysica gemiddeld het best op de FRE-score. Al scoren de wetenschappelijke artikels rond microbiologie gemiddeld het zwakst op de FRE-score, ze scoren gemiddeld beter op de FOG-score.

### 2.2.4. Aanpak voor het lezen van wetenschappelijke artikelen

Als reactie op een satirisch artikel van Ruben (2016), bracht Pain (2016) het onderwerp bij wetenschappers aan het licht om zo verschillende tactieken te verzamelen om wetenschappelijke artikelen te begrijpen. Sommige wetenschappers zoeken direct onbekende woorden op of raadplegen extra informatiebronnen, terwijl andere wetenschappers hoofdstukken overslaan. Het is belangrijk om een balans te vinden tussen het begrijpen van de inhoud en het efficiënt gebruiken van de tijd. Sommige wetenschappers geven toe dat ze het soms opgeven als het te moeilijk wordt of als de literatuur net niet relevant is voor hun onderzoek. Pain2016 bouwt verder op deze adviezen en bouwt een stappenplan op hoe (startende) lezers wetenschappelijke artikelen kunnen aanpakken.

1. Lees de samenvatting en conclusie om een idee te krijgen van het doel en de uitkomst van het onderzoek.
2. De figuren en tabellen in het artikel zijn cruciaal omdat deze een snelle en duidelijke weergave geven van de belangrijkste bevindingen.
3. Focus op de nodige informatie en ga vervolgens terug om de technische details te begrijpen.
4. Let op de beperkingen en interpretatie van de resultaten. Controleer of de onderzoeksvraag en -methode adequaat zijn.
5. Controleer of de referenties relevant zijn en zoek naar andere artikelen over hetzelfde onderwerp.
6. Overweeg welke stukken prikkelend, nieuw en relevant zijn voor eigen onderzoeksvragen en hypothesen.
7. Maak aantekeningen en schrijf tijdens het lezen, zodat de lezer actief betrokken is bij het lezen van het artikel.

Wetenschappelijke artikelen vereisen een selectieve leesstijl volgens de bevraagde onderzoekers in Hubbard en Dunbar (2017). Bepaalde delen van een artikel worden geprioriteerd, zoals de abstract. De abstract en de discussie bepaalt of het artikel de moeite waard is voor de onderzoeker om te lezen. Sommige bevraagden adviseren om de methodologie te negeren en direct over te gaan naar de discussie of resultaten, terwijl andere onderzoekers aanbevelen om eerst de hypothesen van een artikel te achterhalen. Een artikel wordt nadrukkelijk meermaals gelezen, waarbij de lezer steeds in meer detail leest. Kritisch lezen is belangrijk, waarbij de conclusies worden beoordeeld en de data voor zichzelf spreekt. Er is geen standaardaanpak volgens Hubbard en Dunbar (2017), maar de bevraagde onderzoekers bevelen tactieken aan zoals selectief, kritisch en met een specifiek doel voor ogen lezen.

### **2.2.5. Conclusie**

Het lezen van wetenschappelijke artikelen kan overweldigend zijn, vooral bij onbekende vakgebieden, lange artikelen en technisch vakjargon. Nieuwe versies van een wetenschappelijk artikel moeten meer doelgroepen toelaten om over voldoende achtergrondinformatie te beschikken. De gebruikte syntax, woordenschat en compact formaat sluiten aan bij de mogelijke struikelblokken voor een scholier met dyslexie in de derde graad van het middelbaar onderwijs.

## **2.3. Tekstvereenvoudiging**

Vereenvoudigde teksten worden geschreven om leerlingen te ondersteunen bij het begrijpen van specifieke taalkenmerken, het beperken van de hoeveelheid nieuwe

woordenschat en het beheersen van de complexiteit van de tekst. Deskundigen zijn van mening dat vereenvoudigde teksten nuttig zijn voor startende en gevorderde lezers (Louwerse e.a., 2007). Samenvattingen van teksten bieden een oplossing om een snel zicht te krijgen over (lange) documenten, of om de kerninhoud van een tekst die al gelezen is opnieuw te prikkelen (McCombes, 2022). Vereenvoudigen kan handmatig door de docent gebeuren, maar recente technologische ontwikkelingen laten de automatisatie van dit proces toe met een gelijkwaardig eindresultaat. Deze sectie beantwoordt de volgende onderzoeksvraag: "Welke aanpakken zijn er voor geautomatiseerde tekstvereenvoudiging?". Aansluitend hierop wordt de volgende subvraag beantwoordt: "Hoe worden teksten handmatig vereenvoudigd voor scholieren met dyslexie?"

### 2.3.1. Manuele tekstvereenvoudiging

Wetenschappelijke artikelen moeten informatie begrijpelijk weergeven voor een breed publiek, waaronder de scholieren die deze artikelen voorgeschoteld krijgen. Teksten vereenvoudigen heeft volgens Crossley e.a. (2012) drie algemene doelen, namelijk het illustreren van een specifiek taalkenmerk, ongekende woordenschat voor een doelgroep aan te passen en de hoeveelheid gegeven informatie onder controle te houden. Crossley e.a. (2012) wijst op twee soorten van handmatige tekstvereenvoudiging. Intuïtieve tekstvereenvoudiging is een methode waar de auteur die de transformatie uitvoert, wordt beïnvloed door persoonlijke vermoedens over wat een tekst beter leesbaar maakt. Structurele vereenvoudiging daarentegen vervangt vermoedens door het gebruik van woordenlijsten en leesbaarheidsformules zoals Flesch Reading Ease (FRE), Gunning Fog (FOG), SMOG-Cro (SMOG) en de Coleman-Liau Index (CLI).

#### Lengte en formaat

De leesgraad van de woordenschat moet overeenstemmen met de syntactische leesgraad. vereenvoudigd. Vervolgens kan tekst naar een ander formaat worden omgezet, zoals *post-it*notes, *postcards* of emails, om het begrijpelijker te maken. Dit wordt vooral ingezet in het lager onderwijs. De schrijf- en vertelstijl moet consistent blijven in het nieuwe formaat. Ten slotte moeten verwijswwoorden worden aangepast om de tekst toegankelijker te maken voor meertalige lezers. Bijvoorbeeld door eenvoudige verwijswwoorden zoals 'zij' of namen te gebruiken (Rijkhoff, 2022).

Een samenvatting verkort de lengte van een tekst. Kernzinnen en trefwoorden worden eerst in een tekst gemarkeerd en vervolgens op een nieuw blad geschreven. De kernzinnen worden achterhaald door woord- en zoektermfrequentie en anderszijds door het stellen van algemene vragen over het artikel. Trefwoorden achterhalen gebeurt gelijkaardig en deze zijn regelmatig af te leiden uit de inhoudstafel en titels. Voor deze twee methoden moet de persoon die een samenvatting maakt al vooraf de tekst meermaals gelezen hebben. Een alternatief op markeren is het

parafraseren van de tekst. De geparafraseerde tekst blijft semantisch gelijk, maar het neemt een andere syntax, structuur en woordenschat aan (Rijkhoff, 2022).

Volgens Hollenkamp (2020) en McCombes (2022) moet de samenvatting van een wetenschappelijk artikel altijd de volgende drie vragen kunnen beantwoorden:

- Waarom werd het onderzoek verricht? Welke achtergrondinformatie en context nam de onderzoeker in acht. Daarnaast moeten de geformuleerde hypothesen aan bod komen.
- Wat werd er geëxperimenteerd? Alle gebruikte methoden en resultaten moeten in een samenvatting terug te vinden zijn en enkel de noodzakelijke kwalitatieve waarden mogen aan bod komen.
- Welke conclusies trekken de onderzoeker(s) uit het onderzoek? De implicaties en beperkingen tijdens het onderzoek, alsook de aanradingen moeten in de samenvattingen aan bod komen.

### Woordenschat

Moeilijke woorden kunnen op twee manieren beperkt worden. Eerst moet een tekst op maat zijn van het doelpubliek. De gehanteerde (vak)termen moeten begrijpelijk zijn voor iedereen binnen het doelpubliek. Deze leesgraad en voorkennis wisselt sterk af bij scholieren in de laatste graad middelbaar onderwijs. Als het niet mogelijk is om eenvoudigere synoniemen te gebruiken, dan komt het van pas om de woorden uit te leggen. Dit hoeft enkel te gebeuren bij de eerste keer dat deze woorden voorkomen (Bosmans e.a., 2022b, 2022c). Case (2008) haalt aan om homoniemen en homofonen te vervangen.

### Grammatica

Mensen proberen bij het schrijven wel vaak om iets kort en bondig te schrijven, al schrijven veel mensen nog steeds volgens hun gedachten. Dit leidt tot een omslachtige schrijfstijl en dit heeft een nadelig effect op de grammatica in een tekst. Tangconstructies, lange zinsaanlopen en voorzetselketens zijn vaak voorkomende boosdoeners volgens Bosmans e.a. (2022a). Tangconstructies zijn zinsstructuren waarbij een bijwoordelijke bepaling of bijzin tussen het onderwerp en de persoonsvorm van een zin wordt geplaatst. Tangconstructies kunnen minder zwaar worden gemaakt door de twee grijpers van de tang dicht bij elkaar te brengen of door lange zinnen te splitsen. Aanvullende informatie kan beter kort worden gehouden en vooraan of achteraan in de zin worden geplaatst om onderbrekingen te voorkomen. Bosmans e.a. (2022a) en Rijnvis (2020) halen drie aanpakken aan om tangconstructies te vermijden:

- De bijzin naar het begin of het einde van een zin plaatsen.
- De zin splitsen in twee kortere zinnen.

- Het onderwerp en de persoonsvorm dichter bij elkaar plaatsen door minder informatie tussen hen in te plaatsen.

Een lange zinsaanloop kan worden verkort om de aandacht van de lezer niet op de proef te stellen. Voorzetselketens kunnen worden vermeden door informatie over verschillende zinnen te verdelen. Voorzetseluitdrukkingen worden best vervangen door gewone voorzetsels (Bosmans e.a., 2022a).

### Pedagogische en onderwijsgerelateerde kritieken

Al besluit het onderzoek van Crossley e.a. (2012) dat tekstvereenvoudiging een bevorderend effect heeft voor scholieren met afwisselende leesgraden, het onderzoek beklemtoont de unieke aanpakken per docent. Iedere docent heeft een eigen procedure om een vereenvoudigde tekst op maat te kunnen schrijven.

### 2.3.2. Natural Language Processing

Tekstvereenvoudiging is het proces waarin het technisch leesniveau en/of woordgebruik van een geschreven tekst wordt verminderd. Het resultaat van deze fase is een tekst die korter en aangenamer is, zonder het verlies van de kerninhoud. Binnen machinaal leren (ML) is tekstvereenvoudiging een zijtak van natuurlijke taalverwerking. (Siddharthan, 2006) Volgens (Siddharthan, 2014) bestaat een complete en geautomatiseerde tekstvereenvoudiging uit vier verschillende vereenvoudigingen. *Natural Language Processing* (NLP) of natuurlijke taalverwerking is een brede term die zich richt op het verwerken en analyseren van menselijke taal door computers (Eisenstein, 2019). NLP omvat verschillende technieken, zoals tekstanalyse, taalherkenning en -generatie, spraakherkenning en -synthese, en semantische analyse. Computers zijn in staat om op een menselijke manier te communiceren en begrijpen wat er wordt gezegd. De volgende begrippen worden aangehaald in Eisenstein (2019) en Sohom (2019) en zijn fundamenteel voor de concepten die volgen.

#### Tokenisation

Tokenisatie splitst de stam of basisvorm van woorden in een tekst. Gebruikelijk zetten ontwikkelaars deze stap in om een woordenschat voor een taalmodel op te bouwen. Bij tokenisatie wordt er geen rekening gehouden met de betekenis achter ieder woord. Tokeniseren kan volgens Menzli (2023) op vier manieren:

- Word-level tokenisation of WTL splitst de tekst op per woord.
- Character-level tokenisation of CLT splitst de tekst per karakter. 'Slimmer' wordt s-l-i-m-m-e-r. Deze vorm achterhaalt de semantiek van een tekst beter en laat de het. Nadelig hebben de karakters op zich weinig betekenis, alsook maakt deze vorm de inputlengte groter. (Ribeiro e.a., 2018)
- Subword-level tokenisation splitst de tekst op in stukken op basis van de woordfrequentie. Veelvoorkomende woorden worden hele woorden getokeniseerd,



terwijl zeldzamere woorden opgesplitst worden in kleinere stukken die kunnen worden gebruikt. De rest van de woorden in de relevante dataset te creëren. Dit biedt een voordeel ten opzichte van word-level tokenisation omdat het een balans biedt tussen WLT en CLT (Iredale, 2022).

- Sentence tokenization splitst de tekst op per zin. Fardeen- (2021) haalt aan dat de tokenizer ineffectief is tegen afkortingen, maar dit is afhankelijk volgens de gebruikte dataset.

Feature	spaCy	NLTK	Gensim
Word-level	X	X	X
Character-Level	X	X	
Subword-Level	X		X
Sentence-level	X	X	X

**Tabel 2.1:** Vergelijking van methodes voor tokenisatie met Spacy, nltk, en gensim.

### Lemmatiseren en parsen

Lemmatiseren in NLP bouwt verder op *stemming*, maar de betekenis van ieder woord wordt in acht genomen. Voor het lemmatiseren bestaan er Nederlandstalige modellen, waaronder JohnSnow<sup>3</sup>. Bij **omgekeerd lemmatiseren** wordt er een afgeleide achterhaald vanuit de stam. Bijvoorbeeld voor het werkwoord 'zijn' zou dit 'is', 'was' of 'ben' zijn. Voor zelfstandige naamwoorden, zoals 'hond', is dit dan enkelvoud of meervoud (Eisenstein, 2019).

Bij een **parsing**-fase wordt er een label aan ieder woord of zinsdeel toegekend. Voorbeelden van labels zijn zelfstandig naamwoord, bijwoord, werkwoord, bijzin of stopwoord. Het herkennen van zinsdelen wordt *chunking* genoemd. Parsing heeft een dubbelzinnigheidsprobleem, want een 'plant' staat niet gelijk aan de vervoeging van werkwoord 'planten' (Eisenstein, 2019).

### Sequence labeling en part-of-speech tagging

Een machine moet de betekenis achter ieder token kunnen vatten. Hier komt *sequence labeling* aan de pas volgens Eisenstein (2019). Elk woord in een tekst wordt gekoppeld aan een *Part-of-Speech* (PoS) of *Named-Entity-Recognition* (NER) label. Deze NLP-fase achterhaalt de structuur van een tekst. PoS-tagging richt zich op grammaticale categorieën van woorden, terwijl NER-labeling instaat voor het herkennen van specifieke entiteiten in een tekst.

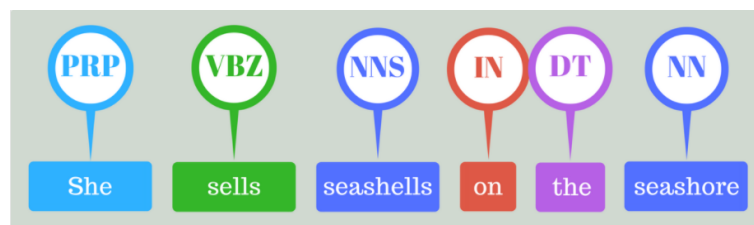
Bij PoS-tagging worden de woorden in een zin geanalyseerd. Elk woord wordt gekoppeld aan een grammaticale categorie, zoals een zelfstandig naamwoord, werkwoord, bijvoeglijk naamwoord of bijwoord. *PoS-tagging* helpt bij het achterhalen

<sup>3</sup>[https://nlp.johnsnowlabs.com/2020/05/03/lemma\\_nl.html](https://nlp.johnsnowlabs.com/2020/05/03/lemma_nl.html)

van de syntactische structuur van een zin. Deze taak komt van pas bij parsing en machinevertaling. *PoS-tagging* wordt aanschouwelijk gemaakt op 2.12. Namen van personen, organisaties en locaties worden herkend en geclassificeerd met NER-labeling. Met NER-labeling wordt volgens Jurafsky e.a. (2014) specifieke informatie uit tekst gehaald, zoals het identificeren van de namen van personen, plaatsen of bedrijven die in nieuwsartikelen worden genoemd, of het extraheren van belangrijke data of getallen uit financiële rapporten. Dit wordt aanschouwelijk gemaakt 2.13. J. Li e.a. (2018) benoemt vier vormen voor NER-labeling:

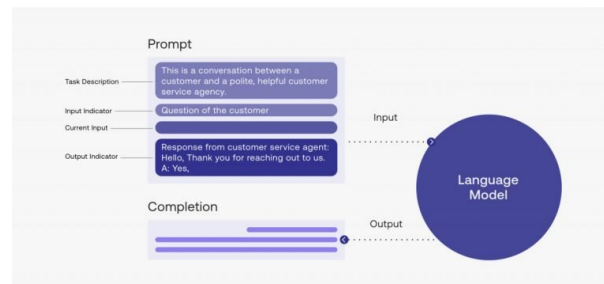
- *Dictionary-based NER labeling* gebruikt vooraf gedefinieerde woordenboeken die de namen van de entiteiten bevatten. Het vergelijkt de woorden in de tekst met de woordenboeken en labelt ze als ze overeenkomen.
- *Rule-based NER labeling* gebruikt vooraf gedefinieerde regels die gebaseerd zijn op syntactische of semantische patronen om de entiteiten te identificeren. Het past de regels toe op de tekst en labelt de woorden die aan de regels voldoen.
- *Machine learning-based NER labeling* gebruikt statistische modellen zoals Hidden Markov Model (HMM) of Conditional Random Field (CRF) om te leren van gelabelde trainingsgegevens hoe ze entiteiten kunnen herkennen. Het gebruikt kenmerken zoals het woord zelf, omliggende PoS-labels en het hoofdlettergebruik om te beslissen welk label aan elk woord moet worden toegekend.
- *Deep learning-based NER labeling* gebruikt neurale netwerken zoals recurrent neural network (RNN) of convolutional neural network (CNN) om te leren van ongelabelde of gedeeltelijk gelabelde trainingsgegevens hoe ze entiteiten kunnen herkennen. Het gebruikt woordvectoren en niet-lineaire representaties om complexe relaties tussen woorden te modelleren.

Poel e.a. (2008) onderzocht *PoS-tagging* met een neuraal netwerk voor Nederlandstalige teksten. Het model behaalde een nauwkeurigheid van 97,88% voor bekende woorden en 41,67% voor onbekende woorden. Het model gebruikte de Corpus Gesproken Nederlands (CGN) als trainingsdata.

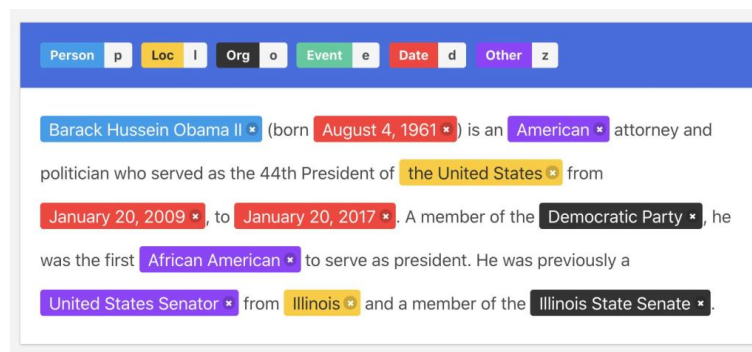


**Figuur (2.12)**

Voorbeeld van PoS-labeling op de Engelstalige zin "She sells seashells on the seashore". Afbeelding van Bilici (2021)

**Figuur (2.14)**

Afbeelding uit McFarland (2023). Een illustratie over de werking en begeleiding van prompt engineering bij een taalmodel.

**Figuur (2.13)**

Voorbeeld van sequence labeling op de Engelstalige zin "She sells seashells on the seashore". Afbeelding van Bilici (2021)

### Prompt engineering

Large Language Models of LLM's zoals GPT-3, BERT en T5 genereren tekst en karakters op basis van de probabiliteit of waarschijnlijke uitkomst van een gegeven input. Deze modellen maken gebruik van een neurale netwerk om patronen in de input te herkennen en deze patronen te gebruiken om voorspellingen te doen over de uitvoer (Liu e.a., 2020). Iedereen kan volgens McFarland (2023) een input of prompt schrijven. Deze tools zoals chatbots zijn ontworpen om zo intuïtief mogelijk te zijn voor een algemeen doelpubliek. Prompt engineering is een steeds belangrijkere vaardigheid die nodig is om effectief te communiceren met LLM's, zoals ChatGPT (Harwell, 2023).

Deze prompts werken volgens Liu e.a. (2020) op dezelfde manier als bij mensen en kunnen worden gebruikt om werk te produceren dat is aangepast aan het doel. Text is momenteel het belangrijkste middel van communicatie tussen mens en AI. Een concrete en geoptimaliseerde prompt omvat een concrete scope, duidelijke vraagstelling, specifieke sleutelwoorden, de context en ten slotte gepersonaliseerde keuzes (McFarland, 2023). Bij een zoekopdracht moeten voldoende parameters in de prompt worden opgenomen, zoals het type. Zo niet zal het model te

algemeen blijven en mogelijks afwijken van de intentie van de gebruiker. Effectieve AI prompt engineering leidt tot hoogwaardige trainingsgegevens die het AI-model in staat stellen om nauwkeurige voorspellingen en beslissingen te maken (Liu e.a., 2020).

Prompt patterns is samen met prompt engineering naar boven gekomen en is vergelijkbaar met software patterns. Deze patronen zijn herbruikbare oplossingen voor veelvoorkomende problemen in een bepaalde context, waaronder vooral de interactie bij het werken met LLM's. White e.a. (2023) haalt vijf verschillende prompt patterns aan.

- Intent-prompts waarbij een LLM een instructie krijgt met een specifiek verwacht antwoord.
- Restriction-prompts die het antwoord van een LLM inperkt. Deze pattern is noodzakelijk om een LLM binnen de lijnen te houden.
- Contextualization-prompts verzekeren dat de output van een LLM relevant is. Een context wordt aan de LLM meegegeven.
- Expansion/reduction-prompts genereren een output dat beknopt is, maar met voldoende details.

### Traditional en contextual word embeddings

NLP-systemen en machines moeten woorden, grammatica en nuancering kunnen begrijpen. Embeddings transformeren woorden tot een numerieke representatie, waarop een machine deze representaties kan aanleren om nadien tekst te verwerken. Traditionele word embeddings bouwen een woordenschat op met unieke woorden. De betekenis achter ieder woord wordt niet opgevolgd. Voorbeelden van traditionele word embeddings zijn Word2Vec en Glove.

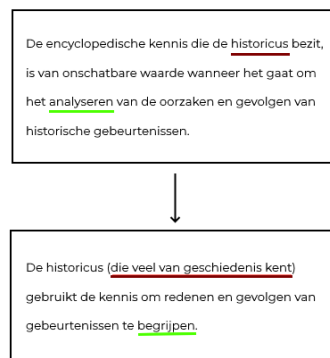
Contextual word embeddings lossen dit probleem op en houden rekening met de context waarin een woord wordt gebruikt. ELMo en BERT zijn voorbeelden van een model voor contextuele embedding. Deze vorm houdt de semantiek bij van een woord in een bepaalde context en is noodzakelijk wanneer een machine polysemantische woorden in een tekst moet begrijpen. Contextuele word embeddings worden verkregen uit transformer-gebaseerde modellen. Ze worden verkregen door een volledige zin door te geven aan een pre-trained model.

BERT is een meertalig LLM, getraind op 110 miljoen parameters uit 104 verschillende talen<sup>4</sup>, waaronder Nederlands. Dit taalmodel kent alternatieven die verderbouwen op het oorspronkelijke BERT-model. Voor de Nederlandse taal zijn er twee, namelijk RobBERT en BERTje. Volgens (..) is RobBERT de krachtigste van de twee modellen, waar BERTje compacter is. Vervolgens bepaalt de *Substitution Ranking* of SR-stap welke vervanging de beste is uit een set van kandidaten. SR gebeurt door gegenereerde substituties op basis van relevantie te rangschikken.

<sup>4</sup><https://github.com/google-research/bert/blob/master/multilingual.mdlist-of-languages>

## 2.4. De verschillende soorten tekstvereenvoudiging

Tekstvereenvoudiging bestaat volgens Siddharthan (2014) uit vier soorten transformaties: lexicale, syntactische en semantische vereenvoudiging en samenvatten.



**Figuur (2.15)**

Voorbeeld van tekstvereenvoudiging. Oorspronkelijke tekst uit Historia 5 bron toe te voegen

### 2.4.1. Lexicale vereenvoudiging

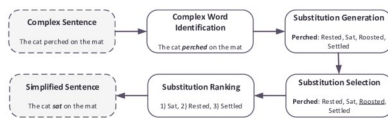
Bij *lexical simplification* (LS) of lexicale vereenvoudiging worden complexe woorden vervangen door eenvoudigere synoniemen. Bijvoorbeeld, het woord 'adhesief' wordt vervangen door 'klevend'. Kandula e.a. (2010) haalt twee manieren aan om lexicale vereenvoudiging mogelijk te maken, namelijk het vervangen door een synoniem en het aanmaken of genereren van extra uitleg. De zinsstructuur verandert niet en er is garantie dat de kerninhoud en benadrukking in een tekst identiek blijft. Het doel van lexicale vereenvoudiging is om de moeilijkheidsgraad van de woordenschat in een zin of tekst te verlagen.

#### Complex Word Identification

*Complex word identification* of CWI is een gesuperviseerde NLP-taak. In een pipeline voor lexicale tekstvereenvoudiging is CWI de eerste stap. Moeilijke woorden of *multi-word expressions* (MWE) in een tekst worden achterhaald (Gooding & Kochmar, 2019; Shardlow, 2013). Na CWI kan LS gebruikt worden om deze woorden te vervangen door eenvoudigere synoniemen of om verdere elaboratie te voorzien met behulp van voorbeelden of definities (Kandula e.a., 2010; Zeng e.a., 2005). CWI is volgens Shardlow (2013) een cruciale stap, want een lage *recall* van dit component zal een uitvoertekst geven waar moeilijke woorden niet worden vereenvoudigd. Het model zal moeilijke woorden laten staan.

#### Substitutiegeneratie en ranking

Substitutiegeneratie wordt gedaan door synoniemen te zoeken voor een doelwoord in lexicale databanken zoals WordNet, BERT, context2vec, nPIC of OOC.

**Figuur (2.16)**

Afbeelding van Althunayyan en Azmi (2021). Deze pipeline wordt in meerdere onderzoeken rond lexicale vereenvoudiging toegepast, zoals Bingel e.a. (2018), Bulté e.a. (2018) en Paetzold en Specia (2016)

### 2.4.2. Syntactische vereenvoudiging

Syntactische vereenvoudiging verlaagt de leesgraad en complexiteit van een zin door de grammatica en zinsstructuur van een tekst aan te passen. Twee afzonderlijke zinnen kunnen samengevoegd worden tot één eenvoudigere zin. Zo worden complexe of onduidelijke zinsconstructies verminderd, terwijl de inhoud en betekenis van de tekst behouden blijft. Dergelijke transformaties zijn het vereenvoudigen van de syntax of door de zinnen korter te maken. Zinnen worden toegankelijker, zonder de kerninhoud of relevante inhoud te verliezen.

Kandula e.a. (2010) ontwikkelde een toepassing om medische informatie bij beschikbare biomedische bronnen te vereenvoudigen. Dit model verlaagt de leesgraad door syntactische vereenvoudiging op zinniveau toe te passen. Zinnen met meer dan tien woorden worden in het onderzoek als complex beschouwd en worden vereenvoudigd door drie modules. Na deze transformatie kan de oorspronkelijke zin ongewijzigd worden behouden of vervangen worden door twee of meer kortere zinnen. De architectuur van dit model omvat drie onderdelen: een *Part of Speech (PoS) Tagger*, een *Grammar Simplifier* en een *Output Validator*.

- Voor de *PoS Tagger*-fase gebruikten Kandula e.a. (2010) beschikbare functies uit het open-source pakket OpenNLP<sup>5</sup>.
- De *Grammar Simplifier* module splitst de lange zin in twee of meer kortere zinnen door POS-patronen te identificeren en een set transformatieregels toe te passen.
- De *Output Validator* module controleert de output van de Grammar Simplifier op grammatica en leesbaarheid.

### 2.4.3. Conceptuele of semantische vereenvoudiging

Conceptuele tekstvereenvoudiging deelt complexe concepten op in eenvoudigere delen, past duidelijke en bondige taal toe en vermijdt technische jargon en abstracte uitdrukkingen. Er wordt meer uitleg of voorbeelden gegeven, of dat niet-relevante delen van de tekst worden weggelaten. Na deze transformatie is de tekst beter te begrijpen, zonder het verlies aan betekenis of nauwkeurigheid. Siddharthan

<sup>5</sup><https://opennlp.apache.org/>

(2014) noemt deze transformatie een vorm van elaboratie of het uiteenzetten van een begrip.

#### 2.4.4. Overige vormen van vereenvoudiging

Metaforen, *short language* of *slang* en idiomen kan de menselijke bedoeling achter een zin of paragraaf in de war brengen voor een machine. Pragmatische vereenvoudiging zet deze constructies om naar een letterlijke en duidelijke tekst (Javourey-Drevet e.a., 2022). Ten slotte is het mogelijk om het formaat van parafrases en alinea's aan te passen. Een opsomming of oplistening benadrukt belangrijke punten en maakt een duidelijke structuur van een mogelijks complexe tekst. Een opsomming beklemtoont dezelfde tekst vergeleken met een doorlopende tekst volgens Hale (2022) en Siddharthan (2014).

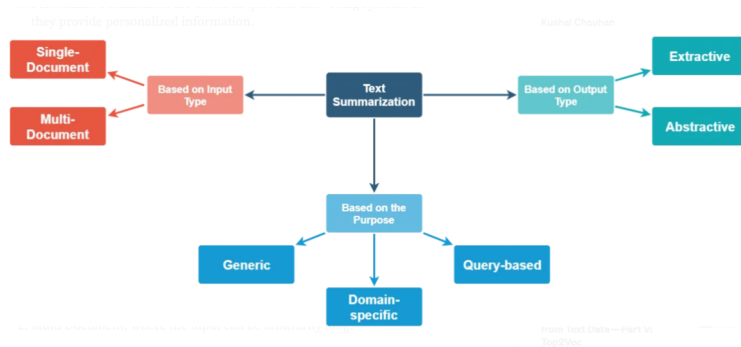
#### 2.4.5. Tekstvereenvoudiging automatiseren

Geautomatiseerde tekstvereenvoudiging is geen nieuw concept. Volgens onderzoeken van Canning e.a. (2000) en Siddharthan (2006) waren de eerste aanpakken op geautomatiseerde tekstvereenvoudiging gebouwd op rule-based modellen. Deze modellen bewerken de syntax door zinnen te splitsen, te verwijderen of de volgorde van de zinnen in een tekst aan te passen. Lexicale vereenvoudiging kwam hier niet aan de pas. Enkel bij recentere onderzoeken van Bulté e.a. (2018) en Coster en Kauchak (2011) werd het duidelijk hoe lexicale en syntactische vereenvoudiging gecombineerd kon worden.

### 2.5. Samenvatten

Teksten vereenvoudigen met lexicale, conceptuele en/of syntactische vereenvoudiging biedt geen garantie dat de tekstinhoud korter zal worden. Bij de drie soorten vereenvoudiging wordt er initieel enkel per zin gekeken. De vereenvoudiging houdt geen rekening met voorafgaande of opvolgende zinnen (DuBay, 2004). Teksten machinaal samenvatten is geen nieuw concept. Het onderzoek van Hahn en Mani (2000) onderzoekt hoe teksten automatisch samengevat kunnen worden. Dit onderzoek haalt onder meer twee aanpakken aan hoe een machine teksten kan samenvatten, namelijk extraherend en abstraherend. Daarnaast reikt Hahn en Mani (2000) drie manieren aan welke inhoud er zeker in de samengevatte versie moet op te merken zijn:

- Informatieve samenvattingen vervangen de oorspronkelijke tekst. Alles wat de lezer nodig heeft, dus hoofd- en bijzaken, zijn betrokken in de samengevatte tekst.
- Indicatieve samenvattingen behouden enkel een tekst met links die een lezer doorverwijzen naar andere bronnen.



**Figuur (2.17)**

Afbeelding van Chauhan (2018). De manier waarop teksten automatisch samengevat kunnen worden, is afhankelijk van drie verschillende domeinen.

- Kritische samenvattingen of *reviews* bestaan uit de kerninhoud van de oorspronkelijke tekst en een opiniestuk over die specifieke kerninhoud.

Verder haalt Hahn en Mani (2000) ook het onderscheid tussen een generieke en een gebruikersgerichte samenvatting. Een generieke samenvatting staat niet stil bij speciale noden of interesses van de eindgebruiker. Daarnaast houdt een gebruikersgerichte samenvatting wel rekening met sleutelwoorden of thema's in een tekst. Hahn en Mani (2000) haalt aan dat technologieën zoals *full-text-search* en gepersonaliseerde informatiefiltering het belang van gebruikersgerichte samenvatting naar voor duwen. Hahn en Mani (2000) omschrijft de architectuur van een samenvattingssysteem aan de hand van drie fases. Allereerst wordt de brontekst geanalyseerd. Daarna worden de *salient points* of kernpunten in een tekst aangeduid. Deze punten zijn zinnen of tokens. Ten slotte worden de punten samengevoegd tot één uitvoertekst. De nadruk is verschillend per samenvattingsmethode.

### 2.5.1. Extraherend samenvatten

Bij deze vorm worden de belangrijkste zinnen gemarkeerd en vervolgens opnieuw neergeschreven. Dit is het equivalent van handmatig zinnen markeren en vervolgens op een blanco papier neerschrijven. Het nadeel hiervan is dat de uitvoertekst niet samenhangend kan zijn na het samenvatten. Dit maakt de uitvoertekst minder aangenaam om te lezen. Verma en Verma (2020) onderzocht de verschillende manieren waarop een tekst extraherend kan worden samengevat. Zij halen drie grote componenten aan, namelijk:

- Graafgebaseerd extraherend samenvatten
- Maximal Marginal Relevance
- Meta-heuristic-gebaseerd



### Graafgebaseerd extraherend samenvatten

Graafgebaseerd extraherend samenvatten is een techniek die een document voorstelt als een graaf, waarbij de knopen zinnen en bogen de relatie tussen de zinnen representeren. Deze algoritmen achterhalen de kernzinnen in een tekst. Bijvoorbeeld kan het PageRank-algoritme, dat vaak wordt gebruikt voor het rangschikken van webpagina's in zoekmachines, worden gebruikt om de zinnen in de grafiek te rangschikken op basis van hun belangrijkheid.

Parveen en Strube (2015) raadt een graafgebaseerd systeem aan voor *unsupervised learning*. Belangrijke zinnen worden met een lokaal minimum bepaald, als ook wordt redundantie vermeden. Deze methode kan significante resultaten opleveren bij het ophalen van kernzinnen uit zowel lange wetenschappelijke artikelen als korte nieuwsartikelen. Daarnaast vermeldt Parveen en Strube (2015) dat het systeem beter presteert wanneer coherentie wordt opgenomen en gecombineerd wordt met positionele informatie.

Abdel-Salam en Rafea (2022) voerden een vergelijkend onderzoek uit rond SqueezeBERT en BERT. De compacte architectuur van SqueezeBERT kan ingezet worden voor real-time samenvatting. Dit is volgens Abdel-Salam en Rafea (2022) een interessant alternatief op het BERT-model. In vergelijking heeft de voorgestelde SqueezeBERT slechts ongeveer 62 miljoen parameters, terwijl het prestatieniveau nog steeds boven de 90% van het BERT-baseline model blijft. De evaluatie gebeurde aan de hand van de ROUGE-score. De onderzoekers besluiten dat SqueezeBERT een goed alternatief is, vooral door het trainen met bijna de helft van de grootte van het oorspronkelijke model en minimale afbreuk in de prestaties bij het samenvatten. Daarnaast kan het gebruik van efficiënte netwerken, zoals *grouped convolutional layers*, de NLP-downstream taken verbeteren.

### Maximal Marginal Relevance

Traditionele extraherende samenvattingssystemen bouwen verder op de door Carbonell en Goldstein (1998) ontworpen architectuur. Deze architectuur gebruikt een maximaal marginale relevantiescore of MMR. Deze architectuur houdt rekening met de diversiteit en de relevantie van de gemarkeerde zinnen. De relevantie van een zin in een tekst wordt bepaald door de mate waarin het taalmodel de belangrijkste informatie overbrengt van de tekst waarvan het afkomstig is. Om diversiteit aan tekstinhoud te waarborgen, wordt er gekeken naar de mate waarin de geselecteerde zinnen verschillen van de eerder geselecteerde zinnen in de samenvatting. Als een zin relevant is maar qua inhoud te veel overlapt met de eerder geselecteerde zinnen, dan heeft deze minder kans om in de geëxtraheerde samenvatting opgenomen te worden. Deze score kan doorgaans berekend worden met KeyBERT<sup>6</sup>.

Extraherend samenvatten met de MMR-methode is de methode bij uitstek voor ML-toepassingen. Onderzoekers bouwen verder op de architectuur die beschre-

<sup>6</sup><https://maartengr.github.io/KeyBERT/api/mmr.html>

ven staat in Carbonell en Goldstein (1998). In McDonald (2007) stelt de onderzoeker voor om het gulzige zoekalgoritme van MMR te vervangen door een globaal optimale formulering, waarbij het MMR-framework wordt uitgedrukt als een knapzakprobleem of NP-volledig probleem. Daarmee wordt er gewezen naar een *integer linear programming* (ILP) solver die gebruikt kan worden om de wiskundige functie van MMR te maximaliseren. De MMR-methode hield voordien enkel rekening met relevantie en diversiteit, maar niet met de optimale combinatie van zinnen die in een samenvatting moet worden opgenomen. De aanpak van McDonald (2007) vereist echter meer rekenkracht en tijd dan de standaard MMR-methode, maar het experiment van McDonald (2007) haalde wel aan dat deze methode leidde tot significant betere resultaten. H. Lin en Bilmes (2010) evalueerde dit MMR-algoritme. Bij de evaluatie van deze architectuur benadrukte zij de significant betere resultaten.

### Metaheuristiek-gebaseerd

Metaheuristische samenvatting maakt gebruik van metaheuristische optimalisatie-algoritmen zoals genetische algoritmen, *simulated annealing* of zwermoptimalisatie om de belangrijkste zinnen in een tekst te achterhalen. Deze algoritmen zoeken volgens Premjith e.a. (2015) en Verma en Verma (2020) naar de beste combinatie van zinnen die de belangrijkste informatie in de tekst bevatten. De evaluatiefunctie in metaheuristische samenvattingsalgoritmen kan gebaseerd zijn op verschillende criteria, zoals zinslengte, -relevantie en -verbanden. Rani en Kaur (2021) benadrukt dat teksten samenvatten met een metaheuristische methode regelmatig vastraakt in een lokaal optimum. Dit is een tekortkoming op andere methoden. Daarnaast wijst het onderzoek uit aan dat metaheuristische methoden geen *steepness* of extremen op een *search space behaviour* aanduiden. Om de convergentie aanzienlijk te versnellen, moet er gebruik worden gemaakt van een optimalisatiestrategie gebaseerd op gradiënten.

### Experimenten over extraherend samenvatten

McKeown e.a. (1999) voerden experimenten uit op extraherende samenvattingen van nieuwsartikelen. De resultaten wijzen erop dat deze vorm vatbaar is op vooroordelen of *bias* van de auteur. De zinnen worden genomen zoals ze zijn. Hahn en Mani (2000) bouwde verder op dit experiment. Zij voerden een experiment uit met een mix van *knowledge-rich* en *knowledge-poor* methoden, met significant positieve resultaten tot gevolg. De nadruk bij extraherend samenvatten ligt in het kiezen van de *salient text units*. Deze punten zijn typisch in de vorm van zinnen. Er is nood aan een manier om de lexicale en statistische relevantie van een zin te kunnen aanduiden. Hiervoor haalt Hahn en Mani (2000) twee manieren aan:

- Een lineair gewicht model. Iedere teksteenheid wordt gewogen op factoren zoals de *location weight* en het aantal voorkomens.
- Een gewicht model op basis van de statistische opvallendheid van een een-

heid. Zo wordt er rekening gehouden met de aanwezigheid van een woord in (sub)titels.

Nallapati e.a. (2017) wilden de nauwkeurigheid van deze modellen overbruggen. Dit doen ze met *SummaRuNNer*<sup>7</sup>, een oplossing voor het extraherend samenvatten van teksten met een neurale netwerk. De toepassing werd opgebouwd met PyTorch in en bestaat uit een combinatie van drie modellen: een recurrent neurale netwerk, een convolutioneel recurrent neurale netwerk en een *hiërarchische attentie netwerk*.

### 2.5.2. Abstraherend samenvatten

Extraherend samenvatten houdt geen rekening met de structuur van een tekst. De uitvoertekst is op het eerste zicht vrij machinaal. Kernzinnen achterhalen gebeurt bij geautomatiseerde samenvatting met zes features volgens Khan (2014), namelijk de woordfrequentie, de plaats van een zin in de tekst, de *cue method* of een woord dat de kerngedachte van een paragraaf benadrukt, titels, de lengte van de zin, de gelijkenissen tussen de zin en de rest van het document, het gebruik van *proper nouns* en ten slotte de afstand tussen *text units* waarin entiteiten voorkomen. Een tekst abstraherend samenvatten kan volgens Khan (2014) op twee manieren: semantisch en structuurgebaseerd. Cao (2022) deed verder onderzoek naar *deep learning* methoden om abstraherende samenvattingen te automatiseren.

#### Structuurgebaseerde benaderingen op abstraherend samenvatten

Structuurgebaseerde benaderingen zoeken naar belangrijke informatie in de tekst en gebruiken regels om samenvattingen te maken. Ze worden vaak gecombineerd met andere methoden. Khan (2014) motiveert dat deze vorm ideaal is voor het samenvatten naar korte, samenhangende teksten met redundante zinnen. De linguïstische kwaliteit ligt laag en samengevatte zinnen kunnen volgens Khan (2014) grammaticale fouten bevatten, omdat een structuurgebaseerde aanpak niet afhankelijk is van de oorspronkelijke tekststructuur.

#### Semantisch-gebaseerde benaderingen op abstraherend samenvatten

De semantisch-gebaseerde benadering gebruikt de betekenis van de tekst om samenvattingen te maken. Deze samenvattingen zijn kort, duidelijk en bevatten veel informatie. Er zijn verschillende manieren om dit te doen. Deze aanpak toont vorderingen op structuurgebaseerd volgens Khan (2014). Zo zijn de samenvattingen beknopt en bevatten voldoende kern- en deelzaken voor de lezer. De linguïstische kwaliteit is beter en er zijn minder redundante zinnen. Ten slotte worden er minder grammaticale fouten gemaakt, al is er wel de nood aan een extra parsingfase.

---

<sup>7</sup><https://github.com/hpzhao/SummaRuNNer>

### Abstraherend samenvatten met deep learning

*Deep learning* voor abstraherend samenvatten kan met verschillende modellen zoals RNN's (terugkerende neurale netwerken), CNN's (convolutionele neurale netwerken) en sequence-to-sequence (Seq2Seq) modellen. Deze modellen kunnen leren om samenvattingen te maken door de betekenis van de tekst te begrijpen en nieuwe tekst te maken die de belangrijkste informatie overbrengt (Suleiman & Awajan, 2020). Om een abstraherende samenvatting met deep learning op te bouwen bestaan er verschillende modellen. Het Pegasus-model beschreven in J. Zhang e.a. (2020) handelt *gap-sentences* af met pre-trained models voor samenvatting met NLP. Dit model werd getraind en beoordeeld op samenvattingstaken zoals emails, patenten, rekeningen en ook wetenschappelijke artikelen (J. Zhang e.a., 2020).

#### 2.5.3. Hybride samenvatten

In het best denkbare geval wordt abstraherende en extraherende samenvatting gecombineerd volgens Hsu e.a. (2018) en Huang e.a. (2019). Zo omvat een pipeline voor hybride samenvatting twee onderdelen: een *content selection* fase waarin de kernzinnen met extraherende samenvatting worden opgehaald en *paraphrasing*-fase waarbij de gemarkeerde kernzinnen abstraherend worden samengevat.

#### 2.5.4. Evaluatie

Samenvattingen van lange documenten handmatig beoordelen vergt tijd en voldoende planning van een mens (Nenkova & Passonneau, 2004). Met behulp van een vooraf geschreven samenvatting als referentietekst zorgen twee metrieken voor ondersteuning om een samenvatting automatisch te laten beoordelen. Samenvattingen beoordelen kan ook zonder referentietekst, al moeten verschillende factoren worden opgevolgd.

#### Evaluatie met referentieteksten

Onderzoeken bij het vergelijken van een teksttransformatie schakelen BLEU en ROUGE in. Deze twee metrieken meten de gelijkheid tussen een machine-gegenereerde tekst en een referentietekst. Deze referentietekst is gemaakt door de mens. ROUGE is recall-gebaseerd en gebaseerd op exacte *token matches*. Deze aanpak impliceert dat er geen rekening wordt gehouden met synonieme zinnen of zinnen met een gelijke betekenis (C.-Y. Lin, 2004). De ROUGE-modellen die verderbouwen houden rekening met deze tekortkomingen. ROUGE-2 van Ganesan (2018) voorziet dictionaries van synoniemen, zodat er rekening wordt gehouden met synonieme zinnen. ROUGE-G van ShafieiBavani e.a. (2018) gebruikt graafalgoritmen om lexicale en semantische matching mogelijk te maken. C.-Y. Lin (2004) beschrijft een Python-bibliotheek<sup>8</sup> dat de berekening van deze metriek vereenvoudigt.

<sup>8</sup><https://github.com/pltrdy/rouge>

BLEU volgt een gelijkaardige werking en beoordeelt de gelijkenis tussen een machine-gegenereerde tekst en een referentietekst, maar deze meting is precision-gebaseerd. BLEU introduceert een *brevity penalty* of strafterm voor te korte teksten, om te voorkomen dat taalmodellen enkel veelvoorkomende woorden genereren die een hoge precisie opleveren met weinig zinvolle zinnen (Chiusano, 2022). Voor de BLEU-metrick bestaat er ook een Python-bibliotheek<sup>9</sup>.

Metriek	ROUGE	BLEU
Soort metriek	Recall-gebaseerd	Precision-gebaseerd
Nut en gebruik	Evalueert hoe goed een samenvatting de belangrijke inhoud van een referentiesamenvatting dekt	Evalueert hoe goed een gegenereerde samenvatting overeenkomt met een referentiesamenvatting
Te meten waarde	De overlap van n-grams (sequentie van n woorden) tussen de gegenereerde samenvatting en de referentiesamenvatting	De overlap van n-grams (sequentie van n woorden) tussen de gegenereerde samenvatting en de referentiesamenvatting
n-gram range	Meet meestal overlap van 1-4 grams	Kan overlap van 1-4 gram meten, of elk ander bereik van n-grams
Weging	Gewogen of ongewogen	Gewogen of ongewogen
Sterke punten	Goed voor het meten van de dekking van inhoud en de algehele kwaliteit van de samenvatting	Goed voor het meten van de vloeiendheid van zinnen en de grammaticale correctheid
Zwakke punten	Kan gevoelig zijn voor verschillen in woordvolgorde en zinsstructuur tussen de gegenereerde samenvatting en de referentiesamenvatting	Vangt de semantische overeenkomst tussen de gegenereerde samenvatting en de referentiesamenvatting niet op
Gebruik	Vaak gebruikt voor samenvattingstaken	Vaak gebruikt voor machinevertalingstaken

**Tabel 2.2:** Comparison of ROUGE and BLEU metrieken

<sup>9</sup><https://github.com/neural-dialogue-metrics/BLEU>

### Evaluatie zonder referentieteksten

Een samengevatte tekst beoordelen zonder een referentietekst vereist volgens Steinberger en Jezek (2009) meer subjectiviteit en menselijke betrokkenheid dan met een referentietekst. Deze soort kan handmatig gebeuren, maar ook semi-automatisch. De type tekst, de lengte en de complexiteit van de oorspronkelijke tekst zijn factoren die in acht moeten worden genomen bij het beoordelen van de samengevatte tekst. Daar moet er worden gekeken naar het doelpubliek en het formaat. De tekst- en inhoudskwaliteit van de samengevatte tekst moet worden beoordeeld. De tekstkwaliteit is de grammaticale correctheid, niet-redundantie van zinnen en woordenschat en coherente structuur (McCombes, 2022). De inhoudskwaliteit wijst op de informatie dat wordt opgenomen in de samengevatte tekst. Dit omvat de relevantie met de doelgroep bij kern- en bijzaken of misleidende informatie door een misinterpretatie van het systeem (McCombes, 2022).

### 2.5.5. Tekstvereenvoudigingstechnieken voor scholieren met dyslexie.

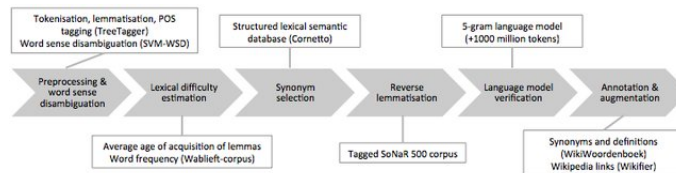
Bingel e.a. (2018) beschrijft een systeem voor dat gericht is op lexicale tekstvereenvoudiging en een embeddings-gebaseerde aanpak voor substitution generation. Tijdens de substitution selection maakt het systeem gebruik van een ongesuperviseerde boundary ranker om de synoniemen te filteren. De geselecteerde synoniemen worden vervolgens gerangschikt met behulp van een gesuperviseerd SR-model. Daarnaast is het systeem in staat om het model te personaliseren op basis van gebruikersfeedback en maakt het gebruik van een seed-dataset van complex-simple overeenkomsten. Deze overeenkomsten houden rekening met de context van een woord in een zin en helpen bij de initiële vereenvoudiging van teksten. Alle gebruikersinformatie wordt met een PostgreSQL databank bijgehouden. Gebruikersinformatie wordt gekoppeld aan de demografische informatie informatie, om zo de ideale uitvoertekst te kunnen genereren.

Het onderzoek van De Belder (2010) richt zich op tekstvereenvoudiging voor kinderen. De doelgroep ligt echter jonger dan deze casus, maar het onderzoek haalt aan hoe de onderzoekers een methode opzetten voor lexicale en syntactische vereenvoudiging. Bulté e.a. (2018) werkte het aspect rond lexicale vereenvoudiging verder uit. Het resultaat van dit onderzoek was een *pipeline* ontworpen om moeilijke woordenschat naar simpele synoniemen te vervangen. Eerst ging de tekstinhoud door een *preprocessing*-fase, samen met het uitvoeren van WSE. Daarna werd de moeilijkheidsgraad van ieder token overlopen. De moeilijkheidsgraad is gebaseerd op de frequentie in SONAR500<sup>10</sup> een corpus met eenvoudige Nederlandstalige woorden, en ook de Wabliet-corpus, een archief van nieuwsartikelen in eenvoudig Nederlands. Synoniemen werden teruggevonden met Cornetto<sup>11</sup>, een lexicale databank met Nederlandstalige woorden, samen met een *reverse lemma*-

<sup>10</sup><https://taalmaterialen.ivdnt.org/download/tstc-sonar-corpus/>

<sup>11</sup><https://github.com/emsrc/pycornetto>

tization fase. Lexicale vereenvoudiging is ingewikkeld wanneer er geen eenvoudigere synoniemen zijn. In dat geval blijft een moeilijk woord voor wat het is. Voor deze toepassing maakten de onderzoekers gebruik van LLM's samen met Wikipedia annotaties. Deze annotaties bevestigden of de gegeven informatie correct is. De resultaten van deze toepassingen vallen in lijn met andere huidige toepassingen van hun soort. De onderzoekers benadrukken dat deze toepassing ook voor andere doelgroepen met een lagere leesgraad kunnen dienen, zoals L2-lezers.



**Figuur (2.18)**

Afbeelding uit Bulté e.a. (2018). Deze pipeline omvat de stappen die de toepassing aflegt.



**Figuur (2.19)**

Afbeelding uit Bulté e.a. (2018). TODO

Al zijn er onderzoeken over lexicale, syntactische en semantische vereenvoudiging voor kinderen en scholieren met dyslexie, het aantal onderzoeken over samenvatten voor deze doelgroep is schaars. Zoals eerder aangehaald is er wel onderzoek gedaan naar de verschillende manieren om een tekst samen te vatten, maar er is geen toepassing of onderzoek dat dit concreet uitwerkt.

### 2.5.6. Conclusie

Wetenschappelijke artikelen volgen een gelijke structuur. De inhoud in PDF- of afbeeldingvorm vergt voldoende cleaning-fasen. Het beoordelen van de samenvatting op basis van een referentietekst met de ROUGE-metriek wordt aangeraden, al kan deze beoordeling niet enkel machinaal gebeuren. Daarnaast is er input en bijsturing nodig van de mens omtrent een samenvatting op maat en de grammaticale, lexicale en semantische correctheid. Tools gericht op het lexicaal en adaptief vereenvoudigen van teksten voor kinderen en scholieren met dyslexie zijn reeds uitgewerkt. Methoden om menselijke en grammatisch correcte samenvatting op te bouwen zijn reeds beschikbaar.



## 2.6. Valkuilen en uitdagingen voor AI-ontwikkelaars bij tekst-vereenvoudiging met AI

AI en ML zijn volop in groei. NLP gebruikt AI en ML om menselijke taal te verwerken, terwijl NLU deze technologieën gebruikt om menselijke taal te begrijpen. Hoewel deze technologieën veelbelovend zijn, moeten AI-ontwikkelaars rekening houden met veelvoorkomende en genegligeerde uitdagingen en valkuilen (Khurana e.a., 2022; Roldós, 2020; Sciforce, 2020). Deze sectie beantwoordt de volgende onderzoeksvraag: "Met welke valkuilen bij taalverwerking met AI moeten ontwikkelaars rekening houden?"

### 2.6.1. Uitdagingen voor softwarebedrijven

Volgens het jaarlijks rapport van IBM behoren Natural Language toepassingen tot de duurste soort om te ontwikkelen. Ongeveer 54% van de bevroagde IT-professionals vond de bijhorende kosten een obstakel bij het starten van de ontwikkeling voor NLP-toepassingen. Professionals halen verschillende pijlers aan, waaronder de kwaliteit en kwantiteit van de data, het trainen van de data, het gebrek van NLP-experten, de integratie en deployment van de taalmodellen en ten slotte de transparantie van het model. Gespecialiseerde modellen zijn use-case afhankelijk wat ze niet voor iedere toepassing bruikbaar maakt. Als oplossing kunnen softwarebedrijven partnerships afsluiten, investeren in NLP-talent, klein starten en stelselmatig opschalen, cloud-gebaseerde oplossingen aanreiken of de transparantie van een model benadrukken in hun specifieke toepassing (IBM, 2022).

### 2.6.2. Ambigüiteit, synoniemen en homoniemen

*Sequence Labeling* koppelt labels aan tokens in doorlopende tekst. Homoniemen kunnen echter roet in het eten gooien. Volgens Roldós (2020) heeft een machine moeite om de context van homoniemen te achterhalen. Bijvoorbeeld bij het woord 'bank' is het niet duidelijk voor de machine of het gaat over de geldinstelling of het meubel. *Word Sense Disambiguation* of WSD is een NLP-taak waarbij de betekenis van een woord wordt bepaald. Deze bepaling gebeurt volgens (Eisenstein, 2019) op basis van de context waarin een woord gebruikt wordt. Deze taak is nodig binnen NLP om rekening te houden met homoniemen. WSD implementeren kan dictionary-gebaseerd, gesuperviseerde, semi-gesuperviseerd of niet-gesuperviseerd. PoS-tagging kan dit probleem aanpakken volgens Liu e.a. (2020) aan als een oplossing op dit probleem, samen met het gebruik van contextual embeddings. Spacy biedt een `sense2vec`<sup>12</sup> aan.

Bij het bouwen van NLP-systemen moeten zo veel mogelijke betekenissen en synoniemen van een woord worden opgenomen. Tekstanalysemodellen zijn niet foutloos, maar hoe meer relevante trainingsgegevens ze ontvangen, hoe beter ze sy-

<sup>12</sup><https://github.com/explosion/sense2vec>



noniemen zullen begrijpen (Roldós, 2020). Dandekar (2016) reikt twee methoden aan: *candidate generation* door gebruik te maken van word embeddings, historical user data of lexicale synoniemen. Daarnaast is er ook een gesuperviseerde methode, namelijk *synonym detection*. Aanvullend kunnen antoniemen volgens Dandekar (2016) op eenzelfde manier worden achterhaald met NLP ML.

Het onderzoek van Sciforce (2020) haalt aan dat het merendeel van NLP-toepassingen Engelstalige invoer gebruikt. Niet-Engelstalige toepassingen zijn zeldzaam. De opkomst van AI technologieën die twee datasets gebruiken, biedt een oplossing voor dit probleem. De software vertaalt eerst de oorspronkelijke tekst naar de gewenste taal, voordat de tekst wordt herwerkt. BERT maakt volgens Roldós (2020) gebruik van meertalige transformers, wat de impact van deze uitdaging kan dempen.

### 2.6.3. Paternalisme en ethische overwegingen

De doelstelling van ondersteunende toepassingen is om gelijke kansen te bieden aan iedere doelgroep. Tekstvereenvoudiging transformeert de oorspronkelijke tekst naar een tekst met een simpelere syntax, kortere zinnen, verminderde lexicale en semantische complexiteit en gereduceerd aantal zinnen. Volgens Niemeijer e.a. (2010) zijn de ethische overwegingen die samenhangen met tekstvereenvoudiging niet gemakkelijk te scheiden van de gebruikte technologie om het resultaat te bereiken. Het onderzoek van Gooding (2022) haalt pijlers aan waarmee ontwikkelaars en softwarebedrijven rekening moeten houden bij de ontwikkeling van adaptieve en ondersteunende leessoftware, voornamelijk toepassingen voor tekstvereenvoudiging. Ontwikkelaars moeten zich meer bewust worden van de behoeften en verwachtingen van de eindgebruiker bij het ontwikkelen van een tekstvereenvoudigingstoepassing. Haar onderzoek benadrukt de paternalistische of afhankelijke aard van assisterende toepassingen. Tekstvereenvoudiging omvat vier transformaties, maar niet iedere transformatie is vereist voor iedere gebruiker. Een adaptieve tekstvereenvoudigingstoepassing moet de eindgebruiker een keuze aanbieden om aan te passen wat vereenvoudigd wordt, afhankelijk van specifieke behoeften.

Software-ontwikkelaars verkiezen volgens Sikka en Mago (2020) voor *black-box* modellen bij de ontwikkeling en finetuning van een NLP-toepassing met AI. AI is het verschil qua nauwkeurigheid minimaal, de afweging wordt gemaakt bij de transparantie van het model. Na een transformatie wordt er niet aangegeven waarom specifieke transformaties werden uitgevoerd, bijvoorbeeld het vervangen van een woord door een eenvoudiger synoniem. Xu e.a. (2015) benadrukt dat toepassingen voor tekstvereenvoudiging meer rekening moeten houden met de doelgroep waarvoor ze worden ontwikkeld. White-box modellen zijn er in schaarse hoeveelheden. Om dit probleem op te lossen, is het belangrijk om de eindgebruiker, in dit geval scholieren met dyslexie in het derde graad middelbaar onderwijs, de keuze te geven. Zoals beschreven in Gooding (2022), zijn er verschillende mogelijkheden. Bij-

voorbeeld, de eindgebruiker moet de mogelijkheid hebben om te kiezen welke synoniemen de tekst lexicaal zullen aanpassen. Een alternatieve aanpak voor syntactische vereenvoudiging is om de scholier zelf zinnen te laten markeren die moeilijk te begrijpen zijn, zodat het systeem alleen de door de eindgebruiker aangegeven zinnen vereenvoudigt.

#### **2.6.4. Valkuilen bij prompt engineering**

Iedereen is in staat om een conversatie met een chatbot op te bouwen. Het gebruik van de API voor een doelgericht en doordacht antwoord vergt echter planning bij de ontwikkelaar. Miszczak (2023) waarschuwt voor 'garbage-in garbage-out'. De kwaliteit van de input kan de kwaliteit van de output bepalen. Jiang (2023) benoemt de misopvatting bij de intentie van de gebruiker als de voornaamste uitdaging voor een taalmodel dat input vereist. Volgens Jiang kan dit te wijten zijn aan de gebruiker die een onvolledige prompt of een prompt met onvoldoende context op een concreet antwoord schrijft. Daarnaast kan een gebrek aan trainingsdata ook aan de oorzaak liggen van een onnauwkeurige uitvoertekst of bias in het taalmodel. Andere factoren die Miszczak (2023) aanhaalt, zijn de afwisselende probabiliteit van de outputtekst en de meegegeven parameters die het model beïnvloeden, zoals de temperature dat de creativiteit van het model beïnvloedt. Als oplossing kan de prompt als een conditionele expressie worden opgebouwd, zodat het taalmodel enkel met zekerheid een antwoord teruggeeft.

#### **2.6.5. Evaluatie en interpretatie**

ROUGE en BLEU zijn twee metrieken die geen rekening houden met de semantiek van de tekst. Twee zinnen met een identieke semantiek kunnen een lage score opleveren als er geen overlap is tussen de oorspronkelijke en de samengevatte tekst. Daarnaast kan een tekst grammaticaal fout zijn en nog steeds een goede tot uitstekende score hebben. Deze beperkingen van de ROUGE en BLEU scores kunnen worden opgelost door gebruik te maken van verwante metrieken, zoals ROUGE-L, ROUGE-SU en METEOR. Deze scores houden wel rekening met aanvullende factoren zoals zinsemantiek -of overeenkomsten. ROUGE, BLEU en de verwante metrieken zijn geen vervanging op menselijke evaluatie. Zowel machinale als menselijke beoordeling moet worden overwogen bij het onderzoeken van een samenvattingsmethode (Raj, 2017; Tatman, 2019).

Er is geen algemene consensus bij onderzoekers over de ideale evaluatiemetriek(en) voor tekstvereenvoudiging -en samenvatting. Er is volgens Fabbri e.a. (2020) nood aan een mix tussen machinale en geautomatiseerde evaluatie, alsook evaluatie door de mens en met voorkeur in die volgorde. Iskender e.a. (2021) zien menselijke evaluatie niet als een gulden standaard bij het beoordelen van een samengevatte tekst. Ze stimuleren verder onderzoek naar nieuwe standaarden en best practices voor betrouwbare menselijke beoordeling op samengevatte teksten. Het onder-

zoek onderzocht niet of er verschillende resultaten zijn bij de beoordeling van extraherende en abstraherende samenvattingen. De beste samengevatte versie van een tekst achterhalen met menselijke feedback vergt de juiste methode. De doelgroep waarvoor een tekst wordt samengevat, moeten nauw in het proces worden opgenomen (Iskender e.a., [2021](#)).

## **2.7. Beschikbare software voor tekstvereenvoudiging**

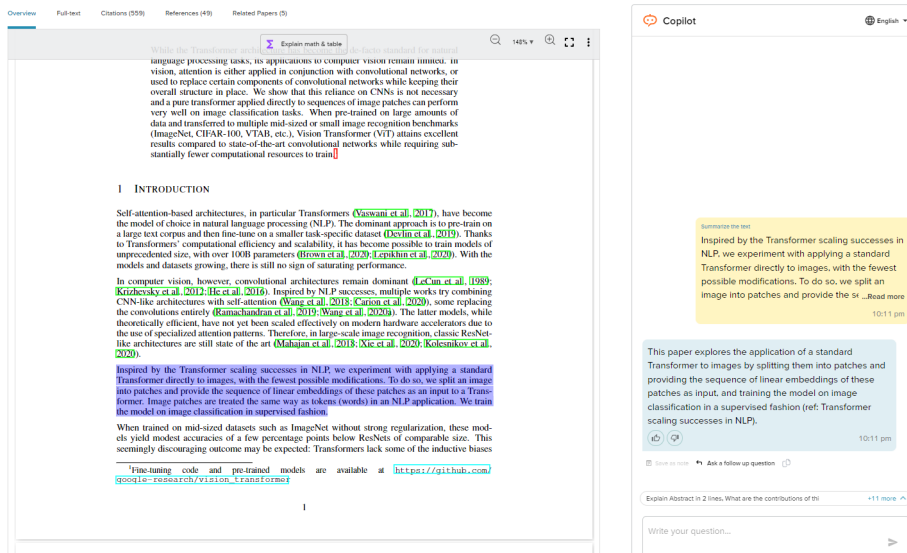
Dyslexie is een veelvoorkomende aandoening die de lees- en schrijfvaardigheden van scholieren kan belemmeren. Om deze scholieren te ondersteunen, worden er verschillende softwareprogramma's en tools ontwikkeld. In dit hoofdstuk zal worden gekeken naar mogelijke nationale en internationale software die specifiek is ontworpen om scholieren met dyslexie te helpen bij het lezen van teksten. Er zal met name worden gekeken naar de beschikbare software in Vlaamse middelbare scholen, chatbots, zoals Bing AI en ChatGPT, en software die speciaal is ontwikkeld om dyslexie te ondersteunen bij het lezen. Deze sectie beantwoordt de volgende onderzoeksvraag: "Welke toepassingen, tools en modellen zijn er beschikbaar om Nederlandstalige geautomatiseerde tekstvereenvoudiging met AI mogelijk te maken?"

### **2.7.1. Momenteel ingezet in het onderwijs**

In het middelbaar onderwijs wordt lees- en studieondersteuning voor scholieren met dyslexie enkel in de vorm van voorleessoftware voorzien (De Craemer e.a., [2018](#); OnderwijsVlaanderen, [2023](#)). OnderwijsVlaanderen ([2023](#)) leent licenties voor de volgende softwarepakketten uit:

- SprintPlus
- Kurzweil3000
- Alinea Suite
- IntoWords
- TextAid

Naast luister- en schrijfopties kunnen scholieren deze toepassingen gebruiken om zinnen te markeren om deze zinnen vervolgens samen te vatten. Enkel de gemarkeerde zinnen worden betrokken in de samengevatte versie, dus de zinnen blijven lexicaal, syntactisch en semantisch identiek. Alle vermelde softwarepakketten bieden echter geen onafhankelijke samenvat- of vereenvoudigfunctie aan. Tops e.a. ([2018](#)) benadrukt de handige aspecten van deze software, maar deze software moet zo vroeg mogelijk in een schoolcarrière worden ingezet. Zo raken de scholieren snel vertrouwd met het gebruik, wat kan leiden tot een optimaal gebruik in



**Figuur (2.20)**  
Schermabbeelding van SciSpace.

verdere studies. Volgens Tops e.a. (2018) is het te laat om deze software pas in het hoger onderwijs te introduceren.

### 2.7.2. Proof-of-concepts en online webapplicaties

Online zijn er tools beschikbaar om teksten generiek samen te vatten. Resoomer, Paraphraser en Scholarcy zijn oorspronkelijk Engelstalige tools, met ondertussen de mogelijkheid om een abstraherende samenvatting te maken van Nederlandstalige teksten. De taalmodellen waar deze applicaties op werken, is niet gekend. Daarnaast zijn er ook geen API's beschikbaar om mee te werken. Gepersonaliseerde toepassingen zijn er in mindere mate. Bingel e.a. (2018) omschrijft een proof-of-concept voor een webtoepassing dat teksten vereenvoudigd, met oog op mensen met dyslexie. Deze software noemt nu Hero en bevindt zich in betafase.

Toepassingen om wetenschappelijke artikelen te vereenvoudigen zijn schaars, maar er zijn enkele gratis en betalende toepassingen beschikbaar. SciSpace<sup>13</sup> is gratis. Scholarcy<sup>14</sup> is betalend.

### 2.7.3. GPT-3

*Generative Pretrained Transformer 3* of GPT-3 is een taalmodel ontworpen door OpenAI. Dit taalmodel werkt met een tweestapsleerparadigma. Pre-training gebeurt ongesuperviseerd met een language modelling objective. Vervolgens wordt het taalmodel gesuperviseerd gefine-tuned. Over drie versies heen is het model aanzienlijk vergroot, van anderhalf miljard parameters bij GPT-2 naar 175 miljard parameters bij GPT-3. Het model is getraind op niet-gecategoriseerde data van

<sup>13</sup><https://typeset.io/>

<sup>14</sup><https://www.scholarcy.com/?ref=theresanaiforthat>

het internet en gebruikt datasets waaronder Common Crawl, WebText2, Books1, Books2, and Wikipedia. Dit taalmodel steunt op *Reinforcement Learning from Human Feedback* of RLHF (C. Li, 2022; Radford e.a., 2019).



Can you simplify the following paragraph:

At some ideas you stand perplexed, especially at the sight of sin. And you must ask yourself whether to combat it with force or with humble love. Always decide to combat it with humble love. If you decide this once and for all, you can conquer the whole world! Humble love is a frightening force, of which there is nothing stronger.



Here is a simpler version of the paragraph:

When you see something bad, like sin, you might be confused. Ask yourself if you should use force or love to fight it. Always choose love. If you do, you can win against anything. Love is the strongest power.

### Figuur (2.21)

Afbeelding van Gooding 2022. De invoertekst is een paragraaf uit een niet-vermeld boek van de Russische schrijver Dostoevsky. Het resultaat van de meegegeven prompt is een transformatie dat iedere vorm van vooraf aangehaalde vereenvoudiging weergeeft. Lexicale, conceptuele en syntactische vereenvoudiging worden op de invoertekst toegepast.



Can you simplify the following paragraph in three different versions:

At some ideas you stand perplexed, especially at the sight of sin. And you must ask yourself whether to combat it with force or with humble love. Always decide to combat it with humble love. If you decide this once and for all, you can conquer the whole world! Humble love is a frightening force, of which there is nothing stronger.



Here are three simplified versions of the paragraph:

1. When you see something bad, like sin, you might be confused. Ask yourself if you should use force or love to fight it. Always choose love. If you do, you can win against anything. Love is the strongest power.
2. When you encounter ideas that confuse you, especially when it comes to sin, you may be unsure how to react. You must ask yourself if it is better to fight back with force or with humble love. It is always better to choose humble love. If you make this decision once and for all, you can overcome anything. Humble love is a powerful and intimidating force that is stronger than anything else.
3. If you see something bad, like sin, and you don't know what to do, remember that love is always the answer. Choose love over force. If you do, you will be able to conquer anything. Humble love is the most powerful force in the world.

### Figuur (2.22)

Afbeelding van Gooding 2022. Gooding haalt verder aan dat modellen zoals ChatGPT op twee vlakken de leesbaarheid van een tekst kan bevorderen. Allereerst door het verlenen van verschillende mogelijke versies van een vereenvoudigingstaak.



Which of the three simplified versions is the simplest?



Of the three simplified versions, the first one is the simplest. It uses the shortest and simplest words and phrases to convey the main idea of the original paragraph, which is that humble love is the strongest force and can be used to overcome anything. The other two versions add more details and examples, which make them slightly more complex than the first version. However, all three versions are simpler than the original paragraph and effectively convey the same idea.

### Figuur (2.23)

Afbeelding van Gooding 2022.

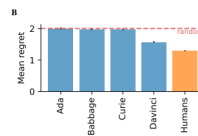
Lisowski (2023) vergelijkt de twee OpenAI taalmodellen met een *mixed-methods* onderzoek. Al blijken de twee heel gelijkaardig, het experiment benadrukt dat het ChatGPT-model gericht is op conversationele doeleinden met voorkeur als chatbot, terwijl GPT-3 een ML-model is bedoeld om met hoogstens één prompt te werken. De grootte van het GPT-3 model met 175 miljard parameters imposanter dan ChatGPT. Daarnaast is de limiet bij het meest recente GPT-3 model is 4000 tokens. Verder haalt Lisowski aan dat de kwaliteit bij beide modellen sterk afhankelijk is van de invoer. De prompts moeten concreet genoeg zijn, om zo niet af te wijken van wat de gebruiker wilt (Lisowski, 2023). Deze twee API's zijn nu vrij beschikbaar voor ontwikkelaars als betalende API (Greg e.a., 2023).

### Beschikbare GPT-3 engines

De documentatie van OpenAI<sup>15</sup> reikt vier verschillende engines voor het GPT-3 taalmodel aan, namelijk Davinci, Curie, Babbage en Ada. In Maart 2023 voegde een vijfde engine zich toe, namelijk GPT-3 Turbo wat de basis is achter Chat-GPT (Greg e.a., 2023).

- Davinci-003 is het meest recente model en is in staat om alles te doen wat de andere engines in staat zijn. Het model maakt gebruik van text de invoertekst eventueel kan aanvullen, mocht de invoertekst afgebroken zijn. Deze aanvulling kan in vraag worden gesteld. Aanvullend geeft aan dat deze engine de meest menselijke antwoorden geeft op basis van foutenmarge. De andere engines kwamen niet in de buurt. Binz en Schulz (2023) vult aan dat deze engine de meest *menselijke* antwoorden teruggeeft. Daarom is deze engine geschikt voor taken zoals essays schrijven, vragen beantwoorden of code genereren.
- Curie is in staat om zinnen met nuance te verwerken, maar maakt een grote sprong terug qua menselijke interpretatie vergeleken met Davinci. Deze engine wordt ingezet voor taken zoals tekstclassificatie, samenvatting of vertaling.

<sup>15</sup><https://platform.openai.com/docs/>

**Figuur (2.24)**

Afbeelding van Binz en Schulz (2023). Dit toont de *mean regret* aan tussen de vier engines en de menselijke antwoorden.

- Ada is niet in staat om zinnen met nuance te verwerken. Deze engine wordt aangeraden voor taken zoals het aanvullen van tekst of het aanmaken van labels.
- Babbage is de minst krachtige en kan ook geen zinnen met nuance verwerken, maar het is de snelste engine. Deze engine toekennen aan eenvoudige taken zoals sentiment analyse of het achterhalen van keywords is volgens de documentatie *best practice*.

### Tools met GPT-3

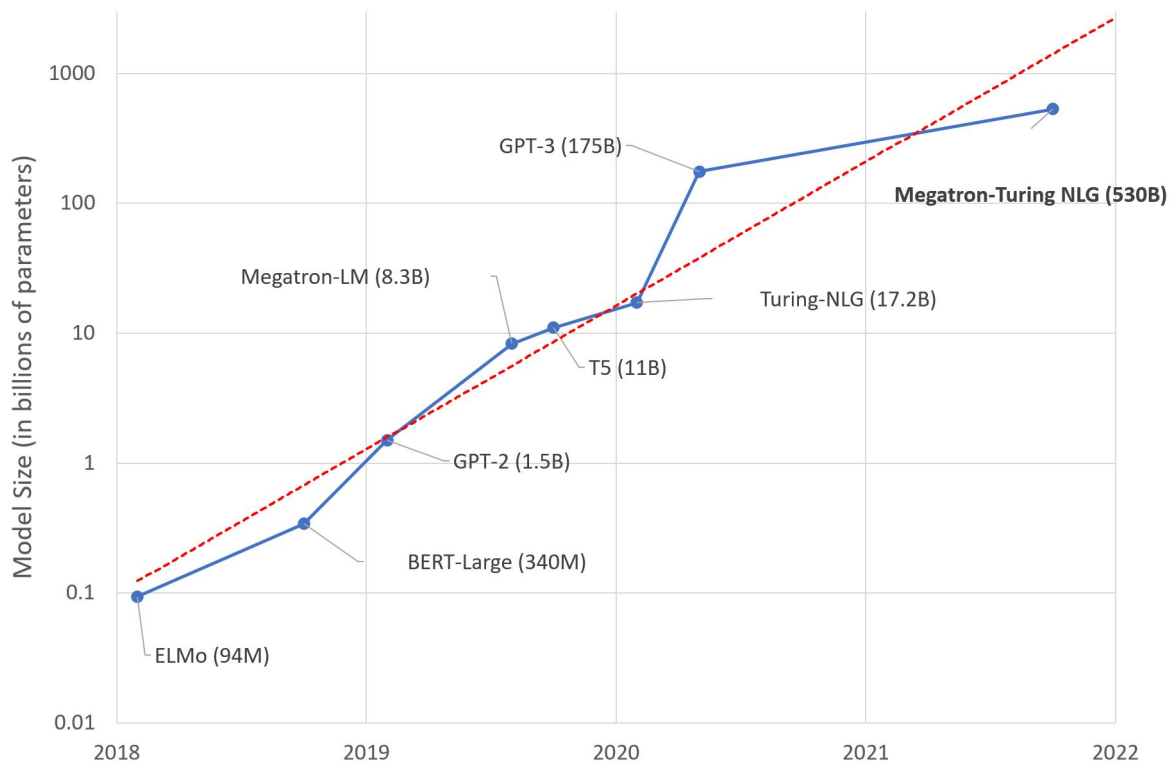
Mottesi (2023) haalt lees- en schrijftools aan die gebruik maken van de GPT-3 API. Jasper AI is een chatbot bestemd voor customer support en een virtueel assistent voor e-commerce. ChatSonic is een tool gericht om social media posts of nieuwsartikelen te genereren op basis van een kernzin of kernwoord.

Verschillende artikels vermelden de mogelijkheden voor het gebruik van GPT-3 en ChatGPT in het onderwijs. Roose (2023) haalt zo de hoge toegankelijkheid, engagement bij scholieren en granulaire personalisatie aan dat het GPT-3 model toe in staat is. Garg (2022) ziet in GPT-3 en ChatGPT een portaalfunctie, om scholieren te helpen bij het opzoeken van nieuwe informatie tijdens de les en bij het instuderen.

### Vergelijking met andere taalmodellen

De architectuur tussen GPT-3 en BERT is volgens Mottesi (2023) het meest opvallende verschil. GPT-3 is een autoregressief model en houdt daarmee enkel rekening met de linkercontext bij het voorspellen of genereren van tekst. BERT daarentegen is bidirectioneel en neemt zowel de linker- als de rechtercontext in overweging. De bidirectionele werking is geschikt voor sentimentanalyse waarbij begrip van de volledige zincontext noodzakelijk is. GPT-3 heeft toegang tot meer informatie (45TB) dan BERT (3TB), wat het een voordeel kan geven bij het samenvatten of het vertalen. Ten slotte zijn er ook verschillen in grootte. Hoewel beide modellen erg groot zijn, GPT-3 is aanzienlijk groter dan de voorganger vanwege de uitgebreide trainingsdatasetgrootte (Brown e.a., 2020).





**Figuur (2.25)**

Afbeelding van Simon (2021). De evolutie van pre-trained taalmodellen wordt hier weergegeven tot eind 2022. De performantie van de modellen ten opzichte van de grootte volgt een lineaire functie.

C. Li (2022) benadrukt dat GPT-3 voor simpele taken *overkill* is. Taken buiten het genereren van teksten, zoals sentimentanalyse en -classificatie, worden beter met een kleinschaliger taalmodel zoals BERT en verwante modellen uitgevoerd. Deze keuze beïnvloedt het budget, want GPT-3 is een API waar per token wordt betaald, terwijl BERT gratis en open-source is. Simon (2021) en Strubell e.a. (2019) halen de ecologische effecten aan van ontwikkelaars die te snel voor deze modellen grijpen. Er is een bewezen effect kleinere modellen, gebruik van Cloud-infrastructuur en ten slotte een geschikte model finetuning bijdragen tot efficiëntere alsook minder klimaatbelastende effect.

### GPT-3 voor samenvattingen

Onderzoek naar OpenAI's ChatGPT en GPT-3 modellen bevindt zich in een vrij vroeg stadium, al zijn er wel enkele vergelijkende onderzoeken die de kracht en zwaktes van deze technologieën aantonen. Het experiment van Tanya Goyal (2022) achterhaalt het gebruik van *zero-shot* samenvattingen buiten generieke samenvattingen. Het onderzoek staat stil bij de impact van prompt-gebaseerde modellen voor het automatisch samenvatten van nieuwsartikelen. Daarnaast maakte het onderzoek gebruik van text-davinci-002 als case study. Uit het experiment besluiten de onderzoekers dat *zero-shot* samenvattingen met GPT-3 beter presteren dan *fine-*



*tuned* modellen, en dat bestaande automatische metrieken zoals BLEU, ROUGE en BERTScore niet geschikt zijn om *zero-shot* samenvattingen te beoordelen. Verder blijkt dat zero-shot samenvattingen meer coherentie en relevantie hebben voor trefwoord-gebaseerde samenvattingen, terwijl aspect-gebaseerde samenvattingen nog vaak blijven te falen.

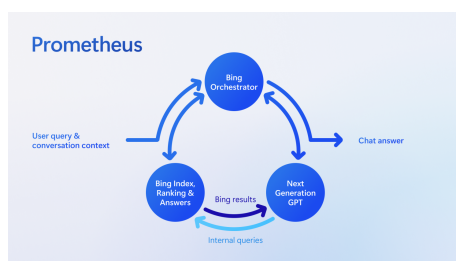
### GPT-3 finetuning

Parameter	Omschrijving	Mogelijke waarden
model	Het GPT-3 model om te gebruiken	davinci, curie, babbage, ada, text-davinci-002, text-curie-001, text-babbage-001, text-ada-001, davinci-codex
temperature	De gulzigheid van een generatief model. Een lagere waarde zal conservatieve en voorspelbare tekst teruggeven. Hogere waarden zullen meer gevarieerde en onverwachtse tekst teruggeven, wat beter werkt bij creatieve toepassingen.	Een kommagetal tussen 0 en 1.
max_tokens	Het maximaal aantal tokens (woorden of subwoorden) dat het generatief model kan teruggeven.	Een getal tussen 1 and 2048.
top_p	Vergelijkbaar met temperature, maar deze waarde onderhoudt de probability distribution voor common tokens. Hoe lager de waarde, hoe waarschijnlijker de woordenschat dat het model zal overwegen bij het genereren van tekst. Een hoge waarde is toepasselijker wanneer een toepassing gericht is op nauwkeurigheid en correctheid.	Een kommagetal tussen 0 en 1.
stop	Een tekstwaarde (woord/symbool) tot waar het model zal genereren. When the model generates a string that matches any of the specified strings, it stops generating text.	Een lijst van stringwaarden, of een enkele string.
presence_penalty	Factor die bepaalt hoe regelmatig woorden voorkomen.	Een kommagetal tussen 0 en 1

### 2.7.4. Bing AI

Microsoft en OpenAI werken nauw samen. Zo maakt het conversationele taalmodel van Bing ook gebruik van GPT-3. Deze chatbot bouwt verder en biedt zo verwijzingen en referenties aan naar andere websites. Deze verwijzingen zijn volgens mogelijk door de Prometheus-technologie van Microsoft (Ribas, 2023).

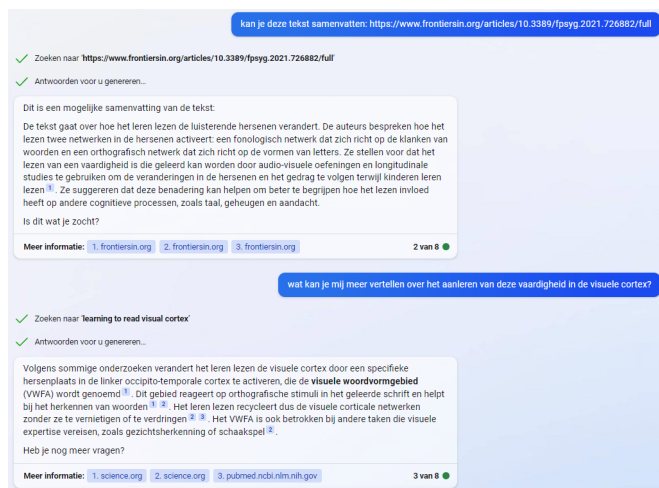
Prometheus is een eigen technologie die door Bing is ontwikkeld. Het AI-model is volgens Ribas (2023) de eerste van zijn soort die de Bing-index-, ranking- en antwoordresultaten combineert met het redeneervermogen van OpenAI's GPT-modellen. Prometheus maakt gebruik van de kracht van Bing en GPT om iteratief via een component genaamd *Bing Orchestrator* een set interne queries te genereren met als doel binnen gegeven gesprekscontext een nauwkeurig antwoord op gebruikersqueries te bieden (Ribas, 2023).



**Figuur (2.26)**

Afbeelding van Ribas (2023).

Bing AI is nu in testfase met wachtlijst en bestaat in de vorm van een webpagina en een browserextensie voor Microsoft Edge. Onderzoek naar deze chatbot staat nog in de kinderschoenen en er is nood aan onderzoek naar de credibiliteit en correctheid van de verwijzingen. Deze chatbot gebruikt een combinatie van extraherende en abstraherende samenvattingen. In tegenstelling tot GPT-3 is er geen officiële API beschikbaar. Daarnaast is de limiet ook lager met 2000 tokens per bericht tijdens een conversatie.

**Figuur (2.27)**

In deze afbeelding wordt er een online wetenschappelijk artikel meegegeven. Er wordt geen titel of onderwerp meegegeven, maar de Bing AI chatbot is in staat om een abstraherende samenvatting te maken van het artikel. Daarna geeft de chatbot verder uitleg over een bepaald onderwerp en geeft het extra referenties mee.

Het bedrijf DuckDuckGo dat instaat voor de gelijknamige zoekmachine probeert een gelijkaardig initiatief. Met *DuckAssist* biedt de onderneming een eigen AI-oplossing aan om een algemene doelgroep te ondersteunen bij het opzoeken van (nieuwe) informatie. Zij halen informatie direct uit enkel Wikipedia pagina's (Weinberg, 2023). Daarnaast maakt dit DuckAssist ook gebruik van het GPT-3 model. Nadelig heeft deze toepassing voorlopig beschikking tot een kleinere zoekruimte dan Bing AI, wat gebruik maakt van meer sites inclusief onderzoekssites zoals ResearchGate (McAuliffe, 2023). Deze beperkte zoekruimte reduceert de kans op incorrecte of foutieve informatie volgens Weinberg (2023), al is dit eerder een intuïtie van het bedrijf.

### 2.7.5. Meta LLaMa

Ten slotte is *Large Language Model Meta AI* of LLaMa) van Meta een generatief taalmodel in dezelfde lijn als de vooraf vermelde taalmodellen. Meta omschrijft LLaMa als een "kleiner foundation model". Het doel van Meta is om een even groot taalmodel als GPT-3 te lanceren met minder rekenkracht en nodige middelen. Dit taalmodel staat in de kinderschoenen en is nog niet beschikbaar in de vorm van online webtoepassing of online API (Hern, 2023; Touvron e.a., 2023). LLaMa toont potentieel, want bij het experiment van Touvron e.a. (2023) is LLaMa sterker dan GPT-3 en soortgenoten, terwijl het van tien keer minder parameters gebruik maakt.

### **2.7.6. Samenvattend schema AI NLP-modellen**

#### **2.7.7. Conclusie**

Experten halen het GPT-3 model en ChatGPT aan als de toekomst voor gepersonaliseerde en adaptieve uitleg aan scholieren. Bing AI biedt een extra dat revolutionair kan zijn bij het opzoeken van uitleg voor zoektermen, zonder het verlies aan bronvermelding. Huidige toepassingen staan mogelijks in een spreekwoordelijke schaduw eenmaal leessoftware voor scholieren met dyslexie worden ontwikkeld met AI. De mogelijkheden van GPT-3 zijn eindeloos en toepassingen die hiervan gebruik maken, kunnen in het onderwijs ingezet worden als ondersteunende software.

### **2.8. Conclusie**

De noden van scholieren met fonologische dyslexie in de derde graad van het middelbaar gaan verder dan gewoon moeizaam lezen. Het ontcijferen en automatiseren van wordeherkenning gebeurt langzaam. Er zijn bewezen voordelen van manuele tekstvereenvoudiging en adaptieve visuele weergaven op kinderen en jongeren met dyslexie.

De leesbaarheid van wetenschappelijke artikelen bevindt zich in een neergaande trend. Het formaat, gebruik van vakjargon en ingewikkelde woordenschat en ten slotte de moeizame syntax en zinsbouw sluiten een algemene doelgroep uit bij het lezen van wetenschappelijke artikelen. Enkel wetenschappelijk geletterden zijn in staat om deze artikelen te lezen. Het uniforme formaat van een wetenschappelijk artikel biedt kansen aan voor een geautomatiseerde aanpak tot het vereenvoudigen van een tekst.

Experten halen meerdere bewezen tactieken aan om teksten automatisch te vereenvoudigen op maat voor een scholier met dyslexie. Handmatig worden teksten vereenvoudigd aan de hand van leesbaarheidsformules of intuïtie. Zinnen moeten lexicaal, syntactisch en semantisch worden vereenvoudigd. Teksten samenvatten maakt de tekst korter zonder het verlies van de kernboodschap. Voor deze vier transformaties zijn er taalmodellen beschikbaar in de vorm van API's of open-source software.

Huidige software dat de overheid uitleent aan scholieren met dyslexie in het middelbaar onderwijs dient vooral als voorleessoftware en bevat nu geen functie dat teksten kan vereenvoudigen. Nieuwe en opkomende technologieën zoals het GPT-3 taalmodel bieden de mogelijkheid wel aan. De ontwikkeling van tools met dit taalmodel is in opmars, maar ontwikkelaars moeten bewust zijn dat andere taalmodellen zoals BERT voor taken zoals semantische analyse minder rekenkracht vereisen voor eenzelfde en soms beter resultaat.

# 3

## Methodologie

### 3.1. Requirementsanalyse

Aan de hand van een vergelijkende studie met beschikbare tools, API's en software wordt gekeken hoe scholieren met dyslexie in de derde graad van het middelbaar onderwijs worden ondersteund. Eerst wordt een shortlist opgesteld van ondersteunde software dat momenteel aan scholieren met dyslexie wordt uitgeleend, alsook nieuwe en opkomende tools die specifiek toegespitst zijn op het vereenvoudigen en samenvatten van teksten. Op basis van deze shortlist wordt een requirementsanalyse opgesteld. Deze requirementsanalyse dient als de basis voor de ontwikkeling van het prototype.

#### 3.1.1. Software in het onderwijs

#### 3.1.2. Tekstanalyse

Geen softwarepakket of tool geeft standaard visueel weer waarom het taal of AI-model een zin moeilijk of belangrijk vindt, of waarom het model een kernwoord koos. Dit sluit aan bij de bevindingen van Gooding (2022). Het GPT3-model en verwante Bing-model doet dit wel wanneer het taalmodel dit expliciet gevraagd wordt. Andere prompt-gedreven tools in de shortlist houden hier geen rekening mee en verwerpen de vraag. Het vragen aan het taalmodel biedt een alternatief aan, maar valt buiten de perken en capaciteiten van de standaardgebruiker. Deze prompt kan worden aangeboden in de vorm van een intuïtieve knop. (Suggestie)

#### 3.1.3. Lexicale vereenvoudiging

##### Syntactische vereenvoudiging

##### Samenvatten

De experimenten met teksten wijzen uit dat Bing AI de nadruk legt op het behouden van bronreferenties. Wanneer expliciet gevraagd, geeft het model bronnen

terug buiten het oorspronkelijke artikel.

### 3.1.4. Conclusie

## 3.2. Ontwikkelen van een prototype

### 3.2.1. Schetsing met Python notebooks

De werking van tekstvereenvoudiging via Python-code wordt optimaal weergegeven in Python notebooks. Deze omgeving belemmert niet het gebruik van API's en voorziet een snelle weergave zonder dat code uitgevoerd moet worden.

#### PDF text mining

De meest gebruikte vorm van tekstbestanden is in de vorm van PDF-bestanden. Methoden om deze bestanden uit te lezen en om te zetten naar tekstdata bestaan reeds in de vorm van Python-bibliotheken. De keuze van de bibliotheek speelt wel een rol. Sommige Python-bibliotheken bieden meer functies aan die dynamischer omgaan met de inhoud van een PDF-document. PDF Miner biedt zo extra functies aan om de metadata van een PDF-bestand op te halen, classificatie op basis van lettertypes of pagina's per titel op te halen. Deze functies bevatten een intuïtieve naamgeving en besparen ontwikkelaars de overbodige werklust.

#### Data pre-processing

##### Extraherende samenvatting

Factoren zoals topic diversity en semantische redundantie moeten overwogen worden bij het kiezen van een taalmodel voor extraherend samenvatten.

### 3.2.2. Integreren naar Flask

Flask biedt een snelle en toegankelijke opzet aan voor Python-ontwikkelaars. De combinatie van een robuuste front-end en back-end maakt dit binnen de schaal van een prototype ideaal.

```
@app.route('/', methods=['GET'])
def home():
    return render_template('index.html')

if __name__ == "__main__":
    app.run()
```

# 4

## Conclusie

### **Voldoet de huidige software aan de noden van scholieren met dyslexie in het middelbaar onderwijs?**

De huidige software uitgeleend aan scholieren met dyslexie in de derde graad van het middelbaar onderwijs voldoet niet aan de noden. De tekstvereenvoudiging - en samenvattingsfuncties gaan niet verder dan het markeren van zinnen, om deze later om te vormen naar een tekst die geen rekening houdt met de coherentie.

### **Advies bij het gebruik van pre-trained taalmodellen bij automatische tekstvereenvoudiging.**

Bestaande taalmodellen vereenvoudigen de ontwikkeling voor een dergelijke toepassing. Het prototype wijst uit dat vrij beschikbare taalmodellen zoals BERT en verwante BERT-varianten een oplossing bieden voor semantische analyse, kernwoorden achterhalen en extraherende samenvatting. Complexere vereenvoudigingstechnieken zoals lexicale en syntactische vereenvoudiging en abstraherende samenvatting kunnen ontwikkeld worden met complexere taalmodellen, zoals GPT-3. Ontwikkelaars moeten bij deze keuze rekening houden met de schaal van de modellen. BERT is mogelijk om abstraherende samenvatting mogelijk te maken, maar zoals aangetoond staat deze in de schaduw van GPT-3 wat voor ontwikkelaars vrij beschikbaar is. Omgekeerd is GPT-3 een *overkill* voor taken zoals het ophalen van kernwoorden of aanduiden van belangrijke zinnen, wat taken zijn die kost- en energiereducerend kunnen aangepakt worden.

Het GPT-3 model maakt complexe en granulaire NLP-transformaties op lexicaal en syntactisch niveau mogelijk voor geautomatiseerde tekstvereenvoudiging. Echter houdt het model geen rekening met referenties buiten de getrainde data, wat tot problemen bij de credibiliteit van data kan leiden. Bing AI daarentegen doet dit wel en vormt een goede fundering voor ontwikkelaars om meer referentiemateriaal aan te bieden in ondersteunende software binnen het onderwijs.

**Nieuwe vragen tot verder onderzoek.**

Er is meer onderzoek nodig naar hoe de deployment van webtoepassingen en browserextensies die tekstvereenvoudiging in het onderwijs kan gebeuren. Met alsmaar grotere taalmodellen, zoals het opkomende GPT-4 en LLaMa moet er ook meer onderzoek zijn naar de verschillen op taalvlak ten opzichte van de toename in parameters. De overstap maken qua taalmodellen is kostelijk voor ontwikkelaars en het eerdere onderzoek van (..) wees uit dat de grootte van taalmodellen alsmaar minder relevant wordt.





# Onderzoeksvoorstel

## Samenvatting

Ingewikkelde woordenschat en zinsbouw hinderen scholieren met dyslexie in het derde graad middelbaar onderwijs bij het lezen van wetenschappelijke artikelen. Adaptieve tekstvereenvoudiging helpt deze scholieren bij hun lees- en verwerkingsnelheid. Daarnaast kan artificiële intelligentie (AI) dit proces automatiseren om de werkdruk bij leraren en scholieren te verminderen. Dit onderzoek achterhaalt met welke technologische en logopedische aspecten AI-ontwikkelaars rekening moeten houden bij de ontwikkeling van een AI-toepassing voor adaptieve en geautomatiseerde tekstvereenvoudiging. Hiervoor is de volgende onderzoeksvraag opgesteld: "Hoe kan een wetenschappelijk artikel automatisch worden vereenvoudigd, gericht op de unieke noden van scholieren met dyslexie in het derde graad middelbaar onderwijs?". Een vergelijkende studie beantwoordt deze onderzoeksvraag en is uitgevoerd met bestaande toepassingen en een prototype voor adaptieve en geautomatiseerde tekstvereenvoudiging. Uit de vergelijkende studie blijkt dat toepassingen om wetenschappelijke artikelen te vereenvoudigen, gemaakt zijn voor een centrale doelgroep en geen rekening houden met de unieke noden van een scholier met dyslexie in het derde graad middelbaar onderwijs. Adaptieve software voor geautomatiseerde tekstvereenvoudiging is mogelijk, maar ontwikkelaars moeten meer inzetten op de unieke noden van deze scholieren.

## A.1. Introductie

Het Vlaams middelbaar onderwijs staat op barsten. Leraren en scholieren worden overspoeld door werkdruk en stress. Bovendien is de derde graad van het middelbaar onderwijs een belangrijke mijlpaal voor de verdere loopbaan van scholieren, al hebben zij volgens Dapaah en Maenhout (2022) dan moeite om grip te krijgen

op de vakliteratuur bij STEM-vakken. Het STEM-agenda<sup>1</sup> van de Vlaamse Overheid moet het STEM-onderwijs tegen 2030 aantrekkelijker te maken, door de ondersteuning voor zowel leerkrachten als scholieren te verbeteren. Toch wordt het aanpakken van de steeds complexere wetenschappelijke taal, zoals beschreven in Barnett en Doubleday (2020), niet opgenomen in het STEM-agenda. Wetenschappelijke artikelen vereenvoudigen, op maat van de noden voor een scholier met dyslexie in het middelbaar onderwijs, is tijds- en energie-intensief voor leerkrachten en scholieren. Automatische en adaptieve tekstvereenvoudiging biedt hier een baanbrekende oplossing om de werkdruk in het middelbaar onderwijs te verminderen. Het doel van dit onderzoek is om te achterhalen met welke technologische en logopedische aspecten AI-ontwikkelaars rekening moeten houden bij de ontwikkeling van een adaptieve AI-toepassing voor geautomatiseerde tekstvereenvoudiging. De volgende onderzoeksvraag is opgesteld: "Hoe kan een wetenschappelijke artikel automatisch vereenvoudigd worden, gericht op de verschillende behoeften van scholieren met dyslexie in de derde graad middelbaar onderwijs?". Het doel wordt bereikt door een antwoord op de volgende deelvragen te formuleren. Eerst geeft de literatuurstudie een antwoord op de eerste vier deelvragen. Daarna vormt het veldonderzoek een antwoord op de vijfde deelvraag. Ten slotte beantwoordt de vergelijkende studie de zesde en laatste deelvraag. De resultaten van dit onderzoek zetten AI-ontwikkelaars aan om een toepassing te maken om scholieren met dyslexie te kunnen ondersteunen in de derde graad middelbaar onderwijs.

1. Welke aanpakken zijn er voor geautomatiseerde tekstvereenvoudiging? Aansluitende vraag: "Hoe worden teksten handmatig vereenvoudigd voor scholieren met dyslexie?"
2. Welke specifieke noden hebben scholieren van de derde graad middelbaar onderwijs bij het begrijpen van complexere teksten?
3. Wat zijn de specifieke kenmerken van wetenschappelijke artikelen?
4. Met welke valkuilen bij taalverwerking met AI moeten ontwikkelaars rekening houden?
5. Welke toepassingen, tools en modellen zijn er beschikbaar om Nederlandstalige geautomatiseerde tekstvereenvoudiging met AI mogelijk te maken?
6. Welke functies ontbreken AI-toepassingen om geautomatiseerde én adaptieve tekstvereenvoudiging mogelijk te maken voor scholieren met dyslexie in de derde graad middelbaar onderwijs? Aansluitende vraag: "Welke manuele methoden voor tekstvereenvoudiging komen niet in deze tools voor?"

<sup>1</sup><https://www.vlaanderen.be/publicaties/stem-agenda-2030-stem-competenties-voor-een-toekomst-en-missiegericht-beleid>

## A.2. State-of-the-art

### A.2.1. Tekstvereenvoudiging

De voorbije tien jaar is artificiële intelligentie (AI) sterk verder ontwikkeld. Vasista (2022) benadrukt dat de toename in kennis voor nieuwe toepassingen zorgde. Tekstvereenvoudiging vloeide hier uit voort. Momenteel bestaan er al robuuste toepassingen die teksten kunnen vereenvoudigen, zoals Resoomer<sup>2</sup>, Paraphraser<sup>3</sup> en Prepostseo<sup>4</sup>. Binnen het kader van tekstvereenvoudiging is er bestaande documentatie beschikbaar waar onderzoekers het voordeel van toegankelijkheid aanhalen, maar volgens Gooding (2022) ontbreken deze toepassingen de extra noden die scholieren met dyslexie in de derde graad middelbaar onderwijs vereisen.

Shardlow (2014) haalt aan dat het algemene doel van tekstvereenvoudiging is om ingewikkelde bronnen toegankelijker te maken. Het zorgt voor verkorte teksten zonder de kernboodschap te verliezen. Siddharthan (2014) haalt verder aan dat tekstvereenvoudiging op één van drie manieren gebeurt. Er is conceptuele vereenvoudiging waarbij documenten naar een compacter formaat worden getransformeerd. Daarnaast is er uitgebreide modificatie die kernwoorden aanduidt door gebruik van redundantie. Als laatste is er samenvatting die documenten verandert in kortere teksten met alleen de topische zinnen. Met deze concepten zijn ontwikkelaars volgens Siddharthan (2014) in staat om ingewikkelde woorden te vervangen door eenvoudigere synoniemen of zinnen te verkorten zodat ze sneller leesbaar zijn.

Tekstvereenvoudiging behoort tot de zijtak van *Natural Language Processing* (NLP) in AI. NLP omvat methodes om menselijke teksten om te zetten in tekst voor machines. Documenten vereenvoudigen met NLP kan volgens Chowdhary (2020) op twee manieren: extract of abstract. Bij extractieve vereenvoudiging worden zinnen gelezen zoals ze zijn neergeschreven. Vervolgens bewaart een document de belangrijkste taalelementen om de tekst te kunnen hervormen. Deze vorm van tekstvereenvoudiging komt volgens (Sciforce, 2020) het meeste voor. Daarnaast is er abstracte vereenvoudiging waarbij de kernboodschap wordt bewaard. Met de kernboodschap wordt er een nieuwe zin opgebouwd. Volgens het onderzoek van Chowdhary (2020) heeft deze vorm potentieel, maar het zit nog in de kinderschoenen.

### A.2.2. Noden van scholieren met dyslexie

Het experiment van Franse wetenschappers

Gala en Ziegler (2016) illustreert dat manuele tekstvereenvoudiging schoolteksten toegankelijker

maakt voor kinderen met dyslexie. Dit deden ze door simpelere synoniemen en

<sup>2</sup><https://resoomer.com/nl/>

<sup>3</sup><https://www.paraphraser.io/nl/tekst-samenvatting>

<sup>4</sup><https://www.prepostseo.com/tool/nl/text-summarizer>

zinsstructuren te gebruiken. Tien kinderen werden opgenomen in het experiment, variërend van 8 tot 12 jaar oud. Verwijswoorden werden vermeden en woorden kort gehouden. De resultaten waren veelbelovend. Het leestempo lag hoger en de kinderen maakten minder leesfouten. Ook bleek er geen verlies van begrip in de tekst bij geteste kinderen. Resultaten van de studie werden gebundeld voor de mogelijke ontwikkeling van een AI-tool.

De visuele weergave van tekst beïnvloedt de leessnelheid bij scholieren met dyslexie. Zo haalt het onderzoek van Rello, Kanvinde e.a. (2012) tips aan waarmee teksten en documenten rekening moeten houden bij scholieren met dyslexie in de derde graad middelbaar onderwijs. Het gaat over speciale lettertypes, spreiding tussen woorden en het gebruik van inzoomen op aparte zinnen. Het onderzoek haalt verder aan dat teksten voor deze unieke noden aanpassen tijdrovend is, dus tekstvereenvoudiging door AI kan een revolutionaire oplossing bieden. De Universiteit van Kopenhagen is met bovenstaande idee aan de slag gegaan. Onderzoekers Bingel e.a. (2018) hebben gratis software ontwikkeld, genaamd Hero<sup>5</sup>, om tekstvereenvoudiging voor scholieren in het middelbaar onderwijs met dyslexie te automatiseren. De software bestudeert met welke woorden de gebruiker moeite heeft, en vervangt die door simpelere alternatieven. Hero bevindt zich nu in beta-vorm en wordt enkel in het Engels en Deens ondersteund. Als alternatief is er Readable<sup>6</sup>. Dit is een Engelstalige AI-toepassing dat zinnen beoordeeld met leesbaarheidsformules.

Roldós (2020) haalt aan dat NLP in de laatste decennia volop in ontwikkeling is, maar ontwikkelaars botsen nog op uitdagingen. Het gaat om zowel interpretatie- als dataproblemen bij AI-modellen. Het onderzoek haalt twee punten aan. Allereerst is het voor een machine moeilijk om de context van homoniemen te achterhalen. Bijvoorbeeld bij het woord 'bank' is het niet duidelijk voor de machine of het gaat over de geldinstelling of het meubel. Daarnaast zijn synoniemen een probleem voor tekstverwerking.

Het onderzoek van Sciforce (2020) haalt aan dat het merendeel van NLP-toepassingen Engelstalige invoer gebruikt. Niet-Engelstalige toepassingen zijn zeldzaam. De opkomst van AI technologieën die twee datasets gebruiken, biedt een oplossing voor dit probleem. De software vertaalt eerst de oorspronkelijke tekst naar de gewenste taal, voordat de tekst wordt herwerkt. Hetzelfde onderzoek bewijst dat het vertalen van gelijkaardige talen, zoals Duits en Nederlands, een minimaal verschil opleverd. Volgens Plavén-Sigray e.a. (2017) houden onderzoekers zich vaak in hun eigen taalbubbel, wat negatieve gevolgen heeft voor de leesbaarheid van een wetenschappelijk artikel. Bovendien vormt de stijgende trend van het gebruik aan acroniemen Barnett en Doubleday (2020) een extra hindernis. Donato e.a. (2022) haalt aan dat onbegrijpelijke literatuur, waaronder studiemateriaal geschreven door de docent

---

<sup>5</sup><https://beta.heroapp.ai/>

<sup>6</sup><https://readable.com/>

en online wetenschappelijke artikelen, één van de redenen is waarom scholieren met dyslexie in het middelbaar onderwijs van richting veranderen.

### A.2.3. Huidige toepassingen

Vlaanderen heeft weinig zicht op de geïmplementeerde AI software in scholen. Dit werd vastgesteld door (Martens e.a., 2021a), een samenwerking tussen de Vlaamse universiteiten en overheid voor AI. Vergeleken met andere Europese landen, maakt België het minst gebruik van leerling-georiënteerde hulpmiddelen. Degenen die wel gebruikt worden, zijn vooral online leerplatformen voor zelfstandig werken. Ook maakt België amper gebruik van beschikbare software die de leermethoden en -noden van leerlingen evalueert (Martens e.a., 2021b).

Verhoeven (2023) haalt aan dat AI-toepassingen zoals ChatGPT, Google Bard en Bing AI kunnen helpen om routinematig werk te verminderen in het onderwijs. Echter haalt Deckmyn (2021) aan dat GPT-3, het model van ChatGPT, sterker staat in het maken van Engelstalige teksten vergeleken met Nederlandstalige teksten. De databank waar het GPT-3 model mee is getraind, bestaat uit 92% Engelstalige woorden, terwijl er 0,35% Nederlandse woorden aanwezig zijn in dezelfde databank. Ontwikkelaars moeten de evolutie van deze modellen opvolgen, voordat er Nederlandstalige toepassingen mee worden gemaakt.

### A.2.4. Ontwikkelen met AI

Python staat bovenaan de lijst van programmeertalen voor NLP-toepassingen. Volgens het onderzoek van Thangarajah (2019) is dit te wijten aan de eenvoudige syntax, kleine leercurve en grote beschikbaarheid van kant-en-klare bibliotheken. Wetenschappelijke berekeningen of statistische analyses kunnen worden uitgevoerd met één lijn code. Malik (2022) haalt de twee meest voorkomende aan, namelijk NLTK<sup>7</sup> en Spacy<sup>8</sup>. *Deep Martin*<sup>9</sup> bouwt verder op het onderzoek van Shardlow (2014) naar een pipeline voor lexicale vereenvoudiging. *Deep Martin* maakt gebruik van *custom transformers* om invoertekst te converteren naar een vereenvoudigde versie van de tekstinhoud.

Voor Germaanse talen zijn er enkele datasets en word embeddings beschikbaar die de complexiteit van woorden bijhouden. Zo zijn er in de Duitse taal Klexikon<sup>10</sup> en TextComplexityDE<sup>11</sup>. Een onderzoek van Suter e.a. (2016) bouwde een rule-based NLP-model met 'Leichte Sprache', wat een dataset is met eenvoudige Duitstalige zinsconstructies. Nederlandstalige datasets zijn in schaarse hoeveelheden beschikbaar, dus het vertalen uit een Germaanse taal is hier een optie.

Volgens Garbacea e.a. (2021) is het belangrijk dat AI-ontwikkelaars niet alleen aan-

---

<sup>7</sup><https://www.nltk.org/>

<sup>8</sup><https://spacy.io/>

<sup>9</sup><https://github.com/chrislemke/deep-martin>

<sup>10</sup><https://github.com/dennlinger/klexikon>

<sup>11</sup><https://github.com/babaknaderi/TextComplexityDE>

dacht besteden aan het aanpassen van woorden en zinnen, maar ook aan de gebruiker meegeven waarom iets is aangepast. De onderzoekers wijzen op twee ethische aspecten. Eerst moet de toepassing duidelijk aangeven waarom een woord of zin is aangepast. Het model moet de moeilijkheidsgraad van de woorden of zinnen bewijzen. Iavarone e.a. (2021) beschrijft een methode met regressiemodellen om de moeilijkheidsgraad te bepalen door een gemiddeld moeilijkheidspercentage per zin te berekenen. Daarnaast benadrukt Garbacea e.a. (2021) het belang van het markeren van de complexere delen van een tekst. Hiervoor haalt hetzelfde onderzoek methoden aan zoals *lexical* of *deep learning*.

Er is een tactvolle aanpak nodig om een vereenvoudigde tekst met AI te beoordelen. De studie van Swayamdipta (2019) haalt aan dat er extra nood is aan NLP-modellen waarbij de tekst zijn kernboodschap behoudt. Samen met Microsoft Research bouwden ze NLP-modellen die gericht waren op de bewaring van zinsstructuur en -context door *scaffolded learning*. Hiervoor maakten de onderzoekers gebruik van een voorspellingsmethode die de positie van woorden en zinnen in een document beoordeelde. De Flesch-Kincaid leesbaarheidstest is volgens Readable (2021) een alternatieve manier om vereenvoudigde tekstinhoud te beoordelen, zonder de nood aan *pre-trained* modellen. Deze score kan eenvoudig worden berekend met de *Python-library textstat*<sup>12</sup>.

### A.3. Methodologie

Een *mixed-methods* onderzoek toont aan hoe toepassingen automatisch een wetenschappelijke artikel kunnen vereenvoudigen, gericht op scholieren met dyslexie in de derde graad middelbaar onderwijs. Het onderzoek houdt vijf grote fases in. De eerste fase is het proces van geautomatiseerde tekstvereenvoudiging beschrijven. Dit gebeurt via een grondige studie van vakliteratuur en wetenschappelijke teksten. Ook blogs van experts komen hier aan bod. Na het verwerven van de nodige inzichten wordt er een verklarende tekst opgesteld.

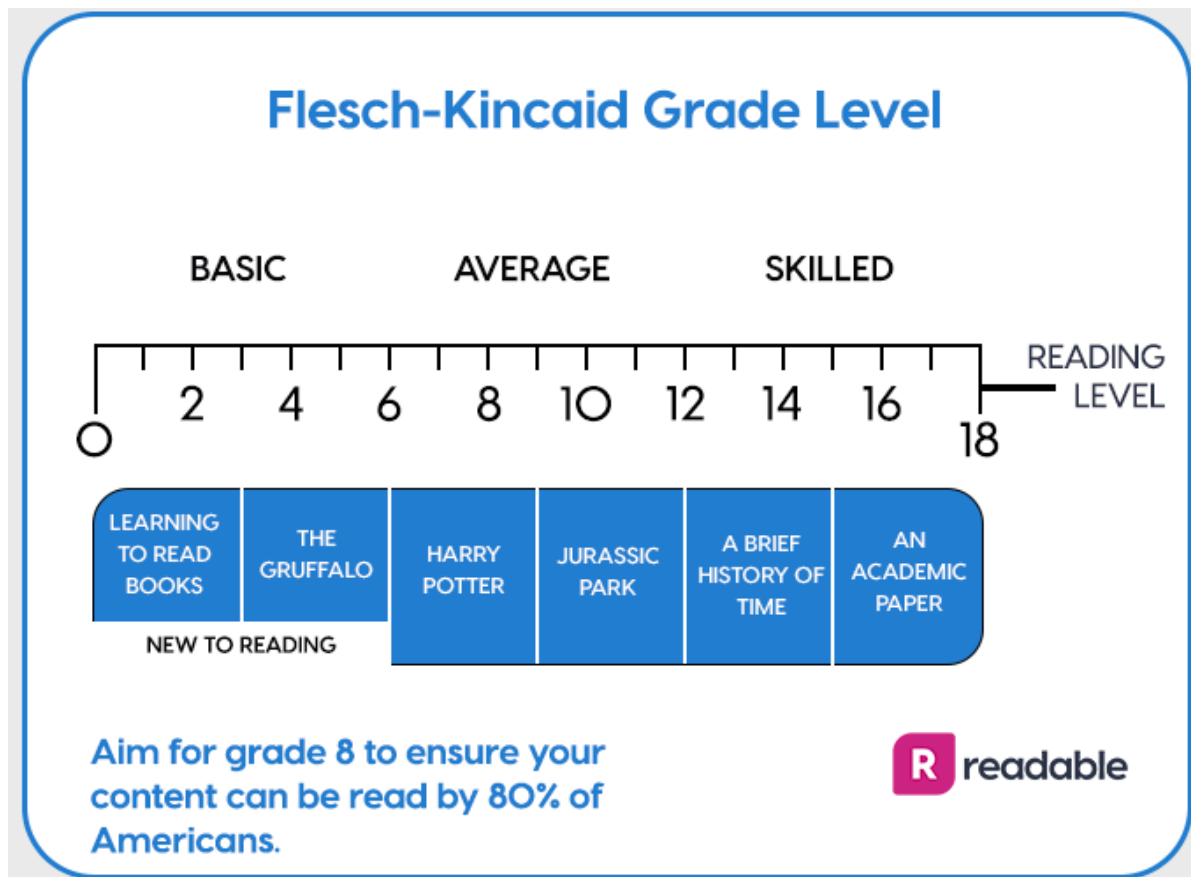
De tweede fase bestaat uit het analyseren van wetenschappelijke werken over de bewezen voordelen van tekstvereenvoudiging bij scholieren met dyslexie van de derde graad middelbaar onderwijs. Hiervoor zijn geringe thesissen beschikbaar, die zorgvuldigheid vragen tijdens interpretatie. De resulterende tekst bevat de voordelen samen met hun wetenschappelijke onderbouwing.

De derde fase is opnieuw een beschrijving. Hier worden de valkuilen bij taalverwerking met AI-software nagegaan. Deze fase van het onderzoek brengt mogelijke nadelen en tekortkomingen van AI-software bij tekstvereenvoudiging aan het licht. Dit gebeurt aan de hand van een technische uitleg.

De vierde fase omvat een toelichting over beschikbare AI toepassingen voor tekstvereenvoudiging. Aan de hand van een veldonderzoek op het internet en bij bedrijven wordt een longlist opgesteld van beschikbare toepassingen voor tekstver-

---

<sup>12</sup><https://pypi.org/project/textstat/>



**Figuur (A.1)**  
(Readable, 2021)

eenvoudiging in het middelbaar onderwijs. Met een requirementsanalyse wordt er een shortlist opgesteld van software. Het toetsen van verschillende tools wordt ook betrokken in deze fase. De shortlist vormt de basis voor de ontwikkeling van een prototype voor geautomatiseerde en adaptieve tekstvereenvoudiging.

De vijfde en laatste fase van het onderzoek bestaat uit het testen en beoordelen van gekozen AI-toepassingen voor tekstvereenvoudiging. In dit experiment proberen scholieren met dyslexie in de derde graad middelbaar onderwijs de shortlisted AI toepassingen en het prototype uit. Het doel van het experiment is om de effectiviteit en gebruikersvriendelijkheid van deze toepassingen te beoordelen. Na een grondige analyse wordt er met de resultaten bepaalt of de toepassingen aan de unieke noden van een scholier met dyslexie in de derde graad middelbaar onderwijs voldoen om wetenschappelijke artikelen te vereenvoudigen voor scholieren in het middelbaar onderwijs.

#### **A.4. Verwacht resultaat, conclusie**

Er wordt verwacht dat de huidige softwareoplossingen voor tekstvereenvoudiging onvoldoende aansluiten bij de noden van scholieren met dyslexie in de derde graad middelbaar onderwijs. Het prototype is moeilijk af te stemmen op de specifieke noden van deze doelgroep. Ontwikkelaars die werken met bestaande modellen moeten *custom transformers* inzetten om bevredigende resultaten te krijgen. Bovendien ontbreken er Nederlandstalige word embeddings die de complexiteit van elk woord bijhouden en aan kant-en-klare modellen die de inhoud van wetenschappelijke artikelen kunnen vereenvoudigen. Word embeddings uit een Germaanse taal gebruiken, gevolgd door vertaling naar het Nederlands is wel een aanvaardbaar alternatief.



# Bibliografie

- Abdel-Salam, S. & Rafea, A. (2022). Performance Study on Extractive Text Summarization Using BERT Models. *Information*, (2).
- Althunayyan, S. & Azmi, A. (2021). Automated Text Simplification: A Survey. *ACM Computing Surveys*, 54, Article no. 43. <https://doi.org/10.1145/3442695>
- Anthony, C. (2020). Website design trends for dyslexia and ADHD. <https://www.verndale.com/insights/accessibility/website-design-trends-for-dyslexia-and-adhd>
- Association, A. P. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (Deel 5). American psychiatric association Washington, DC.
- Ball, P. (2017). It's not just you: science papers are getting harder to read. *Nature*.
- Barnett, A. & Doubleday, Z. (2020). Meta-Research: The growth of acronyms in the scientific literature (P. Rodgers, Red.). *eLife*, 9, e60080.
- Bezem, A. & Lugthart, M. (2016). Visuele Disfunctie een onzichtbare belemmering bij lezen, spelling en concentratie. <https://beeldenbrein.nl/>
- Bilici, Ş. (2021). Sequence labeling.
- Bingel, J., Paetzold, G. & Søgaaard, A. (2018). Lexi: A tool for adaptive, personalized text simplification. *Proceedings of the 27th International Conference on Computational Linguistics*, 245–258.
- Binz, M. & Schulz, E. (2023). Using cognitive psychology to understand GPT-3. *Proceedings of the National Academy of Sciences*, 120(6).
- Bonte, M. (2020). *Bestaat Dyslexie?: En is het een relevante vraag?* uitgeverij SWP.
- Bosmans, A., Croon, S. & Verreycken, V. (2022a). Woordgebruik - Moeilijke constructies. <https://www.vlaanderen.be/taaladvies/taaladviezen/teksten-schrijven/formulering/zinsbouw-moeilijke-constructies>
- Bosmans, A., Croon, S. & Verreycken, V. (2022b). Woordgebruik - Moeilijke Woorden. <https://www.vlaanderen.be/taaladvies/taaladviezen/teksten-schrijven/formulering/woordgebruik-moeilijke-woorden>
- Bosmans, A., Croon, S. & Verreycken, V. (2022c). Woordgebruik - Synoniemen. <https://www.vlaanderen.be/taaladvies/taaladviezen/teksten-schrijven/formulering/woordgebruik-synoniemen>
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). Language Models are Few-Shot Learners.

- Bulté, B., Sevens, L. & Vandeghinste, V. (2018). Automating lexical simplification in Dutch. *Computational Linguistics in the Netherlands Journal*, 8, 24–48. <https://clinjournal.org/clinj/article/view/78>
- Canning, Y., Tait, J., Archibald, J. & Crawley, R. (2000). Cohesive Generation of Syntactically Simplified Newspaper Text. In P. Sojka, I. Kopeček & K. Pala (Red.), *Text, Speech and Dialogue* (pp. 145–150). Springer Berlin Heidelberg.
- Cao, M. (2022). A Survey on Neural Abstractive Summarization Methods and Factual Consistency of Summarization.
- Carbonell, J. & Goldstein, J. (1998). The use of MMR, diversity-based reranking for reordering documents and producing summaries. *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, 335–336.
- Case, A. (2008). More ways of simplifying your classroom language. <https://www.tefl.net/elt/articles/teacher/how-to-simplify-your-classroom-language/>
- Chauhan, K. (2018). Unsupervised text summarization using sentence embeddings. <https://medium.com/jatana/unsupervised-text-summarization-using-sentence-embeddings-adb15ce83db1>
- Chiusano, F. (2022). Two minutes NLP-learn the Rouge metric by examples. <https://medium.com/nlplanet/two-minutes-nlp-learn-the-rouge-metric-by-examples-fl79cc285499>
- Chowdhary, K. (2020). *Fundamentals of Artificial Intelligence*. Springer, New Delhi.
- Coster, W. & Kauchak, D. (2011). Learning to Simplify Sentences Using Wikipedia. *Proceedings of the Workshop on Monolingual Text-To-Text Generation*, 1–9. <https://aclanthology.org/W11-1601>
- Crevits, H. (2022, maart 13). *Kwart van bedrijven gebruikt artificiële intelligentie: Vlaanderen bij beste leerlingen van de klas* (Persbericht). Vlaamse Overheid Departement Economie, Wetenschap en Innovatie.
- Crossley, S. A., Allen, D. & McNamara, D. S. (2012). Text simplification and comprehensible input: A case for an intuitive approach. *Language Teaching Research*, 16(1), 89–108.
- Dandekar, N. (2016). How to use machine learning to find synonyms. <https://medium.com/@nikhilbd/how-to-use-machine-learning-to-find-synonyms-6380c0c6106b>
- Daoud, N. A. (2023). Vandaag Internationale Dag van de Moedertaal: Pleidooi voor Het Nederlands. <https://doorbraak.be/vandaag-internationale-dag-van-de-moedertaal-pleidooi-voor-het-nederlands/>
- Dapaah, J. & Maenhout, K. (2022, juli 8). *Iedereen heeft boter op zijn hoofd* (D. Standaard, Red.). [https://www.standaard.be/cnt/dmf20220607\\_97763592](https://www.standaard.be/cnt/dmf20220607_97763592)
- De Belder, M.-F., Jan; Moens. Text simplification for children. eng. In: ACM; New York, 2010.

- De Craemer, J., Van Beeumen, L., Cooreman, A., Moonen, A., Rottier, J., Wagemakers, I. & Mardulier, T. (2018). Aan de slag met voorleessoftware op school. Een gids met 8 vragen en antwoorden. <https://onderwijs.vlaanderen.be/nl/onderwijspersoneel/van-basis-tot-volwassenenonderwijs/lespraktijk/ict-in-de-klas/voorleessoftware-voor-leerlingen-met-leesbeperkingen/aan-de-slag-met-voorleessoftware-op-school>
- De Meyer, I., Janssens, R. & Warlop, N. (2019). Leesvaardigheid van 15- jarigen in Vlaanderen: Overzicht van de eerste resultaten van PISA2018. <https://data-onderwijs.vlaanderen.be/documenten/bestand.ashx?id=12265>
- Deckmyn, D. (2021, maart 19). *Robot schrijft mee De Standaard* (D. Standaard, Red.). [https://www.standaard.be/cnt/dmf20210319\\_05008561](https://www.standaard.be/cnt/dmf20210319_05008561)
- Desoete, A. (2017). Dyslexie of dyscalculie, niet de schuld van het onderwijs! En een correcte diagnose schaadt niet. *SPRANKEL*, 28(2), 17–31.
- Diels, L. (2022). Over dyslexie, vooroordelen en vriendschap. <https://www.maastrichtuniversity.nl/nl/nieuws/over-dyslexie-vooroordelen-en-vriendschap>
- Donato, A., Muscolo, M., Arias Romero, M., Capri, T., Calarese, T. & Olmedo Moreno, E. M. (2022). Students with dyslexia between school and university: Post-diploma choices and the reasons that determine them. An Italian study. *Dyslexia*, 28(1), 110–127.
- Dronberger, G. B. & Kowitz, G. T. (1975). Abstract readability as a factor in information systems. *Journal of the American Society for Information Science*, 26(2), 108–111.
- DuBay, W. H. (2004). The principles of readability. *Online Submission*.
- Eisenstein, J. (2019). *Introduction to Natural Language Processing*. MIT Press. <https://books.google.be/books?id=72yuDwAAQBAJ>
- F, M. & Ennals, R. (2010). True Enough: Learning to Live in a Post-Fact Society , John Wiley, New Jersey (2008). *International Journal of Information Management - INT J INFORM MANAGE*, 30, 180–180.
- Fabbri, A. R., Kryściński, W., McCann, B., Xiong, C., Socher, R. & Radev, D. (2020). SummEval: Re-evaluating Summarization Evaluation.
- Fardeen-, B. (2021). Complete Guide to Spacy Tokenizer with examples. <https://machinelearningknowledge.ai/complete-guide-to-spacy-tokenizer-with-examples/>
- Filipiak, P. (2020). Leesproblemen en Dyslexie: Leesproblemen -2-. [https://wij-leren.nl/leesproblemen-dyslexie-woordbenoeming-woordherkenning-begripsprobleem-deel-twee.php#\\_edn11](https://wij-leren.nl/leesproblemen-dyslexie-woordbenoeming-woordherkenning-begripsprobleem-deel-twee.php#_edn11)
- Gala, N. & Ziegler, J. (2016). Reducing lexical complexity as a tool to increase text accessibility for children with dyslexia. *Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity (CL4LC)*, 59–66.

- Ganesan, K. (2018). ROUGE 2.0: Updated and Improved Measures for Evaluation of Summarization Tasks.
- Garbacea, C., Guo, M., Carton, S. & Mei, Q. (2021). Explainable Prediction of Text Complexity: The Missing Preliminaries for Text Simplification. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 1086–1097. <https://doi.org/10.18653/v1/2021.acl-long.88>
- Garg, H. (2022). Using GPT-3 for education: Use cases. <https://indiaai.gov.in/article/using-gpt-3-for-education-use-cases>
- Ghesquière, P. (2018). *Als leren pijn doet: Kinderen met een leerstoornis opvoeden en begeleiden*. Acco.
- Gooding, S. (2022). On the Ethical Considerations of Text Simplification. *Ninth Workshop on Speech and Language Processing for Assistive Technologies (SLPAT-2022)*, 50–57. <https://doi.org/10.18653/v1/2022.slpac-1.7>
- Gooding, S. & Kochmar, E. (2019). Complex word identification as a sequence labelling task. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 1148–1153.
- Greg, B., Atty, E., Elie, G., Joane, J., Logan, K., Lim, R., Luke, M. & Michelle, P. (2023). Introducing chatgpt and Whisper Apis. <https://openai.com/blog/introducing-chatgpt-and-whisper-apis>
- Hahn, U. & Mani, I. (2000). The Challenges of Automatic Summarization. *Computer*, 33, 29–36. <https://doi.org/10.1109/2.881692>
- Hale, A. (2022). Bullet points: What, why, and how to use them. <https://www.contentandcommas.com/bullet-points/>
- Hartley, J. (1999). From Structured Abstracts to Structured Articles: A Modest Proposal. *Journal of Technical Writing and Communication*, 29(3), 255–270. <https://doi.org/10.2190/3RWW-A579-HC8W-6866>
- Harwell, D. (2023). Tech's hottest new job: Ai whisperer. no coding required. <https://www.washingtonpost.com/technology/2023/02/25/prompt-engineers-techs-next-big-job/>
- Hayes, D. P. (1992). The growing inaccessibility of science. <https://www.nature.com/articles/356739a0>
- Hern, A. (2023). TechScape: Will meta's massive leak democratise AI – and at what cost? <https://www.theguardian.com/technology/2023/mar/07/techscape-meta-leak-llama-chatgpt-ai-crossroads>
- Hollenkamp, J. (2020). Summary and analysis of Scientific Research Articles - San Jose State ... <https://www.sjsu.edu/writingcenter/docs/handouts/Summary%20and%20Analysis%20of%20Scientific%20Research%20Articles.pdf>

- Hsu, W.-T., Lin, C.-K., Lee, M.-Y., Min, K., Tang, J. & Sun, M. (2018). A Unified Model for Extractive and Abstractive Summarization using Inconsistency Loss.
- Huang, S., Wang, R., Xie, Q., Li, L. & Liu, Y. (2019). An Extraction-Abstraction Hybrid Approach for Long Document Summarization. *2019 6th International Conference on Behavioral, Economic and Socio-Cultural Computing (BESC)*, 1–6.
- Hubbard, K. E. & Dunbar, S. D. (2017). Perceptions of scientific research literature and strategies for reading papers depend on academic career stage. *PLOS ONE*, 12(12), 1–16.
- Iavarone, B., Brunato, D. & Dell'Orletta, F. (2021). Sentence Complexity in Context. *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*, 186–199. <https://doi.org/10.18653/v1/2021.cmcl-1.23>
- IBM. (2022). IBM Global AI Adoption Index 2022. <https://www.ibm.com/downloads/cas/GVAGA3JP>
- Iredale, G. (2022). An overview of tokenization algorithms in NLP. <https://101blockchains.com/tokenization-nlp/>
- Iskender, N., Polzehl, T. & Möller, S. (2021). Reliability of Human Evaluation for Text Summarization: Lessons Learned and Challenges Ahead. *Proceedings of the Workshop on Human Evaluation of NLP Systems (HumEval)*, 86–96. <https://aclanthology.org/2021.humeval-1.10>
- Javourey-Drevet, L., Dufau, S., François, T., Gala, N., Ginestié, J. & Ziegler, J. C. (2022). Simplification of literary and scientific texts to improve reading fluency and comprehension in beginning readers of French. *Applied Psycholinguistics*, 43(2), 485–512. <https://doi.org/10.1017/S014271642100062X>
- Jiang, R. K. (2023). Prompt engineering : Deconstructing and managing intention. <https://www.linkedin.com/pulse/prompt-engineering-deconstructing-managing-intention-jiang/>
- Jones, R., Colusso, L., Reinecke, K. & Hsieh, G. (2019). r/science: Challenges and Opportunities in Online Science Communication. *CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–14. <https://doi.org/10.1145/3290605.3300383>
- Jurafsky, D., Martin, J., Norvig, P. & Russell, S. (2014). *Speech and Language Processing*. Pearson Education. <https://books.google.be/books?id=Cq2gBwAAQBAJ>
- Kandula, S., Curtis, D. & Zeng-Treitler, Q. (2010). A semantic and syntactic text simplification tool for health content. *AMIA annual symposium proceedings, 2010*, 366.
- Khan, A. (2014). A Review on Abstractive Summarization Methods. *Journal of Theoretical and Applied Information Technology*, 59, 64–72.

- Khurana, D., Koli, A., Khatter, K. & Singh, S. (2022). Natural Language Processing: State of The Art, Current Trends and Challenges. *Multimedia Tools and Applications*, 82, 25–27.
- Kleijnen, R. e., Bosman, A., De Jong, P., Henneman, K., Pasman, J., Paternotte, A., Ruijsenaars, A., Struiksma, A., Van den Bos, K., Van der Leij, A. e.a. (2008). Dyslexie: Diagnose en behandeling van dyslexie.
- Li, C. (2022). OpenAI's GPT-3 language model: A technical overview. <https://lambdalabs.com/blog/demystifying-gpt-3>
- Li, J., Sun, A., Han, J. & Li, C. (2018). A Survey on Deep Learning for Named Entity Recognition.
- Lin, C.-Y. (2004). ROUGE: A Package for Automatic Evaluation of Summaries. *Text Summarization Branches Out*, 74–81. <https://aclanthology.org/W04-1013>
- Lin, H. & Bilmes, J. (2010). Multi-document summarization via budgeted maximization of submodular functions. *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 912–920.
- Linderholm, T., Everson, M. G., van den Broek, P., Mischinski, M., Crittenden, A. & Samuels, J. (2000). Effects of Causal Text Revisions on More- and Less-Skilled Readers' Comprehension of Easy and Difficult Texts. *Cognition and Instruction*, 18(4), 525–556.
- Lisowski, E. (2023). GPT-3 vs. CHATGPT. the key differences. <https://addepto.com/blog/gpt-3-vs-chat-gpt-the-key-differences/>
- Lissens, F., Asmar, M., Willems, D., Van Damme, J., De Coster, S., Demeestere, E., Maes, R., Baccarne, B., Robaeyst, B., Duthoo, W. & Desoete, A. (2020). Het stopt nooit...De impact van dyslexie en/of dyscalculie op het welbevinden en studeren van (jong)volwassenen en op de transitie naar de arbeidsmarkt: een bundeling van Vlaamse pilootstudies.
- Liu, Q., Kusner, M. J. & Blunsom, P. (2020). A Survey on Contextual Embeddings.
- Louwerse, M., McCarthy, P. & McNamara, D. (2007). A Linguistic Analysis of Simplified and Authentic Texts. *The Modern Language Journal*, 91, 15–30.
- Malik, R. S. (2022, juli 4). *Top 5 NLP Libraries To Use in Your Projects* (T. Al, Red.). <https://towardsai.net/p/l/top-5-nlp-libraries-to-use-in-your-projects>
- Martens, M., De Wolf, R. & Evens, T. (2021a). *Algoritmes en AI in de onderwijscontext: Een studie naar de perceptie, mening en houding van leerlingen en ouders in Vlaanderen*. Kenniscentrum Data en Maatschappij. Verkregen 30 maart 2022, van <https://data-en-maatschappij.ai/publicaties/survey-onderwijs-2021>
- Martens, M., De Wolf, R. & Evens, T. (2021b, juni 28). *School innovation forum 2021*. Kenniscentrum Data en Maatschappij. Verkregen 1 april 2022, van <https://data-en-maatschappij.ai/nieuws/school-innovation-forum-2021>



- McAuliffe, Z. (2023). DuckDuckGo introduces new AI feature: DuckAssist. <https://www.cnet.com/tech/services-and-software/meet-duckassist-duckduckgos-new-ai-feature/>
- McCombes, S. (2022). How to write A summary: Guide amp; examples. <https://www.scribbr.com/working-with-sources/how-to-summarize/>
- McDonald, R. (2007). A study of global inference algorithms in multi-document summarization. *Advances in Information Retrieval: 29th European Conference on IR Research, ECIR 2007, Rome, Italy, April 2-5, 2007. Proceedings* 29, 557–564.
- McFarland, A. (2023). What is prompt engineering in AI amp; Why It Matters. <https://www.unite.ai/what-is-prompt-engineering-in-ai-why-it-matters/>
- McKeown, K., Klavans, J. L., Hatzivassiloglou, V., Barzilay, R. & Eskin, E. (1999). Towards multidocument summarization by reformulation: Progress and prospects.
- McNutt, M. (2014). Reproducibility. *Science*, 343(6168), 229–229. <https://doi.org/10.1126/science.1250475>
- Menzli, A. (2023). Tokenization in NLP: Types, challenges, examples, tools. <https://neptune.ai/blog/tokenization-in-nlp>
- Minnen, J., Verbeylen, J. & Glorieux, I. (2018). De arbeidsduur en werkcontext van leraren. Onderzoek naar de tijdsbesteding van leraren uit het basis- en secundair onderwijs. <https://torvub.be/wp-content/uploads/2018/09/Samenvatting.pdf>
- Miszczak, P. (2023). Prompt engineering: The ultimate guide 2023 [GPT-3 amp; chatgpt]. <https://businessolution.org/prompt-engineering/>
- Mottes, C. (2023). GPT-3 vs. Bert: Comparing the two most popular language models. <https://blog.invgate.com/gpt-3-vs-bert>
- Murdos, A., Hodges, M., Rubio, D. & Adams, J. S. (2014). Examining the readability of research abstracts to Determine Whether the General Public Can Understand Key Findings in Science. [https://ctsi.ucla.edu/education/files/view/docs/06\\_08\\_2016\\_PALM\\_Communication\\_of\\_Science\\_Poster.pdf](https://ctsi.ucla.edu/education/files/view/docs/06_08_2016_PALM_Communication_of_Science_Poster.pdf)
- Nallapati, R., Zhai, F. & Zhou, B. (2017). SummaRuNNer: A Recurrent Neural Network Based Sequence Model for Extractive Summarization of Documents. *Proceedings of the AAAI Conference on Artificial Intelligence*, 31(1). <https://doi.org/10.1609/aaai.v31i1.10958>
- Nandhini, K. & Balasundaram, S. (2013). Improving readability through extractive summarization for learners with reading difficulties. *Egyptian Informatics Journal*, 14(3), 195–204.
- Nenkova, A. & Passonneau, R. (2004). Evaluating Content Selection in Summarization: The Pyramid Method. *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics: HLT-NAACL 2004*, 145–152.

- Niemeijer, A., Frederiks, B., Riphagen, I., Legemaate, J., Eefsting, J. & Hertogh, C. (2010). Ethical and practical concerns of surveillance technologies in residential care for people with dementia or intellectual disabilities: an overview of the literature. *Psychogeriatrics*, 22(7), 1129–1142. <https://doi.org/10.1017/S1041610210000037>
- OnderwijsVlaanderen. (2023). Voorleessoftware voor Leerlingen met Leesbeperkingen. <https://onderwijs.vlaanderen.be/voorleessoftware-voor-leerlingen-met-leesbeperkingen>
- Paetzold, G. & Specia, L. (2016). SemEval 2016 Task 11: Complex Word Identification. *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, 560–569. <https://doi.org/10.18653/v1/S16-1085>
- Pain, E. (2016). How to (seriously) read a scientific paper. <https://www.science.org/content/article/how-seriously-read-scientific-paper>
- Parveen, D. & Strube, M. (2015). Integrating importance, non-redundancy and coherence in graph-based extractive summarization. *Twenty-Fourth International Joint Conference on Artificial Intelligence*.
- Plavén-Sigray, P., Matheson, G. J., Schiffler, B. C. & Thompson, W. H. (2017). Research: The readability of scientific texts is decreasing over time (S. King, Red.). *eLife*, 6, e27725.
- Poel, M., Boschman, E. & op den Akker, R. (2008). A Neural Network Based Dutch Part of Speech Tagger [<http://eprints.ewi.utwente.nl/14662>; 20th Benelux Conference on Artificial Intelligence, BNAIC 2008, BNAIC ; Conference date: 30-10-2008 Through 31-10-2008]. In A. Nijholt, M. Pantic, M. Poel & H. Hondorp (Red.), *BNAIC 2008* (pp. 217–224). Twente University Press (TUP).
- Premjith, P., John, A. & Wilscy, M. (2015). Metaheuristic Optimization Using Sentence Level Semantics for Extractive Document Summarization, 347–358. [https://doi.org/10.1007/978-3-319-26832-3\\_33](https://doi.org/10.1007/978-3-319-26832-3_33)
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I. e.a. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, 1(8), 9.
- Raj, D. (2017). Metrics for NLG evaluation. <https://medium.com/explorations-in-language-and-learning/metrics-for-nlg-evaluation-c89b6a781054>
- Rani, R. & Kaur, B. (2021). The TEXT SUMMARIZATION AND ITS EVALUATION TECHNIQUE. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12(1), 745–752.
- Readable. (2021). *Flesch Reading Ease and the Flesch Kincaid Grade Level*. <https://readable.com/readability/flesch-reading-ease-flesch-kincaid-grade-level/>
- Rello, L. & Baeza-Yates, R. (2013). Good fonts for dyslexia. *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS 2013*.



- Rello, L., Baeza-Yates, R., Dempere-Marco, L. & Saggion, H. (2013). Frequent Words Improve Readability and Short Words Improve Understandability for People with Dyslexia.
- Rello, L., Baeza-Yates, R. & Saggion, H. (2013). The Impact of Lexical Simplification by Verbal Paraphrases for People with and without Dyslexia. *7817*, 501–512.
- Rello, L. & Baeza-Yates, R. A. (2015). How to present more readable text for people with dyslexia. *Universal Access in the Information Society*, *16*, 29–49.
- Rello, L. & Bigham, J. (2017). Good Background Colors for Readers: A Study of People with and without Dyslexia, 72–80.
- Rello, L., Kanvinde, G. & Baeza-Yates, R. (2012). Layout Guidelines for Web Text and a Web Service to Improve Accessibility for Dyslexics. *Proceedings of the International Cross-Disciplinary Conference on Web Accessibility*.
- Rello, L., Saggion, H., Baeza-Yates, R. & Graells, E. (2012). Graphical schemes may improve readability but not understandability for people with dyslexia, 25–32.
- Ribas, J. (2023). Building the new bing. <https://www.linkedin.com/pulse/building-new-bing-jordi-ribas/>
- Ribeiro, E., Ribeiro, R. & de Matos, D. M. (2018). A Study on Dialog Act Recognition using Character-Level Tokenization.
- Rijkhoff, J. (2022). Tekst Inkorten?: 9 tips om Je Teksten korter Te Maken. <https://dialogtrainers.nl/tekst-inkorten-tips/>
- Rijnvis, D. (2020). Tangconstructies - 5 tips om ze Te Vermijden. <https://www.schrijfvis.nl/tangconstructies/>
- Rivero-Contreras, M., Engelhardt, P. E. & Saldaña, D. (2021). An experimental eye-tracking study of text adaptation for readers with dyslexia: effects of visual support and word frequency. *Annals of Dyslexia*, *71*, 170–187.
- Roldós, I. (2020, december 22). *Major Challenges of Natural Language Processing (NLP)*. MonkeyLearn. Verkregen 1 april 2022, van <https://monkeylearn.com/blog/natural-language-processing-challenges/>
- Roose, K. (2023). Don't ban chatgpt in schools. teach with it. <https://www.nytimes.com/2023/01/12/technology/chatgpt-schools-teachers.html>
- Ruben, A. (2016). How to read a scientific paper. <https://www.science.org/content/article/how-read-scientific-paper-rev2>
- Sciforce. (2020, februari 4). *Biggest Open Problems in Natural Language Processing*. Verkregen 1 april 2022, van <https://medium.com/sciforce/biggest-open-problems-in-natural-language-processing-7eb101ccfc9>
- ShafieiBavani, E., Ebrahimi, M., Wong, R. & Chen, F. (2018). A Graph-theoretic Summary Evaluation for ROUGE, 762–767. <https://doi.org/10.18653/v1/D18-1085>
- Shardlow, M. (2013). A Comparison of Techniques to Automatically Identify Complex Words. *51st Annual Meeting of the Association for Computational Lin-*

- guistics *Proceedings of the Student Research Workshop*, 103–109. <https://aclanthology.org/P13-3015>
- Shardlow, M. (2014). A Survey of Automated Text Simplification. *International Journal of Advanced Computer Science and Applications(IJACSA), Special Issue on Natural Language Processing 2014*, 4(1). <https://doi.org/10.14569/SpecialIssue.2014.040109>
- Siddharthan, A. (2006). Syntactic Simplification and Text Cohesion. *Research on Language and Computation*, 4(1), 77–109. <http://oro.open.ac.uk/58888/>
- Siddharthan, A. (2014). A survey of research on text simplification. *ITL - International Journal of Applied Linguistics*, 165, 259–298.
- Sikka, P. & Mago, V. (2020). A Survey on Text Simplification. *CoRR*, abs/2008.08612. <https://arxiv.org/abs/2008.08612>
- Simon, J. (2021). Large language models: A new moore's law? <https://huggingface.co/blog/large-language-models>
- Snow, C. (2010). Academic Language and the Challenge of Reading for Learning About Science. *Science (New York, N.Y.)*, 328, 450–2.
- Sohom, G., Ghosh; Dwight. (2019). *Natural Language Processing Fundamentals*. Packt Publishing. <https://medium.com/analytics-vidhya/natural-language-processing-basic-concepts-a3c7f50bf5d3>
- Steinberger, J. & Jezek, K. (2009). Evaluation Measures for Text Summarization. *Computing and Informatics*, 28, 251–275.
- Strubell, E., Ganesh, A. & McCallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP.
- Suleiman, D. & Awajan, A. (2020). Deep Learning Based Abstractive Text Summarization: Approaches, Datasets, Evaluation Measures, and Challenges. *Mathematical Problems in Engineering*, 2020.
- Surma, T., Vanhoyweghen, K., Sluijsmans, D., Camp, G., Muijs, D. & Kirschner, P. (2019). *Wijze lessen: twaalf bouwstenen voor effectieve didactiek* (1ste ed.). Ten Brink Uitgevers.
- Suter, J., Ebling, S. & Volk, M. (2016). Rule-based Automatic Text Simplification for German.
- Swayamdipta, S. (2019, januari 22). *Learning Challenges in Natural Language Processing*. Verkregen 1 april 2022, van <https://www.microsoft.com/en-us/research/video/learning-challenges-in-natural-language-processing/>
- Tanya Goyal, G. D., Junyi Jessy Li. (2022). News Summarization and Evaluation in the Era of GPT-3. *arXiv preprint*.
- Tatman, R. (2019). Evaluating text output in NLP: Bleu at your own risk. <https://towardsdatascience.com/evaluating-text-output-in-nlp-bleu-at-your-own-risk-e8609665a213>

- Thangarajah, V. (2019). Python current trend applications-an overview.
- Tops, W., Callens, M., Brysbaert, M. & Schouten, E. L. (2018). *Slagen met Dyslexie in Het Hoger Onderwijs*. Owl Press.
- Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E. & Lample, G. (2023). LLaMA: Open and Efficient Foundation Language Models.
- van der Meer, C. (2022). Dyslexie hebben is Niet Zo Raar: Lezen is iets heel onnatuurlijks. <https://www.demorgen.be/beter-leven/dyslexie-hebben-is-niet-zo-raar-lezen-is-iets-heel-onnatuurlijks~bc608101/>
- Van Vreckem, C. & Desoete, A. (2015). Het ene kind met dyslexie is het andere niet: implicaties uit begrijpend leesonderzoek en spellingonderzoek voor diagnostiek en therapie. *LOGOPEDIE*, 58–66.
- Vasista, K. (2022). Evolution of AI Design Models. *Central Asian Journal of Theoretical and Applied Science*, 3(3), 1–4.
- Vellutino, F., Fletcher, J., Snowling, M. & Scanlon, D. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of child psychology and psychiatry, and allied disciplines*, 45, 2–40.
- Verhoeven, W. (2023, februari 8). *Applaus voor de studenten die ChatGPT gebruiken* (Trends, Red.). [https://trends.knack.be/economie/bedrijven/applaus-voor-de-studenten-die-chatgpt-gebruiken/article-opinion-1934277.html?cookie\\_check=1676034368](https://trends.knack.be/economie/bedrijven/applaus-voor-de-studenten-die-chatgpt-gebruiken/article-opinion-1934277.html?cookie_check=1676034368)
- Verma, P. & Verma, A. (2020). A review on text summarization techniques. *Journal of scientific research*, 64(1), 251–257.
- Weinberg, G. (2023). Duckduckgo launches DuckAssist. <https://spreadprivacy.com/duckassist-launch/>
- Wentink, W., Verhoeven, L. & Druenen, M. (2008). Protocol leesproblemen en dyslexie voor groep 1 en 2. *Clinical Rheumatology - CLIN RHEUMATOL*.
- White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J. & Schmidt, D. C. (2023). A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT.
- Xu, W., Callison-Burch, C. & Napoles, C. (2015). Problems in current text simplification research: New data can help. *Transactions of the Association for Computational Linguistics*, 3, 283–297.
- Zeng, Q., Kim, E., Crowell, J. & Tse, T. (2005). A Text Corpora-Based Estimation of the Familiarity of Health Terminology. In J. L. "Oliveira, V. Maojo, F. Martín-Sánchez & A. S. Pereira (Red.), *Biological and Medical Data Analysis* (pp. 184–192). Springer Berlin Heidelberg.
- Zhang, J., Zhao, Y., Saleh, M. & Liu, P. (2020). Pegasus: Pre-training with extracted gap-sentences for abstractive summarization. *International Conference on Machine Learning*, 11328–11339.

Zhang, M., Riecke, L. & Bonte, M. (2021). Neurophysiological tracking of speech-structure learning in typical and dyslexic readers. *Neuropsychologia*, 158, 107889.