

/instituut voor de
Nederlandse taal/

jaarverslag

2022

verslag aangaande de werkzaamheden,
gebeurtenissen enz. in het afgelopen jaar;
jaarlijks verslag

(Woordenboek der Nederlandsche Taal)

Jaarverslag 2022

Instituut voor de Nederlandse Taal

Inhoud

Inhoud	3
Inleiding: het jaar 2022.....	6
1. Hedendaags Nederlands.....	9
1.1 Infrastructureel	9
1.1.1 <i>GiGaNT-Molex</i>	9
1.1.2 <i>Corpus Hedendaags Nederlands</i>	9
1.2 Lexicale beschrijving van het hedendaags Nederlands.....	11
1.2.1 <i>Onderzoek naar nieuwe wegen</i>	11
1.2.2 <i>Algemene woordenschat</i>	11
1.2.3 <i>Vertaalwoordenboeken</i>	13
2. Historisch Nederlands.....	14
2.1 Infrastructureel	14
2.1.1 <i>Historisch woordenboekportaal</i>	14
2.1.2 <i>GiGaNT-Hilex</i>	15
2.1.3 <i>Semantisch lexicon DiaMaNT</i>	15
2.1.4 <i>Historische corpora</i>	16
2.1.5 <i>Etymologie</i>	16
2.1.6 <i>Lexicon Frisicum</i>	16
3. Beschrijving van de Nederlandse dialecten.....	17
3.1 Elektronische Woordenbank van de Nederlandse dialecten (eWND).....	17
3.2 Database van de Zuidelijk-Nederlandse Dialecten (DSDD)	17
3.3 Gesproken Corpus van de Zuidelijk-Nederlandse Dialecten	18
3.4 Prijzen en projecten	18
4. Terminologie	18
4.1 Termenlijsten.....	18
4.2 Tools.....	19
4.3 Veldondersteuning	19

4.4	Medische vaktaal, juridische vaktaal en Nederlands als wetenschapstaal.....	19
4.5	Terminologisch netwerk	21
5.	Grammatica	21
5.1	Taaladvies	22
5.2	e-ANS.....	22
5.3	Taalportaal.....	23
6.	Taalmaterialen.....	23
6.1	CLARIN-ERIC; het INT als CLARIN-centrum	24
6.2	European Language Resources Coordination Initiative (ELRC).....	24
6.3	IMPACT-centrum en digitization.eu	25
7.	Overige infrastructuur- en netwerkprojecten	25
7.1	European Lexicographic Infrastructure (ELEXIS).....	25
7.2	CLARIAH+ Nederland.....	25
7.3	CLARIAH Vlaanderen	26
7.4	SignOn-project	27
7.5	SABeD – Spoken Academic Belgian Dutch.....	27
7.6	COST – European network for Web-centered linguistic data science	28
7.7	COST - Universality, diversity and idiosyncrasy in language technology (UniDive, 2022-2026).....	28
8.	Plaats van het instituut in Nederland, Vlaanderen en de wereld	29
8.1	Doelgroepenbeleid (inclusief onderwijs).....	29
8.2	Communicatiemiddelen	29
8.2.1	Website	29
8.2.2	Huisstijl	31
8.2.3	Nieuwsbrieven	31
8.2.4	Sociale media	31
8.2.5	Podcasts	32
8.3	Populairwetenschappelijke activiteiten.....	32
Bijlage 1: Raad van Toezicht en Raad van Advies		34
	<i>Raad van Toezicht</i>	<i>34</i>

<i>Raad van Advies</i>	34
Bijlage 2: Medewerkers	35
Bijlage 3: Publicaties, lezingen, media, prijzen etc.	36
<i>Publicaties</i>	36
<i>Lezingen en presentaties</i>	42
<i>Congressen & workshops</i>	46
<i>Onderwijs</i>	46
<i>In de media</i>	46
<i>Diversen</i>	48
<i>Begeleiding thesis, stage, vrijwilligers</i>	51
Bijlage 4: Taalmaterialen	52
Bijlage 5: Besteding lumpsum basistakenpakket	55

Inleiding: het jaar 2022

Na de hervormingen van 2016 is het Instituut voor de Nederlandse Taal (INT) voluit aan de slag gegaan met alle taken die het moet uitvoeren. Het meerjarenbeleidsplan (2018-2022) werd uitgevoerd, in 2022 weer met een aantal nieuwe accenten en dit in nauw overleg met de Nederlandse Taalunie (NTU). We verwijzen verder op in dit verslag naar de Taalunieoverdragtaken die door het INT worden uitgevoerd.

De internationale visitatie van 2021 was een goede toets om te zien hoe het hervormde instituut de waaier van taken had ingevuld. Het eindrapport beoordeelde het INT als excellent en beschreef ons als hét kennisinstituut voor de Nederlandse taal. Ook het Comité van Ministers nam kennis van het rapport en sprak lof uit voor onze activiteiten.

De visitatiecommissie geeft in haar rapport een kwalitatieve en kwantitatieve beoordeling voor het werk van het INT in de afgelopen vijf jaar (2016-2021). Het rapport is duidelijk: het proces om het instituut vlot te trekken na de reorganisatie was een uitdaging; met minder geld en een zeer beperkte staf moesten heel wat taken worden vervuld. Daarbij stipt de commissie aan dat ook de wetenschappelijke inzichten rond lexicografie, taalkunde en de productie van digitale taalmaterialen aan grondige verandering onderhevig waren. We kunnen ons vinden in de aanbevelingen van de commissie. De commissie stelt dat het aantal taken en de prestaties van het INT indrukwekkend zijn, maar dat zij zich zorgen maakt of dit duurzaam kan zijn nu de financiering achterblijft. De structurele financiering en het weerstandsvermogen zijn te laag waardoor een verdere ontwikkeling van het instituut problematisch kan worden. Tegenslagen in projecten kunnen nauwelijks worden opgevangen en de belasting van het personeel kent haar grenzen ondanks dat de commissie vol lof is over de betrokkenheid, bevlogenheid en deskundigheid van het personeel. De aanbeveling om de taken van het personeel goed te bewaken wordt ter harte genomen en de planning van de projecten nauwkeurig gevolgd. De commissie vindt onze benadering van het uitbouwen van de taalinfrastructuur zeer innovatief en moedigt ons aan om op deze weg verder te gaan. Het lexicografische werk dat op deze infrastructuur voortbouwt, heeft zich met succes gediversifieerd in kleinere projecten die op verschillende manieren aan elkaar gekoppeld zijn. Door deze nieuwe benadering en modulaire aanpak kunnen hedendaagse, historische, dialectologische, morfologische, syntactische en semantische informatie enz. voor een bepaald woord bijeengebracht worden in verschillende maar nauw samenwerkende projecten en kan deze informatie afhankelijk van de doelgroep op diverse manieren online aangeboden worden. Op deze manier kan het INT zijn opdracht vervullen en moderne wetenschappelijke onderbouwde lexicografische producten blijven aanbieden.

De beperkte financiële middelen worden meermaals door de commissie onder de aandacht gebracht, ook al is het INT succesvol in het aanvragen van Europese financiering. Zij adviseren om het instituut toegang te geven tot nationale financiering, zoals NWO en FWO middelen, zodat ook jonge onderzoekers bij het instituut aan de slag kunnen. Het instituut zou dit toejuichen omdat we nu alleen als secundaire partner in projecten kunnen worden betrokken. Er zal worden gezocht naar verdere projectfinanciering.

Het advies om de samenwerking met alle belangengroepen te formaliseren via reguliere ontmoetingen, presentatie van casussen etc. wordt ter harte genomen. Tegelijk willen we meer aandacht hebben voor de gebruikers en hun noden en werken aan tevredenheidsenquêtes. We belichten dit thema verder in rubriek 9 van dit jaarverslag.

Het INT heeft als structureel gefinancierd 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 voert daarbij een aantal taken uit het Taalunieverdrag uit. We verwijzen hier naar hoofdstuk 1 uit het Taalunieverdrag, artikelen 2, 3, 4 en 5. 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 deze samen met de eigen taalmaterialen duurzaam ter beschikking via repository's, websites, API's en als open-source software. Het INT promoot die taalinfrastructuur bij onderzoekers, ontwikkelaars en het brede publiek om zo onderzoek en andere activiteiten rond de Nederlandse taal te stimuleren en te ondersteunen. Daarnaast heeft het INT als toegepast onderzoeksinstituut ook de doelstelling de kennis en expertise over taalinfrastructuur verder uit te bouwen door eigen wetenschappelijk onderzoek.

Wat betreft het Taalunieverdrag draagt het INT bij aan de ontwikkeling en bevordering van de kennis van het Nederlands, en ook de studie en verspreiding ervan. We ontwikkelen en updaten de spellingapplicatie woordenlijst.org en zijn verantwoordelijk voor de Algemene Nederlandse Spraakkunst. Door onze rol als CLARIN-instituut voor Nederland én Vlaanderen moedigen we het wetenschappelijk onderzoek aan. Verder wordt door het Expertisecentrum Nederlandstalige Terminologie gewerkt aan de promotie van het Nederlands als wetenschapstaal en de verspreiding ervan in databanken en woordenlijsten. Daarnaast participeert het INT in projecten op nationaal en internationaal niveau en houdt het zo de vinger aan de pols wat betreft de ontwikkelingen in de lexicografie en de computationele taalkunde.

Bij de aanvang van 2022 werkten er 35 werknemers (28,45 fte). Eind december 2022 waren dat 33 personeelsleden (26,85 FTE). Er waren enkele verschuivingen door pensionering van één werknemer en het uit dienst gaan van de dataconversiemedewerker. Daarnaast gingen twee medewerkers van een 80%-aanstelling naar een 100%-opdracht.

In de Raad van Toezicht beëindigde de voorzitter, Paul Rupp, zijn mandaat op 1 december en werd een nieuwe voorzitter aangesteld, Mieke Zaanen.

Het INT is hét kennisinstituut voor het Nederlands en neemt een centrale positie in voor het hele Nederlandse taalgebied (Nederland en de Caribische rijkdelen, Vlaanderen en Suriname) op het vlak van het wetenschappelijk verantwoord ontwikkelen, bewaren en duurzaam beschikbaar stellen van taalmateriaal. Het INT streeft ernaar om het best gesorteerde wetenschappelijk instituut te zijn op het gebied van de Nederlandse

taal en de woordenschat, gecombineerd met hoge kwaliteit en goede toegankelijkheid. Het instituut ontwikkelt en levert data voor woordenboeken, (computationele) lexica, corpora en tools. De woordenboeken zijn online te raadplegen. Software en computerlinguïstische tools zijn open-source beschikbaar. Het instituut speelt in op de nieuwe ontwikkelingen in de geesteswetenschappen, met name op het terrein van de digital humanities. Om deze rol te kunnen vervullen beheert en onderhoudt het INT een digitale infrastructuur voor het Nederlands, met aandacht voor taalvariatie (terminologie, dialecten etc.). Zowel academische als niet-academische partijen kunnen gebruikmaken van deze infrastructuur.

Er stonden in 2022 een heleboel activiteiten op het programma. Er werd vooral ingezet op het verbeteren van de digitale infrastructuur van het Nederlands, zoals in het beleidsplan werd aangekondigd. Dit bouwt voort op de inzichten van de Lorentzworkshop, waardoor een optimale structuur voor de taaldata, zowel historisch als hedendaags, wordt gerealiseerd.

In 2022 werkten we dit uit in een nieuw meerjarenbeleidsplan voor de periode 2023-2027. Hierover werd uitvoerig van gedachten gewisseld met het senior team; ook met de Raad van Advies werden verschillende vergaderingen belegd om de prioriteiten voor de komende vijf jaar vast te leggen.

Eveneens namen we de leiding in het ELRC-consortium (European Language Resources Coordination) voor Nederland en organiseerden we op 11 juni in dit kader een workshop. Daarnaast werden we verantwoordelijk voor het ELE-project (European Language Equality) en deden we een bevraging naar de stand van zaken van de digitale taalmaterialen en softwaretools voor het Nederlands in Nederland en Vlaanderen. De resultaten van dit onderzoek werden in 2022 worden gepubliceerd in een nieuwe whitepaper 'The Dutch Language in the Digital Age'.

De lijn van webinars en online presentaties die we sinds de pandemie hebben uitgewerkt werd nog versterkt door een nieuwe reeks podcasts in samenwerking met Onze Taal, waarin we experts aan het woord laten over één thema binnen de taalkunde, zoals dialectologie, neologismen, terminologie, grammatica, etc. Deze podcasts worden gretig beluisterd door een breed publiek, ook door studenten en docenten Nederlands in het buitenland.

Februari 2023

prof. dr. Frieda Steurs

wetenschappelijk directeur/bestuurder

1. Hedendaags Nederlands

1.1 Infrastructureel

1.1.1 GiGaNT-Molex

GiGaNT (Groot Geïntegreerd Lexicon van de Nederlandse Taal) is het computationele woordvormenlexicon dat de door het INT beschreven woordenschat bevat van het Nederlands van de 6e eeuw tot nu, en de centrale database vormt van het INT. Werkzaamheden aan deze centrale data-infrastructuur zijn modulair opgezet en worden dienovereenkomstig uitgevoerd. GiGaNT bestaat uit twee grote componenten, GiGaNT-Molex, dat de kern is van de formele beschrijving van de hedendaagse woordenschat, en GiGaNT-Hilex, de historische lexiconcomponent, die onderdeel vormt van de lexicografische infrastructuur voor het historisch Nederlands. Beide modules gebruiken het modern Nederlands lemma + woordsoort als ingang.

De moderne lexiconcomponent, GiGaNT-Molex, is de centrale ruggengraat voor de lexicale beschrijving van het modern Nederlands, en wordt voortdurend uitgebreid met nieuwe data, niet alleen in de zin van toegevoegd vocabulaire, maar ook door uitbreiding van de features, zoals in 2020-2021 met toegevoegde werkwoordkenmerken die het project Woordcombinatie ten goede komen en het toevoegen van uitspraak informatie (zie hiervoor het onderdeel Spelling). Het lexicon is beschikbaar als dataset onder een niet-commerciële en commerciële licentie, en via de lexiconservice, een API (Application Programming Interface) die toegang geeft tot het lexicon.

In 2022 is gewerkt aan een nieuwe versie van de Spelling-API (zie bij het onderdeel spelling), aan uitbreiding van de structuur voor betere integratie van meerwoordscombinaties, met name vanuit het project Woordcombinaties en vanuit terminologie, en aan het aanpassen van de woordsoortinformatie aan de Tagset Diachron Nederlands. Laatstgenoemde werkzaamheden, die ook voor GiGaNT-Hilex worden gedaan, hebben de nodige impact gehad op de Molexdata (ruim 18.000 wijzigingen zijn aangebracht in de lemmawoordsoort, bijvoorbeeld) en de daaraan gekoppelde bronnen. De koppelingen worden waar nodig bijgewerkt.

De aanpassing van de GiGaNT-woordsoortinformatie aan de TDN maakt het noodzakelijk de koppeling tussen GiGaNT-Molex en GiGaNT-Hilex opnieuw te bekijken. Deze revisie is in volle gang.

Tot slot is versie 2.0 van het lexicon ter beschikking gesteld via taalmaterialen.ivdnt.org.

1.1.2 Corpus Hedendaags Nederlands

De moderne lexiconbouw en de beschrijving van het hedendaags Nederlands wordt gebaseerd op modern corpusmateriaal. Het Corpus Hedendaags Nederlands (CHN) bevat al het hedendaags Nederlands corpusmateriaal waarover het INT beschikt. Het corpus wordt continu uitgebreid. Het bestaat uit materiaal uit Nederland, Vlaanderen, Suriname en de Antillen. De samenstelling wordt mede bepaald door de projecten die van de corpusdata gebruikmaken. Omdat we de Nederlandse taal systematisch willen monitoren, proberen we data te verwerven die systematisch en continu aangeleverd worden. Kranten, sociale media en ook parlementair materiaal zijn daar een voorbeeld van. De werkzaamheden bestaan enerzijds uit acquisitie, dataprocessing en

taalkundige verrijking, en anderzijds uit het beschikbaar stellen van het materiaal in een corpusapplicatie voor intern en extern gebruik.

Data

Het interne corpus bevat inmiddels ruim 2,4 miljard tokens en ruim 5,3 miljoen documenten, en wordt wekelijks geüpdatet; het externe corpus bevat ruim 1,15 miljard woorden en ruim 3 miljoen documenten en wordt maandelijks geüpdatet. Naast de reguliere updates van het CHN is de belangrijkste realisatie van 2022 de overeenkomst met DPG Media, waardoor nu wekelijks ook Nederlandse kranten als *de Volkskrant*, *Algemeen Dagblad*, *Parool*, *Trouw* en Vlaamse kranten als *De Morgen* en *Het Laatste Nieuws* aan het CHN worden toegevoegd, via automatische conversie en indexering. Samen met de Nederlandse *NRC (Handelsblad)* en de Vlaamse kranten *Het Nieuwsblad*, *De Gazet van Antwerpen*, *Het Belang van Limburg* en *De Standaard* vormen zij een monitorcorpus van kranten. Intern hebben we wekelijks een update van al het materiaal dat de week ervoor is binnengekomen. Voor onze externe gebruikers zijn er maandelijks updates. Omdat de toevoeging van de diverse kranten op diverse tijdstippen is gebeurd, zijn we gestart met het chronologisch opvullen van gaten in de tijd en werken we aan een diachroon monitorcorpus.

Het ParlaMint-project heeft in 2022 een vervolg gekregen, waardoor het materiaal van het Belgisch federaal parlement verder is aangevuld. We hebben nu data van de periode 2014/06/19 - 2022/07/13. Op termijn zullen die data ook aan het CHN worden toegevoegd.. Tot slot is er een afspraak gemaakt voor levering van het Surinaamse tijdschrift *De Parbode*.

Workflow en verrijking

In het kader van de integratie in het lexicale productieproces is verder onderzoek gedaan naar de vereisten op het gebied van syntactische verrijking – daarbij gaat het zowel om de verrijkingslaag zelf als om de zoekmogelijkheden die de corpusapplicatie op dit gebied moet bieden.

Voorts zijn op basis van de in de literatuur gerapporteerde accuratesse een aantal *universal dependency pipelines* geselecteerd en in de praktijk getest (met name [UDify](#), [spaCy](#) en [Trankit](#)). SpaCy en Trankit zijn hertraint op het naar CONLL-U geconverteerde Lassy-Klein-corpus, waarbij Trankit de hoogste accuratesse haalde. Er is een eerste versie van een in de corpusbuilding-workflow inschakelbare webservice ontwikkeld.

Corpusapplicatie

Om efficiënter te kunnen omgaan met de steeds groeiende hoeveelheid materiaal is gewerkt aan het mogelijk maken van een gedistribueerde opzet door de applicatie in Apache Solr te integreren. Dit is onder andere noodzakelijk om het CHN efficiënter te kunnen inzetten in de workflow voor de beschrijving van neologismen.

1.2 Lexicale beschrijving van het hedendaags Nederlands

1.2.1 Onderzoek naar nieuwe wegen

Voortbouwend op de feedback in het visitatierapport en met het oog op het opstellen van het nieuwe meerjarenbeleidsplan voor de periode 2023-2027 werd in de loop van 2022 een grondige analyse gemaakt van de verschillende lexicografische activiteiten aan het instituut zowel voor het hedendaags als het historisch Nederlands en met een speciale aandacht voor nieuwe gebruikersnoden en recente evoluties binnen de internationale lexicografie. Op basis daarvan werd een strategie ontwikkeld om in de komende jaren zowel intern efficiëntiewinsten te realiseren als voor externen innovatieve toepassingen te ontwikkelen dankzij een integratie van de woordenschatbeschrijving in een centrale kennisbank voor de Nederlandse woordenschat en door een verdere optimalisatie van de corpusgebaseerde lexicografische workflow. Een uitgebreide beschrijving van de visie en strategische planning is te vinden in het Meerjarenbeleidsplan 2023-2027, paragraaf 4 (Beschrijving van de woordenschat door de eeuwen heen: naar een centrale kennisbank), bijlage II (Werkzaamheden voor de uitbouw van de kennisbank) en bijlage III (Wetenschappelijke visie op lexicografie).

1.2.2 Algemene woordenschat

De beschrijving van de algemene woordenschat gebeurt op drie verschillende terreinen: de spelling van woorden, de betekenis en de constructies waarin ze voorkomen. De informatie wordt beschikbaar gesteld via respectievelijk *woordenlijst.org*, *anw.ivdnt.org* en *woordcombinaties.ivdnt.org*. In 2021 is daaraan nog *neologismen.ivdnt.org* toegevoegd.

Spelling

Net als de voorbije jaren lag bij de spellingwerkzaamheden in 2022 vooral de focus op kwaliteit. Reguliere updates werden dit jaar niet ingepland, zodat meer tijd en ruimte vrijkwam voor het optimaliseren van de al beschikbare data. Het kwaliteitsaspect werd ook doorgetrokken naar de dienstverlening: (spelling)vragen van gebruikers werden vakkundig beantwoord en waar nodig werden ze doorverwezen naar een meer geschikte persoon of dienst.

Daarnaast is er verder gewerkt aan het toevoegen van uitspraak informatie aan de online woordenlijst.

De Commissie Spelling werd in 2022 verder door het INT ondersteund in haar werkzaamheden. Hiervoor werden o.a. allerlei trefwoorden- en andere selecties aangeleverd. Er waren in 2022 zowel overlegmomenten online (2) als vergaderingen in Breda (3).

Tot slot is de nieuwe publieke zoekmachine voor de lexicaal data (Spelling-API en *woordenlijst.org*) aan de Taalunie voorgelegd. De nieuwe versie maakt het onder andere eenvoudiger nieuwe functionaliteit toe te voegen: zoeken op meerdere kenmerken, flexibeler omgaan met de presentatie van resultaten, relevantere suggesties bij het zoeken, en het gebruik van uitspraak informatie.

Semantische beschrijving

Algemeen Nederlands Woordenboek

Het *Algemeen Nederlands Woordenboek* (ANW) is een corpusgebaseerd, digitaal, multimediaal woordenboek van het eigentijdse Nederlands in Nederland en de Caribische rijkdelen, Vlaanderen en Suriname. De taalperiode die het ANW bestrijkt, loopt van 1970 tot heden. Naast de woorden uit de kernwoordenschat van de genoemde taalgebieden wordt in het ANW ook een selectie van door het INT verzamelde neologismen (nieuwe woorden, nieuwe verbindingen, nieuwe uitdrukkingen, nieuwe betekenissen van al bestaande woorden) opgenomen.

Door de ANW-redactie zijn in 2022 799 volledig bewerkte, nieuwe artikelen online gezet. Daarvan kwamen 333 artikelen alleen in het ANW terecht, 358 in zowel het ANW als het *Woordenboek van Nieuwe Woorden* (WNW) en 108 alleen in het WNW. Uit het ANW zijn in 2022 ook artikelen verwijderd. Het gaat om neologismen met een korte levensduur (de ‘eendagsvliegen’) die zijn overgeheveld naar het WNW.

Eén artikel kan meerdere lemmata beschrijven, bv. als er sprake is van gelijkvormige woorden waarvan de woordsoort of etymologische herkomst verschilt. Alle verschillende lemmata binnen de nieuwe artikelen zijn door de ANW-redactie ook (toegevoegd en) gelinkt aan GiGaNT-Molex, het grote computationele lexicon van het hedendaags Nederlands dat ook de basis vormt van de *Woordenlijst Nederlandse Taal*.

Behalve aan het ANW en WNW werkte de redactie ook mee aan GiGaNT-Molex, o.a. vanuit het perspectief van de inhoudelijke consistentie tussen het ANW en het centrale lexicon. Ze voerde ook webredactionele taken en correctie- en vertaalwerkzaamheden uit. Verder schreef de redactie mee aan de wekelijkse rubriek ‘Nieuw woord van de week’.

In november is Rob Tempelaars, de projectleider van het ANW, met pensioen gegaan. Hij werd opgevolgd door ANW-redacteur Boukje Verheij.

Woordenboek van Nieuwe Woorden

Sinds 2018 wordt er op het INT gewerkt aan het *Woordenboek van Nieuwe Woorden* (WNW): een onlinewoordenboek waarin woorden die vanaf het jaar 2000 zijn ontstaan, worden beschreven. Het WNW neemt niet alleen de beklijvende nieuwe woorden op (bijvoorbeeld *app*, *selfie*, *keuzestress* en *ontspullen*), maar juist ook veel nieuwe woorden die kortere tijd bestaan (zoals *appaire* en *stookschaamte*). Ook in 2022 is er verder gewerkt aan dit woordenboek. Het WNW bevatte eind 2022 ruim 9600 woorden. Het WNW wordt inmiddels dagelijks geüpdatet.

Alle woorden in het WNW zijn op dit moment digitaal gekoppeld aan de bijbehorende lemma’s in het spellingbestand Molex en daarmee ook aan de bijbehorende ingangen in de *Woordenlijst Nederlandse Taal*. Ook afgelopen jaar is er iedere week een neologisme behandeld in de rubriek ‘Nieuw woord van de week’. Verder publiceerde de WNW-redactie enkele artikelen over nieuwe woorden, hield de redactie veel lezingen, webinars en interviews (op radio en tv) over dit onderwerp en begeleidde zij diverse profielwerkstukken over neologismen.

Nederlands als vreemde taal

Woordcombinaties

Woordcombinaties is een onlinetaaltool in ontwikkeling die leerders van het Nederlands als vreemde taal ondersteunt bij het gebruiken van woorden in context. Het project integreert de beschrijving van collocaties, idiomen en syntactische patronen en behandelt in de eerste fase de combinatoriek van werkwoorden. Het project is corpusgebaseerd en gebruikt de Sketch Engine als corpusquerytool (www.sketchengine.co.uk). *Woordcombinaties* toont hoe woorden gebruikt worden in goede voorbeeldzinnen, welke woorden met elkaar gecombineerd worden (collocaties of combinatiemogelijkheden) en hoe valentiepatronen en andere patronen samen met collocaties gebruikt worden voor het bouwen van zinnen. Taalleerders leren niet alleen woorden kennen, maar ook woorden gebruiken in context.

In het verslagjaar zijn patronen bewerkt van werkwoorden uit de schooltaalwoordenschat en het Frequentiewoordenboek Nederlands (Tiberius, C. en T. Schoonheim (2013)). Daarnaast zijn ook voorbeeldzinnen en combinaties bewerkt van meer werkwoorden en is begonnen met de combinatoriek van de substantieven. Het datamodel voor de idiomen is ontwikkeld en zal in 2023 in de editor geïmplementeerd kunnen worden.

Voor de bewerkingsomgeving zijn werkzaamheden uitgevoerd om een efficiëntere bewerking van de idiomen mogelijk te maken. Daarnaast zijn er werkzaamheden uitgevoerd voor een update van de gebruikersapplicatie begin 2023.

Oefenen.nl

Oefenen.nl wil een app *EenvoudigNL* realiseren voor woordenschat en zinnen voor anderstaligen. De inzet is om te werken vanuit een corpus 'Eenvoudige taal' zodat mensen onbepaald kunnen blijven oefenen met woorden en zinnen. De app voorziet in oefeningen op taalniveau A2 van het Raamwerk NT2 en in een plusversie waarin gewerkt kan worden aan woordenschat richting taalniveau B1. De rol van het INT is om de infrastructuur te bouwen die data aan de app kan leveren. Er is in 2021 gestart met een onderzoek naar hoe die infrastructuur eruit moet zien. Er zal een proof of concept worden gemaakt waarin twee thema's worden uitgewerkt. Deze werkzaamheden zijn in 2022 voortgezet. De opschaling van e.e.a., het verder automatiseren en optimaliseren van de workflow en het in productie nemen is een substantieel project, waarvoor extra financiering gezocht moet worden.

1.2.3 Vertaalwoordenboeken

In september 2017 heeft het INT het online platform Vertaalwoordenschat gelanceerd. Via dit platform worden de tweetalige bestanden ontsloten die in de afgelopen decennia, onder meer in opdracht van de Commissie Lexicologische Vertaalvoorzieningen (CLVV, 1993-2003), zijn ontwikkeld voor taalparen die op de commerciële markt niet spontaan aan bod kwamen.

Inmiddels staan het Nederlands-Nieuwgrieks / Nieuwgrieks-Nederlands, het Nederlands-Portugees / Portugees-Nederlands, het Nederlands-Estisch en het Nederlands-Fins / Fins-Nederlands online. De woordenboeken zijn gratis raadpleegbaar op vertaalwoordenschat.ivdnt.org en in de app Vertaalwoordenschat, gratis te downloaden uit de Google Play Store of Apple App Store.

In 2022 is de nieuwe redactieomgeving voor de Vertaalwoordenschat opgeleverd. Deze is in de loop van het jaar gebruikt door twee externe redacteurs voor het verwerken van de opmerkingen over het Nieuwgrieks-Nederlands / Nederlands-Nieuwgrieks die na het verschijnen van het papieren woordenboek in 2008 door de UvA zijn verzameld. Deze correcties zullen in een volgende update aan de applicatie worden toegevoegd. Daarnaast is de applicatie verder uitgebreid zodat extra informatie zoals flexiecodes in de toekomst ook online getoond kan worden. Tot slot is in 2022 de ontsluiting van Nederlands-Deens / Deens-Nederlands voorbereid. Dit nieuwe taalpaar zal in het eerste kwartaal van 2023 aan de applicatie worden toegevoegd.

Op 30 maart 2022 is de *Transcriptor 2.0* gelanceerd bij het INT: dit is een tool die automatisch Russische en Oekraïense personen- en plaatsnamen omzet naar de spelling in het Nederlands en naar enkele andere talen; de Transcriptor staat op: <https://transcriptor.nl/> en <https://transcriptor.be/>.

2. Historisch Nederlands

2.1 Infrastructureel

2.1.1 Historisch woordenboekportaal

De beschrijving van de historische woordenschat van het Nederlands is te vinden in de verschillende historische woordenboeken van het INT. Deze zijn online beschikbaar in het historische woordenboekenportaal (gtb.ivdnt.org). Het betreft het *Oudnederlands Woordenboek* (ONW), het *Vroegmiddelnederlands Woordenboek* (VMNW), het *Middelnederlandsch Woordenboek* (MNW) en het *Woordenboek der Nederlandsche Taal* (WNT).

Om de rijkdom van deze historische woordenboeken en van de webapplicatie waarin ze te raadplegen zijn nog meer onder de aandacht te brengen, werd net als in voorgaande jaren in de webrubriek ‘Terug in de Taal’ een twintigtal bijdragen rond aansprekende historische woorden en begrippen gepubliceerd. In het gelegenheidswoordenboekje *Struinen in Streuvels* is een selectie opgenomen van voornamelijk West-Vlaamse woorden uit het werk van Streuvels die in het WNT terecht zijn gekomen. Uit de loggegevens én uit de reacties en vragen die het instituut regelmatig zowel van specialistische als van niet-specialistische gebruikers ontvangen, blijkt dat de webapplicatie ondertussen een veelgebruikt werkinstrument geworden is.

De woordenboekdata vormen de kern van de historische lexiconmodule van GiGaNT (GiGaNT-Hilex). In GiGaNT-Hilex vinden we de volgende woordenboekelementen: historisch lemma, modern Nederlands equivalent, woordsoort van het lemma, geattesteerde woordvormen van het lemma in de citaten en de citaten zelf. De werkzaamheden die ook in 2022 in dit kader zijn uitgevoerd, leveren correcties op die ook voor de online woordenboeken relevant zijn. Als de conversie van de onderliggende data naar P5 is afgerond – de werkzaamheden daarvoor zijn eind 2022 weer opgepakt – kunnen deze correcties in de woordenboekbestanden worden doorgevoerd. Hetzelfde geldt voor de vertalingen in het Duits en Engels van de definities van het ONW, die dit jaar zijn afgerond.

Citatenmateriaal van het WNT vormt ook onderdeel van het in ontwikkeling zijnde trainingmateriaal voor het verrijken van historisch corpusmateriaal (zie onder).

2.1.2 GiGaNT-Hilex

GiGaNT is het centrale lexicon waaraan alle lexicale producten van het INT gekoppeld worden. Het historische onderdeel GiGaNT-Hilex is gebaseerd op de historische woordenboeken. Het huidige Hilex bevat het materiaal uit het *Woordenboek der Nederlandsche Taal* en het *Middelnederlandsch Woordenboek*. De koppeling met de online woordenboeken is behouden. Het lexicon is voor onderzoekers en applicatieontwikkelaars beschikbaar via de lexiconservice (een API die toegang geeft tot het lexicon) in maandelijkse updates.

Omdat binnen GiGaNT-Hilex met striktere lemmatiseringsprincipes wordt gewerkt dan in de woordenboeken waarop het lexicon is gebaseerd, vinden er allerlei herschikkingen plaats van de lexiconinhoud. Deze werkzaamheden zijn ook in 2022 voortgezet.

De woordsoorttoekenning in Hilex was tot en met 2021 nog gebaseerd op de informatie uit de woordenboeken, die mede bron van onderzoek zijn geweest voor de ontwikkeling van de Tagset Diachron Nederlands (TDN). In 2022 is er, niet alleen voor GiGaNT-Molex maar ook voor GiGaNT-Hilex, gewerkt aan het conform maken van het lexicon aan de TDN. Hierdoor was het noodzakelijk om de koppelingen tussen de modules te reviseren.

Ook in 2022 werd de lexiconservice, waarmee GiGaNT-Hilex toegankelijk wordt gemaakt voor gebruik in bijvoorbeeld corpusapplicaties als Nederlab, Delpher of het Corpus Middelnederlands, maandelijks geüpdatet.

2.1.3 Semantisch lexicon DiaMaNT

DiaMaNT bouwt een betekenislaag op GiGaNT-Hilex en heeft als doel een hulpmiddel te bieden bij tekstontsluiting en bij het onderzoek naar begrippen door de eeuwen heen. Ook voor DiaMaNT vormen de historische woordenboeken de belangrijkste bron. In de afgelopen jaren is gewerkt aan het opzetten

van een infrastructuur voor de lexiconbouw op basis van woordenboek- en corpusdata. De in het lexicon opgenomen data zijn nu merendeels gebaseerd op het oplossen van de synoniemverwijzingen in de MNW- en WNT-data naar uit het WNT afkomstige lemmata en betekenissen. Het resultaat is, mede in het kader van CLARIAH, gepubliceerd als linked open data en gepubliceerd in een eerste versie van een applicatie voor het bredere publiek.

Het voor 2022 voorziene werk aan DiaMaNT is uitgesteld ten gunste van het mede in het kader van het beleidsplan 2023-2027 verrichte onderzoek naar de verdere uitbouw van de lexicale infrastructuur. De betekenisinventaris van DiaMaNT zal een spil zijn binnen het te ontwikkelen betekenisregister en zal in synergie daarmee verder worden ontwikkeld.

2.1.4 Historische corpora

In 2022 is, mede in het kader van CLARIAH+, verder gewerkt aan de infrastructuur voor historisch Nederlands, en met name ook aan het toepassen van de TDN-tagset op training- en evaluatiemateriaal voor de taalkundige verrijking (15e-19e eeuw), zie ook het onderdeel CLARIAH+.

De 17^e-eeuwse kranten die door Nicoline van der Sijs in samenwerking met het Meertens Instituut en een groep vrijwilligers zijn gedigitaliseerd, zijn in 2022 op het INT verrijkt en omgezet naar een corpusapplicatie (ook weer mede in het kader van CLARIAH+ gecureerd): het [Couranten Corpus](#). Om de kranten die ondertussen door de KB zijn gescand, toe te voegen is er een nieuw vrijwilligersproject opgestart waarbij gebruikgemaakt wordt van Transkribus.

Verder is het Corpus Oudnederlands online gepubliceerd, en is gewerkt aan de curatie van het corpus Gekaapte Brieven. Dit zal in 2023 worden gepubliceerd in de corpusapplicatie.

Aan de ontsluiting van de Digitale Pallas, een 18e-eeuws Russisch meertalig woordenboek, hebben in 2022 drie stagiaires meegewerkt, waaronder een Oekraïense en een Russische.

2.1.5 Etymologie

Sinds 2020 zijn de etymologiebank (etymologiebank.nl), en de uitleenwoordenbank (uitleenwoordenbank.ivdnt.org) in het beheer van het INT. In 2022 hebben vier stagiaires een bijdrage geleverd aan de verrijking van de etymologiebank. Verder is het INT verantwoordelijk voor het hosten van de data van het Global Anglicism Database Network (GLAD); hiervoor is een online bewerkingapplicatie met behulp van het Lex'it-platform ontwikkeld. De Nederlandse gegevens van GLAD zijn in 2022 uitgebreid van 2900 tot 9500 trefwoorden.

2.1.6 Lexicon Frisicum

Het Lexicon Frisicum is 148 jaar geleden geschreven door Joast Hiddes Halbertsma, in het Latijn, en was een

eerste aanzet tot een woordenboek van het Fries. Het woordenboek beschrijft 6400 woorden, wat slechts een klein deel is van de totale woordenschat. Halbertsma overleed namelijk voor hij het kon afmaken. Lexicograaf en taalkundige Anne Dykstra vertaalde het Lexicon naar het Fries, samen met zijn collega Bouke Slofstra en verschillende vrijwilligers. Het INT, dat via het historisch woordenboekportaal ook het *Wurdboek fan de Fryske taal / Woordenboek der Friese taal* (WFT) online heeft gezet, is ook verantwoordelijk voor het online zetten van dit unieke woordenboek. In 2022 heeft het woordenboek een data-update gekregen.

3. Beschrijving van de Nederlandse dialecten

3.1 Elektronische Woordenbank van de Nederlandse dialecten (eWND)

In 2022 is met behulp van vrijwilligers een nieuw woordenboek toegevoegd aan de elektronische Woordenbank van de Nederlandse Dialecten (eWND).

3.2 Database van de Zuidelijk-Nederlandse Dialecten (DSDD)

Eind 2022 bevatte de Database van de Zuidelijk-Nederlandse Dialecten (DSDD) ongeveer 30.000 concepten. De database werd aangevuld met de woordenschat van de vaktalen, die slechts beperkt vertegenwoordigd was in het pilotproject (bv. de loodgieter, de smid, de timmerman). Met hulp van vrijwilligers is vooral voor de algemene woordenschat nagegaan of er nog lacunes zijn die gemakkelijk aan te vullen zijn voor een bepaalde regio. Op basis van die lijst zijn al enkele lemmata klaargemaakt, die begin 2023 zullen worden toegevoegd. Er is een start gemaakt met het analyseren en bewerken van de Zeeuwse gegevens uit het *Woordenboek der Zeeuwse dialecten* en het supplement, zodat die in de loop van 2023 kunnen worden gekoppeld. Er is overleg geweest met de redactie van het *Woordenboek van de Overijsselse dialecten* (WOD), om de database uit te breiden met de gegevens van het WOD.

3.3 Gesproken Corpus van de Zuidelijk-Nederlandse Dialecten

Het INT is partner in het project Gesproken Corpus van de Zuidelijk-Nederlandse Dialecten, een project dat loopt van 2020 tot 2024 aan de UGent. De opnames worden volgens een nieuw ontwikkeld transcriptieprotocol getranscribeerd om vervolgens met bestaande tools taalkundig verrijkt te worden. Het INT heeft in de loop van 2022 een Lex'it-omgeving gebouwd om de metadata die bij de opnames behoren doorzoekbaar te maken, en het nam deel aan diverse overlegmomenten.

3.4 Prijzen en projecten

De *Atlas van het dialect in Vlaanderen* werd in 2022 genomineerd voor de LOT-populariseringsprijs, voor de LIA's, een prijs die uitgereikt wordt door De Taalsector in België, en voor de taalboekenprijs van Onze Taal, het ANV en Trouw. Het won uiteindelijk een van de drie, nl. de LIA van De Taalsector.

In 2022 verscheen tweewekelijks 'Uit de Streek', een rubriek over dialectwoorden uit de DSDD en de woordenbank. In het kader van CLARIAH+ (zie aldaar) neemt het INT deel aan een deelproject ten behoeve van de integratie van dialectwoordenboeken in de CLARIAH-infrastructuur.

Het INT zetelt ook in de werkgroep contactvariëteiten, een werkgroep die bekijkt hoe dialectgemeenschappen (met als pilot het Bildts) eigen informatie kunnen samenbrengen en delen. In 2022 werd het pilotproject 'Taalinfrastructuur voor de Bildtse taal' verder voorbereid en opgestart. Het wil een bestaande infrastructurele voorziening (WoordWaark) beschikbaar maken voor de beschrijving van het Bildts, die daarna zal worden gehost door het INT. WoordWaark werd met dat doel in de loop van 2022 vernieuwd.

4. Terminologie

Het Expertisecentrum Nederlandstalige Terminologie heeft zich in 2022 verder toegelegd op zijn traditionele kerntaken: veldondersteuning en positionering in een terminologisch netwerk. Deze missie is in 2022 als volgt gerealiseerd.

4.1 Termenlijsten

In lijn met het streven van het INT om het best gesorteerde en een zeer toegankelijk wetenschappelijk instituut te zijn op het gebied van de Nederlandse taal en haar woordenschat streeft het ENT naar een gestage uitbreiding van de verzameling digitale termenlijsten. Ze ondersteunen communicatiebehoefte in meer specialistische domeinen en zijn door begripsomschrijvingen of vertaalsuggesties in het bijzonder dienstbaar voor de vele taalgerichte beroepsbeoefenaars zoals vertalers en technische schrijvers. Binnen het kader van de zeer uitgebreide onderwerpenclassificatie van de Library of Congress vonden updates plaats van de al ingevoerde termenlijsten en werden 55 nieuwe termenlijsten toegevoegd.

4.2 Tools

In 2022 is er begonnen met werkzaamheden om een nieuwe terminologietool te ontwikkelen. Beslist is om twee bestaande taaltechnologische hulpmiddelen, te weten TermTreffer, geschikt voor termextractie en bepaalde editeermogelijkheden, en de TermBeheerder, ontworpen voor termenbankmanagement, te integreren in een nieuwe online werkomgeving. Deze omgeving zal maximaal gebruikmaken van softwarecomponenten en databanken die reeds door het INT ontwikkeld en onderhouden worden, zodat de ondersteuning en het onderhoud van deze nieuwe terminologietool ook in de toekomst gegarandeerd blijft. Voor deze applicatie zijn de wenselijke functionaliteiten en veldenstructuur geïnventariseerd. De technische ontwikkeling had voorsnog betrekking op de extractiemodule in pilotversie, waarbij ook in kaart werd gebracht hoe de toekomstige productieversie optimaal in de bestaande INT-infrastructuur geïntegreerd kan worden. De verdere uitwerking van die productieversie en de realisatie van de editeeromgeving worden voorzien in 2023. Voor de termextractie is het doel om, naast de bestaande hybride benadering met taalkundige en statistische modules van de oude TermTreffer, nieuwe extractietechnieken aan te bieden die zijn gebaseerd op sequentiële benaderingen, zoals in [D-Terminer](#). Hiervoor wordt samengewerkt met de academische onderzoekers in dit domein.

4.3 Veldondersteuning

Veldondersteuning werd in 2022 deels gerealiseerd vanuit de website van het INT en ENT en deels met projecten die door hun thematiek en online raadpleegbare materiaalverzamelingen een toegepaste waarde hebben.

Voor de eerste realisatievorm mag opnieuw worden gewezen op vier nieuwsbrieven die op de website gepubliceerd en gearchiveerd worden. Ze bevatten een vaste rubriekenreeks over voorbije evenementen, nieuwe ontwikkelingen en publicaties, tools en bronnen, een uitgelichte term, berichten van NL-Term en een agenda. Als veldondersteuning fungeerden ook de geregelde updates van de evenementenrubriek op de ENT-webpagina over terminologiecongressen en -symposia in Europa.

Op de webrubriek Terminologieprojecten is het HOTNeV-project over hogeronderwijsterminologie verder uitgebouwd door de stagebegeleiding van vier stagiairs. De termanalyses en -beschrijvingen hebben geleid tot verdere uitbreiding van het termenbestand. Een update om de resultaten online beschikbaar te maken, heeft intussen plaatsgevonden. Een mogelijkheid voor verdere aanvulling diende zich ook aan in de opdrachtverstrekking voor twee masterscripties over HOTNeV-termen voor het academiejaar 2022-2023.

4.4 Medische vaktaal, juridische vaktaal en Nederlands als wetenschapstaal

Speerpunt 1: de medische vaktaal

Pinkhof Geneeskundig woordenboek is intussen online beschikbaar en is te raadplegen bij de INT-

taalmaterialen.

De database gaat terug op de editie van 2012 met aanvullingen en verbeteringen in de loop van de tijd. De huidige versie evolueert momenteel nog technisch en nieuwe releases zijn in voorbereiding. Werkzaamheden hadden betrekking op de verbetering van de digitale doorzoekbaarheid van het bestand. Het INT werkt samen met de Stichting Beheer Database Pinkhof Geneeskundig Woordenboek, die finaal de beslissingen neemt over de aanpassingen. In het kader van deze samenwerking is het Pinkhof-bestand geanalyseerd en zijn onder meer de encyclopedische trefwoordartikelen of ‘profieltermen’ in kaart gebracht. Deze bevatten expliciet gecodeerde dan wel impliciete vermeldingen van categorieën die in aanmerking komen voor actualisering. Te denken valt dan aan begrippen als ‘behandeling’, ‘prognose’ e.d. De stichting heeft zich voorgenomen deze artikelen daarom te herwerken. Verder zijn die Pinkhof-ingangen geïnventariseerd waar de typografie of artikelindeling nog verder moet worden geoptimaliseerd. Het bestand is in overleg met de stichting ook doorgenomen met het oog op de uitzuivering van taalkundige opmerkingen van algemene aard.

Gerelateerd aan *Pinkhof Geneeskundig woordenboek* is ook een pilotproject opgestart over termen die verband houden met COVID-19. Het doel is een strategie te ontwikkelen en te testen om medische termen – artsenjargon én lekentermen – toe te voegen aan het *Pinkhof Geneeskundig woordenboek*, een strategie die in de toekomst ook toegepast kan worden op andere deeldomeinen van de geneeskunde. Daarom is als casus gekozen voor een deelgebied dat nog niet is vertegenwoordigd in de huidige Pinkhof. Het beoogde eindresultaat van het project is een gevalideerde termenlijst met coronaterminologie, met labels en koppelingen naar de medische woordenboeken Pinkhof en SNOMED CT (Systematized Nomenclature of Medicine Clinical Terms). Voorbereidend overleg vond plaats tussen de leden van de Raad van Advies medische terminologie over aspecten van de taakomschrijving, zoals de verzameling van digitale terminologielijsten rond corona, corpusbouw, automatische termextractie op basis van dit corpus, de labeling van termen en uitvoeringswerkzaamheden door studenten, hun statuut en werving.

Speerpunt 2: de juridische vaktaal

Het *Juridisch Woordenboek Nederlands-Spaans met register Spaans-Nederlands voor rechtspraktijk, handel en bedrijfscommunicatie* van M.C. Oosterveld-Egaz Repáraz en J.B. Vuyk-Bosdriesz behandelt de terminologie van het Nederlandse rechtssysteem wat betreft de uitgangstermen, en de rechtstermen van het Spaanse rechtssysteem als doeltaal. De bestanden van het woordenboek werden geconverteerd naar het standaard XML-formaat voor termenbanken en woordenboeken. Een nieuwe applicatie voor het online brengen van het juridisch woordenboek voor het Nederlands recht werd uitgewerkt en voor feedback voorgelegd aan de redacteuren. Na enkele iteraties is er een versie waarvan de publieke release is voorzien voor voorjaar 2023. Een voorlopige versie werd door Frieda Steurs gedemonstreerd op de TiNT-dag (<https://taalunie.org/Agenda/225/tint-2022-terminologie-in-juridische-taal>)

Er werd geëxploreerd wat de financieringsmogelijkheden zijn om Belgische equivalenten hieraan toe te voegen.

Speerpunt 3: Nederlands als wetenschapstaal

Bij de overstap van voortgezet onderwijs naar het eerste bachelorjaar hebben studenten behoefte aan goede begeleiding op het vlak van onder meer geschreven vaktaal, zo bleek uit onderzoek naar onderwijs en woordenschat. Met dit doel ontwikkelt het 'Proefproject Nederlands als wetenschapstaal – van corpora naar terminologielijsten' taalhulpmiddelen in een samenwerkingsverband dat is opgestart tussen de Stichting Nederlands / Vlaams Platform Taalbeleid Hoger Onderwijs, de Taalunie, KU Leuven, UGent en het INT. Het pilotproject heeft betrekking op zowel alfa-, bèta- als gamma-richtingen. Er is gekozen voor een bredere benadering die de selectie en beschrijving van termen geschikt maakt voor een ruimere studentenpopulatie, zodat bijvoorbeeld scheikundige en wiskundige termen niet louter voor scheikundigen of wiskundigen in opleiding geschikt zijn maar ook voor verpleegkundigen respectievelijk economiestudenten. In 2022 zijn de digitale bestanden voor de vakgebieden scheikunde en wiskunde verder getest en vonden verbeteringen plaats op het vlak van de weergave van bestandscodes en sortering van data. De versies van de ato's (acceptatietestomgevingen) zijn nu geschikt voor publicatie als online webinterface met een zoekapplicatie. Het bestand met natuurkundetermen is eveneens opgeleverd door de externe auteurs en in een eerste versie overgedragen aan het INT, waar een bestandsanalyse is uitgevoerd. Die esulteerde in de gangbare voorstellen voor inhoudelijke en tekstuele verbetering van de data en hun structurering.

4.5 Terminologisch netwerk

Positionering in het terminologische netwerk kreeg onder meer vorm door deelname aan of bijwonen van congressen en symposia (Pan-European Terminology Summit, TiNT 2022). Eveneens werd een cursus Terminologie en ICT met gebruik van de TermTreffer gegeven aan de KU Leuven. Het ENT nam opnieuw deel aan terminologieoverleg tijdens de vergaderingen die de Termraad Nederlands belegde voor Nederlandstalige (ver)taaldiensten van EU-instellingen en van Nederlandse, Belgische en Vlaamse overheden. Eveneens participeerde het ENT in een stagevoorlichting van de Termraad Academy. Voor hogeronderwijsinstellingen in Nederland en België werden vier stagiairs begeleid, waarvan een in samenwerking met het Departement Nederlands van het Directoraat-generaal Vertaling (DGT) van de EU.

Verder werden de nodige voorbereidingen gestart voor het internationale terminologiecongres dat in 2023 zal plaatsvinden in Brussel. Het thema is '[terminology: domain loss and gain](#)'. Het INT is mede-initiatiefnemer, samen met NL-Term, EAFT en Infoterm. De website werd door het INT ontwikkeld.

5. Grammatica

In 2022 is begonnen met de ontwikkeling van het Grammaticaportaal, een webpagina die de spil moet gaan

vormen van de verschillende grammaticaonderdelen van het INT, en die zal fungeren als ontvangstpagina voor geïnteresseerde gebruikers, met informatie over projecten en producten, een zoekfunctie voor alle producten en een loket voor vragen. Een testversie staat nu online.

5.1 Taaladvies

Aan Taaladvies.net, een zeer druk bezochte site die onder beheer is bij het INT, is afgelopen jaar regulier onderhoud gepleegd. Daarnaast zijn zowel de CMS als de site gebruikersvriendelijker gemaakt, met onder meer extra functionaliteiten voor de redacteurs en een alfabet bij pagina's met veel adviezen. Ook is er veel onderling contact geweest met de redactie van de ANS om de kruisverwijzingen te verbeteren.

5.2 e-ANS

In 2022 is het vernieuwde hoofdstuk 12 *Woordvorming en woordstructuur* gepubliceerd. Het is door prof. em. Geert Booij inhoudelijk in overeenstemming gebracht met de huidige inzichten in dit deelgebied van de Nederlandse grammatica. Het hoofdstuk is daarnaast op verschillende manieren verrijkt, volgens de doelstellingen van de herziening van de e-ANS. Zo zijn er uitgebreide literatuurverwijzingen toegevoegd, koppelingen naar andere informatiebronnen, nieuwe voorbeelden en informatie over gebruiksverschillen tussen Nederland en België.

Bovendien zijn in 2022 onderwijsmodules toegevoegd aan het nieuwe hoofdstuk 1 *De klankleer van het Nederlands* en het herziene hoofdstuk 29 *Negatie*. Deze onderwijsmodules zijn speciaal ontwikkeld voor docenten en leerders van de Nederlandse grammatica, voor wie het niveau van de ANS vaak net te hoog ligt. De modules bevatten samenvattingen, termenlijsten, veelgestelde vragen en lesideeën. Ook is hoofdstuk 1 *De klankleer van het Nederlands* verrijkt met talloze audiofragmenten.

Achter de schermen wordt hard doorgewerkt aan de herziening van de e-ANS. De conceptversies van de hoofdstukken over het telwoord, tussenwerpsel, woordvorming en de werkwoordelijke constituent zijn voorgelegd aan externe experts, om de inhoudelijke kwaliteit te waarborgen. Maaïke Beliën houdt als coördinator van de inhoudelijke herziening contact met de externe auteurs, leest mee en bewaakt de toegankelijkheid en interne consistentie van de hoofdstukken. Ook werd er, dankzij een financiële versnellingsimpuls, weer een extra contract afgesloten, voor het herzien van hoofdstuk 7 *Het voornaamwoord*. Op technisch gebied is de ANS uitgebreid met een webservice van Lex'it, waarmee bij taalkundige termen via een INT-database de bijbehorende definities worden getoond. Deze webservice wordt ook gebruikt bij het Grammaticaportaal.

5.3 Taalportaal

Het nieuwe syntaxis-hoofdstuk *Coordination* van Hans Broekhuis is dit jaar geconverteerd naar XML, en ligt nu bij de auteur ter controle. Daarnaast is begonnen met de herziening van de site, niet alleen op technisch gebied, maar ook om die in lijn te krijgen met de opmaak van de overige INT-websites. Daarnaast is de Zuid-Afrikaanse versie van Taalportaal, draaiend op de op het INT ontwikkelde webapplicatie, en met de talen Setswana en isiXhosa, dit jaar online gegaan.

6. Taalmaterialen

Via de website taalmaterialen.ivdnt.org stelt het Instituut voor de Nederlandse Taal (INT) bronnen, data en tools beschikbaar voor taalkundig onderzoek en taal- en spraaktechnologie (TST) binnen het hele Nederlandse taalgebied. In 2022 zijn er wederom een aantal productpagina's van de website voor taalmaterialen verbeterd en is het systeem om tags aan producten toe te kennen verbeterd zodat het makkelijker is om op die manier te zoeken naar producten. Verder is er veel werk gestoken in het onderhoud van het data-archief. Het totaal aantal downloads voor zowel commerciële als niet-commerciële producten is afgenomen in vergelijking met het voorgaande jaar:

- Niet-commerciële downloads 522 (930 in 2021)
- Commerciële downloads 5 (27 in 2021)

Het is niet duidelijk waar dit aan ligt. Wellicht dat de maatschappelijke impact van de coronalockdown hierbij ook een rol speelt.

In het afgelopen jaar werden de volgende producten toegevoegd aan de catalogus:

- Corpus Vlaamse Gebarentaal (Corpus VGT)
- Belgian Covid Sign Language Corpus (BeCos Corpus)
- Annotated Corpus for Term Extraction Research (ACTER)
- BLISS Dialogue Summaries
- Couranten Corpus Online
- AI-Trainingset - Tag de Tekst voor Named Entity Recognition (NER)
- Pinkhof Geneeskundig Woordenboek (Online)
- Corpus Juridisch Nederlands (Online)
- GiGaNT-Molex (update)
- Corpus Middelnederlands (Online)
- Corpus Gysseling (Online)

(Zie bijlage 4 voor een uitgebreide omschrijving van de nieuwe producten en een overzicht van alle commerciële en niet-commerciële downloads.)

Gebruikers dienen zich te registreren alvorens ze producten kunnen downloaden. Daarbij wordt ook hun e-mailadres vastgelegd. Aan de hand van de extensie van die adressen kan een beeld worden verkregen vanuit welke landen belangstelling bestaat voor deze taalmaterialen. De meesten komen uit Nederland (93), daarna België (49) en verder Denemarken (8), Verenigd Koninkrijk (5), en lager.

De e-mailadressen geven niet een volledig beeld van waar de gebruikers zitten, omdat veel een ambigue extensie hebben zoals .com of .org. Daarom werd er ook gekeken naar de geo-locatie van de IP-adressen die zijn geregistreerd bij de downloads. Dat leverde het volgende overzicht op: Nederland (265), België (121), Verenigde Staten (12), Duitsland (10), Ierland (7), Hongkong en Spanje (6) en een groot aantal lagere scores.

6.1 CLARIN-ERIC; het INT als CLARIN-centrum

Voor 2022 vervulde Vincent Vandeghinste voor het INT de rol van Nationaal Coördinator voor CLARIN op Europees niveau, als vertegenwoordiger voor België in het National Coordinators Forum. Daarnaast nam Jesse de Does vanuit het INT de vertegenwoordiging van België op in het Standing Committee on CLARIN Technical Centres. Het INT vervulde een actieve rol als liaison tussen Belgische onderzoekers en de Europese CLARIN-infrastructuur en tussen de Vlaamse onderzoekers in CLARIAH-VL en de CLARIN-infrastructuur van het INT als CLARIN-B centrum voor België.

Als gevolg van CLARIN-BE werden er nieuwe datasets gereleased, zoals het Corpus Vlaamse Gebarentaal (Corpus VGT), het Belgian Covid Sign Language Corpus (BeCoS Corpus), en de Annotated Corpora for Term Extraction Research (ACTER).

K-Dutch, het CLARIN Knowledge Centre for Dutch, werd verder uitgebouwd, waarbij de website (<https://kdutch.ivdnt.org>) verder werd uitgebreid en verfijnd.

6.2 European Language Resources Coordination Initiative (ELRC)

Het INT is betrokken bij het ELRC-initiatief dat als doel heeft tekstdata te verzamelen in alle EU-lidstaten, IJsland en Noorwegen, die gebruikt kunnen worden om eTranslation (de automatische vertaaldienst van de Europese Commissie) verder te ontwikkelen. De kwaliteit van een automatische vertaling hangt onvermijdelijk samen met de kwaliteit en kwantiteit van de taalbronnen die worden gebruikt om het systeem te 'trainen'. Grote hoeveelheden taaldata zijn dan ook nodig om de kwaliteit van de Nederlandse vertalingen te verbeteren.

In 2022 heeft het INT zich ingezet om het belang van het verzamelen van Nederlandse taaldata t.b.v het trainen van eTranslation verder te benadrukken en om de bekendheid en het gebruik van eTranslation binnen Nederland te promoten.

In samenwerking met Emma Hartkamp (*Europese Commissie, directoraat-generaal*) is een flyer over eTranslation ontwikkeld voor het Nederlands en op 14 december is een speciale workshop over eTranslation gehouden voor de ministeries in Den Haag.

Op 9 november vond de laatste LRB Meeting plaats in Praag en in januari 2023 gaat ELRC over in de European Language Data Space. Ook binnen dit initiatief blijft het INT het nationale aanspreekpunt.

6.3 IMPACT-centrum en digitization.eu

Het INT is voorzitter van het IMPACT Centre of Competence (www.digitisation.eu). Dit is een non-profitorganisatie bestaande uit publieke en commerciële organisaties met als doel de digitalisering van historisch materiaal “beter, sneller en goedkoper” te maken. Het centrum voorziet in data, tools, services en expertise op het gebied van document imaging, taaltechnologie en het verwerken van historisch tekstmateriaal. Het IMPACT Centre of Competence is sedert 2019 ook CLARIN Knowledge Centre. De werkzaamheden m.b.t. digitalisering die in de context van CLARIAH+ worden uitgevoerd, worden in samenwerking met het Centre uitgevoerd.

In december 2022 is het tienjarig bestaan van het IMPACT Centre of Competence gevierd. Er is een workshop georganiseerd met als doel een white paper te produceren rondom het thema *Sharing and Sustaining Digitisation Knowledge*. De doelgroep van de white paper zijn onderzoekers en erfgoedinstellingen en de focus ligt op de uitdagingen die het veranderende digitaliseringslandschap met zich meebrengt.

7. Overige infrastructuur- en netwerkprojecten

7.1 European Lexicographic Infrastructure (ELEXIS)

Het INT was partner in [ELEXIS](#), een Europees Horizon 2020-project dat eind juli 2022 afliep.

Binnen ELEXIS is een infrastructuur voor e-lexicografie opgezet a) om zo de lexicografische inspanningen binnen Europa zoveel mogelijk te harmoniseren door het formuleren van best practices en het ontwikkelen van [open-source tools](#), en b) om lexicografische data open en toegankelijker te maken, zodat deze hoogwaardige semantische data ook kunnen worden gebruikt in andere vakgebieden, zoals AI, NLP en Digital Humanities. Tevens draagt ELEXIS bij aan training en educatie van toekomstige lexicografen dankzij een speciaal ontwikkeld ELEXIS-curriculum dat via [DARIAH-CAMPUS](#) beschikbaar wordt gesteld. De belangrijkste resultaten van het project zijn te vinden op de website en in de [deliverables](#).

Om de duurzaamheid van de infrastructuur te waarborgen is een nieuwe associatie in het leven geroepen, de ELEXIS-associatie, en is er een aanvraag ingediend bij CLARIN voor een CLARIN Knowledge Centre voor lexicografie.

Op 7 en 8 juni vond in Florence het ELEXIS Showcase Event plaats: *ELEXIS from A to Ω: Outcomes, Sustainability & Afterlife of a new European Lexicographic Infrastructure*. Tot slot zijn in de eerste helft van 2022 nog 3 gastonderzoekers met een reisbeurs via ELEXIS trans-national access (TNA) ontvangen op het INT.

7.2 CLARIAH+ Nederland

Het CLARIAH (Common Lab for Research in the Arts and Humanities) CORE-project (2015-2018) was erop gericht een gemeenschappelijke infrastructuur tot stand te brengen voor data-intensief wetenschappelijk

onderzoek in de geesteswetenschappen. Vanaf begin 2019 tot en met 2023 loopt het vervolgproject CLARIAH+, waarin het accent nog meer gericht is op het concreet ondersteunen van de onderzoeker door middel van het tot stand brengen van (virtuele) onderzoeksomgevingen. Het INT houdt zich onder andere bezig met een verbetering van de infrastructuur voor historisch Nederlands, een uitbreiding op de corpuszoekmachine BlackLab naar parallelle corpora en treebanks, hulpmiddelen voor het aanbrengen van persistente gebruikersannotaties in corpuszoekresultaten, een gebruikersvriendelijker digitalisatieworkflow en curatie van dialectwoordenboekdata.

In het kader van de infrastructuur voor historisch Nederlands is voortgewerkt aan het platform GALAHAD voor betere en gebruikersvriendelijke taalkundige verrijking van historisch Nederlands, en aan de handmatige annotatietool COBALT. Eerste versies van training- en evaluatiecorpora voor de 15^e, 16^e, 17^e, 18^e en 19^e eeuw zijn met behulp van COBALT volgens de TDN (Tagset Diachron Nederlands) taalkundig verrijkt. De corpora zijn al gebruikt om met behulp van het [HuggingFace's Transformers platform](#) Deep Learning-gebaseerde taggers te trainen. Ondanks de nog niet al te grote omvang van de corpora zijn behoorlijke resultaten gehaald dankzij het gebruik van voorgetrainde taalmodellen, waarbij over het algemeen het in het [MacBerth](#)-project ontwikkelde [GysBERT](#)-model voor historisch Nederlands de beste resultaten gaf.

Binnen de infrastructuur voor digitalisatie en conversie richten we ons voornamelijk op het inrichten van een workflow met gebruikmaking van het Transkribus-platform, waarmee vrijwilligers op een makkelijke manier aan de digitalisatie van historische kranten kunnen werken. Door de tekstherkenning periodiek te hertrainen wordt de kwaliteit steeds beter.

Het werk aan de uitbreidingen van BlackLab met zoekmogelijkheden voor syntactisch verrijkte en parallelle corpora is uitgesteld tot 2023 ten gunste van de BlackLab-Solr-integratie; zie het onderdeel 'Corpusapplicatie' boven.

Binnen de *CLARIAH use cases* is gewerkt aan een prototype voor de uitbreiding van het corpus-frontend, met zoekmogelijkheden voor combinaties van concepten, en annotatie van zoekresultaten via het Lex'it-platform. Dit prototype wordt in 2023 als add-onmodule geïntegreerd met het corpus-frontend.

Use case 1 (gebruikmaking van VOC-gerelateerde data ten behoeve van historisch onderzoek) is afgerond.

Als resultaat van use case *digitalisatie 17^e eeuwse kranten* is de eerste set online gepubliceerd als het [Couranten Corpus](#), en is het bovenvermelde vernieuwde digitalisatieproces voor de later gescande pagina's volop in gang.

7.3 CLARIAH Vlaanderen

Binnen het FWO-IRI-project 'CLARIAH-VL: Advancing the open humanities service infrastructure' is de hoofdtaak van het INT het voorzien van de benodigde infrastructuur voor het opzetten van de Digital Text Analysis Dashboard & Pipeline. Deze pipeline werd verder uitgebouwd en [gepresenteerd](#) op de jaarlijkse CLARIN-conferentie.

Er werd in het kader van CLARIAH-VL ook verdergewerkt aan een pilotproject in samenwerking met de

Vlaamse Super Computer (VSC), waarbij het plan is om een groot contextueel taalmodel te trainen op basis van de corpora hedendaags Nederlands waarover het INT beschikt. Er werd meegewerkt aan een High Performance Cluster (HPC) Tier I-aanvraag van de UGent en deze aanvraag werd gehonoreerd. Het INT stelt voor de onderzoekers van de UGent data rechtstreeks ter beschikking op de HPC. De uitvoering van dit plan is voorzien voor 2023.

7.4 SignOn-project

Het INT is als consortium betrokken bij het SignON-project, dat vanaf voorjaar 2021 voor drie jaar gefinancierd wordt binnen het kader van het Horizon 2020-programma van de Europese Commissie. Het hoofddoel van dit project is het opzetten van automatische vertaalservices tussen gebarentalen en zogenaamde gesproken talen. De gebarentalen die bovenaan de agenda staan van deze Research and Innovation Action zijn Vlaamse Gebarentaal (VGT), Nederlandse Gebarentaal (NGT) en Ierse Gebarentaal. Gesproken talen zijn in eerste instantie het Nederlands en het Engels. Het consortium van dit project heeft een sterk Belgisch-Nederlandse component, met als consortiumpartners uit België: VRT, KU Leuven, UGent, Vlaams Gebarentaalcentrum en European Union for the Deaf. Vanuit Nederland nemen deel: INT, de Taalunie, Radboud Universiteit Nijmegen, Tilburg University, en als derde partij Beeld en Geluid. Het project wordt geleid door Dublin City University.

De taak van het INT bestaat hoofdzakelijk uit het opzetten van de infrastructuur voor dit onderzoek. Er wordt eveneens gewerkt aan het verzamelen van gebarentaalcorpora, zowel voor VGT als voor NGT. Dit heeft onder meer geleid tot het BeCoS Corpus, een verzameling van de videos van de officiële persconferenties van de Belgische overheid, die automatisch geannoteerd zijn voor het gesproken Nederlands als voor Vlaamse Gebarentaal. Daarnaast werd meegewerkt aan een aanvraag van de Universiteit Tilburg voor extra financiering vanuit European Language Equality om een dataset te maken waarbij online hotelrecensies door dove vertalers vertaald worden naar NGT. Deze aanvraag werd gehonoreerd en wordt uitgevoerd in 2023. Een andere aanvraag van de Universiteit Tilburg voor extra financiering vanuit de European Association for Machine Translation voor het vertalen naar het Nederlands van reeds bestaande video's met VGT als brontaal werd eveneens gehonoreerd.

Een andere taak van het INT is om de infrastructuur op te zetten om Vertalen-als-een-service aan te kunnen bieden, die dan aangesproken kan worden binnen de Android- en iPhone-apps die ontwikkeld worden in de use cases, die in samenspraak met de doelgroepen ontwikkeld worden. Die infrastructuur werd grotendeels uitgebouwd in 2022.

7.5 SABeD – Spoken Academic Belgian Dutch

Het INT werkte mee aan de supervisie van de ontwikkeling van het corpus Spoken Academic Belgian Dutch van de KU Leuven, en bouwde eveneens een aantal tools om om te gaan met gesproken data. Een van die tools is een automatische segmentatietool op zinsniveau, die aan punctuatie-insertie doet. Deze tool werd in 2022

gebouwd, geëvalueerd en toegepast op de SABeD-delen waar geen manuele transcriptie van gebeurt.

7.6 COST – European network for Web-centered linguistic data science

Het INT neemt deel aan de COST-actie NexusLinguarum. Het thema van deze actie is ‘linguistic data science’, een deelgebied binnen de opkomende 'data science'. Taalkundige data vormen een specifiek geval en zijn tot nu toe nog grotendeels onontgonnen in een big data-context. Het hoofddoel van NexusLinguarum is om taalkundigen, computerwetenschappers, terminologen en andere belanghebbenden in één netwerk bij elkaar te brengen om zo samenwerking en kennisdeling op het gebied van ‘linguistic data science’ te bevorderen. De actie is eind oktober 2019 van start gegaan en heeft een looptijd van 4 jaar die eindigt op 27 april 2024. In april 2022 heeft het INT een STSM gehost van Ilan Kernerman over ‘Linking lexicographic resources and CEFR-graded vocabulary lists’. De daaruit resulterende gemeenschappelijke paper ([Heylen, Kernerman & Tiberius 2022](#)) is gepresenteerd tijdens de conferentie ‘LLOD approaches for language data research and management’ in Vilnius in september. Hier vond ook de jaarlijkse plenaire bijeenkomst van NexusLinguarum plaats.

7.7 COST - Universality, diversity and idiosyncrasy in language technology (UniDive, 2022-2026)

Het INT neemt deel aan het Europese onderzoeksnetwerk (COST) UniDive18 (Universality, diversity and idiosyncrasy in language technology). Het doel van deze COST-actie is om te onderzoeken hoe taaltechnologie verbeterd kan worden door betere kennis van wat talen gemeenschappelijk hebben en van waarin ze zich onderscheiden. Met de actie beoogt men aan de theoretische kant een beter begrip van taaluniversalia te krijgen en, aan de praktische kant, de beschikking te zullen hebben over taaltechnologie die om kan gaan met een grotere verscheidenheid aan taalverschijnselen in een groot aantal talen, waaronder talen met weinig middelen en bedreigde talen. Op 23 september 2022 vond de eerste bijeenkomst van het Management Committee van deze COST-actie plaats in Brussel.

8. Plaats van het instituut in Nederland, Vlaanderen en de wereld

8.1 Doelgroepenbeleid (inclusief onderwijs)

Het INT richt zich als toegepast wetenschappelijk instituut traditioneel op onderzoekers en taalkundigen. Bestaande contacten met onderzoekers, al dan niet verbonden aan wetenschappelijke instituten en universiteiten, worden onderhouden en waar mogelijk geïntensiveerd en uitgebreid. Aan de KU Leuven is het vak Computatieve Lexicografie gegeven in de Master Taalkunde en Taal- en Letterkunde en aan de Universiteit Leiden de collegereeks Corpus Lexicography, onderdeel van de Research Master in Linguistics. Tijdens gastlessen aan de Universiteit Leiden is het gebruik van de online Historische Woordenboeken uitgelegd aan studenten Nederlands.

Daarnaast verschuift het INT zijn werkterrein, gezien de brede taakomschrijving, nadrukkelijk naar docenten en studenten. In dat verband was het Instituut aanwezig op en profileert het zich op beurzen, (online) conferenties (Student Conference on Linguistics, HSN-conferentie), festivals (Week van het Nederlands, Letterlijk Leiden) en wetenschappelijke bijeenkomsten. Ook heeft het INT een bescheiden bijdrage geleverd aan de jaarlijkse Taalkundeolympiade van de Universiteit Leiden.

Op de website van het INT is een apart menu-item voor onderwijs ingericht, onderverdeeld in informatie voor docenten, leerlingen en studenten. Het INT heeft een webinar georganiseerd over Onderzoek in de digitale bronnen van het INT. Tijdens de Week van het Nederlands presenteerde het INT een Taalvragenvuur waarin taalvragen werden voorgelegd aan deskundigen van Onze Taal en Team Taaladvies.

Ook het algemene publiek wordt niet uit het oog verloren. Zo heeft het INT in het kader van het festival Letterlijk Leiden de talige wandeling 'Weg van woorden' aangeboden, die inmiddels ook als audiotour beschikbaar is. Op onze website verschenen wekelijks populairwetenschappelijke rubrieken van eigen medewerkers over dialectwoorden, nieuwe woorden en historische woorden. We zijn actief op verschillende sociale media en maakten nieuwe afleveringen van onze podcastserie *Over taal gesproken*. Tien keer per jaar werden twee soorten nieuwsbrieven verstuurd aan geïnteresseerden. Daarnaast hielden medewerkers voordrachten, waren zij te horen in radioprogramma's en droegen zij bij aan publicaties voor een algemeen publiek dat belangstelling heeft voor taal in het algemeen en Nederlands in het bijzonder. Zie daarvoor het uitgebreide overzicht in bijlage 3.

8.2 Communicatiemiddelen

8.2.1 Website

De website is vaak het eerste wat het publiek ziet van het instituut, en dat maakt het een van de belangrijkste communicatiemiddelen. Bijna alle communicatie-uitingen zijn erop gericht om bezoekers naar de website te trekken. In 2022 telde ivdnt.org een half miljoen paginaweergaven; dat aantal is gelijk gebleven met het jaar ervoor. 75% van de bezoekers komt voor het eerst op de website, 25% is terugkerende bezoeker.

Het grootste deel van de bezoekers (61%) komt via Google op de website terecht, 17% tikt direct ivdnt.org in de browser in, 14% komt via diverse nieuwsbrieven (eigen nieuwsbrief, Taalpost, Taalunienieuwsbrief) en 3,6% via de sociale media: de meesten daarvan vinden de website via Twitter, gevolgd door Facebook, LinkedIn en Instagram.

Van de statische content is de pagina van het *Woordenboek der Nederlandsche Taal* (WNT) het vaakst bezocht met 19.613 paginabezoeken. De best bezochte rubriek is opnieuw ‘Nieuw woord van de week’, met in totaal 76.119 bezoeken. Ook de (sinds 2020 niet meer bijgehouden) rubriek ‘Woordbaak’ wordt nog veel geraadpleegd met in 2022 83.158 bezoeken in totaal. Het onderdeel Terminologie kreeg dit jaar 5.487 bezoeken; 25% daarvan raadpleegt de termenlijsten.

Paginabezoeken Woorden van de week

Woordbaak	83.158
Nieuw woord van de week	76.119
Uit de streek	13.768
Terug in de taal	10.340
Uitgelichte term	2.293
Taal materiaal uitgelicht	1.027

Paginabezoeken columns

WoordHoek (Ewoud Sanders)	36.929
Woorden weten alles (Ludo Permentier)	12.268

Opvallende ‘hits’ waren het Woordbaakartikel *Waar komt marathon vandaan?* (7.320 bezoeken), wat waarschijnlijk te maken heeft met de Amsterdam Marathon op 16 oktober, en ‘Nieuw woord van de week’ [stoplichtregel](#) (2.862 bezoeken). Op 4 januari verscheen dit nieuwe woord in Taalpost.

Meest bezochte onderwerpen (in deze volgorde):

- Historische woordenboeken
- Nieuwe woorden
- Woorden van de week
- Etymologische woordenboeken
- Spelling en grammatica
- Couranten Corpus
- Podcast
- Algemeen Nederlands Woordenboek
- Transcriptor

- Groene Woordenboekje
- Taalmaterialen
- Terminologie

Eind 2021/begin 2022 liep er een korte gebruikersenquête op de website die in totaal 2.560 keer ingevuld is. Het overgrote deel gaf aan taalliefhebber te zijn, gevolgd door professionele taalgebruiker en taaldocent/taalstudent. 177 van de invullers waren onderzoeker.

De deelnemers gaven aan geïnteresseerd te zijn in de volgende onderwerpen (in die volgorde): hedendaags Nederlands, historisch Nederlands, spelling & grammatica, terminologie en Woorden van de week. Dat komt aardig overeen met de in praktijk meest bezochte onderwerpen op de website.

48% van de bezoekers raadpleegt ivdnt.org op een smartphone, 48% maakt gebruik van een desktop of laptop en 4% bezoekt de website op een tablet.

Aan de achterkant van de website wordt continu gewerkt aan verbetering. Zo worden fouten hersteld, wordt de veiligheid verbeterd, is het termenlijstbeheer geoptimaliseerd en wordt op verschillende manieren gewerkt aan het versnellen van de laadtijd. Daarnaast is er een taalagenda toegevoegd, is er een koppeling gemaakt met evenementen.ivdnt.org en is er een begin gemaakt met het tweetalig maken van de website.

Het uiterlijk van de website heeft geen grote veranderingen ondergaan. Wel nieuw zijn de vier illustraties die elke week bij het nieuwe woord gemaakt worden door vier externe tekenaars.

8.2.2 Huisstijl

Een team bestaande uit de webmaster, softwareontwikkelaars en de communicatieadviseur heeft een standaardwebapplicatie ontwikkeld die als vertrekpunt dient voor de ontwikkeling van nieuwe INT-applicaties. De gelijkvormigheid en herkenbare INT-huisstijl zorgen ervoor dat alle door het INT ontwikkelde applicaties door de gebruikers ook als zodanig herkend worden.

8.2.3 Nieuwsbrieven

In 2021 zijn er zes algemene INT-nieuwsbrieven verschenen en vier nieuwsbrieven over terminologie. Daarnaast zijn er nog diverse aparte mailings en persberichten verzonden. De algemene nieuwsbrief telde eind 2022 4.436 abonnees, van wie per editie 42% de nieuwsbrieven opent; van de 3.272 geadresseerden van de nieuwsbrief terminologie doet 37% dat.

8.2.4 Sociale media

Alle berichten die verschijnen op ivdnt.org worden direct doorgeplaatst op Twitter (5.646 volgers), LinkedIn (2.101 volgers) en Facebook (738 vind-ik-leuks). Voor Instagram (1.170 volgers) wordt voor elke post een aparte afbeelding gemaakt zodat de huisstijl er goed uitspringt.

Van alle grote platforms groeit Instagram het hardst; het dagelijks gebruik van Facebook loopt juist terug. Het gebruik van Twitter blijft redelijk stabiel. Voor het INT is het daarom nuttig om op al deze platforms actief te blijven.

Via het eigen [YouTube-kanaal](#) waren er diverse livestreams, die ook achteraf terug te zien waren, zoals *Taalvragenvuur* (1.200 keer bekeken) en *Wat mag je ook zeggen? Een gesprek over nieuwe, inclusieve taal* (725 keer bekeken).

8.2.5 Podcasts

In het najaar van 2022 is het tweede seizoen gestart van de podcast ‘Over taal gesproken’, die het INT samen met Onze Taal maakt. Laura van Eerten (INT) en Raymond Noë (Onze Taal) spreken elke aflevering een deskundige over het Nederlands. De drie uitgebrachte afleveringen gingen over inclusieve taal (3.316 downloads), computertaalkunde (2.786 downloads) en diertaal (2.026 downloads). ‘Over taal gesproken’ staat in de top 5% (979 downloads in de eerste zeven dagen) van alle podcasts die gehost worden op het platform Buzzsprout.

In totaal zijn er nu elf afleveringen verschenen van ‘Over taal gesproken’, en de drie best beluisterde in de eerste zeven dagen na publicatie zijn ‘Thiu werilt is hiera vilo hardo! Hoe klonk het vroegste Nederlands?’ (2.054 downloads), ‘Hoe intelligent is Google Translate nu echt?’ (1.529 downloads) en ‘Grammatica is leuker als je denkt’ (1.508 downloads). De eerste aflevering, over het Nederlands in Nederland en België, heeft in totaal al bijna tienduizend downloads, maar die is ook het langst beschikbaar. Over alle afleveringen verspreid waren er in totaal 30.755 downloads van de podcast in 2022. Het publiek komt voor 71% uit Nederland en voor 13% uit België.

In 2022 is er geen nieuw seizoen verschenen van ‘Waar komt pindakaas vandaan?’, maar alle afleveringen samen genomen is de podcast toch 70.127 keer gedownload. Voor de festivalmaand Letterlijk Leiden is er binnen deze serie een enkele aflevering gemaakt over het woord *letterlijk*. Deze aflevering is inmiddels 2.513 keer gedownload. Het publiek is voornamelijk Nederlands (87%, tegen 4% uit België).

8.3 Populairwetenschappelijke activiteiten

Na een proef in 2021 werd in november 2022 de eerste editie van de festivalmaand Letterlijk Leiden georganiseerd, een maand vol activiteiten op het gebied van taal en literatuur. Het INT was betrokken bij de ontwikkeling van het programma en organiseerde zelf ook een aantal activiteiten. Zo stelden we een talige stadswandeling samen, maakten we een audiotour, organiseerden we een pubquiz over taal en maakten we een podcastaflevering over het woord *letterlijk*.

Tijdens de Week van het Nederlands in oktober zonden we twee livestreams uit op YouTube. Ook brachten we

weer een populairwetenschappelijke publicatie uit aan het einde van het jaar. Dit keer het boek *Leukerlands. Taal in stukjes* van Rob Tempelaars, die in 2022 met pensioen ging.

In het kader van de herdenking van het Rampjaar 1672 hebben vrijwilligers twaalf geschiedenisboekjes gedigitaliseerd over vreemde heerschappij in de Lage Landen, gepubliceerd tussen 1674 en 1814; naar aanleiding daarvan is een gelegenheidswoordenboekje verschenen: [Spiegels voor de jeugd](#).

Het INT was in 2022 verantwoordelijk voor de techniek achter de website Neerlandistiek.nl en zal in 2023 ook de hosting op zich nemen.

Bijlage 1: Raad van Toezicht en Raad van Advies

Raad van Toezicht

Drs. Paul Rüpp, voorzitter (tot 1 december 2022)

Mr. Mieke Zaanen, voorzitter (vanaf 1 december 2022)

Prof. mr. Jan Cerfontaine

Drs. Gertine van der Vliet

Raad van Advies

Prof. dr. Hans Bennis (voorzitter)

Prof. dr. Willy Vandeweghe (vicevoorzitter)

Prof. dr. Paula Fikkert

Prof. dr. Dirk Geeraerts

Prof. dr. Jack Hoeksema

Prof. dr. Veronique Hoste

Drs. Jan Jaap Knol

Prof. dr. Reinhild Vandekerckhove

Lic. Wim Vanseveren

Dr. Eric Mijts

Bijlage 2: Medewerkers

Prof. dr. Frieda Steurs – directeur/bestuurder

Dr. Maaïke Beliën – onderzoeker/taalkundige

Marjolijn van Bennekom – taalkundig assistent

Dr. Bob Boelhouwer – computerlinguïst

Dr. Roland de Bonth – onderzoeker/taalkundige

Tim Brouwer – softwareontwikkelaar

Lic. Lut Colman – onderzoeker/taalkundige

Lic. Griet Depoorter – onderzoeker/taalkundige

Lic. Katrien Depuydt – onderzoeker/taalkundige

Lic. Veronique De Tier – onderzoeker/taalkundige

Dr. Jesse de Does – computerlinguïst

Job van Doeselaar – webmaster

Laura van Eerten MA – communicatieadviseur

Drs. Mathieu Fannee – systeemontwikkelaar

Drs. Thomas Haga – taalkundig assistent

Dr. Kris Heylen – onderzoeker/taalkundige

Ruud de Jong – systeemontwikkelaar

Dr. Dirk Kinable – onderzoeker/taalkundige

Drs. Marco van der Laan – systeembeheerder

Dr. Frank Landsbergen – computerlinguïst

Koen Mertens – systeemontwikkelaar

Drs. Jan Niestadt – systeemontwikkelaar

Henk van der Pol – systeemontwikkelaar (tot 1 december 2022)

Dr. Kathy Rys – onderzoeker/taalkundige

Prof. dr. Nicoline van der Sijs –

onderzoeker/taalkundige

Paulette Tacx – managementassistent

Drs. Rob Tempelaars – onderzoeker/taalkundige
(pensioen per 1 november 2022)

Dr. Carole Tiberius – computerlinguïst

Dr. Vincent Vandeghinste – computerlinguïst en
CLARIN-coördinator

Lic. Katrien Van pellicom –

onderzoeker/taalkundige

Drs. Boukje Verheij – onderzoeker/taalkundige

Rayvano van Vliet – systeembeheerder

Machteld de Vos, MPhil – onderzoeker in
opleiding

Drs. Vivien Waszink – onderzoeker/taalkundige

Karin van Weerlee – managementassistent

Laetitia de Winter – administrateur

Bijlage 3: Publicaties, lezingen, media, prijzen etc.

Publicaties

- Beelen, H. & Van der Sijs, N. (2022). Tolken en taalmannen. In: *Onze Taal* 4, 30.
- Beelen, H. & Van der Sijs, N. (2022). Vintage. In: *Onze Taal* 4, 30.
- Beelen, H. & Van der Sijs, N. (2022). “Injurieuze expressien”. Scheldwoorden in Amsterdamse notarisakten 1583-1796. In: *Alle Amsterdamse Akten, Jaarboek Amstelodamum 2022*, pp. 166-181.
- Beliën, Maaïke (2022). Woordvorming is overal! Een vernieuwd hoofdstuk in de e-ANS. Op: www.neerlandistiek.nl.
- Bonth, Roland de (2022). Achter het achtervoegsel 13: Onderwonder. Op: www.neerlandistiek.nl (6 januari)
- Bonth, Roland de (2022). Plofklas en preteaching. Op: www.neerlandistiek.nl (18 januari)
- Bonth, Roland de (2022). Terug naar de bron: verwijzen naar het WNT. Op: www.neerlandistiek.nl (13 februari)
- Bonth, Roland de (2022). Etymologica: Alle hens aloud? Op: www.neerlandistiek.nl (21 februari)
- Bonth, Roland de (2022). Podcasts over taal. Op: www.neerlandistiek.nl (15 februari)
- Bonth, Roland de (2022). Corpus Oudnederlands Online! Op: www.neerlandistiek.nl (23 februari)
- Bonth, Roland de (2022). Grâce aux corsaires anglais, nous savons comment écrivaient les Néerlandais aux XVIIe et XVIIIe siècles. Op: les-plats-pays.com
- Bonth, Roland de (2022). Achter het achtervoegsel 15: Past u precies! Op: www.neerlandistiek.nl (3 maart)
- Bonth, Roland de (2022). Ich hauo Oudnederlands gescriuen! Op: www.neerlandistiek.nl (17 maart)
- Bonth, Roland de (2022). Achter het achtervoegsel 16: Van retinette naar flexilette. Op: www.neerlandistiek.nl (7 april)
- Bonth, Roland de (2022). Etymologica: Er zijn meer hondjes die Luna heten. Op: www.neerlandistiek.nl (18 april)
- Bonth, Roland de (2022). Het juiste, andere woord. Op: www.neerlandistiek.nl (19 april)
- Bonth, Roland de (2022); Achter het achtervoegsel 17: Van Ensignette tot Haponette. Op: www.neerlandistiek.nl (5 mei)
- Bonth, Roland de (2022). Twee Taalportalen, voor onderwijs en onderzoek. Op: www.neerlandistiek.nl (16 mei)
- Bonth, Roland de (2022); Het ‘modern lemma’ maakt het verleden toegankelijk. Op: www.neerlandistiek.nl (11 juni)
- Bonth, Roland de (2022). Voyant Tools: snel (eigen) teksten analyseren. Op: www.neerlandistiek.nl (17 juni)
- Bonth, Roland de (2022). Van Cd-Rom Middelnederlands naar Corpus Middelnederlands. Op: www.neerlandistiek.nl (21 juni)
- Bonth, Roland de (2022). Onderzoek in de digitale bronnen van het INT. Op: www.neerlandistiek.nl (22 juni)

- Bonth, Roland de (2022). Achter het achtervoegsel 19: De verleden tijd van stofzuigers. Op: www.neerlandistiek.nl (7 juli)
- Bonth, Roland de (2022). Roland de Bonth kiest Willem Frederik Hermans. Op: www.neerlandistiek.nl (29 juli)
- Bonth, Roland de (2022). Achter het achtervoegsel 21: Eten met -ette. Op: www.neerlandistiek.nl (1 september)
- Bonth, Roland de (2022). Etymologica: Kaesijoleert? Op: www.neerlandistiek.nl (5 september)
- Bonth, Roland de (2022). Samen Lezen. Op: www.neerlandistiek.nl (14 september)
- Bonth, Roland de (2022). Struinen in Streuvels. Op: www.neerlandistiek.nl (3 oktober)
- Bonth, Roland de (2022). Het plusje in het Vroegmiddelnederlands Woordenboek. Op: www.neerlandistiek.nl (7 oktober)
- Bonth, Roland de (2022). neerLanDistIek Is eCht onMIsbaar Voor eLke taaL- en VIsLiefhebber Uit CaLLants oog. Op: www.neerlandistiek.nl (20 oktober)
- Bonth, Roland de (2022). Heilig en veilig: barbette. Op: www.neerlandistiek.nl (3 november)
- Bonth, Roland de (2022). Hij/zij/het brengt ons Sint-Nicolaas. Op: www.neerlandistiek.nl (10 november)
- Bonth, Roland de (2022). Jantje zag eens een hangende pruim. Op: www.neerlandistiek.nl (17 november)
- Bonth, Roland de (2022). Etymologica: chowchow. Op: www.neerlandistiek.nl (12 november)
- Bonth, Roland de (2022). *Struinen in Streuvels* [gelegenheidswoordenboekje ter gelegenheid van de 151^e geboortedag van Stijn Streuvels]. Op: ivdnt.org (vgl. de aankondiging op www.neerlandistiek.nl)
- Bonth, Roland de (2022). "Zwakheden des hoogen ouderdoms?" Korte aanmerkingen op *Bilderdijs Korte aanmerkingen*. In: *Jaarboek Bilderdijsk 2022*. Hilversum: Verloren, 118-135.
- Bonth, Roland de (2022). Elke twee weken een bijdrage aan de webrubriek 'Terug in de Taal'. Op: ivdnt.org
- Bonth, Roland de, Laura van Eerten (2022). Weg van woorden. Een talige stadswandeling en [audiotour](#) gemaakt voor de festivalmaand Letterlijk Leiden.
- Eerten, Laura van (2022). [Hoe letterlijk is het woord 'letterlijk'?](#) Podcastaflevering *Waar komt pindakaas vandaan?* voor de festivalmaand Letterlijk Leiden.
- Eerten, Laura van (2022). Ik wil dansen. Hedendaagse Nederlandstalige pop. In: *Onze Taal* 9, 36-37.
- Eerten, Laura van (2022). In je moerstaal. Nederpop, nederhop en luisterliedjes van nu. In: *Onze Taal* 9, 34-35.
- Eerten, Laura van (2022). Jaaroverzicht 2022 in nieuwe woorden. Op: ivdnt.org.
- Eerten, Laura van (2022). [Klimaatwoordenquiz](#). Op: kennislink.nl.
- Eerten, Laura van (2022). [Over taal gesproken](#). 3 podcastafleveringen i.s.m. Genootschap Onze Taal.
- Fannee, M. (2022). *Atlas van Laatmiddeleeuws Warmond – Reconstructie van het dorp en zijn omgeving (1250-1574)*. Historisch Genootschap Warmelda, Warmond. (456 pp)
- Fannee, M. (2022). *Warmonds toponymisch woordenboek – Middeleeuwen & Vroegmoderne tijd*. Historisch Genootschap Warmelda, Warmond. (260 pp)
- Fannee, M. (2022). Het Hemmeer, een middeleeuwse turfafgraving. In: *De Hekkenluiter*, Historisch

- Genootschap Warmelda, jaargang 19, nr. 1, 44-56.
- Hubers, F., Cucchiari, C. & Van der Sijs, N. (2022). Knowledge of idiomatic expressions in the native language: do emigrants lose their touch? In: *Lingua* 271, pp. 1-18.
- Kuhn, Tanara Zingano, Špela Arhar Holdt, Rina Zviel-Girshin, Rina, Ana R. Luís, Carole Tiberius, Carole, Kristina Koppel, Branislava S. Todorović, Iztok Kosem. (2022). Introducing Crowll – the Crowdsourcing for Language Learning game. In *Book of Abstracts of the XX Euralex International Congress*, 12-16 July 2022, Mannheim, Germany.
- Kuhn, Tanara Zingano, Carole Tiberius, Špela Arhar Holdt, Iztok Kosem, Kristina Koppel, Rina Zviel-Girshin, Ana R. Luís, Ana. (2022). *The CrowLL project - Manually-annotated corpora for teaching and learning purposes of Brazilian Portuguese, Dutch, Estonian, and Slovene*. [Poster](#) presented at the CLARIN Bazaar 2022. CLARIN Annual Conference 2022, 10-12 October 2022, Prague, Czech Republic.
- Landsbergen, F. (2022). *Das dicke Buch der Alphabete*. Impian.
- Lemmens, A. & Vandeghinste, V. (2022). A Lightweight NLP Workflow Engine for CLARIN-BE. In: *Proceedings of the CLARIN Annual Conference*. Prague, Czech Republic.
- Martelli, Federico, Marco Maru, Cesare Campagnano, Roberto Navigli, Paola Velardi, Rafael-J Ureña-Ruiz, Francesca Frontini, Valeria Quochi, Jelena Kallas, Kristina Koppel, Margit Langemets, Jesse de Does, Rob Tempelaars, Carole Tiberius, Rute Costa, Ana Salgado, Simon Krek, Jaka Čibej, Kaja Dobrovoljc, Polona Gantar, Tina Munda. (2022). ELEXIS deliverable 3.8 Lexical-semantic analytics for NLP - final report. Op: elex.is
- Martelli, Federico, Roberto Navigli, Simon Krek, Jelena Kallas, Polona Gantar, Svetla Koeva, Sanni Nimb, Bolette Sandford Pedersen, Sussi Olsen, Margit Langemets, Kristina Koppel, Tiiu Üksik, Kaja Dobrovoljc, Rafael Ureña-Ruiz, José-Luis Sancho-Sánchez, Veronika Lipp, Tamás Váradí, András Gyórfy, Simon László, Valeria Quochi, Monica Monachini, Francesca Frontini, Carole Tiberius, Rob Tempelaars, Rute Costa, Ana Salgado, Jaka Čibej, Tina Munda. (2022). Parallel sense-annotated corpus ELEXIS-WSD 1.0, Slovenian language resource repository CLARIN.SI, ISSN 2820-4042. Op: clarin.si
- Montes, M. & Heylen, K. (2022). Visualizing distributional semantics. In: D. Tay & M. Pan (Eds.), *Data Analytics in Cognitive Linguistics. Methods and Insights*. Mouton De Gruyter. (geaccepteerd)
- Norré, M., Vandeghinste, V., François, T. & Bouillon, P. (2022). Investigating the Medical Coverage of a Translation System into Pictographs for Patients with an Intellectual Disability. In: *Proceedings of the 9th Workshop on Speech and Language Processing for Assistive Technologies (SLPAT)*. Dublin, Ireland.
- Romein, C. Annemieke, et al. [onder wie Depuydt, Katrien] (2022). Exploring Data Provenance in Handwritten Text Recognition Infrastructure: Sharing and Reusing Ground Truth Data, Referencing Models, and Acknowledging Contributions. Starting the Conversation on How We Could Get It Done. Op:

zenodo.org

- Rys, K. & Schaffel-Bremenkamp, E. (2022). Using data of Zeelandic Flemish in Espírito Santo, Brazil for historical reconstruction. In: Matt Coler & Andrew Nevins (eds.), *Contemporary research in minoritized and diaspora languages of Europe (Contact and Multilingualism 6)*. Berlin: Language Science Press. 341-384. DOI: 10.5281/zenodo.7442323. Op: zenodo.org
- Shterionov, D. et al. [onder wie Vandeghinste, V.] (2022). Sign Language Translation: Ongoing Development, Challenges and Innovations in the SignON Project. In: *Proceedings of the 23rd Annual Conference of the European Association for Machine Translation*, pp. 325–326, Ghent, Belgium. European Association for Machine Translation.
- Van der Sijs, N. (2022). 1602. De Nederlandse taal gaat overzee'. In: *Nog meer Wereldgeschiedenis van Nederland*, red. L. Heerma van Voss e.a., Amsterdam, pp. 183-188.
- Van der Sijs, N. (2022). 481. Het Frankische rijk als kraamkamer van het Nederlands'. In: *Nog meer Wereldgeschiedenis van Nederland*, red. L. Heerma van Voss e.a., Amsterdam, pp. 81-86.
- Van der Sijs, N. (2022). Couranten Corpus staat online. Op: www.neerlandistiek.nl (18 mei)
- Van der Sijs, N. (2022). De Oekraïense graanschuur. In: *Onze Taal* 6, 36.
- Van der Sijs, N. (2022). Dialectverschillen binnen Amsterdam. In: *Amstelodamum* 1, 41-51.
- Van der Sijs, N. (2022). Een kolossaal bouwwerk uit losse brokstukken. Recensie van Jelle Stegeman, Grote geschiedenis van de Nederlandse taal. In: *De Lage Landen*.
- Van der Sijs, N. (2022). Etymologica: de Oekraïense graanschuur. Op: www.neerlandistiek.nl (4 april)
- Van der Sijs, N. (2022). Etymologica: de huzarensalade. Op: www.neerlandistiek.nl (10 januari)
- Van der Sijs, N. (2022). Etymologica: kiekeboe! Op: www.neerlandistiek.nl (27 juni)
- Van der Sijs, N. (2022). Etymologica: prinsessenbonen. Op: www.neerlandistiek.nl (14 februari)
- Van der Sijs, N. (2022). Geheimtalen vroeger en nu. In: *Dialectendoeboek. De schatkamer van 90 jaar Meertens Instituut*, Van Oostendorp, M. & Wolff, S. (reds.), pp. 141-146
- Van der Sijs, N. (2022). Gelegenheidswoordenboekje 'Spiegels voor de jeugd'. Op: ivdnt.org
- Van der Sijs, N. (2022). Herkomst en gebruik van het woord wolf in het Nederlands. Op: www.historici.nl
- Van der Sijs, N. (2022). How the World Views Dutch and Dutch Speakers. Op: www.the-low-countries.com
- Van der Sijs, N. (2022). Kiev of Kyjiv? De spelling van Oekraïense plaatsnamen. In: *Onze Taal* 4, 22-23.
- Van der Sijs, N. (2022). L'image des Pays-Bas et du néerlandais dans d'autres langues. Op: www.les-plats-pays.com
- Van der Sijs, N. (2022). Losgehen wie ein Holländer. Het imago van Nederland(s) in andere talen. In: *De Lage Landen*.
- Van der Sijs, N. (2022). Manager. In: *Onze Taal* 8, 34.
- Van der Sijs, N. (2022). Standaardnederlands is minder zuidelijk dan gedacht. In: *De Lage Landen*
- Van der Sijs, N. (2022). Vaccinatie. In: *Onze Taal* 2/3, 30.
- Van der Sijs, N. (2022). Visagist. In: *Onze Taal* 7, 34.

- Van der Sijs, N. (2022). Vlegeljaren. In: *Onze Taal* 1, 30.
- Van der Sijs, N. (2022). Welke talen gebruik jij? Oproep aan docenten en studenten in het hoger en universitair onderwijs in Nederland en België van alle vakgebieden. Op: www.neerlandistiek.nl (16 juni)
- Van der Sijs, N. (2022). À la recherche du temps perdu: het WNT als Nederlands taalmuseum. In: *Boeken die geschiedenis schreven*, red. Kasper van Ommen & Garrelt Verhoeven, Amsterdam, pp. 237-243.
- van der Sijs, N. (2022). Oude vogelnamen. En wat we daarvan kunnen leren. In: *Gevleugelde geschiedenis van Nederland. De Nederlanders en hun vogels*, red. Jan Luiten van Zanden, Helena Čordašev en Erik de Bruin, 89-103.
- Van der Sijs, N. & Weduwen, A. der, m.m.v. Feis, B., Lammers, M. & Landzaat, E. (2022). *Franse Tirannie. Het Rampjaar 1672 op school*, Zwolle: Waanders Uitgevers.
- De Sisto, M., Vandeghinste, V., Egea Gómez, S., De Coster, M., Shterionov, D. & Saggion, H. (2022). Challenges with Sign Language Datasets for Sign Language Recognition and Translation. In: *Proceedings of the 13th International Conference on Language Resources and Evaluation (LREC)*. Marseille, France.
- Steurs, F. (2022). The Dutch language: research and policy in the digital age. In: E.J. Jozic, S. Kirchmeier (Eds.), *The Role of National Language Institutions in the Digital Age*, (81-87). Presented at the 18th EFNIL Conference, Cavtat, 06 Oct 2021-08 Oct 2021. Budapest. ISBN: 978-963-9074-92-7. Op: www.youtube.com
- Steurs, F. & Heylen, K. (2022). De Waagmeester van de woordenschat. In: *Nederlandse Taalkunde* 27, 2, p. 163-174.
- Steurs, F. & Kinable, D. (2022). Het Expertisecentrum Nederlandstalige Terminologie (ENT): een actief en attractief platform voor terminologie. In *Nieuwe Reeks Terminologie in het Nederlandse Taalgebied – 2. TiNT-dag 2021*, pp. 18-25. Op: nlterm.eu
- Tasovac, T., Carole Tiberius, Claudia Bamberg, Andrea Bellandi, Thomas Burch, Rute Costa, Matej Ďurčo, Francesca Frontini, Julia Hennemann, Kris Heylen, Miloš Jakubiček, Fahad Khan, Anne Klee, Iztok Kosem, Vojtěch Kovář, Ondřej Matuška, John McCrae, Monica Monachini, Karlheinz Mörth, Tina Munda, Valeria Quochi, Andraž Repar, Christophe Roche, Ana Salgado, Henrike Sievers, Tamáz Váradi, Sandra Weyand, Anna Woldrich, Susanne Zhanial. (2022). ELEXIS deliverable 5.3 Overview of Online Tutorials and Instruction Manuals. Op: <https://elex.is/>
- Tempelaars, Rob (2022). *Leukerlands. Taalplezier in stukjes*. Uitgeverij Scriptum/Instituut voor de Nederlandse Taal.
- Tempelaars, R. & Colman, L. (2022). In memoriam Dick Wortel (1946-2022). Op: www.neerlandistiek.nl (19 augustus)
- Tempelaars, R. & Colman, L. (2022). In memoriam Dick Wortel (1946-2022). In: *Trefwoord*.
- Tiberius, Carole, Simon Krek, John McCrae, Michal Měchura, Toma Tasovac (2022). Deliverable 1.5 Best practices for lexicography – final report. Op: elex.is

- Tiberius, Carole, Tina Munda, Andraž Repar, Simon Krek (2022). Deliverable D1.6 Lexicographic data in ELEXIS. Op: elex.is
- Tiberius, Carole, Jelena Kallas, Svetla Koeva, Margit Langemets, Iztok Kosem (2022). An insight into lexicographic practices in Europe. Results of the extended ELEXIS Survey on User Needs. In Klosa-Kückelhaus, A., Engelberg, S., Möhrs, C and P. Storjohann (eds), *Dictionaries and Society. Proceedings of the XX EURALEX International Congress*. Mannheim: IDS-Verlag, 509-521. Op: euralex.org
- Tiberius, Carole (2022). Country Profile The Netherlands. In Eileen Marra, Andrea Lösch, Stefania Racioppa, Hélène Mazo, Maria Giagkou, *AI for a Multilingual Europe: Why language data matters*. ELRC White Paper, pp.172-176. Op: www.lr-coordination.eu
- De Tier V.(2022). 't Zwien deur de beeëtn jaagn en andere leuke West-Vlaamse woorden en uitdrukkingen (1). In: *Diejalektgazette Bachtn de Kuupe*, jg. 17, nr.1, pp. 7-9.
- De Tier, V. (2022). Preus lijk veertig zijn en andere vergelijkingen. In: *Diejalektgazette Bachtn de Kuupe*, jg. 17, nr. 4, pp. 7-9.
- De Tier, V. (2022). Sjchuufelt joen gat ook voe weg? West-Vlaamse spreuken van Djoos Utendoale. In: *Diejalektgazette Bachtn de Kuupe*, jg. 17, nr. 3, pp. 7-9.
- De Tier, V. (2022). Spiegel van de taalgeschiedenis. Dialecten van het Nederlands. In: *Geschiedenis*, nr. 33, dertiende jaargang, mei 2022, pp. 42-44.
- De Tier, V. (2022). 't Zwien deur de beeëtn jaagn en andere leuke West-Vlaamse woorden en uitdrukkingen (2). In: *Diejalektgazette Bachtn de Kuupe*, jg. 17, nr. 2, pp. 7-9
- De Tier, V. & Sterkenburg, P. van (2022). De lisdodde in de Zeeuwse dialecten. In: *Nehalennia*, 215, pp. 20-22.
- De Tier, V. & Sterkenburg, P. van (2022). Vernielende nikkestaart in Zierikzee. In: *Nehalennia*, 217, pp. 16-17.
- De Tier, V. & Sterkenburg, P. van (2022). Zeeuwse namen voor de okkernoot. In: *Nehalennia*, 218, pp. 19-20.
- De Tier, V. & Sterkenburg, P. van (2022). Zeeuwse wentelteefjes. In: *Nehalennia*, 216, pp. 15-16.
- Vandeghinste, V. (2022). Een persoonlijk stukje geschiedenis van de computerlinguïstiek. In: *Liber Amicorum Francisci Affinii. Festschrift ter gelegenheid van het emeritaat van Frank Van Eynde*, pp. 41-65.
- Vandeghinste, V. & Van de Cruys, T. (2022). Voorgetrainde Grote Taalmodellen en Automatische Vertaling. In: *Dixit. Tijdschrift over toegepaste taal- en spraaktechnologie*, december 2022, pp. 18-19.
- Vandeghinste, V. (ed.) (2022). *Liber Amicorum Francisci Affinii. Festschrift ter gelegenheid van het emeritaat van Frank Van Eynde*. Op: www.ccl.kuleuven.be
- Vandeghinste, V., Van Dyck, B., De Coster, M., Goddefroy, M., & Dambre, J. (2022). BeCoS Corpus: Belgian Covid-19 Sign Language Corpus. A Corpus for Training Sign Language Recognition and Translation. In: *Computational Linguistics in the Netherlands*, 12, 7–17.
- Vandeghinste, V., Van Eerten, L. & Noë, R. (2022). Hoe intelligent is Google Translate nu echt? Podcast. Instituut voor de Nederlandse Taal en Genootschap Onze Taal. Op: www.neerlandistiek.nl

- Vos, M. de (2022). In Between Description and Prescription: Analysing Metalanguage in Normative Works on Dutch 1550–1650. In: *Languages* 2022, 7, 89.
- Vos, M. de (2022). ‘Deze verscheydenheit der Voornamen’. Codification of third-person pronouns in Early Modern Dutch. In: *Taal en Tongval* 74 (1), pp. 107-146.
- Waszink, Vivien & Alex Reuneker (2022). Nederlandstalig in de Top-40. In: *Onze Taal* 9.
- Waszink, Vivien (2022). Alscuus, dickpic en psst-terreur. Nieuwe woorden in tijden van grensoverschrijdend gedrag. In: *Onze Taal* 4.
- Waszink, Vivien (2022). ‘Je taal zodanig onder de knie hebben dat je er zelf mee gaat beginnen dansen’. Rapper Zwangere Guy: een uit de 1000. In: *Onze Taal* 7.
- Waszink, Vivien (2022). *Dat mag je óók (al niet meer) zeggen*. Boek, uitgegeven door *Onze Taal*.
- Waszink, Vivien (2022). Schiet het inclusieve taalgebruik te ver door? In: *Kerk en Leven*.
- Waszink, Vivien (2022). Je mag ook niets meer zeggen. Over taal, vrijheid en controle. In: [rekto:verso](#).

Lezingen en presentaties

- Beliën, Maaïke en Frank Landsbergen. *Gel-ANS-eerd! De nieuwe website van de e-ANS, en hoe verder*. Presentatie op de Grote Taaldag. Utrecht, 4 februari 2022.
- Beliën, Maaïke en Frank Landsbergen. *De e-ANS vernieuwt!* Presentatie op het 21^e Colloquium Neerlandicum. Nijmegen, 23 augustus 2022.
- Bonth, Roland de. *Oude kranten lezen voor Proteusianen. Het Couranten Corpus*. Presentatie voor renaissance-dispuut Proteus. Leiden, 16 november 2022.
- Depuydt, Katrien en Nicoline van der Sijs. *Lessons learned from the 17th-century CourantenCorpus*. Presentatie op de CLARIAH & NDE Annual Conference. Utrecht, 24 november 2022.
- Depuydt, Katrien. *Towards a Diachronic Semantic Lexicon of Dutch*. Presentatie op de conferentie *Historical Lexicography in a Digital Age*. Kopenhagen, Royal Academy of Sciences and Letters, 17-18 November 2022.
- Does, Jesse de & Nicoline van der Sijs (2022). *A new website for the GLAD database*. Presentatie op First Asia Symposium for GLAD contributors, 26 februari 2022.
- Eerten, Laura van & Vivien Waszink. *Bieterballenquiz*. The Wooden Barrel, Leiden, 8 december 2022.
- Eerten, Laura van. [Taalvragenvuur](#). Presentatie i.s.m. Genootschap Onze Taal, Team Taaladvies en de Taalunie, 6 oktober 2022.
- Eerten, Laura van. *Wat mag je ook zeggen? Een gesprek over nieuwe, inclusieve taal*, met Raymond Noë (Onze Taal) en Vivien Waszink, 5 oktober 2022. Op: [youtube.com](https://www.youtube.com/watch?v=...)
- Sijs, Nicoline van der. Moderator bij Eastsplainers #1: Language, lecture on Linguicide by the Ukrainian linguist Oksana Kononchuk, 30 november 2022. Op: www.uva.nl
- Sijs, Nicoline van der. *Spiegels der Jeugd: de oudste geschiedenisleerboekjes*. Studiedag Stichting Geschiedenis Kinder- en Jeugdliteratuur (SGKJ), 11 november 2022.

- Sijs, Nicoline van der. *Enquête Taalattitudes in de Lage Landen*. Lancering tijdens studiemiddag. Luik, 10 mei 2022.
- Sijs, Nicoline van der. *Het beeld van talen en volken in uitdrukkingen*. Publiekslezing, Luik, 26 april 2022.
- Sijs, Nicoline van der. *De Transcriptor*. Bij: De taalsector in actie voor slachtoffers van de oorlog in Oekraïne, DRONGO, 6 april 2022, via zoom.
- Steurs, F. *Toegang tot de bronnen en de taaltechnologie voor het Nederlands*. Presentatie op uitnodiging van IVN Colloquium Neerlandicum op het Voortvarend Nederlands Colloquium Neerlandicum. Nijmegen, 22 Aug 2022-26 Aug 2022.
- Steurs, F. & K. Tryczyńska. *Europa, lidstaten en het arbeidsrecht: uitdagingen in de meertalige terminologie*. Presentatie op uitnodiging van IVN Colloquium Neerlandicum op het Voortvarend Nederlands Colloquium Neerlandicum, Nijmegen, 22 Aug 2022-26 Aug 2022.
- Steurs, F. *Waarom Google Translate geen woordenboek is*. Presented at the Studiedag Google Translate, Leiden, 15 Jun 2022-15 Jun 2022.
- Tiberius, Carole. *Elexifier*. Presentatie op ELEXIS Showcase Event: ELEXIS from A to Ω: Outcomes, Sustainability & Afterlife of a new European Lexicographic Infrastructure. Florence, 7 juni 2022.
- Tiberius, Carole. *Elexifier*. Presentatie op ELEXIS-workshop op EURALEX 2022. Mannheim, 11 juli 2022.
- Tiberius, Carole. *An insight into lexicographic practices in Europe. Results of the extended ELEXIS Survey on User Needs*. Presentatie op EURALEX 2022. Mannheim, 14 juli 2022.
- Tiberius, Carole. *eTranslation en Taaldata*. Presentatie op de bijeenkomst eTranslation voor overheden. Den Haag, 14 december 2022.
- De Tier, V. *Dialecten op de kaart*. Presentatie als salongast in het Gents Universitair Museum (GUM) op de erfgoeddag. Gent, 24 april 2022.
- De Tier, V. *Limburgse dialecten op het internet*. Presentatie op VLDN-Congres. Kinrooi, 30 april 2022.
- De Tier, V., Depuydt, K., De Does, J. e.a. *Towards a new lexicographical Infrastructure for the Dutch Dialects: the Database of Southern Dutch Dialects project*. Presentatie op Methods XVII, (Dia)Lects in the 21st Century. Mainz, 2-5 augustus 2022.
- De Tier, V. *Hoe zoek je in de Database voor de Zuidelijk-Nederlandse Dialecten (DSDD)*. Presentatie op de vrijwilligersbijeenkomst Zeeuwse Dialect Vereniging (Zeeuws-Vlaanderen). Biervliet, 19 mei 2022.
- De Tier, V. *De geschiedenis van de Zeeuws-Vlaamse dialecten*. 3-delige cursus. Clinge, 5-12 en 19 oktober 2022.
- De Tier, V. *Zeeuwse dialecten op de kaart*. Presentatie op de vrijwilligersbijeenkomst Zeeuwse Dialect Vereniging (Zeeuws-Vlaanderen). Biervliet, 7 november 2022.
- De Tier, V. *Het Oudenaardse dialect*. Lezing voor ACW Oudenaarde. Oudenaarde, 8 november 2022.
- De Tier, V. *Ons moeder zei altijd*. Lezing voor InnerWheel Oudenaarde. Oudenaarde, 5 december 2022.
- De Tier, V. *De totstandkoming van de Atlas van het dialect in Vlaanderen*. Lezing voor Library Lunch @FLW bibliotheek. Gent, 13 december 2022

- De Tier, V. *De geschiedenis van de Zeeuws-Vlaamse dialecten*. Lezing voor Past Rotary. IJzendijke, 20 december 2022.
- Vandenbergh, R., De Tier, V. & Hellebaut, L. *De database van de Zuidelijk-Nederlandse Dialecten (DSDD): een toegankelijke databank als sluitstuk van een eeuw dialectonderzoek*. Posterpresentatie op de 3e LW Research Day. Gent, 28 april 2022.
- Waszink, Vivien. *Havermelkelite, vertiktokking en cheugy. De kracht van nieuwe woorden*. Lezing Universiteit Leiden City of Science, 21 januari 2022.
- Waszink, Vivien en Alex Reuneker. *Hiphop hoera, doe de hiphopquiz*. Expeditie Next, wetenschapsfestival voor kinderen. Franeker, 6 mei 2022.
- Waszink, Vivien. *Beyoncéfeminisme, alsuus en ruilfeestdag. Nieuwe woorden zijn er altijd*. Webinar KTV Kennisnet, 19 mei 2022.
- Waszink, Vivien. *Dat mag je óók (al niet meer) zeggen*. Boekpresentatie bij Ledenvergadering Onze Taal, 4 juni 2022.
- Waszink, Vivien. *Dat mag je óók (al niet meer) zeggen*. Lezing en inspiratielunch Taalunie, 23 juni 2022.
- Waszink, Vivien. *Dat mag je óók (al niet meer) zeggen*. Lezing IVN-congres Nijmegen, 24 augustus 2022.
- Waszink, Vivien. *Beyoncéfeminisme, alsuus en omhoogfalen. Hedendaagse woordenschat*. Lezing voor burgemeester en wethouder, 24 augustus 2022.
- Waszink, Vivien. *Dat mag je óók (al niet meer) zeggen*. Lezing voor de redactie van *de Volkskrant*, 12 september 2022.
- Waszink, Vivien. *Goede huisvader of voorzichtig en redelijk persoon? Inclusief taalgebruik*. Kennotelezing en paneldiscussie voor vertaalopleiding Vrije Universiteit Brussel (VUB), 30 september 2022.
- Waszink, Vivien. *Mag je nou echt niks meer zeggen? Inclusief taalgebruik*. Lezing lerarenopleiding Hogeschool Rotterdam, 4 oktober 2022.
- Waszink, Vivien. *Mag je nou echt niks meer zeggen? Inclusief taalgebruik*. Lezing VRT Brussel, 6 oktober 2022.
- Waszink, Vivien. *Mag je nou echt niks meer zeggen? Inclusief taalgebruik*. Lezing Radboud Universiteit, 18 november 2022.
- Waszink, Vivien. *Gespreksnarcisme, energiearmoede en plaktivist*. Lezing Fryske Akademy, 22 november 2022.
- Waszink, Vivien. *Dat mag je óók al niet meer zeggen! Keynotelezing congres Overheidsacademie communicatie*, 24 november 2022.
- Waszink, Vivien. *Over mooie beentjes, othering en pssst-terreur. De taal van seksueel grensoverschrijdend gedrag*. Lezing voor Mariëtte Hamer, regeringscommissaris seksueel grensoverschrijdend gedrag, 1 november 2022.
- Wermuth, M-C., Steurs, F., Buyschaert, J., Tiberius, C. (contr.), Van Gent, I. (contr.). *Terminology data sharing in the Dutch language area*. Presentatie op de FedTerm Terminology Summit. Reykjavik, 28

maart 2022-29 maart 2022.

Congressen & workshops

- Bonth, Roland de. *Onderzoek in de digitale bronnen van het INT*. Profielwerkstukmiddag Nationaal Platform van de Talen, 9 juni 2022. Op: [youtube.com](https://www.youtube.com)
- Depuydt, Katrien en Jesse de Does. Deelname op uitnodiging aan een workshop over historische lexicografie bij Det Danske Sprog- og Litteraturselskab, DSL (The Society for Danish Language and Literature). Kopenhagen, 16 november 2022.
- Tier, V. de, G. Dehaes, G. de Jong en A.S. Ghyselen. Streektaalconferentie: Taalvariatiebeleid in tijden van globalisering: overbodig of net cruciaal. Stichting Nederlandse Dialecten i.s.m. Variaties vzw, Brussel! en de Taalunie, 23 september 2022.

Onderwijs

- Bonth, Roland de. College PAO Taalkunde & taalbeheersing 'Kranten lezen voor gevorderden?'. Radboud Universiteit, 13 oktober 2022
- Bonth, Roland de. Drie maal gast(werk)college aan studenten Nederlands van de Universiteit Leiden over het zoeken in de Online historische woordenboeken (gtb.ivdnt.org), 16 en 17 november 2022.
- Tempelaars, Rob. *Woordbetekenissen, betekenisrelaties en definities*. College computationele lexicografie aan de KU Leuven. Leuven, 28 maart 2022.
- Tempelaars, Rob (2022). *Word senses, sense relations and definitions*. College computationele lexicografie aan de Universiteit Leiden. Leiden, 13 april 2022.
- Tiberius, Carole. MA-collegereeks Corpus Lexicography. Universiteit Leiden, februari-mei 2022.
- Tiberius, Carole. College over Dictionary Writing Systems in de MA-collegereeks Computationele Lexicografie. KU Leuven, 28 maart 2022.
- Tier, V. de. College over de Database voor de Zuidelijk-Nederlandse Dialecten (DSDD) in de MA-collegereeks Computationele Lexicografie. KU Leuven, 23 maart 2022.
- Waszink, Vivien. College over inclusieve taal, BA Nederlandse Taal en Cultuur Universiteit Leiden, 6 april 2022.
- Waszink, Vivien en Jaap de Jong. Collegereeks MA-scriptie inclusieve taal, MA Nederlandse Taal en Cultuur Universiteit Leiden, december 2022-april 2023.

In de media

- Bonth, Roland de. Interview over gelegenheidswoordenboekje *Struinen in Streuvels* met Lieven Vandenhoute in Radio 1-programma [Nieuwe Feiten, 5 oktober 2022](#).
- Sijs, Noline van der. [Villa VdB over het Couranten Corpus](#) op NPO Radio 1, 9 augustus 2022.
- Steurs, Frieda. Interview voor programma Pointer o.a. over de pannenkoekenzaak, op NPO Radio 1, 18 september 2022.

Tempelaars, Rob. '[Over o.a. palindromen en woordenboeken](#)'. Interview op Radio Weetlust, programma in het kader van Leiden als European City of Science, 2 februari 2022.

Tempelaars, Rob. 'Over hoogfrequente woorden'. Interview op NPO radio 1, rubriek 'Feit of Fictie', 7 oktober 2022.

Tier, V. de. 'de Zeeuwse dialecten'. Interview op Radio Centraal Antwerpen, rubriek 'Babilonië', 16 april 2022.

Tier, V. de. Interview in *Boekenspoiler* van de Taalsector. 18 oktober 2022.

Waszink, Vivien. '[Nieuwe Dikke van Dale wordt genderinclusief. 'Samenleving verandert'](#)'. Interview door RTL Nieuws (ook uitzending op Editie.nl, RTL 4), 11 januari 2022.

Waszink, Vivien. [Neologismen](#). Interview voor *Radio Weetlust*, 20 januari 2023.

Waszink, Vivien. '[Bodyshaming, influencer en airfryer: bijna tienduizend 'nieuwe' woorden in Dikke van Dale](#)'. Interview met *Hart van Nederland*, 8 februari 2022.

Waszink, Vivien. '[Van Japans tot Turks: nieuwe Dikke Van Dale vol buitenlandse woorden](#)'. Interview door RTL Nieuws (ook uitzending op Editie.nl, RTL 4), 10 februari 2022.

Waszink, Vivien. '[Zorgen de versoepelingen ook voor het verdwijnen van coronataal?](#)' Interview in *De Taalstaat*, NPO Radio 1, 19 februari 2022.

Waszink, Vivien. '[Vivien Waszink over 'Dat mag je óók \(al niet meer\) zeggen': Taalkundige schrijft boek over inclusief woordgebruik](#)'. In tijdschrift *Visie* (EO), 1 juni 2022.

Waszink, Vivien. Interview over boek [Dat mag je \(ook al niet meer\) zeggen](#)'. In *De Nieuws BV* op NPO Radio 1, 3 juni 2022.

Waszink, Vivien. '[Van 'vrouw' tot 'mens met baarmoeder': Wat mag je dan nog wél zeggen?](#)'. Interview in *Trouw*, 5 juni 2022 (ook doorgeplaatst in *Knack*).

Waszink, Vivien. '[Weg met app of airbag: moeten Engelse leenwoorden in de ban?](#)' Interview door RTL Nieuws (ook uitzending op Editie.nl, RTL 4), 7 juni 2022.

Waszink, Vivien. '[Mag je straks niets meer zeggen? "Nee, er komen integendeel veel woorden bij"](#)'. Interview in *Het Nieuwsblad*, 7 juni 2022 (ook doorgeplaatst in *Gazet van Antwerpen*).

Waszink, Vivien. '[Femanisme' en 'transmigrant' - Taalkundige Vivien Waszink schrijft boek over gevoelig taalgebruik: 'We mogen juist heel veel wél zeggen'](#)'. Interview in *Leidsch Dagblad*, 7 juni 2022.

Waszink, Vivien. '[Taalkundige Vivien Waszink: 'We mogen net méér zeggen vandaag, niet minder'](#)'. Interview in *De Morgen*, 7 juni 2022.

Waszink, Vivien. '[Een boek over uitspraken die niet meer kunnen? Daar moet Gerard meer over weten](#)'. Interview met *Gerard Ekdorf*, radio 10, 8 juni 2022.

Waszink, Vivien. 'Cursus taalgebruik in woke-tijden hard nodig'. Interview in *Nieuwsweekend*, NPO Radio 1, 11 juni 2022.

Waszink, Vivien. '[Verboden woorden? Je kunt vooral steeds meer zeggen](#)'. Interview op: *Kennislink*, 29 juni 2022.

Waszink, Vivien. Interview voor YouTube-platform *De Nieuwe Wereld*, 1 juli 2022.

Waszink, Vivien. [Interview Lunchroom](#), Radio NH, 20 juli 2022.

Waszink, Vivien. ‘[Over het woord ‘gezellig’](#)’. Interview op RTL Nieuws (ook in de uitzending van *Editie.nl*, RTL 4), 6 augustus 2022.

Waszink, Vivien. ‘[Noem mij maar ‘hen’](#): hoe de taal inclusiever kan worden en wat dat doet met het denken’. Interview in *de Volkskrant*, 12 augustus 2022.

Waszink, Vivien. ‘[Onderzoeker en taalkundige: ‘Als er iets verandert, hoeft dat echt niet negatief uit te pakken’](#)’. Interview op: [neerlandistiek.nl](#), 15 augustus 2022.

Waszink, Vivien. ‘[Politieagent of -persoon? Taalunie zoekt uitweg in gender-taalworsteling](#)’. Op: [nos.nl](#), 24 augustus 2022.

Waszink, Vivien. Interview in *Boekenspoiler* van de Taalsector, 20 september 2022.

Waszink, Vivien. ‘[Woorden als klimaatwappie zijn typisch voor deze tijd](#)’. Interview op Kennislink, 5 oktober 2022.

Waszink, Vivien. [Podcastinterview WOMEN Inc](#) over inclusieve taal, 6 oktober 2022.

Waszink, Vivien. Webinar Week van het Nederlands over inclusief taalgebruik (Onze Taal/INT), 6 oktober 2022.

Waszink, Vivien. ‘Mogen we dat ook al niet meer zeggen?’. Interview in *De Standaard*, 8 oktober 2022.

Waszink, Vivien. ‘[Waarom jongeren steeds meer therapietaal gebruiken](#)’. In: *Elsevier*, 14 november 2022.

Waszink, Vivien. ‘Taal bepaalt ons perspectief’. Interview voor tijdschrift *Lezen*, december 2022.

Waszink, Vivien. Interview over Woord van het Jaar, *Editie.nl*, december 2022.

Waszink, Vivien. ‘[Verikeäisering, kunstbooster en groenflatie: het Instituut voor de Nederlandse taal maakt een taaljaaroverzicht](#)’. Interview *Nieuwe Feiten*, radio 1 België, 20 december 2022.

Diversen

Beliën, Maaïke

- Redactielid *Nederlandse Taalkunde*

Bonth, Roland de:

- Oefening Woordspelingen maken. Op: [Schrijfkademie Pilot 1. Domein Taal](#)
- Auteur en jurylid van het Groot Leiderdorps Dictee (12 oktober 2022)
- Wandeling Weg van woorden in het kader van Letterlijk Leiden (18 november 2022)
- Coördinator en jurylid Profielwerkstukwedstrijdprijs van Levende Talen Nederlands (LTN)
- Opdracht over Nederlands voor de Taalkundeolympiade 2022 van de Universiteit Leiden
- Redacteur *Trefwoord*, het digitale tijdschrift voor lexicografie en lexicologie

- Kolomleider literatuuronderwijs HSN 35, 20 & 21 mei, Universiteit Antwerpen
- Redacteur Taalbeschouwing van het Handboek Didactiek Nederlands
- Eindredacteur van Ewoud Sanders, *Met de paplepel. Beeldvorming over joden in Nederlandstalige jeugdverhalen, 1782-heden*. Zutphen: Walburg Pers, 2022

Depuydt, Katrien

- Voor Isabeau De Smet en Freek Van de Velde, ten behoeve van een studie over lexicaal verlies in het Nederlands op basis van de woordenschat in het Oudnederlands woordenboek: voor in het Oudnederlands aangetroffen lemmata de ONW-frequentie, samen met (indien aanwezig) lemma-id, eerste en laatste attestatie, en aantal citaten in het WNT voor het betreffende lemma.
- Met het Nationaal Archief in Den Haag is een overeenkomst afgesloten voor het gebruik van het historisch lexicon (GiGaNt-Hilex) in hun applicatie *zoekintranscripties.nl*.

De Tier, Veronique

- Voorzitter Stichting Nederlandse Dialecten
- Bestuurslid Variaties vzw. Koepelorganisatie voor dialecten en oraal erfgoed in Vlaanderen
- Bestuurslid be.brusseleir
- Lid werkgroep infrastructuur Fries-Nederlandse contactvariëteiten
- Redactielid Tijdschrift Nehalennia

Eerten, Laura van

- Programmapartner Letterlijk Leiden

Sijs, Nicoline van der:

- Koning Willem-Alexander Leerstoel voor Lage Landen-studies aan de Faculteit Letteren en Wijsbegeerte van de Universiteit van Luik
- Lid van de jury voor de LOT-dissertatieprijs 2022
- Manuscriptcommissie Lucas van der Deijl, UvA, 2022, opponeren op 29-9-2022
- Mentor for the University of New Europe (UNE)
- Lid van het Taalstaatpanel voor de verkiezing van Taalstaatmeester 2022

Steurs, Frieda:

- Tijdens de TiNT-dag op 21 oktober 2022 ontving Frieda Steurs de Simon Stevinpenning: een onderscheiding voor een wetenschapper die zich zeer verdienstelijk maakt voor de Nederlandstalige

terminologie en vaktaal.

- National Anchor Point ELE (European Language Equality)
- Secretary-General CIPL (Comité International Permanente des Linguistes)
- Stichter Leerstoel 'Ethnolinguistic Vitality and Diversity' Universiteit Leiden

Tempelaars, Rob:

- Redacteur en redactiesecretaris van het tijdschrift *Trefwoord*
- Reviewer voor het tijdschrift *Neerlandica Wratislaviensia*
- Namens het INT betrokken bij een door diverse partijen opgezet project over geuren, dat uiteindelijk moet leiden tot het boek *NeusWijzer* (te verschijnen in 2023)

Tiberius, Carole

- Board of Trustees Adam Kilgarriff Prize
- Technical National Anchor Point ELRC (European Language Resource Coordination)
- Secretary-Treasurer EURALEX (European Association of Lexicography)

Verheij, Boukje

- Redacteur van Rob Tempelaars, *Leukerlands. Taalplezier in stukjes*. Uitgeverij Scriptum/Instituut voor de Nederlandse Taal, 2022

Waszink, Vivien

- Lid adviesgroep Grensoverschrijdend gedrag en seksueel geweld van regeringscommissaris Mariëtte Hamer
- Auteur lesmethode *Nieuw Nederlands*, Noordhoff, methodes vmbo 3 gt, 4 havo, 5 havo, 4 vwo, 5 vwo en 6 vwo

Begeleiding thesis, stage, vrijwilligers

Colman, Lut:

- Begeleiding werkstudenten Esther Simons en Mila van Nieuwenhuizen

Kinable, Dirk:

- Begeleiding stagiairs Madelief Meijl, Mees Drijgers, Anouk de Wit

Sijs, Nicoline van der:

- Begeleiding stagiairs Martijn Knapen (Digitale Pallas), Merel Balt, Janne Mans (etymologiebank), Irisz Szigali (etymologiebank, vrijwilliger), Els Faasse (vrijwilliger), Yuliia Kroviakova (Pallas, GLAD), Olga Rodzina (Pallas, GLAD), Laura Broersen (etymologiebank)

Tempelaars, Rob:

- Begeleiding stagiair Anouk de Wit voor het ANW (lexicografie)

De Tier, Veronique:

- Begeleiding stagiair Pieter-Jan Ranson (dialectologie)

Verheij, Boukje:

- Begeleiding stagiair Anouk de Wit (vertalen webteksten Nederlands-Engels)

Waszink, Vivien:

- Begeleiding stagiair Charlotte Waszink (maatschappelijke stage)
- Begeleiding stagiair Anouk de Wit voor het WNW

Bijlage 4: Taalmaterialen

Nieuwe producten

In het afgelopen jaar hebben we de volgende producten toegevoegd aan de catalogus.

1. **Corpus Vlaamse Gebarentaal (Corpus VGT)**. Het [Corpus VGT](#) is een verzameling van video's met Vlaamse Gebarentaal. Informanten (sprekers VGT) kregen per twee een reeks thema's om over te praten: vertellen van een verhaal, maken van afspraken, discussiëren over een thema, vertellen over hun schooltijd enz. Bij het selecteren van de informanten werd rekening gehouden met leeftijd, regio en gender. Deze conversaties werden op video vastgelegd en gemonteerd per opdracht. De gesprekken in Vlaamse Gebarentaal werden vervolgens geannoteerd door ze te glosseren. Omdat het corpus nog in bewerking is bieden we alleen de video's aan die volledig zijn geannoteerd.
2. **Belgian Covid Sign Language Corpus (BeCos Corpus)**. Het Belgische Federale COVID-19-corpus, genaamd het [BeCoS \(Belgian Covid Sign Language\) Corpus](#), bestaat uit het volledige archief van officiële persconferenties van de Belgische federale overheid betreffende de COVID-19-pandemie. De sprekers spreken meestal Nederlands of Frans en een enkele keer Duits, en alle spraak wordt getolkt door een dove gebarentaaltolk die live tolkt wat er wordt gezegd.
3. **Annotated Corpus for Term Extraction Research (ACTER)**. [ACTER](#) is een handmatig geannoteerde dataset voor termextractie, die drie talen omvat (Engels, Frans en Nederlands), en vier domeinen (corruptie, dressuur, hartfalen en windenergie).
4. **BLISS Dialogue Summaries**. Deze [dataset](#) bestaat uit Nederlandse spraakopnames van deelnemers die spreken met het BLISS-dialoogsysteem (v1) over alledaagse bezigheden en activiteiten waar ze plezier aan beleven. De dataset bevat 55 opnames met een gemiddelde duur van 2 minuten en 34 seconden.
5. **Couranten Corpus (Online)**. Het [Couranten Corpus](#) bevat dertien zeventiende-eeuwse Nederlandse kranten uit de periode 1619-1700 die momenteel op Delpher beschikbaar zijn. Er zijn 109.532 artikelen in te vinden met een totaal van 18.926.425 woorden. Het corpus is doorzoekbaar via een webapplicatie.
6. **AI-Trainingset - Tag de Tekst voor Named Entity Recognition (NER)**. De [AI-Trainingset](#) voor NER is in 2020 gemaakt door de circa 150 vrijwilligers van crowdsourcingsproject 'Tag de tekst' op VeleHanden.nl. Persoonsnamen, locaties en tijdsaanduidingen zijn geannoteerd in al eerder ontwikkelde Ground Truth (GT)-transcripties van 10.567 scans en gecontroleerd door drie ervaren superusers. De Nederlandstalige teksten komen uit de 17e eeuw tot en met de 19e eeuw. Het gaat om notariële teksten uit Amsterdam, Haarlem en uit zeven andere provincies en archieven van de Verenigde Oost-Indische Compagnie (VOC). Ze zijn afkomstig uit het Stadsarchief Amsterdam, het Nationaal Archief, het Noord-Hollands Archief, en zeven andere Regionaal Historische Centra: Tresoar, het Gelders Archief, de Groningen Archieven, het Brabants Historisch Informatie Centrum, het Zeeuws Archief, het Historisch Centrum Limburg, Het Utrechts Archief en de Collectie Overijssel. De AI-trainingset is ontwikkeld i.h.k.v. de projecten 'De IJsberg zichtbaar maken'

(zoekintranscripties.nl) en 'Slimmer zoeken in archieven' (archieveninbeeld.nl).

7. **Pinkhof Geneeskundig Woordenboek (Online)**. De papieren versie van het *Pinkhof Geneeskundig Woordenboek* werd voor het eerst uitgegeven in 1923. Het beroemde naslagwerk was oorspronkelijk samengesteld door de Amsterdamse arts Herman Pinkhof en is door de jaren heen vele malen herzien en herdrukt. Deze [online versie](#) van de medische woordenlijst bevat ruim 52000 termen met hun betekenissen en/of verwijzingen en is gratis toegankelijk voor iedereen.
8. **Corpus Juridisch Nederlands (Online)**. Het [Corpus Juridisch Nederlands](#) omvat een verzameling van 5.856 wetsteksten uit de periode 1814 tot 1989, die per jaar zijn samengevoegd. Aanvankelijk maakte dit corpus deel uit van het *38 miljoenwoordencorpus* en later van het Corpus Hedendaags Nederlands.
9. **GiGaNT-Molex (update)**. Het [GiGaNT-Molex-lexicon](#) bevat Nederlands materiaal uit Nederland, Vlaanderen, de Antillen en Suriname afkomstig uit hedendaags corpusmateriaal van het Instituut voor de Nederlandse Taal (INT). Alle lemmata en paradigmata zijn handmatig nagekeken en conform de officiële spelling. Versie 2.0 bevat nu 223153 lemmata en 751448 woordvormen met afbrekingen.
10. **Corpus Middelnederlands (Online)**. Het [Corpus Middelnederlands](#) is een verzameling van ruim 350 Middelnederlandse (literaire) rijm- en prozateksten uit de periode 1300-1550. Het corpus is doorzoekbaar via een webapplicatie.
11. **Corpus Gysseling (Online)**. Het [Corpus Gysseling](#) dat hier ter beschikking wordt gesteld is de verzameling van alle dertiende-eeuwse teksten die als bronnenmateriaal hebben gediend voor het Vroegmiddelnederlands Woordenboek. Het is de digitale uitgave, verrijkt met woordsoort en lemma, van het dertiende-eeuwse materiaal uit het Corpus van Middelnederlandse teksten (tot en met het jaar 1300), uitgegeven in de periode 1977-1987 door de Gentse taalkundige Maurits Gysseling.

Overzicht downloads commercieel

In totaal is er vijfmaal een product afgenomen voor commerciële toepassingen.

CGN	2
Jasmin	1
BasiLex Lexicon	1
Duoman	1

Overzicht downloads niet-commercieel

Er zijn in totaal 522 producten gedownload voor niet-commercieel gebruik.

Corpus Gesproken Nederlands (CGN)	83
SoNaR-corpus	41
GiGaNT-Molex	30
Dutch Parallel Corpus (DPC)	23
Referentiebestand Nederlands (RBN)	22
CGN-annotaties	21
CHN N-grams	20
INT Historische Woordenlijst	16
DuOMAn Subjectivity Lexicon	16
Frequentielijsten corpora	16
SoNaR Nieuwe Media Corpus	15
Lassy Klein-corpus	14
Cd-rom Middelnederlands	13
Corpus Middelnederlands (Data)	13
e-Lex	11
Corpus Gysseling (Data)	11
BLISS Dialogue Summaries	10
Corpus Pathologische en Normale Spraak (COPAS)	9
Hoger Onderwijs Terminologie in Nederland en Vlaanderen (HOTNeV)	9
AI-Trainingset - Tag de Tekst voor Named Entity Recognition (NER)	8
RND Woordenlijsten	8
PAROLE-lexicon	7
Medische Termen Belgisch-Nederlands (MedTermBN)	6
Jasmin	6
Eindhoven-corpus	6
Wablieft-corpus	6

WAI-NOT Corpus	6
IFA-corpora	6
INT IMPACT NE-lexicon	5
Corpus Ondertitelde UVN-Colleges (COUC)	5
Dutch Idiom Database: Native Speakers (DID-NS)	5
Medische Pilot (MedPilot)	4
Afrikaans Custom Dictionary for Government Domain	4
DuELME	4
SumNL-samenvattingencorpus	4
Referentiebestand Belgisch-Nederlands (RBBN)	3
DAESO-corpora: parallele Nederlandstalige monolinguale treebank	3
COREA-coreferentiecorpus	3
Brieven als Buit - Gouden Standaard	3
NAMES Corpus	3
OMBI Nederlands-Deens	3
BasiLex Lexicon	2
Afrikaans Genre Classification Corpus	2
Paco-MT Parallele Corpora	2
Moroccorp	2
BLISS Spoken Dialogue Dataset	2
BasiScript Corpus	1
BasiScript Lexicon	1
Chorec	1
IFA Dialoog-videocorpus	1
D-TUNA-corpora	1
Corpus Vlaamse Gebarentaal (Corpus VGT)	1
CombiLex	1
Belgian Covid Sign Language Corpus (BeCoS Corpus)	1
AUTONOMATA-namencorpus	1
Annotated Corpora for Term Extraction Research (ACTER)	1
Xitsonga Genre Classification Corpus	1

Bijlage 5: besteding lumpsum basistakenpakket

Onderbouwing lumpsum 2022

		# fte		Kosten
1	Hedendaags Nederlands incl. spelling	6	€	482.237
2	Historisch Nederlands	4	€	321.491
3	Nederlandse Dialecten	1,2	€	96.447
4	Terminologie	1,2	€	96.447
5	Grammatica	2,1	€	168.783
6	Taalmaterialen incl. Clarin	2,5	€	200.932
7	systeembeheer + IT-diensten	5,3	€	425.976
8	Communicatie	1,9	€	152.708
A	Algemene dienst, secretariaat, financiële dienst, directie	3,8	€	305.416
B	Algemene kosten (Huisvesting, afschrijvingslasten, accountant- & advieskosten, kantoorkosten)	n.v.t.	€	352.966
				<hr/>
	Lumpsum 2022 (28 fte / 35 medewerkers)	28	€	2.603.403