

HOE MEETBAAR IS LEESBAARHEID?

Orphée De Clercq & Véronique Hoste

In een maatschappij waar communicatie centraal staat en we dagelijks bestookt worden met tekstmateriaal allerhande speelt leesbaarheid een grote rol. Maar wat zorgt er nu precies voor dat een bepaalde tekst vlot of eerder moeizaam leest; wat is leesbaarheid? In het verleden verscheen de ene definitie na de andere, verschillend al naargelang van hun finaliteit. Was het de bedoeling om leesmateriaal voor kinderen te selecteren dan werden concepten bovengehaald zoals “de leesvaardigheid die nodig is om tot tekstbegrip te komen”, iets wat zeker van toepassing was bij vroegere werken rond leesbaarheidsformules (Dale & Chall, 1948; Gunning, 1952; Kincaid et al., 1975; Staphorsius, 1994). Maar leesbaarheid kan evenzeer gedefinieerd worden als “de tekstkarakteristieken die ervoor zorgen dat een lezer wil verder lezen” (McLaughlin, 1974). Een adequate definitie van leesbaarheid houdt dus zowel rekening met tekstinterne kenmerken als met de cognitieve vaardigheden van de lezer. In wat volgt, schetsen we eerst de evolutie van het klassieke leesbaarheidsonderzoek en de uitgebreide kritiek die daarop is gekomen. Vervolgens gaan we dieper in op het hedendaags complexer georiënteerd leesbaarheidsonderzoek en de specifieke bijdragen van het Language and Translation Technology Team (LT3, de afdeling Taaltechnologie aan de vakgroep Vertalen, Tolken en Communicatie).

Hoe kan leesbaarheid worden gemeten? In de vorige eeuw waren de automatische leesbaarheidsformules bijzonder succesvol. We kunnen daarbij bijvoorbeeld denken aan de Flesch Reading Ease (Flesch, 1948) of CLIB/CILT-formules, die aan de basis liggen van de bekende AVI-niveaus (nu vervangen door de E- en M-niveaus) en die worden gehanteerd in het lager onderwijs in Vlaanderen en Nederland. Deze leesbaarheidsformules werden vooral gebruikt om leesmateriaal te

selecteren, aangepast aan het leesvaardigheidsniveau van bepaalde lezers. Op basis van een aantal kenmerken van een tekst, zoals de gemiddelde zinslengte of het gemiddelde aantal lettergrepen per woord, geven ze een zogenaamd objectief oordeel van de leesbaarheid van een tekst. Een formule ziet er als volgt uit: $HELDERHEID = gewicht1 * kenmerk1 + gewicht2 * kenmerk2 + \dots$

De kritiek op deze klassieke aanpak was niet gering: zo geven de formules geen absolute score weer (Bailin & Grafstein, 2012), zijn ze enkel gebaseerd op oppervlakkige tekstkenmerken (DuBay, 2004, 2007; Feng et al., 2009; Kraf & Pander Maat, 2009) en suggereren ze dat er een onderliggende regressie bestaat tussen leesbaarheid en de gemodelleerde karakteristieken (Heilman et al., 2008). Bovendien heeft eigen onderzoek uitgewezen dat deze formules, hoe divers ze ook lijken, nauwelijks van elkaar verschillen, zelfs over de taalgrenzen heen (van Oosten et al., 2010). Naar aanleiding van dat onderzoek binnen de afdeling Taaltechnologie werd een demo ontwikkeld waar iedereen mee aan de slag kan[1]. Hij is zowel op Nederlandse als op Engelse teksten toepasbaar en bevat twaalf leesbaarheidsformules (zeven voor het Engels, vier voor het Nederlands en één voor het Zweeds). Ondanks de herhaalde kritiek op deze klassieke leesbaarheidsformules worden ze, bij gebrek aan beter, tot op heden nog altijd frequent gebruikt. Zo zitten de Flesch Reading Ease en Flesch Kincaid Grade Level formule standaard in elke MS Word-tekstverwerker.

De recente interesse van publieke en private organisaties voor het produceren van leesbare documenten en de gigantische vooruitgang die er in de laatste decennia geboekt is in domeinen zoals automatische tekstverwerking, heeft geleid tot een forse opleving van leesbaarheidsonderzoek. Dankzij deze technologische vooruitgang is het nu ook mogelijk om meer grip te krijgen op bepaalde kenmerken die niet tot het oppervlakteniveau van een tekst behoren maar wel de leesbaarheid beïnvloeden, zoals syntactische patronen en referentiële relaties. Dat was het

uitgangspunt van het door het Hogeschoolfonds gefinancierde HENDI-project.

Een van de voornaamste uitgangspunten binnen de taaltechnologie is dat computeralgoritmes gebruikt kunnen worden om karakteristieken van natuurlijke taal uit een tekstcorpus af te leiden. Het HENDI-project beperkt zich daarbij niet tot een bepaalde tekstsoort maar selecteert teksten uit verschillende genres. Deze teksten werden vervolgens beoordeeld op leesbaarheid. Via twee online applicaties hebben we daarvoor zowel een beroep kunnen doen op de mening van taalexperten[2] als op die van het grote publiek, 'de crowd'[3]. Aangezien de aanmaak van gelabelde data tijdsintensief en dus duur is, kwam binnen het domein van de taaltechnologie recent een tendens op gang om 'crowdsourcing' te gebruiken voor allerlei linguïstische annotatietaken (van Oosten & Hoste, 2011). Het uitgangsprincipe voor dit crowdsourcen is dat iedereen met internettoegang een mogelijke annotator is. Aangezien we binnen HENDI toch vooral wilden te weten komen hoe het grote publiek leesbaarheid ervaart, hebben we voor deze weg gekozen. De hoge correlatie tussen de datasets die we via de beide applicaties hebben verkregen, toont ook duidelijk aan dat leesbaarheid door de beide groepen, taalexperten én leken, op eenzelfde manier wordt ervaren (van Oosten et al., 2011; De Clercq et al., 2013).

Dankzij deze referentiedataset konden we op zoek gaan naar de eigenlijke tekstkenmerken die de perceptie van leesbaarheid sturen. In totaal hebben we vier grote groepen kunnen onderscheiden die sterk correleren met de leesbaarheid van teksten: naast de klassieke kenmerken uit het 'traditionele' leesbaarheidsonderzoek, ook, lexicale, syntactische en semantische kenmerken. Een adequate selectie en weging van deze tekstkenmerken heeft uiteindelijk geleid tot een nieuwe holistische leesbaarheidsvoorspelling. In de HENDI-metriek worden dus oppervlakkige kenmerken met complexere en dieperliggende tekstelementen gecombineerd om uiteindelijk tot een goed onderbouwd leesbaarheidsoordeel te komen.

In het Stylene project[4], gefinancierd door het Departement EWI, werd een interface ontwikkeld, die het mogelijk maakt om teksten op leesbaarheid te laten beoordelen. Dat kan gebeuren via de link www.stylene.be/leesbaarheid. Hiervoor werden zowel de klassieke leesbaarheidsformules als de nieuwe HENDI-metrick geïmplementeerd.



Afbeelding 1: Beginscherm van de Stylene leesbaarheidsinterface

De gebruiker kan ervoor kiezen tekst rechtstreeks in de interface te ‘plakken’ of een tekstbestand te uploaden (afbeelding 1). Daarna gebeurt de analyse en vervolgens krijgt de gebruiker een aantal resultaten te zien met enige toelichting. Eerst worden enkele resultaten van klassieke leesbaarheidsformules weergegeven, berekend op basis van oppervlakkige tekstenkenmerken (afbeelding 2).

KLASSIEKE LEESBAARHEIDSFORMULES

Leesbaarheidsonderzoek kent een lange en rijke geschiedenis. Tot voor kort werden eenvoudige leesbaarheidsformules gebruikt om de leesbaarheid van een tekst te beoordelen.

Zo een formule berekent aan de hand van een aantal oppervlakkige tekstenkenmerken een score of niveau die dan overeenkomt met een vooraf bepaalde moeilijkheidsgraad. Denk bijvoorbeeld aan de AVI-niveaus.

Hieronder vind je de scores die jouw tekst krijgt wanneer we deze klassieke formules berekenen. Wanneer je met je muis over de naam beweegt krijg je meer informatie.

Leesbaarheidsformules ontwikkeld voor het Nederlands

NAAM	SCORE	SCHAAL
Leesindex A Brouwer	20.88	0 - ca. 120
Flesch-Douma	30.79	0 - ca. 120
CILT (Cito Leesindex Technisch Lezen)	58.94	50 - 100
CLIB (Cito Leesindex voor het Basis onderwijs)	41.14	0 - 100

Leesbaarheidsformules ontwikkeld voor andere talen

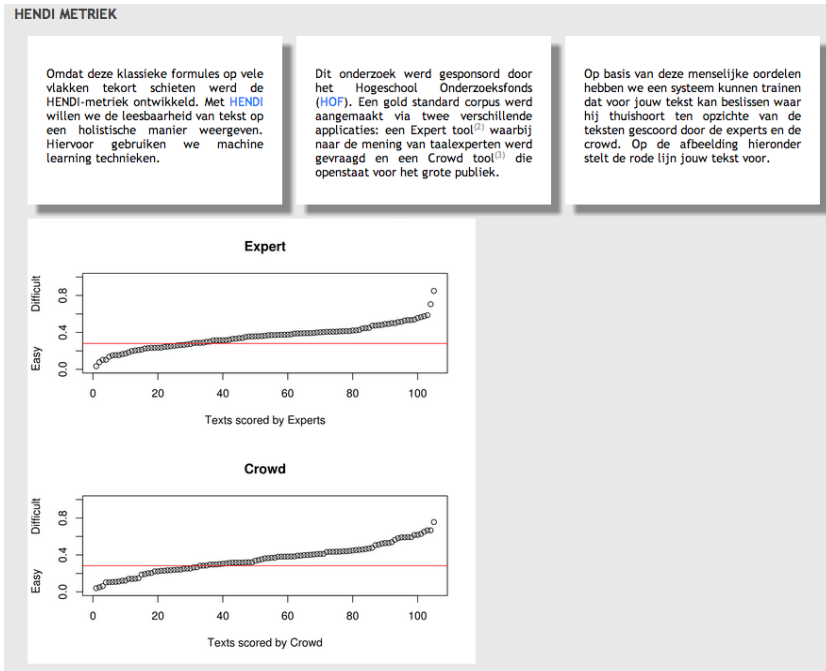
NAAM	SCORE	SCHAAL
Flesch Reading Ease	13.54	0 - ca. 120
Dale-Chall Reading Grade Score	19.16	0 - 16
Coleman-Liau Index	18.7	0 - 15
Flesch-Kincaid Grade Level	15.87	1 - 15
Gunning Fog Index	19.02	0 - 15
ARI (Automated Readability Index)	16.52	1 - 15
SMOG (Simple Measure of Gobbledygook)	16.32	1 - 15
LIX (Lasbarhetsindex Bjornsson)	54.31	0 - 100

Afbeelding 2: leesbaarheidsoordeel volgens verschillende klassieke leesbaarheidsformules

Een handig kenmerk van deze interface is dat het ook mogelijk is om bepaalde items te visualiseren.

Ten slotte, krijgt de gebruiker ook een indicatie van de leesbaarheid op basis van de HENDI-metriek (afbeelding 3). Dit oordeel is gebaseerd op een vergelijking met de referentiedataset die we

hebben verkregen door ofwel taalexperten (Expert) of het grote publiek (Crowd) te raadplegen. De horizontale lijn in de grafieken stelt de ingevoerde tekst voor.



Afbeelding 3: de HENDI metriek

We kunnen besluiten dat in de lange traditie van het leesbaarheidsonderzoek het domein van de taaltechnologie er vooral toe kan bijdragen dat niet alleen oppervlakkige, maar ook meer complexe tekstkenmerken in overweging worden genomen voor het automatisch meten van de leesbaarheid van een tekst. Hoewel het HENDI-project ondertussen is afgelopen, zal de ontwikkelde software nog veelvuldig gebruikt worden binnen de vakgroep Vertalen, Tolken en Communicatie. In het HOF-project ROBOT[5] bijvoorbeeld zal de HENDI-metriek de onderzoekers helpen om te analyseren in hoeverre leesbaarheid overeenkomt met vertaalbaarheid. Ook buiten de muren van onze vakgroep werd al gebruik gemaakt van de HENDI-

tool, bijvoorbeeld voor het meten van de leesbaarheid van juridische teksten.

Referenties

Bailin, A., & Grafstein, A. (2001). The linguistic assumptions underlying readability formulae: a critique. *Language & Communication*, 21(3), 285–301.

Dale, E., & Chall, J. S. (1948). A formula for predicting readability. *Educational Research Bulletin*, 27, 11–20.

De Clercq, O., Hoste, V., Desmet, B., van Oosten, P., De Cock, M. & Macken, L. (2013). Using the Crowd for Readability Prediction. *Natural Language Engineering*, 1-33. Cambridge Journals Online.

DuBay, W. H. (2004). *The Principles of Readability*. Costa Mesa, CA: Impact Information.

DuBay, W. H. (Ed.). (2007). *Unlocking Language: The Classic Readability Studies*. Costa Mesa, CA: BookSurge.

Feng, L., Elhadad, N., & Huenerfauth, M. (2009). Cognitively motivated features for readability assessment. In *Proceedings of the 12th Conference of the European Chapter of the ACL*, Boulder, CO, USA, pp. 229–37.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32(3), 221–33.

Gunning, R. (1952). *The Technique of Clear Writing*. New York: McGraw-Hill.

Heilman, M., Collins-Thompson, K., & Eskenazi, M. (2008). An analysis of statistical models and features for reading difficulty prediction. In *The Third Workshop on Innovative Use of NLP for Building Educational Applications*, Columbus, OH, USA.

Kincaid, J. P., Jr., R. P. F., Rogers, R. L., & Chissom., B. S. (1975). Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for navy-enlisted personnel. Technical Report, Naval Technical Training Command Millington Tenn Research Branch, Department of Navy, Washington, DC.

Kraf, R., & Pander Maat, H. (2009). Leesbaarheidsonderzoek: oude problemen, nieuwe kansen. *Tijdschrift voor Taalbeheersing* 31(2), 97–123.

McLaughlin (1974). Temptations of the Flesch. *Instructional Science* 2(4), 367-383.

Staphorsius, G. (1994). *Leesbaarheid en Leesvaardigheid. De Ontwikkeling van een Domeingericht Meetinstrument*. Arnhem, Netherlands: Cito.

van Oosten, P., Tanghe, D., & Hoste, V. (2010). Towards an improved methodology for automated readability prediction. In *Proceedings of the seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta

van Oosten & Hoste, V. (2010). Readability Annotation: Replacing the Expert by the Crowd. In *Proceedings of the Sixth Workshop on Innovative Use of NLP for Building Educational Applications*, Portland, Oregon.

Projectwebsites

[1] <http://www.lt3.ugent.be/en/tools/classical-readability-formula-calculator/>

[2] <http://www.lt3.ugent.be/en/tools/expert-readers-nl/>

[3] <http://www.lt3.ugent.be/en/tools/sort-by-readability-nl/>

[4] <http://www.lt3.ugent.be/en/projects/stylene/>

[5] <http://www.lt3.ugent.be/en/projects/robot/>