

TAAL IN TRANSFORMATIE

Copyright 2024 © Frieda Steurs | /instituut voor de Nederlandse taal/

ILLUSTRATIES Frank Landsbergen

VORMGEVING BINNENWERK & OMSLAG www.igraph.be

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any other means, electronic or mechanical, including photocopying, recording or by any other information storage and retrieval system, without permission from the publisher.

Insofar as possible, we have requested permission from the rightful claimants to use the illustrations and photographs contained in this book. Anyone who is of the opinion that they have not been approached on this matter while feel they should have, is requested to contact the publisher.

ISBN 978 94 6319 304 7

NUR 600

ivdnt.org

scriptum.nl

Taal in transformatie

Technologie,
economie,
en de kracht van
generatieve AI

Frieda Steurs

/instituut voor de Nederlandse taal/

SCRIPTUM

Inhoud

Voorwoord	9
1 Taal en economie: een onverbrekelijk team!	11
De taalindustrie: communicatie heeft een prijs	12
Wat omvat de taalindustrie allemaal?	14
Data stewards & language engineers: Taalberoepen 2.0	16
De taalindustrie als groeifactor	17
Van steenkool naar kennis: de nieuwe economie	18
Nederland, Vlaanderen en de vitaliteit van het Nederlands	21
Van bouwbedrijf tot voedingsindustrie: een variatie van markten voor vertalers	23
Toerisme & taal, een belangrijke economische pijler	26
De communicatieparadox: wat zegt u?	28
Wablieft? Het belang van helder taalgebruik	31
Simplificatie? Waarom is dat nodig?	36
Samengevat	36
2 Taal + economie = meertaligheid	38
'English only' is geen optie	39
Buitenlandse markten winnen met taalstrategie	42
Meertaligheid in Europa: wat zijn de sterke en de zwakke punten?	46
E-commerce: de rijzende ster in de economie	47
De impact van AI op onlineverkopen	51
Taal en communicatie: noodzakelijk voor een succesvolle commerciële activiteit	52
Welkom in de taalsector!	53
Modellen van interne en externe communicatie in het bedrijfsleven	54

Meertaligheid in Europa: het Europees handvest voor Regionale Talen of Talen van Minderheden	57
Besluit	60
3 Over het lexicon, woordenboeken, nieuwe woorden en vaktaal	61
De neerslag van een taal en de woordenschat: woordenboeken	62
Het grootste woordenboek ter wereld	63
Het Nederlands van nu	67
Zijn nieuwe woorden blijftertjes of slechts beperkt houdbaar?	71
Afzetten tegen het Engels	72
Andere types woordenboeken	75
Woordenschat uit de streek	77
Domeinverlies: kunnen we nog alles benoemen in onze moedertaal?	78
Taalpolitiek en creativiteit: de groei en ontwikkeling van de Afrikaanse talen	80
E-lexicografie, wiki en crowdsourcing	83
Vakexterne en vakinterne woorden	86
Algemene taal en vaktaal: een continuüm	87
Grondslag voor de terminologieleer	89
Drie perspectieven op termen	90
Algemene woordenschat en de terminologie	91
Laten we deze canapé even degusteren? Euh, u zei?	91
De Nederlandse overheid en de pensioenfondsen: terminologische verwarring.	92
Inzicht in de wereld van de visserij: begrippenstelsels	94
Medische classificaties en nomenclaturen	95
Terminologische databanken als bouwplaatsen van kennis	98
Hoe ondersteunt het <i>/instituut voor de Nederlandse taal/</i> bedrijven en taalprofessionals op het vlak van de terminologie?	100
De Termraad en Termraad Academy	101

Termraad Academy	103
ISO: internationale afspraken over terminologie	103
Besluit	104
4 Taal- en Spraaktechnologie en de wereld van de generatieve AI	106
NLP, Machine Learning, Neurale Netwerken en AI: wat betekent dit allemaal?	107
Wie is BERT?	110
Wat doet BERT?	110
Hoe is BERT voorgetraind?	111
Spraaktechnologie: kan de computer spreken en verstaan?	111
Deep Learning-technologie van Nuance	113
Voicebots en chatbots: hé Alexa!	114
Domotica: (spraak)technologie in huis	116
Toepassingen in de zorg: praten tegen je rolstoel	116
Succesvolle bedrijven in de taal- en spraaktechnologie	117
Spraakherkenning versus sprekersherkenning	118
Taal: het moeilijkste paradigma	119
Trends op sociale media: hoe denkt de buitenwereld over mijn bedrijf of mijn product?	120
Taal- en spraaktechnologie voor het Nederlands	121
Hoe talen digitaal dood kunnen gaan	122
De digitale taalinfrastructuur van het Nederlands	123
Boventiteling, voice-over en dubbing: vertaling en massacommunicatie	125
Ondertiteling, closed captions en SDH	127
Mediatogankelijkheid: meer dan alleen kijken	129
Gaming en lokale markten	130
Vertaaltechnologie: de vertaler krijgt hulp	131
Software in de cloud: samenwerken met vertalers over de hele wereld	137
Beveiliging: gevoelige documenten vertalen? Nee, liever niet online!	138
Machinevertaling: neemt de computer het vertalen over?	140

Neurale netwerken: machinevertaling van nu	142
Bias: <i>hij</i> is een dokter, <i>zij</i> is een verpleegster	142
Maar waarom toch die drang naar machinevertaling?	144
Is een menselijke vertaler niet veel beter?	
Het resultaat van machinevertaling: zo publiceren of nog	146
reviseren? Het dilemma van het postediteren	
Besluit	146
5 Levenslang leren: hoe de computer helpt bij	148
het (taal)onderwijs	
E-learning en blended learning	149
Een Europees raamwerk: DigCompEdu	151
Ben je digitaal geletterd? En wat met AI?	152
ChatGPT: gaat de leerling nog origineel werk leveren,	154
of neemt AI alles over?	
Andere AI-toepassingen in het taalonderwijs	156
Een chatbot specifiek voor het onderwijs: Khanmigo	157
Nog meer ondersteuning voor de leerkracht: KlasCement	157
MOOCs: de hele wereld volgt online les	158
Webinars en kennisclips: lessen uit de COVID-19-pandemie	159
M-learning: Mobiel leren, altijd en overal	161
Hogeschooltaal: bereken je startniveau en verbeter je	162
kennis van het Nederlands	
De brede schatkamer van het Nederlands:	163
het INT en onderwijs	
Uitdrukkingen, patronen en idiomen in het Nederlands	164
Hoe spel je nu ook weer het werkwoord <i>updaten</i> in de	165
verleden tijd?	
Wat is het verschil tussen <i>in het bos</i> en <i>het bos in</i> ?	166
Welke klanken heeft het Nederlands? Is <i>zij maakt nooit</i>	
<i>geen fouten</i> nu goed of fout Nederlands?	
Nog meer over grammatica op de INT-website	167
Is er veel materiaal te vinden voor e-learning?	168
Besluit	169

6 Technische communicatie	171
1 Wat is technische communicatie?	172
2 De toekomst is digitaal	175
3 Impact van AI op de technische communicatie	178
4 Duitsland, een sterke (auto-)industrie	179
5 Technische schrijvers: sterke industrie, sterke opleidingen	182
TEKOM: het brandpunt van de technische communicatie	184
DITA: Een architectuur voor technisch schrijven	185
Meertalig terminologiebeheer in de industrie	187
De automobielsector: terminologische verwarring te lijf	188
SAE en de foutenmatrix	190
De luchtvaartsector: Aviation English en Airspeak	191
'Gecontroleerde taal', wat is dat?	194
Kennisbeheer: de kern van elk bedrijf!	199
De minimale handleiding, wat is dat?	200
Besluit	201
7 Europa neemt het voortouw: alle talen digitaal	203
Taalgelijkheid in het digitale tijdperk	204
Grote Europese onderzoeksprojecten effenen de weg voor een digitaal meertalig Europa: ELRA, CLARIN en META-NET	206
De ambities van Europa: Het digitaal kompas voor 2030	209
ELRC en eTranslation: het antwoord op Google Translate	211
Europees beleid: EFNIL is aan zet!	213
Taalkundige rechtvaardigheid	215
Wat zijn de concrete doelstellingen voor 2030 en wat kan het INT hierin betekenen?	216
Besluit	220
Lijst van gebruikte afkortingen	221
Eindnoten	223
Index	230

Voorwoord



Sinds de lancering van ChatGPT in 2022 lijkt het wel alsof plots alle aandacht gaat naar taalmodellen en kunstmatige intelligentie in het digitale tijdperk. Voor het grote publiek is het soms wat beangstigend: neemt de computer ons taalvermogen over; is het nog nodig dat kinderen vreemde talen leren en moeten studenten nog wel een masterproef leren schrijven?

De realiteit is anders: de menselijke taal wordt al gedurende vele decennia bestudeerd in de (computationele) taalkunde en er zijn al heel wat taalmodellen gelanceerd. Ook de zoektocht naar betere machinale vertaalsystemen en andere taaltoepassingen is al vele jaren bezig. Zo werden alle grote Nederlandstalige woordenboeken al jaren geleden gedigitaliseerd. In 2016 werd het vroegere Instituut voor de Nederlandse Lexicologie (INL) omgevormd tot het Instituut voor de Nederlandse Taal (INT), dat een bredere taak kreeg. In opdracht van het Comité van Ministers van de Nederlandse Taalunie werd er werk gemaakt van het beschikbaar maken van zoveel mogelijk taalmaterialen voor het Nederlands, en dit voor het Nederlands in Nederland, Vlaanderen, Suriname en het Caribisch gebied.

Daarbij kon er gebouwd worden op de rijke collectie historische woordenboeken die door het INL waren gemaakt, maar daarnaast werd er verdere uitbreiding voorzien van de digitale taalinfrastructuur voor het hedendaags Nederlands, de spelling, de grammatica en de terminologie.

Het Nederlands is een belangrijke taal, de taal van de lage landen, een sterke economische regio. Taal en economie zijn sterk met elkaar

verbonden, dat toonde ik al aan in een eerder boek, *Taal is Business* (2016).

In dit boek maken we een stand van zaken op: Waar staan we in 2024, in een sterk veranderende wereld waarin de technologische vernieuwingen zeer snel na elkaar gelanceerd worden. Europa wil het voortouw nemen en streeft tegen 2030 naar een sterke digitalisering van alle Europese talen. Samen met de Taalunie werkt het instituut voor de Nederlandse taal daar intensief aan mee en proberen we het Nederlands zo sterk mogelijk in de markt te zetten.

Ik neem u graag mee in een fascinerend verhaal over het belang van taal in onze economische en technologische wereld.

Frieda Steurs

Leiden, november 2024

1

Taal en economie: een onverbreekelijk team!



Vaak wordt taal en de studie van de taal- en letterkunde of toegepaste taalkunde als iets 'softs' gezien, dat weinig nuttig is en nauwelijks kansen biedt op economisch succes. Taal en economie lijken voor velen niet echt met elkaar verbonden. Evenzeer lijkt de taalindustrie een wat vreemde term en de meeste mensen vragen zich af wat taal nu met industrie of met economie te maken heeft? We geven een eerste schets van de internationale markten voor taalprofessionals en van het effect van economische groei op taal, en omgekeerd, van het effect van taal op de economie. Zowel in Nederland als in Vlaanderen zijn er de laatste jaren studies uitgevoerd die aantonen dat taalberoepen wel degelijk in de lift zitten en dat er een grote behoefte is aan goede (meertalige) taalkundigen¹. De maatschappij vraagt naar goed opgeleide taalprofessionals. De derde en vierde industriële golf veroorzaken in onze landen een verschuiving naar een kenniseconomie, die draait om informatie-uitwisseling. Kennis en data zijn op dit ogenblik in onze hoog technologische en geavanceerde maatschappij het meest waardevolle kapitaal van elk bedrijf en elke instelling. Die informatie en data zijn jammer genoeg niet voor iedereen bereikbaar. Heel wat mensen dreigen uit de boot te vallen. Wat kunnen we doen om de nodige informatie voor iedereen toegankelijk te maken?

De taalindustrie: communicatie heeft een prijs

Taal werd vroeger en tot voor kort vooral geassocieerd met het leren van een taal of met het bestuderen van de literatuur ervan. De klassieke filologische benadering van taal werd aan de meeste universiteiten op een zeer traditionele manier ingevuld. Maar de laatste jaren ligt het accent veel meer op communicatie. Daarbij komt ook het enorme aanbod aan (taal)data, die verwerkt moeten worden. De groei en bloei van opleidingen die de taal meer vanuit een economisch standpunt benaderen, is een logische evolutie. Onder economisch standpunt verstaan we binnen deze context de markt. In de eerste plaats de nationale markt, maar onlosmakelijk daarmee verbonden de Europese en ten slotte de uitbreiding naar de markten van andere continenten. Voorbeelden daarvan zijn: opleidingen communicatiemanagement, vertalen, tolken, meertalige communicatie, vertaal-



technologie, computertaalkunde, enzovoort. In al deze opleidingen wordt meer en meer aandacht besteed aan dataverwerking en technologische toepassingen.

Daarnaast is er een belangrijk aantal maatschappelijke ontwikkelingen die bijdragen aan een nieuwe visie op het belang van taal. We noemen:

- De snelle evolutie van wetenschap en technologie die ervoor zorgt dat talloze nieuwe begrippen ontstaan. Die moeten een talige voorstelling krijgen en dat in meerdere talen.
- De invloed en veranderingen door de snelle technologische groei in alle geledingen van onze samenleving heeft tot gevolg dat er snellere en betrouwbare gespecialiseerde informatie nodig is. De internationale uitwisselingen, de samenwerking op politiek, economisch en cultureel vlak zorgen voor een enorm groeiend aantal meertalige contacten.
- De massaproductie van consumptiegoederen die leidt tot een steeds globalere economie, waarbij bedrijven rekening moeten houden met internationale en nationale spelregels, op eenduidige wijze vastgelegd in specificaties en termen.
- Om in dit digitale tijdperk zinvol met grote hoeveelheden gegevens te werken, moet men sterk staan in het analyseren van de menselijke taal. Big data op zich zijn geen troef, het gaat erom hoe deze data kunnen worden verwerkt en geïnterpreteerd.



De term *taalindustrie* refereert aan deze nieuwe ontwikkelingen en is een verzamelnaam voor allerlei talige processen en diensten. *Industrie* verwijst hier naar de economische activiteit, maar ook naar de harde cijfers: de kostprijs van taal en meertaligheid en de prijs van correcte informatie.

De taalindustrie blijft wereldwijd groeien

Artificiële intelligentie (AI) wordt vandaag de dag gezien als dé toekomst in heel wat sectoren. Voor elke sector betekent dit weer andere mogelijkheden, maar ook uitdagingen. De wereld is door digitalisering een stuk kleiner geworden. Meer bedrijven hebben internationale ambities. Goed nieuws voor bedrijven die (ver)taaldiensten aanbieden. Vertaling van bedrijfsdocumentatie en -communicatie is meer nodig dan ooit, nu bedrijven vanuit alle uithoeken van de aardbol zaken met elkaar kunnen doen².

De wereldwijde markt voor taaldiensten bereikte in 2021 een waarde van 62,6 miljard dollar, met een verwachte groei van 7,5 procent tussen 2022–2027. Ondanks de onzekerheden veroorzaakt door de COVID-19-pandemie en de toenemende wereldwijde economische problemen, blijft de taaldienstensector sterk en laat een veerkrachtige groei zien. In feite is de vraag naar taaldiensten, waaronder vertaling en lokalisatie, alleen maar toegenomen tijdens de moeilijkheden van de afgelopen jaren. Als gevolg daarvan maken taaldienstverleners zich op om de komende jaren een nog grotere rol te spelen.³

Wat omvat de taalindustrie allemaal?

Het gaat over heel wat activiteiten: vertalen, tolken, software en websites lokaliseren, audiovisuele vertaling (ondertitelen en dubben van beeldmateriaal) en talen leren via e-learning. Er zijn tal van taalberoepen, van accountmanager in een vertaalbureau tot webredacteur, van adviseur taalbeleid tot woordvoerder of copywriter. Talenkennis wordt steeds belangrijker in steeds meer domeinen van de samen-

leving. Taalgediplomeerden zetten hun kennis en vaardigheden in bij scholen, uitgeverijen, redactie bureaus en vele andere organisaties in de zorg, in migratie en integratie, en in de cultuursector.

In dit boek zal ik verschillende aspecten van technologische en niet-technologische oplossingen voor communicatie behandelen. Niet-technologische oplossingen zijn bijvoorbeeld vertalen, tolken, editeren, lokaliseren, talen leren. Technologische oplossingen zijn machinevertaling, computerondersteund vertalen, computerondersteund talen leren, videoconferenties met tolken, meertalige zoekmachines, datamining (het analyseren van grote hoeveelheden gegevens), meertalige terminologiebanken voor specifieke bedrijven of vakgebieden enzovoort. De recente ontwikkelingen in de kunstmatige intelligentie (generatieve AI) zijn meebepalend voor de vlucht die de technologische toepassingen nu nemen.

Om al deze taaltoepassingen goed te ondersteunen, moet elke taal een sterke digitale infrastructuur hebben; dat wil zeggen, de nodige betrouwbare taal- en spraakdata die in de technologische applicaties gebruikt kunnen worden. Deze taaldata zijn essentieel om bedrijven en onderzoekers in staat te stellen toepassingen, zoals het omzetten van moeilijke teksten in een begrijpelijker versie, in een bepaalde taal uit te bouwen. Hoe gebeurt dat voor het Nederlands?

De digitale taalinfrastructuur

Het /instituut voor de Nederlandse taal/ (INT) in Leiden heeft als kennisinstituut een unieke positie en opdracht om voor het hele Nederlandse taalgebied (Nederland en de Caribische rijkdelen, Vlaanderen en Suriname) op een wetenschappelijk verantwoorde wijze de digitale taalinfrastructuur uit te bouwen. Het INT ontwikkelt enerzijds zelf corpusdata, linguïstische databanken en taalsoftware voor een aantal specifieke domeinen of ondersteunt de ontwikkeling ervan; anderzijds verzamelt het INT ook taalmaterialen en taalsoftware van andere kennisinstellingen en stelt het die samen met de eigen taalmaterialen duurzaam ter beschikking.

/instituut
voor de
Nederlandse
taal/