

Frieda STEURS

ORCID: 0000-0002-2884-5713

Instituut voor de Nederlandse Taal (Leiden) / KU Leuven

Nederlands, een grote taal? Een kwestie van meten

Abstract

In this article we want to explore the status of the Dutch language. Is Dutch an important language, a ‘major’ one? How do we measure the importance of a language? Is this about the number of native speakers? Or are there other parameters that have to be taken into account? We can check the global influence of a language, the number of translations made, and the presence on the internet, amongst other elements. Ethnologue and UNESCO have developed a classification system to measure the importance and vitality of a language. Whether languages continue to be dynamic is also related to their digital power. In this article we will expand on the status of Dutch and the power of the Dutch digital language infrastructure.

Keywords: language infrastructure, endangered languages, CLARIN, language technology, European Language Equality (ELE).

1. Het Nederlands: groot of klein?

Er zijn ongeveer 24 miljoen Nederlandstaligen. Ongeveer 17 miljoen van hen wonen in Nederland, 6,5 miljoen in België, en 400.000 in Suriname. Daarmee is het Nederlands een van de 40 meest gesproken talen in de wereld. Als je bedenkt dat er wereldwijd meer dan 6000 talen gesproken worden, dan scoort het behoorlijk hoog. Niet alleen daarom, ook om allerlei andere redenen is het Nederlands een wereldtaal. Zo is het ook de twaalfde taal op internet en een belangrijke taal in de sociale media. De taaltechnologie voor het Nederlands is hoog ontwikkeld. En de Nederlandse literatuur en cultuur scoren hoog op internationaal vlak.¹

Deze feiten worden vermeld op de website van de Taalunie. Het Nederlands is een officiële taal in de Europese Unie en de Unie van Zuid-Amerikaanse Naties. De Unie van Zuid-Amerikaanse Naties (Unasur) is een politieke en economische unie van de 12 onafhankelijke staten in Zuid-Amerika. Verder wordt het Nederlands in delen van Duitsland en kleinere gebieden in Noord-Frankrijk gesproken.

¹ Zie <http://taalunieversum.org/inhoud/feiten-en-cijfers#feitencijfers>.

We lezen het wel vaker: “het Nederlands scoort hoog”, “Nederlands is een wereldtaal”, “Nederlands is een belangrijke taal”,² “een van de meest gesproken talen”... maar wat betekent dat?

In de Europese Unie spreken ongeveer 24 miljoen mensen Nederlands als eerste taal en een bijkomende acht miljoen als tweede taal. Verder is het Nederlands ook een officiële taal van de Caraïbische eilanden Aruba, Curaçao en Sint-Maarten, terwijl de Franse Westhoek en de regio rondom de Duitse stad Kleef van oudsher Nederlandstalige gebieden zijn, waar Nederlandse dialecten mogelijk nog gesproken worden door de oudste generaties. Ook in de voormalige kolonie Indonesië spreken in bepaalde gebieden sommige leden van de oudere generaties nog Nederlands. Het aantal sprekers van het Nederlands in de Verenigde Staten, Canada en Australië wordt op ruim een half miljoen geschat.

In Europa zijn de drie grootste Germaanse talen het Duits (95 miljoen), Engels (70 miljoen) en Nederlands (25 miljoen). Het Nederlands alleen wordt door meer mensen gesproken dan de Noord-Germaanse (Scandinavische) talen bij elkaar.

Nederland en België waren in het verleden allebei koloniale machten en zo ging het Nederlands ook een rol spelen in deze gebieden. Zo was het Nederlands ook een officiële taal in Nederlands-Indië, Nederlands Nieuw-Guinea, een stukje van Brazilië, Bernice, Belgisch Congo, in Rwanda en Burundi, en in de Nederlandse Antillen; ook in Zuid-Afrika hebben de van oorsprong Nederlandse kolonisten gepleit voor de erkenning van de Nederlandse taal. Het Nederlands heeft echter geen stand gehouden in deze landen in de postkoloniale periode, wat wel het geval is met het Spaans, Frans, Portugees en Engels.

2. Is het Nederlands een wereldtaal?

Om deze vraag te beantwoorden moeten we eerst en vooral definiëren wat er verstaan wordt onder het begrip ‘wereldtaal’. Er zijn tal van parameters die men kan gebruiken om het ‘belang’ of de ‘status’ van een taal te meten. Een wereldtaal wordt vaak gezien als een taal die in grote delen van de wereld als communicatiemiddel wordt gebruikt, en in die zin is het aantal moedertaalsprekers zeker een belangrijke factor. Maar zij zijn niet de enige factoren die de status van een taal bepalen.

In een artikel in de *NRC* van 2014³ concludeert journalist Bruno van Wayenberg dat het Nederlands een grote taal is. Hij ziet het Nederlands toch een beetje als wereldtaal, ook al blijft de taal in de schaduw van het Frans, Duits en Engels. Hoe komt hij tot deze bewering?

Hij verwijst hiervoor naar een nieuwe methode om de wereldwijde invloed van talen te meten. Amerikaanse en Franse onderzoekers (onder wie taalkundige

² Zie https://www.nt2.nl/nl/dossier/nederlands_leren/6_redenen_om_nederlands_te_leren.

³ Zie <https://www.nrc.nl/nieuws/2014/12/16/nederlands-is-een-grote-kleine-taal-1448396-a1078891>.

Steven Pinker van Harvard) gingen na hoe je de invloed van een taal kan meten.⁴ Wat bepaalt of een taal als *global language* kan worden gezien? Het gaat daarbij niet om het aantal moedertaalsprekers van een taal, maar om de ‘globale invloed’ van een taal. De taalkundige David Crystal deed hier ook al uitspraken over (7). Volgens hem heeft de vraag of een taal al dan niet een wereldtaal is niet zoveel te maken met het “aantal” moedertaalsprekers, maar wel met “wie de sprekers van die taal zijn”. Zo was het Latijn een wereldtaal, een pan-Europese taal, niet omdat het de moedertaal was van de meeste Europeanen, maar omdat het de taal was van het Romeinse Rijk, en later de taal van de katholieke kerk, en professoren. Latijn is zo eeuwenlang een universele taal gebleven, omdat het elites kon verbinden. Kunnen we dit idee van een verbinding tussen internationale elites ook toepassen in onze moderne wereld? Welke parameters kunnen we gebruiken om te meten of een taal inderdaad zoals het Latijn mensen op internationale schaal verbindt? Te denken valt dan aan boekvertalingen, meertalige Wikipediapagina’s en internetpagina’s in het algemeen, maar ook aan tekstberichten zoals Twitter. Dit is door verschillende experts onderzocht.

We verwijzen hier naar onderzoekers zoals Rohen, Pinker en hun team. Zij definiëren het *global language network* zo: een netwerk van taalkundige publicaties en uitingen, dat uiteenvalt in drie grote delen: boekvertalingen, Wikipediapagina’s en Twitteksten. Voor wat betreft boekvertalingen werd er gekeken naar wat het Unesco-project *Index Translationum*⁵ aan gegevens beschikbaar stelde. Ook de vraag naar de markt voor boekvertalingen in bepaalde talen werd meegenomen in de berekeningen. Van en naar welke talen wordt er veel vertaald? Verder werden de meertalige verbanden in Wikipedia in kaart gebracht: gebruikers die een Wikipediapagina redigeren en die ook in een andere taal een pagina bewerken, werden zo in kaart gebracht. Hetzelfde werd dan met Twitter gedaan: talen zijn met elkaar verbonden als bepaalde gebruikers, die in een bepaalde taal tweets sturen, dat ook significant veel in een andere taal doen. Op deze wijze verkregen de onderzoekers een soort van netwerk van paden van directe en indirecte communicatie tussen sprekers van verschillende talen.

Wat het Nederlands betreft, kunnen we zien dat relatief veel Nederlandse twitteraars ook af en toe in het Engels twitteren. De onderzoekers zien dat als een sterke band tussen het Nederlands en Engels. Op dezelfde manier versterken twitterende Turkse Nederlanders de band tussen het Turks en het Nederlands. Een ander voorbeeld: veel boeken in het Russisch zijn vertaald van of naar het Tataars, Oezbeeks of Kirgizisch. En veel schrijvers van Wikipedia-artikelen in het Hindi dragen ook bij aan Engelstalige artikelen. Een overzicht van alle verbindingen tussen die talen levert mooie plaatjes van ‘mondiale taalnetwerken’ op. Wereldtalen als Engels, Spaans en Russisch zijn daarin centrale knooppunten,

⁴ Zie <https://www.pnas.org/content/111/52/E5616>.

⁵ Zie <http://www.unesco.org/xtrans/>.

kopstations van ideeën. Minder prominente talen als het Welsh, Litouws en Sami zitten meer in de periferie. Het onderzoek toont aan dat het Nederlands met zijn 25 miljoen sprekers, van wie er veel meertalig zijn, zich in dit netwerk aardig overeind weet te houden. Sterke verbindingen zijn er met het Engels, Afrikaans en Turks (Twitter), en met het Fries (Wikipedia) en Duits (boekvertalingen). Het meest centraal in het Twitternetwerk staat natuurlijk het Engels, maar op de zesde plaats, net onder het Spaans en Filipijns,⁶ komt het Nederlands.

Ook andere metingen bevestigen het belang en de status van het Nederlands: met name de aanwezigheid van het Nederlands als taal op internationale websites en het surfgedrag van Nederlanders en Vlamingen. Nederland behoort tot de kopgroep van de EU-28-landen met de meeste gezinnen die toegang hebben tot internet. In 2017 had 98 procent van de Nederlandse gezinnen internet thuis, het Europese gemiddelde was 87 procent. Ook met snelle breedbandverbindingen behoort Nederland tot de top. Dat blijkt uit een analyse van cijfers van Eurostat door het CBS.⁷ Het aantal Belgische gezinnen met een basisbreedbandverbinding van ten minste 30 Mbps is 98,99 %, ofwel de op drie na beste prestatie van de EU.⁸ Vlamingen en Nederlanders zijn dus heel sterk online aanwezig, en dat vertaalt zich ook in online koopgedrag. De cijfers van de e-commerce stijgen ieder jaar significant. Belgen gaven vorig jaar 10,6 miljard euro uit via webwinkels; de e-commerce-omzet in Nederland komt uit op 25,7 miljard euro.⁹ Een cijfer dat zal blijven stijgen, onder andere door de opkomst van de online supermarkt. Dit was al een trend voor de coronacrisis, maar wordt nu nog versterkt. Het blijkt dat klanten die online winkelen een sterke voorkeur hebben voor websites waarin ze in hun eigen taal kunnen kiezen en kopen; het Nederlands wordt dus, door de sterke markt, ook op een enorm aantal websites gebruikt. Zo is het Nederlands een prominente taal op het internet. De status van een taal hangt samen met een bloeiende economie.

3. De internationale positie van een taal: *Ethnologue* en Unesco

*Ethnologue: Languages of the world*¹⁰ is een overzicht van alle bekende talen op de wereld en dat zijn er ruim 6900. Het overzicht groeit nog steeds omdat door taalkundig onderzoek nog steeds nieuwe talen geïdentificeerd worden. Daarnaast

⁶ Ook Tagalog of Filipino genoemd.

⁷ Zie <https://www.cbs.nl/nl-nl/nieuws/2018/05/nederland-koploper-in-europa-met-internettoegang>.

⁸ Zie <https://tweakers.net/nieuws/89661/nederland-en-belgie-lopen-voorop-met-supersnelbreedband.html>.

⁹ Zie <https://www.logistiek.nl/supply-chain/nieuws/2019/07/nederlandse-e-commerce-stijgt-naar-257-miljard>.

¹⁰ Zie <https://www.ethnologue.com/>.

is er ook het *Observatoire linguistique*,¹¹ een taalkundig onderzoeksinstituut dat zich ten doel heeft gesteld alle talen en dialecten van de wereld op een innovatieve manier in kaart te brengen. De werkwijze van dit instituut is erop gericht de wereld als eenheid zonder staatkundige of andere grenzen in kaart te brengen, puur op basis van de taalvarianten die gesproken worden en de relaties daartussen.

Hiervoor is een classificatiesysteem ontwikkeld op basis van lexicale overeenkomst tussen de basiswoordenschat van verschillende taalvarianten. Dit systeem geeft daarbij historisch taalkundige (phylozones) en geografische verbanden (geozones) aan. Een ander interessant overzicht vinden we in het boek *The Linguasphere Register of the World's Languages and Speech Communities* (Dalby 2000) met een overzicht van de talen in de twintigste eeuw om o.a. taalverschuivingen te kunnen volgen.

Wat ook interessant is, is de lijst van criteria vastgesteld door Unesco voor het meten van de vitaliteit van een taal:

- de mate waarin een taal van generatie op generatie wordt doorgegeven;
- het absolute aantal sprekers van die taal;
- het relatieve aantal sprekers van de taal in verhouding tot de totale groep waarmee de sprekers zich identificeren;
- de positie die de taal inneemt in traditionele maatschappelijke domeinen, bijvoorbeeld in het onderwijs;
- de mate waarin de taal gebruikt wordt in nieuwe maatschappelijke domeinen, bijvoorbeeld in sociale media;
- de mate waarin de taal wetenschappelijk beschreven is in materiaal voor het onderwijs (bijvoorbeeld spellinggidsen, woordenboeken, grammatica's, enz.) én de mate waarin er literatuur geschreven wordt in die taal. Met andere woorden: de mate waarin het mogelijk is om de taal aan te leren én om te lezen in die taal;
- de mate waarin de taal institutioneel en beleidsmatig ondersteund wordt, bijvoorbeeld of de taal een officiële taal is;
- de attitudes van de sprekers ten opzichte van hun taal;
- het soort materiaal en de kwaliteit van het materiaal waarmee de taal kan worden gedocumenteerd (zijn er bijvoorbeeld naast teksten in de bewuste taal in de vorm van grammatica's, woordenboeken, enz., ook audio-en video-opnames van de gesproken variant van die taal?).

De overzichten van *Ethnologue* en het *Observatoire linguistique* geven ook aan wanneer een taal bedreigd wordt. Zo kunnen we stellen dat een bedreigde taal een levende taal is die nog maar door erg weinig mensen als moedertaal wordt gesproken, of waarvan het aantal moedertaalsprekers zo snel afneemt, dat zij binnen afzienbare tijd niet meer gesproken zal worden.

Er zijn vier categorieën voor met uitsterving bedreigde talen. De talen met cijfer 5 hebben de status *Safe* en zijn dus (momenteel althans) niet bedreigd.

¹¹ Zie <http://olst.ling.umontreal.ca/>.

0. *Extinct* (uitgestorven)
1. *Critically endangered* (kritiek)
2. *Severely endangered* (zeer ernstig)
3. *Definitively endangered* (ernstig)
4. *Unsafe* (onveilige toestand)
5. *Safe* (veilig)

Van de circa 6900 talen die momenteel worden gesproken, zijn er 2498 die niet de status *Safe* hebben.

Voor het Nederlands, een taal die in categorie 5 zit, vinden we volgende gegevens in *Ethnologue*:¹²

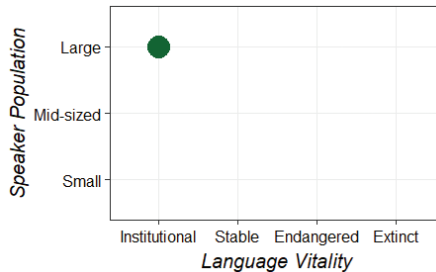


Fig. 1. Language vitality (de grootte van het Nederlands versus de taalvitaliteit); bron: www.ethnologue.com/size-and-vitality/nld

Figuur 1 toont de gegevens voor het Nederlands wat betreft de grootte van de taal (verticale as) en de taalvitaliteit (horizontale as). De verticale as geeft het aantal sprekers van het Nederlands (zowel moedertaalsprekers als NT2 sprekers) aan en maakt gebruik van drie niveaus:

- Large: groot, de taal heeft meer dan 1,000,000 sprekers;
- Mid-sized : gemiddeld, de taal heeft 10,000 tot 1,000,000 sprekers;
- Small: klein, de taal telt minder dan 10,000 sprekers.

De horizontale as laat de graad van vitaliteit zien. Daar worden niveaus en kleuren gebruikt:

- Donkergroen = institutioneel. De taal is zodanig ontwikkeld dat deze gebruikt en ondersteund wordt door formele publieke instellingen zoals ministeries, onderwijs, etc. en is niet alleen een thuistaal of taal van een gemeenschap;

- Lichtgroen = stabiel. De taal wordt niet ondersteund door formele instellingen maar is nog altijd de standaardtaal thuis en in de gemeenschap; kinderen leren de taal nog actief aan;

- Rood = bedreigd. De taal wordt niet langer als vaste waarde beschouwd die door kinderen wordt aangeleerd en gebruikt;

¹² Zie <https://www.ethnologue.com/size-and-vitality/nld>.

— Zwart = uitgestorven. De taal wordt totaal niet meer gebruikt en niemand heeft nog een ethische identiteit die verbonden is met deze taal.

Het Nederlands staat hoog links in de grafiek: het is volgens deze methode een grote en vitale taal.

Er zijn nog andere grafieken die het Nederlands meten tegenover andere talen binnen *Ethnologue*. Zo is er een grafiek die alle levende talen tegenover elkaar zet, de zogenaamde ‘taalwolk’ (*the language cloud*).

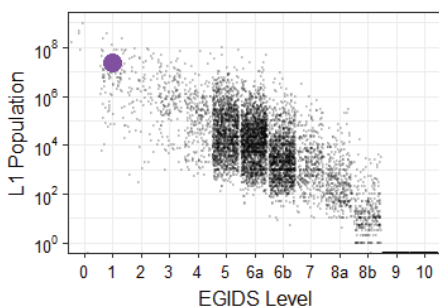


Fig. 2. De verhouding van het Nederlands t.o.v. andere levende talen volgens de *EGIDS*-schaal; bron: <https://www.ethnosproject.org/expanded-graded-intergenerational-disruption-scale/>

Figuur 2 toont hoe het Nederlands zich verhoudt tot de andere levende talen. De verticale as geeft het aantal sprekers weer, en de horizontale as gebruikt de *EGIDS*¹³-schaal om de bedreigde status te meten. De grootste en sterkste talen staan in de linkerbovenhoek in het diagram, de kleinste en zwakste talen staan onderaan rechts. De *EGIDS* schaal gebruikt ook kleuren, en paars, de kleur voor het Nederlands, is de hoogste sterktegraad: de taal is institutioneel ondersteund in alle geledingen van de maatschappij.

4. Is een taal succesvol en blijft ze dat of niet?

Een taalkundige organisatie die de vinger aan de pols houdt over de status van bedreigde talen is *CIPL*,¹⁴ het internationaal comité van linguïsten, samen met Unesco. In het artikel “Language Death and Disappearance: Causes and Circumstances” levert Stephen Wurm (president van *CIPL* van 1997 tot 2001) een interessante bijdrage aan het debat rond bedreigde talen. Hij stelt dat in de loop der jaren vele talen zijn verdwenen, zonder dat zelfs werd opgemerkt hoe dat nu precies is gebeurd. Soms blijven slechts enkele fragmenten over die geschreven of overgele-

¹³ *EGIDS* – *Expanded Graded Intergenerational Disruption Scale*.

¹⁴ Zie www.ciplnet.com.

verd werden door moedertaalsprekers of mensen die de taal observeerden. Dit geeft ons soms nog een beeld van hoe die taal gestructureerd was en wat de taalkundige eigenschappen waren. Van sommige talen rest er evenwel helemaal niets. Wurm merkt op dat er grote studies zijn over hoe planten of diersoorten verdwijnen of bedreigd zijn. Helaas kunnen ook talen verdwijnen, en daar weten we veel minder over. Talen kunnen eerst en vooral verdwijnen door oorlog en genocide, waarbij een hele taalgemeenschap wordt uitgeroeid. Ook een natuurramp kan aanleiding geven tot het uitsterven van een bepaalde taalgroep. Maar pas veel recenter zijn taalkundigen tot het inzicht gekomen dat vele talen moeten worden ondersteund om te overleven; een zekere vorm van ‘management’ en taalplanning is nodig (Wurm 3).

Wurm noemt drie belangrijke factoren die bepalen of een taal al dan niet succesvol is: economie, cultuur en politiek. We gaan kort even in op deze drie factoren:

1. Economische invloed: talen die gesproken worden door groepen die economisch succesvol zijn, zullen overleven en zullen sterker worden; meer nog, ze zullen talen van zwakkere, minder succesvolle gemeenschappen wegduwen. Dat is onder meer het geval in de grensgebieden tussen Rusland en China, waar minderheidstalen verdwijnen omdat de sprekers kiezen voor het Russisch of het Chinees, talen die een economische grootmacht vertegenwoordigen.

2. Culturele suprematie: een taal die alleen mondeling wordt overgeleverd, staat minder sterk dan een taal met een geschreven traditie. Literatuur en religieuze teksten, een rijke woordenschat en (geschreven) onderwijs in een taal maken die cultureel superieur. Zo kunnen minder ontwikkelde talen in onbruik geraken en verdwijnen.

3. Politieke invloed: er zijn talloze voorbeelden van hoe in het verleden een volk als overwinnaar na een oorlog een heel gebied annexeerde en er zijn taal oplegde als officiële taal, zowel in de administratie als de scholen. Zo is in een aantal Afrikaanse landen in het postkoloniale tijdperk nog altijd het Engels of het Frans, de taal van de vroegere overheerser, de officiële taal. Taalpolitiek kan er ook toe leiden dat een taal niet in het onderwijs mag worden gebruikt, en zo sterft een taal dan ook een stille dood.

Een van de speerpunten van *C IPL* als internationale taalkundige organisatie is dan ook het stimuleren van onderzoek naar bedreigde talen en het organiseren van workshops over dit thema.

5. Kunnen talen digitaal dood gaan?

Al in paragraaf 2 van dit artikel verwezen we naar onderzoek waar de digitale voetafdruk en vitaliteit van een taal worden gemeten aan de hand van het aantal Wikipediapagina's, e-commercewebsites, etc.

Taalkundigen in Nederland en Vlaanderen volgden vanaf de ingebruikname van de eerste computersystemen in de jaren '60 de mogelijkheden om er taalkun-

dig onderzoek mee te doen. F. De Tollenaere, die bij het Instituut voor de Nederlandse Lexicologie (INL) redacteur van het *Woordenboek der Nederlandsche Taal* (*WNT*) was en later hoofd van de afdeling thesaurus van het INL, wees reeds heel vroeg op het belang van de digitalisering van voor de taalkunde belangrijke bronnen. Dat lijkt evident, maar wat verstaan we eronder?

In een publicatie uit 1963 (!), “Nieuwe wegen in de lexicologie”, kaartte De Tollenaere al aan dat de automatisering een belangrijke rol zou gaan spelen in het bewerken en verwerken van taalgegevens. Hij beschreef ook de angst die er heerste bij de humane wetenschappen als zou de komst van de computer afbreuk doen aan de eigenheid van het vakgebied en zette zich er kordaat tegen af. Hij vergeleek de revolutie van de boekdrukkunst in de 15^e eeuw met wat er te gebeuren stond door de digitale verwerking van gegevens; dus niet iets om bang van te zijn, maar om op een positieve manier gebruik van te maken in de lexicologie. Hij bepleitte toen al de oprichting van een centrum voor lexicologie in Leiden, dat meteen ook gebruik zou kunnen maken van de troeven van de digitalisering om zowel de te bouwen thesaurus als het *WNT* en andere materiaalverzamelingen te digitaliseren. In 1977, wanneer hij een rede uitspreekt bij de opening van het studiejaar van de school voor Taal- en Letterkunde in Den Haag, gaat hij verder met het uitwerken van zijn ideeën over automatisering in de taalkunde. Zijn ideeën en voorspellingen zijn nu in het nieuwe Instituut voor de Nederlandse taal helemaal realiteit geworden. Het Nederlands is springlevend, en heeft beschikking over enorme hoeveelheden digitale taaldata.

Zijn er dan talen die niet meegaan in deze digitale (r)evolutie? Zijn er talen die digitaal dood gaan? In 2018 maakte een aantal kranten gewag van de digitale dood van het IJslands. Of dat waar of onwaar is, valt te bezien. Feit is in elk geval dat het IJslands het flink moeilijk krijgt. Hoe oud en inventief de taal ook is, amper 340.000 mensen spreken IJslands, en virtuele assistenten zoals Siri (Apple) en Alexa (Amazon) horen daar niet bij. We noemen dat verschijnsel ook ‘digitale marginalisering’, met name de evolutie waarbij een meerderheidstaal in de echte wereld een minderheidstaal wordt in de digitale wereld. Voor het IJslands heeft de regering beslist om sterk te investeren in taaltechnologie, zodat de taal ook in nieuwe digitale toepassingen kan worden gebruikt. Grote talen zoals Chinees, Spaans en Engels hebben de nodige ondersteuning voor het uitbouwen van digitale hulpmiddelen, maar hoe zit dat nu met het rijke talenpalet in het meertalige Europa, waar heel wat ‘kleinere’ talen als officiële taal worden gebruikt? Hoe overleven deze in de digitale wereld?

De Europese instellingen, waar meertaligheid hoog op de agenda staat, geven antwoord op deze uitdagingen door de uitbouw van een specifieke Europese onderzoeksinfrastructuur voor alle talen. Zo werd er eerst het *CLARIN*-netwerk opgericht. Het acroniem *CLARIN* staat voor *Common Language Resources and Technology Infrastructure*. Het is een Europees programma dat in 2009

is gestart om een duurzame gemeenschappelijke taalinfrastructuur op te zetten voor onderzoek in de geesteswetenschappen. Daarbij gaat het tegelijk om softwareprogramma's en om de inhoud, de digitale taalmaterialen. Sinds 2012 is het project uitgegroeid tot *CLARIN ERIC (European Research Infrastructure Consortium)*. Inmiddels nemen er 19 landen aan deel; Nederland speelde vanaf het begin een belangrijke rol als trekker en telt vier *CLARIN B*-centra: het Meertens Instituut, het Huygens Instituut, het Max Planck Instituut en het Instituut voor de Nederlandse Taal (INT). In Vlaanderen zijn er geen *CLARIN*-centra, maar het INT vervult de rol van *CLARIN B*-centrum¹⁵ voor Vlaanderen en legt contacten met Vlaamse academische instellingen en onderzoeksinstituten. Naast alle voorzieningen voor onderzoekers levert de *CLARIN ERIC*-infrastructuur voor de deelnemende landen en instellingen vele voordelen op. Die structuur overschrijdt disciplines, talen en grenzen. Het zet de talen en culturen van een land op de kaart. En doordat er centrale diensten zijn en iedereen dezelfde standaarden gebruikt, worden er veel kosten bespaard.

Hoe staat het dan met het Nederlands in deze digitale taalinfrastructuur? Om te weten of talen klaar zijn voor de digitale uitdagingen is er een Europees consortium opgericht: META-NET (Multilinguale Europese Technologie Alliantie). Het is een netwerk dat de technologische fundamenteën van een meertalige Europese informatiemaatschappij wil bestuderen en verder versterken. Als talen ook digitaal goed ontwikkeld zijn, dan kunnen deze taaltechnologische hulpmiddelen communicatie en samenwerking over talen heen mogelijk maken en sprekers van om het even welke taal verzekeren van gelijke toegang tot informatie en kennis (via bijvoorbeeld apps op de mobiele telefoon, websites in de moedertaal, spraakgestuurde toepassingen, taalgebaseerde interfaces voor technologie van huishoudelijke elektronica, etc.). META-NET heeft in 2012 onderzoek gedaan naar de stand van zaken van de digitale bronnen voor de verschillende Europese talen. Het Nederlands komt daar helemaal niet slecht uit: er is een sterke onderzoeksgemeenschap in de spraak- en taaltechnologie in Nederland en Vlaanderen en daardoor hebben we veel digitale data voor het Nederlands zoals databanken voor gesproken en geschreven taal, bijhorende software voor de analyse en verwerking van taaldata etc. Al deze taalmaterialen zijn via *CLARIN* en ook via de website van het INT vrij toegankelijk beschikbaar; zo kunnen bedrijven bij het ontwikkelen van nieuwe digitale toepassingen ook het Nederlands meenemen als een van de beschikbare talen. De details van de stand van zaken voor het Nederlands zijn te vinden in een witboek: *The Dutch Language in the Digital Age / Het Nederlands in het Digitale Tijdperk*. META-NET wordt momenteel min of meer

¹⁵ Een *CLARIN B*-centrum is een centrum dat beantwoordt aan de vereiste technische criteria voor de opslag en het beschikbaar stellen van taalmaterialen. Het is de hoogste kwalificatie die er momenteel is.

voortgezet door de *European Language Grid*,¹⁶ waar software en taalbronnen (al dan niet tegen betaling) beschikbaar gesteld worden door bedrijven en andere organisaties. Een en ander wordt ook toegelicht in het rapport *Language Equality in the Digital Age*,¹⁷ dat goedgekeurd werd door het Europees Parlement.

6. De schatkamer van het Nederlands: de Taalmaterialen van het INT

Het Instituut voor de Nederlandse Lexicologie werd in 2016 hervormd tot een breder Instituut voor de Nederlandse Taal. Dit ging samen met de vernieuwde taakstelling van het Algemeen Secretariaat van de Taalunie, dat zich ging toespitsen op beleidsmatige taken, en de uitvoerende taken aan het INT overliet. Voor het Instituut voor de Nederlandse Taal biedt deze overdracht van taken de kans om de eigen werkzaamheden te verbreden. Zo werd het INT het centrale aanspreekpunt voor de bewaring en het wetenschappelijk onderhoud van digitale taalmaterialen en het veiligstellen van allerlei materiaalverzamelingen die met alle variatie van het Nederlands te maken hebben. Deze evolutie weerspiegelt het sterk veranderde landschap van het taalkundig onderzoek: grote taalinfrastructuren worden digitaal opgebouwd, en omvatten corpora, woordenboeken, andere gespecialiseerde lexica en databanken, grammatica en zo veel meer. Het INT streeft ernaar om het best gesorteerde en daarmee zeer goed toegankelijke wetenschappelijk instituut te zijn op het gebied van de Nederlandse taal en haar woordenschat. Het instituut ontwikkelt en levert data voor woordenboeken, (computationele) lexica, corpora en tools. De woordenboeken zijn online te raadplegen. Software en computerlinguïstische tools zijn als open source beschikbaar. Het instituut heeft een centrale positie in het hele Nederlandse taalgebied (Nederland, Vlaanderen, Suriname en de voormalige Antillen) op het vlak van het wetenschappelijk verantwoord ontwikkelen, onderhouden, bewaren en duurzaam beschikbaar stellen van taalmateriaal. Het instituut is daartoe goed uitgerust en beschikt over een groot internationaal netwerk waarbij informatie kan worden uitgewisseld met gelijkaardige instellingen.

Nederland en Vlaanderen hebben een sterke traditie van taalkundig en computerlinguïstisch onderzoek, en dat vertaalt zich in een goed uitgeruste materiaalverzameling.

Via de website van het INT¹⁸ vindt de gebruiker onder het kopje *Taalmaterialen* een bijzonder rijke verzameling. Deze catalogus bevat bronnen, data en tools voor taalkundig onderzoek en taal- en spraaktechnologie (TST) binnen het

¹⁶ Zie <https://www.european-language-grid.eu/>.

¹⁷ Zie http://www.europarl.europa.eu/doceo/document/TA-8-2018-0332_EN.html.

¹⁸ Zie <https://ivdnt.org/>.

Nederlandse taalgebied. Ik noem hier tekstverzamelingen, woordenlijsten, wetenschappelijke woordenboeken, spraakcorpora en taal- en spraaktechnologische software. Bedrijven, onderzoekers, docenten, taalliefhebbers: iedereen kan met deze materialen kennismaken en er gebruik van maken om nieuwe toepassingen te ontwikkelen.

7. Conclusie

De status van een taal wordt bepaald door een waaier aan factoren. Het gaat beslist niet alleen om het aantal moedertaalsprekers; de parameters om een taal te beoordelen als groot of klein zijn veel genuanceerder. Het *global language network* gaat uit van taalkundige publicaties, en uitingen, met name de vertaalmarkt (boekvertalingen), Wikipediapagina's en Twitteksten. De classificaties van *Ethnologue* en Unesco kunnen helpen om talen in kaart te brengen. Of een taal succesvol blijft, heeft dan weer te maken met economische invloed, culturele suprematie en politieke keuzes. Talen kunnen ook digitaal uitsterven; een digitale taalinfrastructuur is belangrijk als ondersteuning voor de vitaliteit van elke taal.

Bibliografie

- Crystal, David. *English as a Global Language*. Cambridge University Press, 2012.
- Dalby, David. *The Linguasphere Register of the World's Languages and Speech Communities Linguasphere*. Press/Gwasg y Byd Iaith, 2000.
- De Tollenaere, Felicien. *Den grooten dictionaris: problematiek der historische lexicografie*. School voor Taal- en Letterkunde, 1977.
- . "Nieuwe wegen in de lexicologie". *Verhandelingen der Koninklijke Nederlandse Akademie van Wetenschappen, afd. Letterkunde*, deel 70, n° 1. Noord-Hollandsche Uitgevers Maatschappij, 1963.
- Shahar, Ronen, et al. *Links that speak: The global language network and its association with global fame*. *PNAS*, 30 december 2014, no. 111 (52) E5616-E5622; first published 15 december 2014, <https://doi.org/10.1073/pnas.1410931111>.
- Ulasziuk, Iryna, et al. *Language Policy and Conflict Prevention*. Brill, 2018.
- Wurm, Stephen. "Language Death and Disappearance: Causes and Circumstances". *Endangered Languages*, geredigeerd door Robert H. Robins en Eugenius M. Uhlenbeck. Berg, 1991, pp. 1–18.

Internetbronnen

- European Language Grid*. <https://www.european-language-grid.eu/>. Geraadpleegd 17 augustus 2021.
- Instituut Voor de Nederlandse taal*. <https://ivdnt.org/>. Geraadpleegd 17 augustus 2021.
- UNESCO*. <http://www.unesco.org/xtrans/>. Geraadpleegd 17 augustus 2021.

- Van Wayenberg, Bruno. “Nederlands is een grote kleine taal”. *NRC*, 16 december 2014, <https://www.nrc.nl/nieuws/2014/12/16/nederlands-is-een-grote-kleine-taal-1448396-a1078891>. Geraadpleegd 17 augustus 2021.
- z.a. “European Parliament resolution of 11 September 2018 on language equality in the digital age (2018/2028(INI))”. *European Parliament*, http://www.europarl.europa.eu/doceo/document/TA-8-2018-0332_EN.html. Geraadpleegd 17 augustus 2021.
- z.a. “Feiten en cijfers. Wat iedereen zou moeten weten over het Nederlands”. *Taaluniversum*, <http://taaluniversum.org/inhoud/feiten-en-cijfers#feitencijfers>. Geraadpleegd 17 augustus 2021.
- z.a. “Huishoudens met een breedbandconnectie”. *Statistiek vlaanderen*, <https://www.statistiekvlaanderen.be/nl/huishoudens-met-een-breedbandconnectie>. Geraadpleegd 17 augustus 2021.
- z.a. “Nederland koploper in Europa met internettoegang”. *CBS*, <https://www.cbs.nl/nl-nl/nieuws/2018/05/nederland-koploper-in-europa-met-internettoegang>. Geraadpleegd 17 augustus 2021.