

/instituut voor de  
Nederlandse taal/

# jaarverslag

2024

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

*(Woordenboek der Nederlandsche Taal)*

Jaarverslag 2024  
Instituut voor de Nederlandse Taal

## Inhoudsopgave

1. Inleiding: het jaar 2024 .....	6
2. Language Resources Repository en CLARIN-centrum.....	9
Language Resources Repository .....	9
Nieuwe producten.....	9
Updates van bestaande producten.....	10
Geografische verspreiding van gebruikers .....	10
Het INT als CLARIN-centrum .....	11
3. Corpusinfrastructuur .....	11
Monitorcorpus.....	11
Historische corpora.....	11
Verrijking .....	12
Informatie-extractie.....	13
Datamodel voor de centrale kennisbank van de woordenschat.....	14
Het Centrale Lexicon (GiGaNT).....	14
Betekenisregister .....	14
Lexicografische eindproducten, API's en datasets.....	15
Woordenlijst.org .....	15
Algemeen Nederlands Woordenboek (ANW) .....	15
Woordenboek van Nieuwe Woorden (WNW).....	15
Woordcombinaties.....	16
Historische woordenboeken .....	16
Vertaalwoordenschat .....	17
Woordpeiler .....	17
API's en datasets.....	18
5. Beschrijving van de Nederlandse dialecten .....	18
Elektronische Woordenbank van de Nederlandse dialecten (eWND) .....	18
Database van de Zuidelijk-Nederlandse Dialecten (DSDD).....	18
Database van de Nederlandse Dialecten .....	18
Digitale infrastructuur voor het Bildts en het Overijssels.....	18
Overig .....	19
6. Expertisecentrum voor Nederlandstalige Terminologie .....	19
Veldondersteuning.....	19

Terminologisch netwerk.....	21
7. Grammatica .....	21
Grammaticaportaal.....	21
Taaladvies.....	22
e-ANS.....	22
Taalportaal.....	22
Stage DAGENTA / e-ANS.....	22
8. Nationale en internationale samenwerkingsverbanden.....	23
Netwerken .....	23
IMPACT Centre of Competence .....	23
European Language Data Space.....	23
Elexis Association.....	23
Netwerkprojecten .....	24
Universality, diversity and idiosyncrasy in language technology (UniDive, 2022-2026) .....	24
European Network on Lexical Innovation (ENEOLI, 2023-2027) .....	24
Onderzoeks- en infrastructuurprojecten .....	24
CLARIAH-Vlaanderen (2021-2024) .....	24
CLARIAH Plus-Nederland (2019-medio 2024) .....	24
SSHOC-NL (Social Science and Humanities Open Cloud for the Netherlands; 2024-2028) .....	25
SignON (2021-2024).....	25
SABeD (2021-2024) .....	26
Gesproken Corpus van de Zuidelijk-Nederlandse Dialecten (2020-2024).....	26
Geparset Corpus van de Zuidelijk-Nederlandse Dialecten (2024-2028).....	26
Pilotproject Duidelijke Taal (2023-2024).....	26
Spread the News (2020-2025).....	27
Using CoBaLT and GaLAHaD for historical corpus annotation (2024) .....	27
Buitenlandse Aardrijkskundige Namen – fase 1 .....	27
Namenloket (1-03-2025 – 28-02-2027).....	27
Macroscope Large-Scale Research Infrastructure (National Roadmap consortia 2024).....	28
Overige infrastructurele dienstverlening.....	28
Infrastructuur Papiamento/u.....	28
Etymologiebank.....	29
GLAD .....	29
DAGENTA.....	29
Pallas.....	29
9.Systeembeheer en ICT-capaciteit.....	29

10. Disseminatie.....	30
Doelgroepenbeleid.....	30
Onderzoekers en taalkundigen .....	30
Onderwijs .....	30
Algemeen publiek .....	30
Communicatiemiddelen .....	31
Website .....	31
Huisstijl.....	33
Nieuwsbrieven & persberichten .....	33
Sociale media .....	34
Podcasts.....	34
Andere populairwetenschappelijke activiteiten.....	35
Bijlage 1: Raad van Toezicht en Raad van Advies.....	36
Bijlage 2: Medewerkers.....	36
Bijlage 3: publicaties, lezingen, media, prijzen etc.....	38
Publicaties.....	38
Lezingen en presentaties .....	43
Organisatie van congressen en workshops.....	48
Onderwijs .....	48
In de media.....	49
Diversen.....	50
Bijlage 4: Taalmaterialen.....	52
Downloads van taalmaterialen .....	52

# 1. Inleiding: het jaar 2024

Het eerste meerjarenbeleidsplan (2018-2022) werd succesvol uitgevoerd; aansluitend werd er een nieuw meerjarenplan geschreven dat in 2023 van start ging en ook verder in 2024 werd uitgevoerd. Dit in nauw overleg met de Nederlandse Taalunie (NTU).

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 eerder gemelde zorgen van de visitatiecommissie blijven bestaan: zij stelden dat de prestaties van het INT indrukwekkend zijn, maar zij maken zich zorgen 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, bevoegdheid 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.

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 wetenschappelijk onderbouwde lexicografische producten blijven aanbieden.

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 opensource 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 2024 werkten er 34 personeelsleden (28,15 fte). Eind december 2024 waren dat 32 personeelsleden (26,2 fte). Er waren enkele verschuivingen door pensionering van drie werknemers en er kwam één nieuwe medewerker bij, met name een computationeel taalkundige (65%). Verder werden er twee vacatures uitgeschreven: een taaldata-engineer (80%) en een taalkundig assistent (100%) voegen zich bij het team en starten op 1 januari 2025.

In de raad van advies beëindigden vier leden hun mandaat, met name Dirk Geeraerts, Reinhild Vandekerkhove, Veronique Hoste en Wim Vanseveren. Zij werden vervangen door Freek Van de Velde, Dirk Pijpops, Els Lefever en Willem Bongers-Dek.

Een nieuwe algemeen secretaris trad aan bij de Taalunie; Gunther van Neste is sinds 1 juli 2024 in dienst en hij bouwde meteen een goede band op met het INT.

Het INT is hét kennisinstituut voor het Nederlands en neemt een centrale positie in voor het hele Nederlandse taalgebied 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 opensource 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 gebruik maken van deze infrastructuur.

Er stonden in 2024 een heleboel activiteiten op het programma. Er werd vooral ingezet op het verbeteren van de digitale infrastructuur van het Nederlands, met de koppeling van het hedendaagse met het historische lexicon. Er werd een nieuwe softwaretool ontwikkeld: TermWerk, waarbij zowel termextractie als termbeheer in één applicatie mogelijk zijn. De tool werd voorgesteld op de TiNT-dagen in Breda op 8 november.

Verder werden we meer en meer betrokken bij het project GPT-NL van TNO en Surf. We leveren data en adviezen. De ontwikkeling van een groot taalmodel voor het Nederlands is een prioriteit.

Het thema van het jaar was “de digitale taalinfrastructuur”, wat resulteerde in een aantal activiteiten en een boek van de hand van Frieda Steurs: *Taal in Transformatie. Technologie, economie en de kracht van de generatieve AI*. Dit boek werd op 26 november gepresenteerd in Den Haag en de algemeen secretaris van de Taalunie ontving het eerst exemplaar.

Op Europees vlak werkten we verder binnen de European Language Data Space, waar wij de verantwoordelijkheid voor Nederland nemen. Een LRD-workshop voor Nederland en België werd voorbereid; deze zal in januari 2025 plaatsvinden.

Internationaal werden de banden met het Caribisch gebied aangehaald in de vorm van een samenwerking met het pas opgerichte Nationale Taalinstituut in Curaçao. We bereidden een intensieve samenwerking voor om te helpen het Papiamentu/u te beschrijven en te documenteren zoals we dat ook voor het Nederlands doen. Het officiële samenwerkingsakkoord is in 2024 ondertekend.

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 2025

prof. dr. Frieda Steurs

wetenschappelijk directeur-bestuurder



## 2. Language Resources Repository en CLARIN-centrum

### Language Resources Repository

Via de website [taalmaterialen.ivdnt.org](https://taalmaterialen.ivdnt.org) stelt het Instituut voor de Nederlandse Taal (INT) bronnen, data en tools beschikbaar voor taalkundig onderzoek en taal- en spraaktechnologie (TST), hoofdzakelijk gericht op het Nederlandse taalgebied. De Taalmaterialenwebsite kreeg 26.953 pageviews van 11.936 verschillende gebruikers.

Een aantal van die tools en resources zijn ook beschikbaar vanuit het INT CLARIN portal, op <https://portal.clarin.ivdnt.org>. Deze site kreeg 5142 pageviews van 2569 verschillende bezoekers. Door het gebruik van een nieuwe methode om deze pageviews en bezoekers te meten, is dit niet helemaal te vergelijken met de 3900 websitebezoeken van 3000 bezoekers in 2023.

Het totaal aantal downloads voor zowel commerciële als niet-commerciële producten is gestegen in vergelijking met de voorgaande jaren:

- Niet-commerciële downloads: 1091 (816 in 2023, 522 in 2022)
- Commerciële downloads: 44, (20 in 2023, 5 in 2022)

### Nieuwe producten

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

- "**Menselijke evaluatie van geautomatiseerde tekstvereenvoudiging: resultaten van crowdsourcing**" is samengesteld in het kader van het project Duidelijke Taal. De dataset bestaat uit zinnen uit het SoNaR-corpus, een door GPT-4 vereenvoudigde versie daarvan en de menselijke beoordelingen van die vereenvoudigingen met betrekking tot eenvoud, accuraatheid en vlotheid.
- De **Dataset Synthetische Simplificatie** werd binnen het project Duidelijke Taal samengesteld en is gebaseerd op de WR-P-E-I-component (websites) van het SoNaR-corpus. De dataset bestaat uit drie delen: 6.986 zinnen uit het SoNaR-corpus, een synthetische vereenvoudiging van de SoNaR-zinnen die gemaakt werd door GPT-4 en zinsparen bestaande uit telkens een SoNaR-zin en de vereenvoudigde versie daarvan.
- De termenbank **4-Language Finance, Economy & Business Terminology** — NL-EN-FR-DE (version 2.0) bevat begrippen, afkortingen en namen van instanties uit de financieel-economische wereld. Ze is viertalig met vertaalequivalenten in het Nederlands, Engels, Frans en Duits, en bestaat uit ruim 16.000 records.
- Het **Woordenboek van Nieuwe Woorden (WNW)** is een online woordenboek waarin woorden die vanaf het jaar 2000 zijn ontstaan, worden beschreven. Het WNW neemt niet alleen de beklivende nieuwe woorden op (app, selfie, keuzestress, ontspullen), maar juist ook veel nieuwe woorden die een kortere tijd bestaan.
- Het **Spoken Academic Belgian Dutch Corpus (SABeD)** bestaat uit 200 colleges gegeven op Vlaamse hogescholen en universiteiten. De eerste 25 en de laatste 5 minuten van elk college

werden getranscribeerd met behulp van een ASR-systeem dat is afgestemd op Belgisch-Nederlands. Vervolgens werd de spraak manueel opgedeeld in afzonderlijke uitingen (utterance segmentation) en werden de transcripties manueel geverifieerd. De resulterende teksten werden geanalyseerd door de Frog language analyser.

- **GoSt-ParC-Sign** is een multimodaal corpus van VGT met een vertaling in geschreven Nederlands als doeltaal. Alle VGT-materiaal in dit corpus bestaat uit reeds bestaande video's die werden gemaakt door authentieke VGT-gebaarders voor een VGT-publiek.
- **Hotel Review Corpus in Vlaamse Gebarentaal.** Een multimodaal parallel corpus met de talen Nederlands en Vlaamse Gebarentaal (VGT). 297 geschreven hotelbeoordelingen werden vertaald uit het Nederlands in VGT door 6 professionele, dove vertalers. Het resulterende corpus bevat 21.825 woorden in het Nederlands en ongeveer 4 uur aan VGT-videomateriaal.
- **Hotel Review Corpus in Nederlandse Gebarentaal.** Een multimodaal parallel corpus met de talen Nederlands en Nederlandse Gebarentaal (NGT). 297 geschreven hotelbeoordelingen werden vertaald uit het Nederlands in NGT door 6 professionele, dove vertalers. Het resulterende corpus bevat 22.197 woorden in het Nederlands en zo'n 3 uur en 50 minuten aan NGT-videomateriaal.
- **Hotel Review Corpus in Spanish Sign Language.** Een multimodaal parallel corpus met de talen Spaans en Spaanse Gebarentaal (Lengua de Signos Española – LSE). 283 geschreven hotelbeoordelingen, oorspronkelijk in het Nederlands werden vertaald in het Spaans en vervolgens door 6 professionele, dove vertalers in het LSE.
- Het corpus **Gekaapte brieven** bestaat uit de transcripties van 5862 brieven en andere documenten, zoals rekeningen, geschreven aan en van zeelieden en anderen uit het buitenland in de 17e en 18e eeuw. Deze brieven en documenten waren aanwezig op Nederlandse schepen die door de Engelsen waren gekeerd tijdens een van de vier oorlogen die in deze periode werden uitgevochten tussen Groot-Brittannië en de Republiek der Zeven Verenigde Nederlanden. Deze brieven kwamen terecht in de archieven van het High Court of Admiralty (nu onderdeel van The National Archives at Kew), waar de Nederlandse historicus S. Braunius ze in 1980 ontdekte.

## Updates van bestaande producten

- Het **Lassy Groot-corpus Commercieel 7.0** is een corpus van ongeveer 476 miljoen woorden met automatisch gegenereerde syntactische annotaties.

## Geografische verspreiding van gebruikers

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 onze taalmaterialen. De meesten komen uit Nederland (166), daarna België (83) en verder Duitsland (14), en de Verenigde Staten (.edu) (9), en lager.

Een volledig overzicht van downloads per product is in bijlage 4 te vinden.

## Het INT als CLARIN-centrum

Voor 2024 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 (NCF). Vincent is ook lid van het Knowledge Infrastructure Committee van CLARIN-ERIC. Daarnaast nam Jesse de Does vanuit het INT de vertegenwoordiging van België op zich 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. Vincent was Programme Committee Chair voor de CLARIN-conferentie in Barcelona. K-Dutch, het CLARIN Knowledge Centre for Dutch, werd verder uitgebreid, waarbij de website verder werd uitgebreid en verfijnd, en, in samenwerking met de Taalunie, werd er een Nederlandstalige variant van gemaakt. K-Dutch kreeg 1794 unieke bezoekers in 2024.

## 3. Corpusinfrastructuur

Zorgvuldig samengestelde en wetenschappelijk onderbouwde corpora vormen een essentieel onderdeel van de taalinfrastructuur voor het Nederlands. Op basis van deze primaire taaldata kan de Nederlandse taal worden gedocumenteerd en kunnen taalapplicaties ontwikkeld worden. De corpusinfrastructuur van het INT bevat naast corpora ook gereedschappen voor dataprocessing en ontsluiting. Een belangrijk deel van de werkzaamheden aan de corpusinfrastructuur worden uitgevoerd ten behoeve van de verdere uitbouw van de centrale kennisbank voor de Nederlandse woordenschat, maar resulteert ook in corpusinfrastructuur voor de brede onderzoeksgemeenschap. Daarnaast is het INT betrokken in diverse projecten waarin corpora worden gebouwd, waarbij het INT, naast expertise, ook infrastructurele ondersteuning biedt voor het bouwen, het gebruik dan wel het ter beschikking stellen van het corpusmateriaal.

### Monitorcorpus

Het Corpus Hedendaags Nederlands (CHN) bevat al het hedendaags Nederlandse corpusmateriaal waarover het INT beschikt. Het corpus is in 2024 continu uitgebreid met krantenmateriaal uit Nederland, Vlaanderen, Suriname en de Caribische rijkdelen. Voor de interne versie gebeurt deze update wekelijks, voor de externe versie elke maand. Het interne corpus bevat momenteel ruim 10 miljoen documenten en bijna 4,25 miljard woorden. Afgezien van de reguliere updates zijn voorbereidingen getroffen voor het uitbreiden van de huidige corpusprocessingomgeving met syntactische annotatie. Komend jaar zal, mede dankzij de aanwerving van een taaldata-engineer eind 2024, gewerkt worden aan een verbetering van de huidige corpusprocessingworkflow. Het krantencorpus is de basis voor de *Woordpeiler* waarmee trends in de woordenschat kunnen worden vastgesteld en de frequentie van woorden door de jaren gevisualiseerd kunnen worden (cf. infra).

### Historische corpora

In 2024 is verder gewerkt aan de uitbreiding van het krantenmateriaal van het *Couranten Corpus*. Het afgelopen jaar heeft een groep van ongeveer 25 enthousiaste vrijwilligers onder leiding van

Nicoline van der Sijs en met ondersteuning van het INT ervoor gezorgd dat de resterende 60% van de meer dan 5200 pagina's van de *Amsterdamse Courant* die in 2023 automatisch zijn verwerkt met Transkribus, gecontroleerd is en van metadata voorzien. Daarna zijn deze vrijwilligers hetzelfde gaan doen voor de *Oprechte Haerlemsche Courant*. Van de ruim 5400 pagina's van die krant is inmiddels ruim 20% voltooid. De data moeten uiteindelijk onderdeel gaan vormen van een groot diachroon corpus van kranten. Daarnaast is er verder gewerkt aan de uitbreiding van het trainings- en evaluatiemateriaal voor de verrijking met woordsoort en lemma (cf. infra). Er zijn voorbereidingen getroffen voor een update van het huidige online *Couranten Corpus* en het online brengen van de eerste release van het *Historical Corpus of Dutch*, een gebalanceerd corpus dat ontwikkeld is door onderzoekers van de Vrije Universiteit Brussel en de Universiteit Leiden. Tot slot is het materiaal van het diachroon bijbelcorpus *EDGEs* verder voorbereid om te dienen als use case bij het uitbreiden van BlackLab ten behoeve van het doorzoeken van parallelle corpora (cf. infra).

## Verrijking

De INT-corpora worden samengesteld uit bestaand digitaal materiaal of, waar nodig, door digitalisering. De brondata worden geconverteerd naar eenzelfde XML-standaard (TEI) en zorgvuldig van metadata voorzien en daarna automatisch taalkundig verrijkt. Metadata en taalkundige verrijking bieden een nadrukkelijke meerwaarde om zinvolle informatie uit de corpora te kunnen extraheren. De werkzaamheden van het afgelopen beleidsjaar worden hieronder gespecificeerd.

### Verrijking met woordsoort en lemma

Voor de verrijking van historisch Nederlands is verder gewerkt aan de ontwikkeling van gold standard datasets voor de verrijking met woordsoort en lemma. Inmiddels zijn ruim 500.000 tokens, gespreid over de 13e tot 19e eeuw, volledig nagekeken. Ervaring met het werken aan de datasets heeft geleid tot het publiceren van versie 2.0 van de *TDN* met aangescherpte beschrijvingen van de richtlijnen, meer voorbeelden en aanpassingen aan het voornaamwoordenoverzicht. Daarnaast is een uitgebreide en herziene versie van de *lemmatiseerprincipes* verschenen.

### Syntactische annotatie

Op grond van reeds in 2023 verricht voorwerk is besloten de *Universal Dependencies*-richtlijnen aan te houden voor syntactische annotaties, met enerzijds een aantal extensies die in 2025 geïmplementeerd worden bij de omzetting van het trainingsmateriaal voor syntactische verrijking van het Lassy-formaat naar *Universal Dependencies* (mede in het kader van het GCND+-project, zie aldaar). Een aantal opties (*spaCy*, *Trankit*, *Stanza*, *UDPipe*) is geëvalueerd en getoetst op technische haalbaarheid door middel van integratie in het GaLAHaD-platform en implementatie binnen het TermWerk-project, waarbij al een verrijkingspipeline in de workflow geïmplementeerd is en een pilotreferentiecorpus van 500 miljoen woorden verrijkt is. Er is vastgesteld dat met name *Stanza* en *UDPipe* elkaar niet veel ontlopen, maar dat de lemmatiseercomponenten voor verbetering vatbaar zijn, met name door gebruik van het *Molex*-lexicon. Deze aanpassing gebeurt in 2025 in het kader van SSHOC-NL.

### Metadata

De corpora hebben een gemeenschappelijk metadataformaat, met ruimte voor subcorpus-specifieke metadata. Ten behoeve van verdere uniformering en systematisering is gewerkt aan een CMDI-

tekstprofiel dat toepasbaar is voor alle corpusdata van het INT. Het profiel voor corpora van geschreven taal is in 2024 vastgesteld en de implementatie ervan zal in 2025 worden gestart, beginnend met het *Corpus Hedendaags Nederlands*. Er ligt ook een eerste versie van het CMDI-tekstprofiel voor corpora van gesproken taal.

## Informatie-extractie

Het corpusmateriaal wordt toegankelijk gemaakt via een applicatie waarmee in de corpora gezocht kan worden. Wanneer de IPR (Intellectual Property Rights) het toelaten, wordt het corpusmateriaal ook als dataset beschikbaar gesteld in de language resource repository. De software is opensource beschikbaar.

Voor de ontwikkeling van de corpusapplicatie die bestaat uit de corpuszoekmachine BlackLab en de corpus frontend, ligt de prioritering bij de ondersteuning van de diverse INT-taken. Deze werkzaamheden worden ondersteund door het samenwerken met diverse partijen die de software gebruiken, en door de mogelijkheden die externe projecten bieden om deze software verder te ontwikkelen. Voor wat betreft de backend van de corpusretrievalomgeving (BlackLab, BlackLab Server) is gewerkt aan:

- Ondersteuning van steeds grotere corpora door middel van optimalisaties en gedistribueerd zoeken. Dit is de voortzetting van werkzaamheden die in 2022 zijn opgestart.
- Ondersteuning van zoeken in parallele corpora,
- Ondersteunen van het zoeken op metadata van tekstfragmenten

Daarnaast is de BlackLab Frequency Tool in productie genomen, waardoor bij systematische updates van het CHN de frequentie-informatie opgeslagen kan worden in een daarvoor ingerichte PostgreSQL-database met TimescaleDB-plugin. Deze database is onder andere de basis voor de in januari 2025 gereleaste *Woordpeiler*.

Voor wat betreft de corpus user interface is er gewerkt aan het kunnen ondersteunen van zoeken in parallele corpora. Daarnaast is het nu technisch mogelijk om de user interface meertalig te maken.

In het kader van het GCND-project (cf. infra), is gewerkt aan een versie van GrETEL die toegang biedt tot met Alpino geparseerd dialectmateriaal.

## 4. Beschrijving van de woordenschat door de eeuwen heen

### Datamodel voor de centrale kennisbank van de woordenschat

Zoals uiteengezet in het Meerjarenbeleidsplan 2023-2027 is de koppeling van verschillende lexicografische databanken op zowel vorm- als betekenisniveau de belangrijkste innovatie en de grootste uitdaging voor de woordenschatbeschrijving in huidige beleidsperiode. Voor het uitwerken van de complexe data-infrastructuur voor de centrale kennisbank is in 2024 verder gewerkt aan de uitbreiding van het datamodel en het opzetten van een betekenisregister.

Voor de uitbreiding van het datamodel werd voortgebouwd op de al bestaande onderdelen in de huidige infrastructuur, met name het centrale lexicon GiGaNT (cf. infra), de koppeling op lemmaniveau met de lexicografische databanken en het Diachroon seMantisch lexicon van de Nederlandse Taal (DiaMaNT). Er werd in kaart gebracht welke bijkomende datacategorieën nodig zijn, welke koppelingen binnen de kennisbank wenselijk zijn en welke daarbuiten (bv. met corpusdata). In 2024 zijn de datamodellen van GiGaNT en van DiaMaNT geëvalueerd en aangepast. De implementatie van die aanpassingen zal plaatsvinden in 2025.

In wat volgt beschrijven we de activiteiten die in 2024 voor de verschillende onderdelen van de centrale kennisbank en de eraan gekoppelde lexicografische producten uitgevoerd werden.

### Het Centrale Lexicon (GiGaNT)

In 2024 zijn in eerste instantie de reguliere werkzaamheden aan het centrale lexicon voortgezet. Deze houden in: het onderhoud aan zowel de historische component (GiGaNT-Hilex) als de moderne component (GiGaNT-Molex), het verder werken aan de integratie van deze componenten en de uitbreiding van GiGaNT-Molex ten behoeve van de diverse lexicografische eindproducten die een koppeling met GiGaNT hebben (cf. infra). Afgezien van de toevoeging van nieuwe lexemen aan GiGaNT Molex, is ook verder gewerkt aan het toevoegen van uitspraakinformatie, met aandacht voor de uitspraak in zowel Nederland als Vlaanderen. Ruim 124.000 lexemen in Molex zijn voorzien van handmatig geverifieerde uitspraakinformatie, goed voor ruim 162.000 verschillende uitspraken. Tot slot is een revisie gedaan van het datamodel van GiGaNT om de verdere integratie van de beide lexiconcomponenten te kunnen realiseren.

### Betekenisregister

In 2024 werd de proefversie van het betekenisregister verder uitgewerkt: Op basis van een eerste automatische datalinking werd voor een selectie lemmata een manuele koppeling gemaakt tussen het betekenisregister enerzijds en het ANW, Woordcombinaties en RBN anderzijds. Deze test liet toe om het datamodel te verfijnen maar bracht ook aan het licht dat de functies van de databankeditieromgeving duidelijk uitgebreid moeten worden. Daarnaast werden op basis van ANW en WNW ook datasets aangemaakt die als “ground truth” fungeren voor experimenten met Large Language Models (LLMs) voor het automatisch koppelen op betekenisniveau. Voor het

finetunen van LLMs met het oog op het automatische genereren van kandidaat-definities voor neologismen werd samengewerkt met de masteropleiding Artificial Intelligence van de KU Leuven.

## Lexicografische eindproducten, API's en datasets

### Woordenlijst.org

Ook dit jaar is er weer gewerkt aan de uitbreiding van woordenlijst.org. Er is een technische update uitgevoerd in mei, een data-update met nieuw materiaal in juni en de update met het materiaal dat bewerkt is in de tweede helft van 2024 is gepland voor begin februari 2025. Dat zal betekenen dat er dan 229.857 lemmata en 848.178 paradigmavormen online zullen staan. Tot slot heeft het INT ook dit jaar weer ondersteuning geboden aan de Commissie Spelling bij haar werkzaamheden.

### Algemeen Nederlands Woordenboek (ANW)

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 ons verzamelde neologismen (nieuwe woorden, nieuwe verbindingen, nieuwe uitdrukkingen, nieuwe betekenissen van al bestaande woorden) opgenomen.

Door de redactie zijn in 2024 654 volledig bewerkte, nieuwe trefwoorden ('artikelen') online gezet. Daarvan kwamen 218 trefwoorden alleen in het ANW terecht en 436 trefwoorden zowel in het ANW als in het Woordenboek van Nieuwe Woorden (WNW). Eén trefwoord kan meerdere gelijkvormige woorden beschrijven. Het gaat dan bv. om woorden waarvan de woordsoort of etymologische herkomst verschilt. In de huidige versie van het ANW zijn in totaal bijna 294.000 woorden opgenomen die behandeld zijn in ruim 87.000 trefwoorden. De trefwoorden bevatten in totaal ruim 46.000 betekenissen.

Alle nieuwe woorden zijn door de 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 en het Groene Boekje.

De redactie is ook begonnen met het selecteren van betekenissen uit Woordcombinaties, het Referentiebestand Nederlands (RBN) en de Vertaalwoordenschat om alvast toe te voegen aan ANW-lemmata die nog geen definitie hebben.

De redactie schrijft mee aan de webrubriek 'Nieuw woord van de week', waarin elke week een actueel neologisme wordt besproken.

### Woordenboek van Nieuwe Woorden (WNW)

Het INT onderzoekt de vorming van nieuwe woorden en houdt die woorden onder andere bij in woordenboeken. 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 2024 is er verder gewerkt aan dit woordenboek. Het WNW bevatte eind 2024 ruim 19.000 woorden. Het WNW wordt dagelijks geüpdatet.

Alle nieuwe woorden zijn door de 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 en het Groene Boekje.

Ook afgelopen jaar is er iedere week een neologisme behandeld in de rubriek 'Nieuw woord van de week'.

## Woordcombinaties

Woordcombinaties is een online taaltool waarin het gebruik van woorden in context wordt beschreven en beschikbaar gesteld aan taalgebruikers en leerders van het Nederlands. Het project integreert de beschrijving van collocaties, idiomen en semantisch gemotiveerde syntactische patronen en behandelt de combinatoriek van werkwoorden en substantieven. Het project is corpusgebaseerd en gebruikt de Sketch Engine als corpusquerytool. 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. In Woordcombinaties worden ook idiomen en conversatieformules geïnventariseerd en systematisch bewerkt.

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 werkwoorden en substantieven. Het datamodel voor de idiomen is verder getest en ontwikkeld en zal in 2025 gebruikt kunnen worden. In de idioomomgeving kunnen nu ook de tot nu toe geregistreerde conversatieformules geïntegreerd worden. Op computationeel gebied is verkennend onderzoek gedaan naar de mogelijkheden van AI voor het ondersteunen van de manuele annotatie en identificatie van patronen. Het project is in 2024 gepresenteerd op verschillende internationale bijeenkomsten, zoals Euralex en Phrasalex, en op een werkvloeroverleg met de Taalunie. Er is overleg geweest met projectgroepen van het Digitaal Wörterbuch der deutschen Sprache (DWDS) en het Leibniz-Institut für Deutsche Sprache (IDS) in Mannheim over de behandeling van meerwoordexpressies.

## Historische woordenboeken

De beschrijving van de historische woordenschat van het Nederlands is te vinden in de verschillende historische woordenboeken van het INT. Deze woordenboeken zijn online beschikbaar in het historische woordenboekenportaal ([gtb.ivdnt.org](http://gtb.ivdnt.org)): 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. Daarnaast is in 2024 het historisch gelegenheidswoordenboekje *Ach en wee* online gekomen vanwege de 200e sterfdag van Rhijnvis Feith.



Op verschillende manieren zijn de Historische Woordenboeken – en andere historische taalmaterialen van het INT – in de schijnwerpers geplaatst. Zo is er voor erfgoedmedewerkers in Vlaanderen een webinar over het gebruik van de historischewoordenboekenapplicatie georganiseerd, dat door deelnemers in een enquête positief werd beoordeeld. Daarnaast zijn tijdens presentaties de woordenrijkdom van deze woordenboeken en de onderzoeksmogelijkheden ervan voor het voetlicht gebracht. Uit de loggegevens én uit de reacties en vragen die we regelmatig zowel van specialistische als 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 een lemma, geattesteerde woordvormen van het lemma in de citaten en de citaten zelf. De werkzaamheden die ook in 2024 in dit kader zijn uitgevoerd, leveren correcties op die voor de online woordenboeken eveneens relevant zijn.

Citatenmateriaal van zowel MNW en WNT zijn gebruikt als trainings- en evaluatiemateriaal voor het verrijken van historisch corpusmateriaal (zie boven).

## Vertaalwoordenschat

De *Vertaalwoordenschat* is een onlineapplicatie voor tweetalige woordenboeken. Via deze onlineapplicatie worden de tweetalige bestanden ontsloten, die 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.

In september 2017 heeft het INT de Vertaalwoordenschat-applicatie gelanceerd. Inmiddels staan het Nederlands-Nieuwgrieks / Nieuwgrieks-Nederlands, het Nederlands-Portugees / Portugees-Nederlands, het Nederlands-Estisch, het Nederlands-Fins / Fins-Nederlands, en het Nederlands-Deens online. In 2024 zijn, in samenwerking met externe redacteurs, de taalparen Nederlands-Arabisch / Arabisch-Nederlands en het Deens-Nederlands toegevoegd aan de Vertaalwoordenschat. De officiële release staat gepland voor februari 2025.

De woordenboeken zijn gratis raadpleegbaar op de website en via de app *Vertaalwoordenschat*.

## Woordpeiler

Om de boven vermelde nieuwe tijdreeksdatabank met systematisch geüpdatete woordfrequenties uit het Corpus Hedendaags Nederlands te ontsluiten, werd de nieuwe applicatie *Woordpeiler* ontwikkeld. Voor externe gebruikers laat *Woordpeiler* toe om voor elk woord te visualiseren hoe het gebruik in de verschillende taalvariëteiten van het Nederlands door de tijd heen geëvolueerd is. De grafieken zijn gebaseerd op het krantenmateriaal van het CHN en onderzoekers met een CLARIN-login kunnen vanuit *Woordpeiler* ook de onderliggende gegevens raadplegen. Daarnaast is er, voorlopig alleen voor intern gebruik, ook een module voor trendanalyses ontwikkeld die toelaat om nieuwe of kenmerkende woorden in kaart te brengen voor een willekeurig tijdsinterval uit de periode 2000 tot nu. Deze module zal in 2025 gebruikt worden zowel voor de eigen woordenschatbeschrijving, o.a. neologismendetectie en het koppelen van frequentiegegevens aan GiGaNT-MoLex, als voor communicatie over trends in de woordenschat naar het brede publiek toe (nieuwe rubriek *Woordpeiling*).

## API's en datasets

Ook in 2024 werd de lexiconservice waarmee GiGANT-Hilex toegankelijk wordt gemaakt voor gebruik in bijvoorbeeld corpusapplicaties als Nederlab, Delpher, het Corpus Middelnederlands of de Gekaapte Brieven maandelijks geüpdatet. In 2025 wordt een nieuwe release van de data van Molex voorzien.

## 5. Beschrijving van de Nederlandse dialecten

### Elektronische Woordenbank van de Nederlandse dialecten (eWND)

In 2024 hielpen vrijwilligers enkele nieuwe woordenboeken klaar te maken voor de elektronische Woordenbank van de Nederlandse Dialecten (eWND).

### Database van de Zuidelijk-Nederlandse Dialecten (DSDD)

In 2024 zijn in de Database van de Zuidelijk-Nederlandse Dialecten (DSDD) 3000 bestaande concepten (10%) nagekeken op onregelmatigheden. Op basis daarvan zijn de nodige correcties aangebracht, of zijn dubbele concepten opgespoord, verwijderd of ontdebeld.

In 2024 zijn nieuwe gegevens uit de Landbouwterminologie uit het Woordenboek van de Vlaamse Dialecten (het paard, aardappeloogst, landbouwvoertuigen, ...) gekoppeld aan bestaande concepten in de DSDD. Het Woordenboek der Zeeuwse dialecten is verder bewerkt voor opname in de DSDD.

Eind 2024 bevatte de DSDD 31.118 concepten (met ongeveer 600.000 trefwoorden).

### Database van de Nederlandse Dialecten

Op termijn zal de Database van de Zuidelijk-Nederlandse Dialecten uitgebreid worden naar het noorden. In dat kader zijn de gegevens van het Woordenboek van de Gelderse dialecten (WGD – in het kader van Clariah+) en het Woordenboek van de Overijsselse dialecten (WOD) bewerkt en in Lex'it geplaatst, zodat ze in 2025 gekoppeld kunnen worden aan de DSDD-conceptenlijst.

### Digitale infrastructuur voor het Bildts en het Overijssels

Het INT zetelt in de werkgroep contactvariëteiten, een werkgroep die kijkt hoe dialectgemeenschappen eigen informatie kunnen samenbrengen en delen. In 2024 hebben de vrijwilligers van het Bildts Aigene het materiaal voor het Bildts toegevoegd en verder bewerkt. Eind 2024 is ook een begin gemaakt aan een vergelijkbare infrastructuur voor het Overijssels. Het doel is op basis van bestaande gekoppelde woordenboeken een zakwoordenboek te maken voor het West-Overijssels. In het kader hiervan werden bestaande Overijsselse amateurdialectwoordenboeken gedigitaliseerd en waren er enkele overleg- en startvergaderingen.

## Overig

In 2024 verscheen tweewekelijks 'Uit de Streek', een rubriek over dialectwoorden uit de DSDD en de woordenbank. De projectleider werkte ook mee aan de tweemaandelijks rubriek over streektalen in het tijdschrift *Onze Taal*. Het INT heeft in 2024 ook Limburgs cursusmateriaal toegevoegd en beschikbaar gesteld via de website [Limburgsluisteren.ivdnt.org](http://Limburgsluisteren.ivdnt.org). op verzoek van Levende Talen Limburgs. Nieuw cursusmateriaal voor twee nieuwe plaatsen is in de website geïntegreerd.

## 6. Expertisecentrum voor Nederlandstalige Terminologie

Het Expertisecentrum Nederlandstalige Terminologie (ENT) richt zich primair op veldondersteuning en op positionering in een terminologisch netwerk. Voor beide kerntaken is in 2024 het volgende gerealiseerd.

### Veldondersteuning

#### Termenlijsten

Termenlijsten zijn een essentieel onderdeel van het lexicon van een taal. Ze dienen concrete doelen waar ze voorzien in de omschrijving en afbakening van begrippen voor de beoefening van een vak, in onderwijsverband bij de instructie van leerstof of waar ze in meertalig verband vertaalsuggesties aanleveren voor beoefenaren van taalberoepen. Op een algemener en abstracter niveau vormen ze ook een graadmeter die uitdrukt in hoeverre een taal beschikt over expressiemiddelen buiten de algemene communicatiebehoeften. De ENT-webpagina voorziet daarom in een rubriek die het mogelijk moet maken om Nederlandstalige eentalige of meertalige termenlijsten binnen een enkele verzamelplaats te raadplegen. De ordening gebeurt volgens de onderwerpcatalogus van de Library of Congress, een internationaal systeem dat ook vele Nederlandse universiteitsbibliotheken hanteren om hun collectie te classificeren. De uitbreiding gebeurt gestaag. In 2024 werden ruim 50 nieuwe termenlijsten opgespoord die in 2025 worden toegevoegd.

#### Fineco

Naast deze collectie van termenlijsten die bestaat uit een groeiende verzameling links, heeft het INT ook termenverzamelingen in eigen beheer. Het hosten van externe bestanden is een succesvolle manier gebleken om vaktaal voor de vele vakgebieden te documenteren. Een aanwinst gerealiseerd in 2024 is het Fineco-bestand (4-Language Finance, Economy & Business Terminology — NL-EN-FR-DE (version 2.0)). Deze omvangrijke termenbank bevat begrippen en namen van instanties uit de financieel-economische wereld. Ze bestaat uit ruim 16.000 records en biedt vertaalequivalenten in het Nederlands, Engels, Frans en Duits. De termencollectie komt deels voort uit de communicatie- en vertaalactiviteiten van het bedrijf van de auteur met betrekking tot meer dan 35.000 documenten op het gebied van financiële berichtgeving. Eveneens zijn deels diverse, vrij beschikbare bronnen gebruikt die door de auteur met de hulp van zijn team over de jaren heen zijn verzameld, aangepast, gewijzigd of aangevuld met vertaalequivalenten. Naast termen, verbindingen en uitdrukkingen, biedt deze databank nog ca. tweeduizend afkortingen, gebruikelijk in financiële

communicatie. Om deze collectie publiek beschikbaar te maken heeft het INT het verwerkingsproces zowel redactioneel als computationeel gestroomlijnd en geïmplementeerd.

### **Nederlands als wetenschapstaal**

Andere extern samengestelde termenbestanden die het INT ter beschikking stelt, komen voort uit het terminologieproject 'Nederlands als wetenschapstaal'. Onderzoek naar onderwijs en woordenschat heeft duidelijk gemaakt dat de opstap naar het eerste bachelorjaar een betere begeleiding wenselijk maakt voor zowel geschreven vaktaal als gesproken academisch Nederlands. Aansluitend bij deze behoefte werd een samenwerkingsverband opgestart tussen de Stichting Nederlands/Vlaams Platform Taalbeleid Hoger Onderwijs, de Taalunie, KU Leuven, UGent en het INT. Het proefproject 'Nederlands als wetenschapstaal - van corpora naar terminologielijsten' heeft als doel om taalhulpmiddelen te ontwikkelen die deze overstap van het middelbaar naar het hoger onderwijs makkelijker maken. Dit project leverde drie grote termenverzamelingen op voor scheikunde, wiskunde en natuurkunde. Ook hier zorgde het INT voor terminologische en computationele begeleiding van de publicatie. Het scheikundebestand staat intussen al online maar in 2024 is een script ontwikkeld om de vele interne verwijzingen tussen ingangen met een semi-automatische aanpak te optimaliseren. De toepassing daarvan begon eind 2024 en loopt door in 2025. Voor het wiskundebestand vond een eerste doorlichting plaats van de TBX-structurering, synoniemenvermeldingen en de interne verwijzingen. Het is momenteel al beschikbaar als pdf maar zal, evenals later het natuurkundebestand, als termenbank worden gepubliceerd naar het voorbeeld van het online scheikundebestand.

### **HOTNeV**

Ten slotte was het ENT in 2024 ook zelf actief in het beschrijven van termen. In het HOTNeV-project (Hoger Onderwijs Terminologie in Nederland en Vlaanderen) worden onderwijstermen geanalyseerd en beschreven volgens de principes zoals die zijn vastgelegd in de internationale terminologie-standaard ISO 704. De resultaten van het onderzoek van de termen, hun vertaal-equivalenten en beschrijving worden opgeslagen in een database. Het onderzoek komt tot stand in het kader van stagebegeleiding en de resultaten worden opgeslagen in een tweetalige termenbank Nederlands en Brits-Engels met de terminologietool QTerm. Een export van de data was ook in 2024 beschikbaar in het INT-websiteluik 'Taalmaterialen'.

### **Terminologierubrieken op het INT/ENT-web**

Opnieuw zijn in 2024 vier uitgebreide nieuwsbrieven verstuurd die vervolgens op de website ter beschikking komen en vervolgens worden gearchiveerd. Ze informeren over terminologie in vaste rubrieken zoals een terugblik naar voorbije evenementen, nieuws over ontwikkelingen en publicaties, tools en bronnen, berichten van NL-Term en een agenda. In de loop van het jaar is ook de ENT-webpagina met de evenementenrubriek over Europese terminologiecongressen, -symposia en -workshops bijgehouden. Veldondersteunend is ook het onderhoud en de uitbouw van het terminologiegedeelte van de INT-website. Zo is in 2024 de rubriek voor terminologische hulpmiddelen aanzienlijk uitgebreid. Nieuwe vermeldingen van applicaties zijn toegevoegd en bestaande links werden geactualiseerd. De indeling in deelrubrieken helpt in combinatie met de informatie van de gelinkte websites om zich een beeld te vormen van een aantal belangrijke soorten terminologiegerelateerde hulpmiddelen en subcategorieën daarbinnen. Achtereenvolgens komen aan de orde: alignering, corpusanalyse, online concordantie, termextractie, terminologiebeheer,

vertaalgeheugens met geïntegreerde terminologiemodule, vertaalprojectbeheer met geïntegreerde terminologiemodule en zoekapplicaties.

### **TermWerk**

Op het vlak van terminologische applicaties heeft het INT in 2024 een nieuwe tool als pilot ontwikkeld onder de naam TermWerk. Het gaat hierbij om de integratie van de vroegere termextractor TermTreffer en de termbank-editor TermBeheerder, zodat één tool ontstaat die geschikt is voor de invoer, de verwerking en het beheer van termmateriaal. Het resultaat is een applicatie die in een breed, generiek verband en afgestemd op het Nederlands en zijn taaleigenschappen, de termextractiemodule en de editeerfunctionaliteiten van beide systemen combineert. Deze is door het INT zelf volledig ontwikkeld, rekening houdend met de vereisten van de gebruikers. Het tweede luik van dit project werd in 2024 afgerond, zodat een volledige testapplicatie beschikbaar zal zijn in 2025. Die zal mee worden getest door studenten en de bestuursleden van NL-TERM. TermWerk zal voor het grote publiek operationeel zijn in 2025.

## Terminologisch netwerk

Om zich te positioneren in het terminologische netwerk nam het ENT in 2024 deel aan verschillende evenementen en samenwerkingsvormen. Zo participeerde het INT aan de driemaandelijke bijeenkomsten van de Termraad Nederlands. Op dit overlegplatform voor terminologische afstemming binnen het Nederlandse taalgebied werken EU-terminologen samen met Belgische en Nederlandse partners uit overheidsdiensten, terminologieverenigingen en vertaalopleidingen. Het doel van de bijeenkomsten is terminologische ondersteuning voor deze instellingen door de verzameling, beschrijving en uniformering van terminologie in specifieke vakgebieden. Ook werkte het INT samen met NL-Term voor het TiNT-congres (Terminologie in het Nederlandse Taalgebied) op 7 en 8 november in Breda. Deze tweedaagse editie had als thema 'Taalbeleid in het hoger onderwijs en de impact op professioneel taalgebruik en Nederlandstalige terminologie.' Verder droeg het INT centraal bij aan de organisatie en editie van het vierde *volume* in de reeks *Handbook of Terminology*, dat in 2025 verschijnt. Het boek biedt per taalgebied een overzicht van de terminologische stand van zaken onder de titel 'Terminology planning in Europe'. Voor het Nederlands leverde het ENT de bijdrage 'Efforts and challenges in translating concept to reality: The Dutch case'. Ten slotte vond stagebegeleiding ditmaal plaats voor een masterstudent van de Universiteit Utrecht.

## 7. Grammatica

### Grammaticaportaal

Van het grammaticaportaal, de hub voor de grammaticaonderdelen van het INT, is dit jaar een testversie online gezet. In deze testversie kan worden gezocht in de ANS, Taaladvies.net en Taalportaal. Ook zijn de verschenen Grammaticasafari's erin na te lezen, en kan de digitale termenbank worden doorzocht. Twee afleveringen van de online rubriek 'Grammaticasafari' zijn door uitgeverij Malmberg aangekocht om te worden gebruikt in het lesboek 'Talent' voor havo/vwo.

## Taaladvies

Aan Taaladvies.net, een zeer druk bezochte site die onder beheer is bij het INT, is afgelopen jaar regulier onderhoud gepleegd. Er is onderling contact tussen de redacties van Taaladvies.net en de ANS om de kruisverwijzingen bij herziene hoofdstukken te corrigeren.

## e-ANS

In 2024 zijn de vernieuwde hoofdstukken 6 Het adjectief en 11 Het tussenwerpsel gepubliceerd. Het hoofdstuk over het adjectief is door Kathy Rys (INT) herzien, het hoofdstuk over het tussenwerpsel door Ton van der Wouden (Meertens Instituut). Beide hoofdstukken zijn inhoudelijk in overeenstemming gebracht met de huidige inzichten in de betreffende onderdelen van de Nederlandse grammatica. Daarnaast zijn ze op verschillende manieren verrijkt, volgens de doelstellingen van de herziening van de e-ANS, met actuele literatuurverwijzingen, koppelingen naar taaladviesbronnen, en voorbeelden uit het Corpus Hedendaags Nederlands en het Corpus Gesproken Nederlands. Vooral het hoofdstuk over het tussenwerpsel bevat nu uitgebreide informatie over variatie, in het bijzonder over de gebruiksverschillen tussen Nederland en België.

Aan deze vernieuwde hoofdstukken zijn bovendien onderwijsmodules toegevoegd. Deze modules bevatten samenvattingen, termenlijsten, veelgestelde vragen en lesideeën. Ze zijn speciaal ontwikkeld voor docenten en leerders van de Nederlandse grammatica, voor wie het niveau van de ANS vaak net te hoog ligt. Ook verscheen in 2024 tweewekelijks de online rubriek Grammaticasafari, waarin op luchtige toon onderwerpen uit de ANS voor een breed publiek uitgelicht worden.

Er is een begin gemaakt met de koppeling van DagentA aan de ANS, door links aan te brengen tussen twee historische grammatica's en de relevante locaties in de ANS. Daarnaast is er regulier onderhoud gepleegd en is de toegankelijkheid van de website verbeterd door toevoeging van een onderwerpenregister.

## Taalportaal

Het onderdeel Nederlandse syntaxis van Taalportaal bestaat uit het werk Syntax of Dutch. Van dit werk is een nieuwe editie in de maak, die ook op Taalportaal zal gaan verschijnen. Hiervoor is er gewerkt aan de conversie van de bronbestanden in Word naar Taalportaal-xml.

## Stage DAGENTA / e-ANS

Voor de e-ANS en DAGENTA heeft stagiaire Julia Schoppers koppelingen aangebracht tussen onderwerpen in de ANS en in twee historische grammatica's uit DAGENTA.

## 8. Nationale en internationale samenwerkingsverbanden

### Netwerken

#### IMPACT Centre of Competence

Het INT is voorzitter van het IMPACT Centre of Competence ([www.digitisation.eu](http://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. In september 2024 is bij de TPDL-conferentie een workshop georganiseerd rondom het thema “sustaining and Sharing Digitization Knowledge.”

#### European Language Data Space

De European Language Data Space (LDS) is door de EU officieel aangesteld om een Europees dataplatform en Europese datamarkt te ontwikkelen voor het verzamelen, creëren, uitwisselen en (her)gebruiken van meertalige en multimodale taaldata. De basis hiervoor is gelegd in eerdere initiatieven zoals het European Language Resource Coordination-initiatief (ELRC) en de European Language Grid (ELG). Ook binnen de European Language Data Space is het INT het nationale aanspreekpunt. In het laatste kwartaal van 2024 is de organisatie van de LDS Country Workshop voor België en Nederland (23 januari 2025) voorbereid. In overleg met het nationale aanspreekpunt van België, en na goedkeuring door de LDS, is ervoor gekozen om een gezamenlijke Country Workshop te organiseren voor beide landen.

#### Elexis Association

Het INT was partner in ELEXIS, een Europees Horizon 2020-2022-project, waarbinnen een infrastructuur voor e-lexicografie is opgezet. Om de duurzaamheid van de infrastructuur te waarborgen is een nieuwe associatie in het leven geroepen, de ELEXIS Association. Doel van de associatie is het organiseren en coördineren van lexicografisch gerelateerde activiteiten en NLP-activiteiten voor zover die relevant zijn voor de lexicografie. Het INT is lid geworden van de associatie. Carole Tiberius is Deputy van de President van de ELEXIS Association. In 2024 is gewerkt aan een Europese projectaanvraag omtrent LLM's en lexicografie en is er een speciale workshop georganiseerd over dit onderwerp op de EURALEX-conferentie in oktober in Kroatië.

## Netwerkprojecten

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

Het INT neemt deel aan het Europese onderzoeksnetwerk UniDive (Universality, diversity and idiosyncrasy in language technology). Het doel van deze COST-actie is om te onderzoeken hoe taaltechnologie verbeterd kan worden door te focussen op universaliteit, maar tegelijkertijd ook de diversiteit van talen in het oog te houden. Dit om te voorkomen dat talen met weinig middelen en bedreigde talen buitengesloten worden in de digitale wereld. Op 8 en 9 februari 2024 vond de tweede algemene bijeenkomst van deze COST-actie plaats in Napels. Het INT is actief binnen twee werkgroepen (Corpusannotatie en Lexicon-Corpus Interface). In dit kader is in 2024 een begin gemaakt met het annoteren van verbale meerwoordsuitdrukkingen in een corpus volgens de [PARSEME](#)-richtlijnen om zo te komen tot een PARSEME-corpus voor het Nederlands.

### European Network on Lexical Innovation (ENEOLI, 2023-2027)

Het INT neemt deel aan het Europese onderzoeksnetwerk (COST) European Network On Lexical Innovation (ENEOLI). De belangrijkste doelstellingen van het netwerk kunnen als volgt worden samengevat: 1) Een gemeenschappelijke, meertalige kernterminologie voor lexicale innovatie definiëren; 2) Digitale methodes en instrumenten in kaart brengen die gebruikt worden om lexicale innovaties in Europese talen te identificeren en te verklaren; 3) Vergelijkende studies uit te voeren naar lexicale innovaties in Europese talen, met speciale aandacht voor ontleningen en hun equivalenten; 4) Specifieke training in neologie aan te bieden aan vertalers, redacteurs, journalisten, technisch schrijvers en docenten. Het INT heeft in 2024 vooral bijgedragen aan het in kaart brengen van internationaal gangbare lexicografische selectiecriteria voor neologismen met het oog op het verbeteren van de eigen neologismenworkflow.

## Onderzoeks- en infrastructuurprojecten

### CLARIAH-Vlaanderen (2021-2024)

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. In samenwerking met KU Leuven werd hiervoor de automatische corpusannotatieservice TextLens gebouwd (<https://portal.clarin.ivdnt.org/textlens/>).

Er werd ook meegewerkt aan de vervolgaanvraag CLARIAH-VL+, die eind december 2024 gehonoreerd werd door het FWO, en waarbij het INT dezelfde rol blijft behouden tot 2028.

### CLARIAH Plus-Nederland (2019-medio 2024)

Het vervolgproject van CLARIAH (Common Lab for Research in the Arts and Humanities) liep van 2019 tot en met 2023, met een extensie tot medio 2024. Het INT hield zich onder andere bezig met een verbetering van de infrastructuur voor historisch Nederlands, uitbreiding op de



corpuszoekmachine BlackLab naar parallelle corpora en dependency treebanks, hulpmiddelen voor het aanbrenen van persistente gebruikersannotaties in corpuszoekresultaten, een gebruikersvriendelijkere digitalisatieworkflow en curatie van dialectwoordenboekdata. Het werk aan het project is medio 2024 afgerond. In 2024 is voornamelijk gewerkt aan het afronden van het GaLAHaD-platform voor taalkundige annotatie en evaluatie, de LAnCeLoT-tool voor handmatige correctie van taalkundige annotatie, het ontwikkelen van trainingdata voor tools (zie hierboven bij xxx [taalkundige verrijking]) en het trainen van state-of-the-art modellen voor taalkundige verrijking van historisch Nederlands. Aan GaLAHaD wordt verder gewerkt in de context van SSHOC-NL.

## SSHOC-NL (Social Science and Humanities Open Cloud for the Netherlands; 2024-2028)

Dit vervolproject van CLARIAH Plus beoogt te komen tot een consortium van onderzoeksinfrastructuren, gericht op het creëren van een ecosysteem van diensten, gegevens en instrumenten voor de sociale wetenschappen en menswetenschappen. Het consortium wordt geleid door ODISSEI, de Nederlandse nationale infrastructuur voor sociale wetenschappen en CLARIAH, de Nederlandse nationale infrastructuur voor geesteswetenschappen.

Het INT werkt samen met de groep van Antske Fokkens van de VU aan infrastructuur voor het methodologisch verantwoord inzetten van machine learning en AI voor dataverrijking.

In 2024 zijn de evaluatiemetrieken in het GaLAHAD-platform uitgebreid met precision, recall en F1-scores (micro en macro), en het platform wordt in het voorjaar van 2025 officieel gereleased met tools voor het verrijken van historisch Nederlands. Er is daarnaast inmiddels gewerkt aan het uitbreiden van het annotatiemodel waardoor het naast lemma en woordsoort ook annotaties als dependentierelatie en named-entitylabeling ondersteunt. GaLAHAD kan nu UD-annotaties uit CoNLL-U en TSV lezen en exporteren. Ook is een begin gemaakt aan een nieuwe evaluatievisualisatie op documentniveau om annotatielagen te vergelijken. UD- en NER-taggers (spaCy, Stanza, UDPipe, Flair) worden nu ondersteund, er is geoptimaliseerd voor multi-threading, en flexibeler afhandeling van TEI-invoer en CoNLL-U-uitvoer is mogelijk gemaakt.

Om de behoeften van de doelgroep in kaart te brengen, is er een enquête uitgestuurd naar alle consortiumpartners en een externe contactlijst. De vragen waren gericht op het verkrijgen van inzicht in het gebruik van tools, de (technische) achtergrond en de behoeften van onderzoekers. Er is onderzoek gedaan naar geschikte repositories voor tools en trainingmateriaal (o.a. met EduSources en SSHOC Marketplace), en er zijn nieuwe samenwerkingen opgezet, waaronder met ATRIUM Task 7. Een overzicht van veelgebruikte CLARIAH-tools helpt bij het bepalen van prioriteiten voor het ontwikkelen van trainingmateriaal, en er zijn trainingmodules over evaluatiemethoden in GaLAHAD.

Tot slot is er gewerkt aan een website over verrijking van data, die in 2025 online komt.

## SignON (2021-2024)

Het INT was als consortiumpartner betrokken bij het SignON-project, dat vanaf voorjaar 2021 voor drie jaar gefinancierd werd 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 sterke 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. In 2024 werd dit project succesvol afgerond. Het INT houdt de opgezette infrastructuur voor dit project nog minstens vijf jaar beschikbaar.

Er werd in 2024 een nieuwe aanvraag ingediend met de Vlaamse partners van SignON voor een FWO-SBO-project. In verband met de eerste versie, ingediend in 2023, kregen we de vraag om deze verfijnd en verbeterd opnieuw in te dienen.

### SABeD (2021-2024)

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 met gesproken data om te gaan. Het SABeD-project werd afgerond en de resultaten zijn downloadbaar via de Taalmaterialenwebsite

### Gesproken Corpus van de Zuidelijk-Nederlandse Dialecten (2020-2024)

Op 24 oktober 2024 is het eindresultaat van project ‘Gesproken Corpus van de Zuidelijk-Nederlandse Dialecten’ gepresenteerd, een project dat van 2020 tot 2024 liep en gerealiseerd is aan de UGent. Het project ontsluit een collectie van dialectopnames uit 768 plaatsen in België, Frankrijk en het zuiden van Nederland, opgenomen tussen 1963 en 1976. Het INT is verantwoordelijk voor de ontsluiting en beschikbaarstelling van dit corpus. Het corpus is gepubliceerd in een combinatie van een BlackLab-instantie waarin de basisannotatielagen van het corpus bevroegd kunnen worden (zwaar en licht genormaliseerde transcriptielagen, lemma, woordsoort) en de opnames beluisterd kunnen worden, en een daaraan gekoppelde GrETEL-instantie waarin de syntactische annotatielaag benut kan worden.

### Geparset Corpus van de Zuidelijk-Nederlandse Dialecten (2024-2028)

Dit project is een vervolg op het project ‘Gesproken Corpus van de Zuidelijke Nederlandse Dialecten’ en voorziet in een aanzienlijke geografische uitbreiding waardoor het corpus het hele Europees-Nederlandse gebied zal bevatten. Het project gebruikt de bestaande audio-gealigneerde transcripties en annotaties uit de eerste fase van het GCND om NLP-tools te hertrainen en zo de syntactische annotatie van de nieuwe data te versnellen. De rol van het INT is om de voor het GCND gebouwde infrastructuur om de data te ontsluiten hierop aan te passen en uit te breiden. In 2024 is gewerkt aan de opzet van een verbeterde workflow voor de dataprocessing en aan de principes voor de taalkundige verrijking binnen het vervolgproject.

### Pilootproject Duidelijke Taal (2023-2024)

In dit project wordt een dataset gevalideerd die als goudstandaard kan dienen bij automatische tekstvereenvoudiging. Bestaande data en automatisch gegenereerde data laten we beoordelen door middel van crowdsourcing, waarbij het publiek zinnen kan beoordelen op drie dimensies:

accuraatheid, vlotheid en complexiteit. In 2024 werd dit project afgerond, waarbij de eindresultaten beschikbaar gesteld zijn via de Taalmaterialenwebsite. Een grondige analyse van de resultaten is gepland voor 2025.

## Spread the News (2020-2025)

Het project ‘Spread the new(s). Understanding standardization of Dutch through 17th-century newspapers’, dat wordt uitgevoerd met subsidie van NWO Open Competition SSH, heeft ook in 2024 goede vorderingen gemaakt. Promovenda Machteld de Vos, die gedeeltelijk op het INT en gedeeltelijk bij de Radboud Universiteit werkt, heeft met co-auteurs twee artikelen geschreven die onder review zijn. In het laatste kwartaal lag het promotiewerk stil vanwege zwangerschapsverlof, waardoor de einddatum van het project verschoven is naar eind 2025.

## Using CoBaLT and GaLAHaD for historical corpus annotation (2024)

In het CLARIAH-project ‘Using CoBaLT and GaLAHaD for historical corpus annotation’ is CoBaLT (herdoopt tot LAnCeLoT) een tool voor interactieve corpusannotatie en het GaLAHaD-platform voor taalkundige annotatie van historisch Nederlands geëvalueerd. Het rapport is te vinden op de website van het [INT](#).

## Buitenlandse Aardrijkskundige Namen – fase 1

Het is de bedoeling om de website <https://namen.taalunie.org> te vernieuwen. Op deze website zijn alle buitenlandse aardrijkskundige namen waarvoor een Nederlandse vorm bestaat, te vinden. De schrijfwijze en het gebruik van buitenlandse aardrijkskundige namen in het Nederlands is gestandaardiseerd en de Commissie Aardrijkskundige Namen van de Nederlandse Taalunie beheert en actualiseert de inhoud van de website. In fase 1 van het project wordt de data van de BAN-lijst in een databaseomgeving ondergebracht zodat de BAN-lijst gemakkelijk onderhouden en geüpdatet kan worden. In 2024 is de conversie van de website-gebaseerde datastructuur naar een meer database-gerichte structuur opgezet. In de nieuwe structuur scheiden we *termen* van *geografische concepten*. Verschillende benamingen van eenzelfde geografisch object (exoniemen, endoniemen) zijn op deze manier eenvoudiger bij elkaar te zoeken, en het faciliteren van de koppeling van geografische informatie zoals coördinaten is zo ook meer voor de hand liggend. In 2025 bouwen we de bewerkingsomgeving.

## Namenloket (1-03-2025 – 28-02-2027)

Het project Namenloket werd in 2024 voorbereid door de vakgroep Taalkunde van de UGent en het INT, en het werd goedgekeurd. Het project van Variaties vzw Koepelorganisatie voor Dialecten en Oraal Erfgoed in Vlaanderen wil een digitaal interactief informatieportaal bouwen over het Vlaams namenerfgoed, de eigennamen die Vlamingen vanouds zichzelf, maar ook o.a. plaatsen en dieren geven. Bezoekers zullen er een antwoord vinden op vaakgestelde vragen over eigennamen en zullen ook vragen kunnen stellen. Via een dataplatform zullen erfgoedvrijwillers en –onderzoekers hun eigen namenverzamelingen digitaal kunnen ontsluiten. Het portaal zal ook didactisch materiaal bevatten. Partners in het project naast het INT zijn Histories vzw, de Koninklijke Commissie voor Toponymie en Dialectologie, Didactica, (Antwerp School of Education, Universiteit Antwerpen), Meemoo (Vlaams instituut voor het archief) en de Vakgroep Taalkunde van de Universiteit Gent. Het

INT is naast de bouw van de website verantwoordelijk voor het opzetten van een dataplatform met zoekmodule waarop erfgoedvrijwilligers en onderzoekers hun eigen dataverzamelingen (bv. namenwoordenboeken en -glossaria) kunnen uploaden. Op die manier kunnen ze hun werk digitaliseren, verduurzamen, beter doorzoekbaar maken en voor een breder publiek beschikbaar stellen. Dit project ontwikkelt de digitale infrastructuur en protocollen nodig voor die integratie en maakt bij wijze van aanzet één dataverzameling – het monumentale Woordenboek van Familienamen in België en Noord-Frankrijk van Frans Debrabandere – voor het eerst volledig digitaal beschikbaar en doorzoekbaar voor het brede publiek.

## Macroscopic Large-Scale Research Infrastructure (National Roadmap consortia 2024)

In 2024 is bij NWO een aanvraag ingediend voor een nieuw groot infrastructuurproject, Macroscopic. Het project beoogt de allereerste “Macroscoop” ter wereld te creëren, een krachtig hulpmiddel om complexe maatschappelijke problemen op grote schaal te onderzoeken. Er wordt voortgebouwd op bestaande Nederlandse infrastructuren zoals ODISSEI en CLARIAH. De Macroscoop moet onderzoekers in staat stellen om op een veilige manier toegang te krijgen tot grote hoeveelheden gevoelige, onderling verbonden data en deze te analyseren. Het INT zal betrokken worden bij de bouw van een Nationaal Media Corpus en bij de ontwikkeling van tools voor analyse en annotatie.

## Overige infrastructurele dienstverlening

### Infrastructuur Papiamentu/u

Op 1 februari 2024, de Internationale Dag van de Moedertaal, hebben het INT en het Nationaal Taalinstituut Curaçao (NTIC) een samenwerkingsverband gesloten dat vastgelegd is in een Memorandum of Understanding. Het doel is om samen te werken aan de uitbouw van de digitale lexicaal infrastructuur voor het Papiaments. Het INT brengt haar kennis en expertise in op het gebied van (lexicale) databases en bouwt en host de infrastructuur.

In de eerste fase zijn een databasestructuur en een bewerkingsomgeving voor de morfosyntactische basislagen ontwikkeld. De online bewerkingsomgeving bevat de data van de officiële spellinglijst Buki di Oro, en daarnaast de woordenlijsten die voortkomen uit door FPI ontwikkelde schoolmaterialen, de basiswoordenschat uit de leerboeken Papiamentu A1/A2 en de Standaardisatiebestanden tot 1994. Er is een aanvang gemaakt met het automatisch genereren van paradigma's. Medewerkers van NTIC (directeur Marta Dijkhoff en Ellen Meulens) hebben het INT bezocht voor een workshop van 7 t/m 10 oktober 2024. In een volgende fase zal gewerkt worden aan het koppelen van semantische informatie, waarvoor het INT een bewerkingsomgeving zal ontwikkelen, en aan het koppelen van Bonairiaanse en Arubaanse varianten. We verkennen de mogelijkheden om tot een corpus te komen om dekking van het lexicon te toetsen en een evidentiegebaseerde aanpak van de beschrijving van de Papiamentse woordenschat te ondersteunen.

## Etymologiebank

Sinds 2020 zijn de etymologiebank (etymologiebank.nl), en de uitleenwoordenbank (uitleenwoordenbank.ivdnt.org) in het beheer van het INT. In 2024 hebben twee stagiairs een bijdrage geleverd aan de verrijking van de etymologiebank.

## GLAD

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. Twee stagiaires uit een uitwisselingsprogramma Leuven-Italië hebben aan GLAD meegewerkt. De Nederlandse gegevens van GLAD zijn in 2024 uitgebreid tot ruim 10.000 trefwoorden. Op basis van de GLAD-database is een Routledge Anthology on Anglicisms gepland voor 2025, getiteld *Anglicisms around the Globe: Cross-linguistic Studies on the Impact of English*; in 2024 is hiervoor een hoofdstuk 'A cross-linguistic study on pseudo-Anglicisms in six languages' geschreven aan de hand van o.a. de Nederlandse gegevens.

## DAGENTA

Het INT host de 'Database Geschiedenis Nederlandse Taalkunde' (DAGENTA) (<https://dagenta.ivdnt.org>). Een Nijmeegse stagiaire heeft links gelegd tussen enkele grammatica's uit DAGENTA en de e-ANS.

## Pallas

Aan de ontsluiting van de Digitale Pallas, een 18e-eeuws Russisch meertalig woordenboek, heeft in 2024 een stagiair vele Noord- en Zuid-Amerikaanse woordenlijsten toegevoegd, verzameld door Spaanse missionarissen en bijeengezocht door onderzoekster Rebeca Fernandez Rogriguez. Aan de database, die de inhoud van Peter Simon Pallas, *Comparative Dictionary of All Languages and Dialects* (1790-1791) bevat (<https://pallas.ivdnt.org/>), is in 2024 een Wordpress-site toegevoegd die gevuld zal worden met achtergrondinformatie en artikelen. Deze wordt gereleased in 2025.

## 9. Systeembeheer en ICT-capaciteit

2024 stond voor de infrastructuur vooral in het teken van het up-to-date brengen van verschillende softwareonderdelen en het vervangen en verbeteren van de opslagsystemen. Een nieuwe laag extra snelle opslag is toegevoegd om applicaties die gebruik maken van I/O-intensieve zoeksystemen een *performance boost* te kunnen geven. Voor de toekomst is onderzocht of er mogelijkheden bestaan om de groeiende licentiekosten onder controle te brengen en zijn er verschillende pilots gedaan met systemen om SSO-functionaliteit te bieden aan onze bezoekers. Ook wordt gewerkt aan verdere structurering van de processen tussen ontwikkeling en onderhoud, en wordt onderzocht in hoeverre het INT NIS2- en ISP 270002-compliant is/kan worden. Dit zullen ook in 2025 belangrijke activiteiten blijven. In 2024 is er één keer een incident geweest binnen de beveiligingsstructuur: een certificaat van een derde partij bleek gecompromitteerd en moest direct vervangen worden. Er is bij dit incident geen schade berokkend.

Het INT verwelkomde 13 nieuwe applicaties op het netwerk, waaronder het beheer over Woordenlijst.org van de Taalunie, het DATKD-project van CLARIN-Vlaanderen, het Dative-project van CIPL, het *Jiddisch Nederlands Woordenboek* van Stichting Jiddische Lexicografie Amsterdam en verder onder andere GLAD Global Anglicism Database, Duidelijke Taal AI-taalgeneratiechecker, DAGENTA-literatuurdatabase en GCND Gesproken Corpus Zuid-Nederlandse Dialecten.

## 10. Disseminatie

### Doelgroepenbeleid

#### Onderzoekers en taalkundigen

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. Taalkundigen van het INT hebben aan de KU Leuven het vak Lexicologie en morfologie gegeven in de Master Taalkunde en Taal- en Letterkunde en aan de Universiteit Leiden de collegereeks Corpus Lexicography en Computational Corpus Analysis. Daarnaast heeft het INT zijn werkterrein actief verbreed naar docenten en studenten. In dat verband was het instituut aanwezig op en profileerde het zich op conferenties (HSN-conferentie), festivals (Week van het Nederlands, Letterlijk Leiden, Het Taalfeest) en wetenschappelijke bijeenkomsten. Ook heeft het INT opnieuw een opgave gemaakt voor de jaarlijkse Taalkundeolympiade van de Universiteit Leiden.

#### Onderwijs

De tweede belangrijke doelgroep is het onderwijs: leerlingen, studenten en docenten. De onderwijspagina's op de website van het INT zijn geüpdatet en verder uitgebreid. Daarnaast zijn er nieuwe educatieve posters gemaakt voor het project 'Posters in de klas'. In het kader van de webinarreeks 'Werken met online taalbronnen' organiseerde het INT in samenwerking met de Taalunie twee webinars gericht op het onderwijs: 'Nieuwe woorden in het Nederlands' en 'Sneller spellen met Woordenlijst.org'. In opdracht van de Taalunie en het Instituut voor de Nederlandse Taal zijn er nieuwe lesmaterialen ontwikkeld over vertaaltools voor de internationale neerlandistiek. De resultaten zijn gepubliceerd op de website mijnNederlands.org. Voor deze doelgroep is ook de onderwijsmodule van de e-ANS verder uitgebreid met samenvattingen en lesideeën.

#### Algemeen publiek

Een derde doelgroep is het algemene publiek, de taalgeïnteresseerden en taalliefhebbers. Die doelgroep wordt op verschillende manieren bediend. Jaarlijks neemt het INT deel aan het festival Letterlijk Leiden, in 2024 met een pubquiz over taal. Op onze website verschenen wekelijks populairwetenschappelijke rubrieken over grammatica, dialectwoorden, nieuwe woorden en historische woorden. We zijn actief op verschillende sociale media en maken op regelmatige basis nieuwe afleveringen van onze podcastseries 'Over taal gesproken' en 'Waar komt pindakaas vandaan?' (die ook gebruikt worden als lesmateriaal). Tien keer werden twee soorten nieuwsbrieven verstuurd aan abonnees. Daarnaast hielden medewerkers lezingen voor een divers publiek, waren zij

te gast in programma's op radio en tv 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 ook het uitgebreide overzicht in bijlage 3.

## Communicatiemiddelen

### Website

De website [ivdnt.org](http://ivdnt.org) is vaak het eerste wat het publiek ziet van het INT, en dat maakt het een van de belangrijkste communicatiemiddelen. Veel communicatie-uitingen zijn erop gericht om bezoekers naar de website te trekken. Sinds 2023 worden de statistieken bijgehouden met Plausible. Daarvóór werd Google Analytics gebruikt om statistieken te verzamelen. In 2024 telde de website 626.482 paginaweergaven (423.944 unieke bezoekers), een stijging van zo'n 9% ten opzichte van het voorgaande jaar (570.699 paginaweergaven en 379.620 unieke bezoekers in 2023).

Het grootste deel van de bezoekers (250.000) komt via Google op de website terecht. 78.700 bezoekers tikken [ivdnt.org](http://ivdnt.org) direct in de browser in, 57.100 bezoekers komen via diverse nieuwsbrieven (eigen nieuwsbrief (1.900), Taalpost, Taalunienieuwsbrief) en 7.000 via de sociale media: de meesten daarvan vinden de website via LinkedIn, gevolgd door Facebook, X (voorheen Twitter), Instagram en Bluesky (Mastodon komt niet terug in de statistieken).

Van de statische content is de informatiepagina over het *Woordenboek der Nederlandsche Taal* (WNT) met 29.255 paginabezoeken (24.024 unieke bezoekers) het meest bezocht in 2024. De pagina *Spelling & Groene Boekje* kreeg 20.167 paginabezoeken (18.219 unieke bezoekers). Een opvallende stijging van zo'n 300% ten opzichte van 2023. Dit is waarschijnlijk het gevolg van de website [Woordenlijst.org](http://Woordenlijst.org) die begin 2024 vernieuwd is qua uiterlijk en zoekmogelijkheden, en het daarop volgende webinar over [Woordenlijst.org](http://Woordenlijst.org).

Al jaren is 'Nieuw woord van de week' de best bezochte actuele rubriek, met in totaal 127.487 bezoeken (88.434 unieke bezoekers) in 2024. Gemiddeld blijven de bezoekers zo'n 3,5 minuten op de pagina. Ook de (sinds 2020 niet meer bijgehouden) rubriek 'Woordbaak' wordt nog regelmatig geraadpleegd met in 2024 93.886 bezoeken in totaal (84.425 unieke bezoekers).

### Totaal aantal paginabezoeken per categorie

Actueel	351.416
Woordenboeken	140.117
Spelling & grammatica	23.455
Corpora & lexica	7.981
Over ons	7.433
Terminologie	5.024
Onderwijs	4.616
Themapagina Hedendaags Nederlands	2.355
Themapagina Historisch Nederlands	3.941
Taalmaterialen	1.764
Onderzoek & projecten	1.401



## Totaal aantal paginabezoeken Woorden van de week

Nieuw woord van de week	127.487
Woordbaak	93.886
Uit de streek	28.885
Terug in de taal	12.221
Grammaticasafari	12949

48% van de bezoekers raadpleegt ivdnt.org op een smartphone, 40% maakt gebruik van een desktop, 9% van een laptop, en 3% van een tablet. 17,6% van de bezoekers scrolt door tot het einde van de homepage.

De webredactie en de webmaster werken continu aan de verbetering van de website. Zo worden fouten hersteld, worden kapotte links verwijderd of vervangen, wordt de veiligheid verbeterd, en wordt op verschillende manieren gewerkt aan het versnellen van de laadtijd en aan zoekmachineoptimalisatie (SEO). Wanneer je via een incognitovenster in Google zoekt op termen als 'Nederlandse taal' (tweede resultaat), 'nieuwe woorden' (eerste resultaat), 'hedendaags Nederlands' (tweede resultaat), 'historisch Nederlands' (tweede resultaat) of 'terminologie' (eerste resultaat), verschijnt de INT-website steeds hoog in de zoekresultaten. Daarnaast is er een serverupgrade geweest om verder te kunnen optimaliseren.

## 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 huisstijl zorgen ervoor dat alle door het INT ontwikkelde websites door de gebruikers ook als zodanig herkend worden. Die vormgeving, die zelf ook continu verbeterd en geoptimaliseerd wordt, wordt zo veel mogelijk geïmplementeerd in nieuwe en bestaande websites. In 2024 is dat onder andere gebeurd voor Duidelijke Taal, een project van de Taalunie en het INT. Op die website kon je in een spelomgeving zinnen beoordelen op een aantal punten, zoals juistheid en complexiteit.

## Nieuwsbrieven & persberichten

In 2024 zijn er zes algemene INT-nieuwsbrieven verschenen en vier nieuwsbrieven over terminologie. De algemene nieuwsbrief telde eind 2024 4.662 abonnees, van wie per editie 54% de nieuwsbrieven opent; van de 3.531 geadresseerden van de nieuwsbrief terminologie doet 41% dat.

In de algemene nieuwsbrieven zijn het vooral de Woorden van de week en de nieuwsberichten waarop doorgelinkt wordt naar [ivdnt.org](http://ivdnt.org).

Er is één persbericht verstuurd: Samenwerkingsverband Nationaal Taalinstituut Curaçao en Instituut voor de Nederlandse Taal.

## Sociale media

Alle berichten die verschijnen op [ivdnt.org](http://ivdnt.org) worden met behulp van Blog2Social doorgeplaatst op LinkedIn (4.264 volgers), Facebook (853 vind-ik-leuks), Mastodon (519 volgers) en Bluesky (364 volgers). Voor Instagram (1.793 volgers) wordt voor elke post een aparte afbeelding opgemaakt in de huisstijl. Op LinkedIn zijn het vooral vacatures en nieuwsberichten die veel interactie opleveren en bezoekers naar de website trekken. Op Facebook zijn het de 'Woorden van de week' die het meest gedeeld worden. Op de Facebookpagina van het INT zelf is er vrij weinig interactie. Ook op Instagram scoren de 'Woorden van de week' het beste, en dan met name 'Nieuw woord van de week'. Van de verschillende typen posts zijn het de reels die het grootste publiek bereiken. Het gaat in de meeste gevallen dan om fragmenten uit de podcast.

Bluesky is in 2024 nieuw toegevoegd aan de socialemediakanalen ter vervanging van X. Het INT-account op X bestaat nog wel maar is sinds eind 2024 inactief. Vooralsnog blijven LinkedIn en Instagram de twee kanalen waar het meest op ingezet wordt en waar de meeste activiteit plaatsvindt.

## Podcasts

### Over taal gesproken

In 2024 zijn er zeven afleveringen verschenen 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. In totaal zijn er nu dertig afleveringen verschenen en de twee best beluisterde in de eerste zeven dagen na publicatie zijn 'Etymologie – waar komen onze woorden vandaan?' (2.216 downloads) en 'De uitspraak van de R' (2.298 downloads). Over alle afleveringen verspreid waren er in 2024 in totaal zo'n 52.000 downloads van de podcast. 71% van de luisteraars komt uit Nederland en 11% uit België. 'Over taal gesproken' heeft eind 2024 6.588 abonnees op Spotify en 2.081 abonnees op Apple Podcasts. Verschillende fragmenten zijn opgenomen in educatieve uitgaven. Daarnaast werd de podcast genomineerd voor een Language Industry Award.

Naar aanleiding van het luisteronderzoek in 2023 zijn er enkele wijzigingen aangebracht in de podcast. Zo is de introductie ingekort en wordt er minder (aanwezige) muziek doorheen gemonteerd.

### Waar komt pindakaas vandaan?

Er is een nieuw seizoen met negen afleveringen gemaakt van de podcast 'Waar komt pindakaas vandaan?', dit keer samen met Onze Taal en in eigen beheer geproduceerd. Dit is een volledig

gescripte podcast over de verrassende etymologie van Nederlandse woorden. Het nieuwe seizoen startte in de Week van het Nederlands. In 2024 zijn er in totaal 54.100 afleveringen gedownload. 79% van de luisteraars komt uit Nederland en 7% uit België. Ook uit deze podcast zijn verschillende fragmenten opgenomen in lesmethodes Nederlands.

## Andere populairwetenschappelijke activiteiten

Er zijn in 2024 twee gelegenheidswoordenboekjes gepubliceerd: een geactualiseerde versie van het '3 oktoberwoordenboekje' en 'Ach en Wee', naar aanleiding van de tweehonderdste sterfdag van schrijver Rhijnvis Feith.

De redactie van het Woordenboek van Nieuwe Woorden publiceerde enkele artikelen over nieuwe woorden, hield veel lezingen, webinars en interviews (op radio en tv) over dit onderwerp en begeleidde diverse profielwerkstukken over neologismen.

In samenwerking met de Taalunie heeft het INT twee webinars georganiseerd in de reeks 'Werken met online taalbronnen': 'Sneller spellen met Woordenlijst.org' en 'Nieuwe woorden in het Nederlands'. De webinars vonden plaats in de Week van het Nederlands.

Aan het einde van het jaar brachten we weer een populairwetenschappelijke publicatie uit die als eindejaarsgeschenk verstuurd werd naar INT-relaties. Dit keer was dat het boek *Taal in transformatie* van Frieda Steurs. Tijdens het festivalweekend Letterlijk Leiden presenteerden Vivien Waszink en Laura van Eerten een (uitverkochte) pubquiz over taal.

Het INT was al verantwoordelijk voor de techniek achter de website [Neerlandistiek.nl](http://Neerlandistiek.nl) en zal ook in 2025 de hosting en het onderhoud op zich nemen.

# Bijlage 1: Raad van Toezicht en Raad van Advies

## **Raad van Toezicht**

Mr. Mieke Zaanen (voorzitter)

Drs. Erik Boels

Mr. Frank Judo

## **Raad van Advies**

Prof. dr. Hans Bennis (voorzitter)

Prof. dr. Dirk Geeraerts

Prof. dr. Veronique Hoste

Prof. dr. Reinhild Vandekerckhove

Lic. Wim Vanseveren

Dr. Eric Mijts

Prof. dr. Lisa Cheng

# Bijlage 2: Medewerkers

Prof. dr. Frieda Steurs – directeur/bestuurder

Dr. Maaïke Beliën – onderzoeker/taalkundige

Marjolijn van Bennekom – taalkundig assistent (in dienst t/m december 2024)

Dr. Bob Boelhouwer – computerlinguïst (in dienst t/m juli 2024)

Dr. Roland de Bonth – onderzoeker/taalkundige

Tim Brouwer – softwareontwikkelaar (in dienst t/m juli 2024)

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  
Uma Kraus - managementassistent  
Drs. Marco van der Laan – systeembeheerder  
Dr. Frank Landsbergen – computerlinguïst  
Koen Mertens – systeemontwikkelaar  
Drs. Jan Niestadt – systeemontwikkelaar  
Vincent Prins, MSc - systeemontwikkelaar  
Dr. Kathy Rys – onderzoeker/taalkundige  
Prof. dr. Nicoline van der Sijs – onderzoeker/taalkundige (in dienst t/m 31 januari 2024)  
Prof. 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  
Drs. Vivien Waszink – onderzoeker/taalkundige  
Karin van Weerlee – managementassistent  
Laetitia de Winter – administrateur  
Bram Vanroy – onderzoeker/computerlinguïst (in dienst per april 2024)

## Bijlage 3: publicaties, lezingen, media, prijzen etc.

### Publicaties

- Barbu Mititelu, V., Giouli, V., Evang, K., Zeman, D., Osenova, P., Tiberius, C., Krek, S., Markantonatou, S., Stoyanova, I., Stanković, R. & Chiarcos, Ch. (2024). Multiword Expressions between the Corpus and the Lexicon: Universality, Idiosyncrasy, and the Lexicon-Corpus Interface. In: *Proceedings of the Joint Workshop on Multiword Expressions and Universal Dependencies (MWE-UD) @ LREC-COLING 2024*, pp. 147–153, Torino, Italia. ELRA and ICCL. [link](#)
- Bonth, R. de (2024). ‘Over de operette’. Op: [www.neerlandistiek.nl](http://www.neerlandistiek.nl) (4 januari).
- Bonth, R. de (2024). ‘Salopette: een smeerpoets met een tuinbroek?’. Op: [www.neerlandistiek.nl](http://www.neerlandistiek.nl) (1 februari).
- Bonth, R. de (2024). ‘Alles willen weten over woorden’. Op: [www.neerlandistiek.nl](http://www.neerlandistiek.nl) (18 februari).
- Bonth, R. de (2024). ‘Probleempjes met -ette’. Op: [www.neerlandistiek.nl](http://www.neerlandistiek.nl) (7 maart).
- Bonth, R. de (2024). ‘Vooruitstrevend voertuig: de voiturette’. Op: [www.neerlandistiek.nl](http://www.neerlandistiek.nl) (4 april).
- Bonth, R. de (2024). ‘Afko’s in Beatrijs’. Op: [www.neerlandistiek.nl](http://www.neerlandistiek.nl) (16 april).
- Bonth, R. de (2024). ‘Boter-kaas-en-eieren met -ette’. Op: [www.neerlandistiek.nl](http://www.neerlandistiek.nl) (2 mei).
- Bonth, R. de (2024). ‘Een wolf in de klas’. Op: [www.neerlandistiek.nl](http://www.neerlandistiek.nl) (19 mei).
- Bonth, R. de (2024). ‘Stoppen met roken door -ette’. Op: [www.neerlandistiek.nl](http://www.neerlandistiek.nl) (6 juni).
- Bonth, R. de (2024). ‘Schreef Marten Toonder onzin?’. Op: [www.neerlandistiek.nl](http://www.neerlandistiek.nl) (19 juni).
- Bonth, R. de (2024). ‘Een zwak theoretje: de theorette’. Op: [www.neerlandistiek.nl](http://www.neerlandistiek.nl) (4 juli).
- Bonth, R. de (2024). ‘Etymologica: het medezeggenschap’. Op: [www.neerlandistiek.nl](http://www.neerlandistiek.nl) (15 juli).
- Bonth, R. de (2024). ‘Mmm: mitraillette en matinette’. Op: [www.neerlandistiek.nl](http://www.neerlandistiek.nl) (1 augustus 2024).
- Bonth, R. de (2024). ‘Veilig fietsen en stallen’. Op: [www.neerlandistiek.nl](http://www.neerlandistiek.nl) (5 september).
- Bonth, R. de (2024). ‘Een achtervoegselestafette’. Op: [www.neerlandistiek.nl](http://www.neerlandistiek.nl) (3 oktober).
- Bonth, R. de (2024). ‘Een kansspel met -ette’. Op: [www.neerlandistiek.nl](http://www.neerlandistiek.nl) (7 november).
- Bonth, R. de (2024). ‘“Voert dezen man niet ad bestias”. De uitleveringskwestie-De Vreese’. Op: [www.neerlandistiek.nl](http://www.neerlandistiek.nl) (12 november).
- Bonth, R. de (2024). ‘Schilderen met -ette’. Op: [www.neerlandistiek.nl](http://www.neerlandistiek.nl) (5 december).
- Bonth, R. de, ‘Afko’s in Beatrijs’. Opdracht voor de Taalkundeolympiade 27 januari 2024.
- Bonth, R. de, ‘Van Sterne naar Wuwestert en terug. De eerste en tweede druk van het Middelnederlandsch Handwoordenboek’. In: *Trefwoord*, online tijdschrift over lexicografie, 2024.
- Bonth, R. de (2024). Ach en wee. Een gelegenheidswoordenboekje voor Rhijnvis Feith. Op: [ivdnt.org](http://ivdnt.org).

- Bonth, R. de (2024). Elke twee weken een bijdrage aan de webrubriek Terug in de Taal. Op: [ivdnt.org](http://ivdnt.org).
- Coussé, E., Bouma, G. & Sijs, N. van der (2024), 'Auxiliaries in Old Dutch. A diachronic parallel corpus exploration'. In: *Journal of Historical Linguistics*.
- Depuydt, K., Haga, T., Mooijaart, M., m.m.v. Bonth, R. de, Depoorter, G., van pellicom, K., & Verheij, B. (2014). *Lemmatiseerprincipes voor GiGaNT, het centrale lexicon van het INT* (versie 1.0; INT Working Papers 2). [https://ivdnt.org/wp-content/uploads/2024/11/lemmatiseerprincipesV2\\_combi.pdf](https://ivdnt.org/wp-content/uploads/2024/11/lemmatiseerprincipesV2_combi.pdf)
- De Sisto, M. et al. [o.a. Vandeghinste, V.] (2024). XSL-HoReCo and GoSt-ParC-Sign: Two New Signed Language - Spoken Language Parallel Corpora. CLARIN2023: Selected papers.
- Fanee, M. (2024). 'Een beer voor de vrouwe van Voorne', in: *De Hekkenluiter*, Historisch Genootschap Warmelda, jrg. 21, nr. 2, pp. 26-27.
- Fanee, M. (2024). 'Sporen van 3 oktober 1574 in Warmond', Tekst bij de tentoonstelling & Verantwoording, Historisch Genootschap Warmelda.
- Fanee, M. (2024). "Het oude gerecht van Warmond: Over de doodstraf, de galg en het rechthuis". In: *Geschiedenis van Zuid-Holland*, Erfgoedhuis Zuid-Holland.
- Fanee, M. (2024). "Vuile lusten ende verfoeilijck drincken". Verslag van een reeks zedenzaken in 18de-eeuws Warmond. In: *De Hekkenluiter*, Historisch Genootschap Warmelda, jaargang 21, nr. 1, pp. 35-51.
- Haga, T., Depuydt, K., & Does, J. de. (2024). *Tagset voor Diachroon corpusmateriaal van het Nederlands (TDN)* (versie 2.0; INT Working Papers 2). [https://ivdnt.org/wp-content/uploads/2024/11/TDNV2\\_combi.pdf](https://ivdnt.org/wp-content/uploads/2024/11/TDNV2_combi.pdf)
- Klosa-Kückelhaus, A. & Tiberius, C. (2024). The Lexicographic Process Revisited. In: *International Journal of Lexicography*, 1 November 2024. [link](#)
- Klosa-Kückelhaus, A. & Tiberius, C. (2024). 3 The Lexicographic Process. In: Klosa-Kückelhaus, A. (red.), *Internet Lexicography: An Introduction*. Berlin, Boston: De Gruyter, pp. 61-96. [link](#)
- Kosem, I., Zingano Kuhn, T., Arhar Holdt, Š., Koppel, K., Tiberius, C., Zviel-Girshin, R., Waszink, V. & Zgaga, K. (2024). Can AI Assist in Selecting Dictionary Examples? A Case Study in Four Languages. In: Štrkalj Despot, K., Ostroški Anić, A. & Brač, I. (red.) *Lexicography and Semantics Book of Abstracts of the XXI EURALEX International Congress*. 8–12 oktober 2024 Cavtat, Kroatië, Institut za hrvatski jezik, pp. 128-130. [link](#)
- Martelli, F., Perrella, S., Campolungo, N., Munda, T., Koeva, S., Tiberius, C. & Navigli, R. (2024). DiBiMT: A Gold Evaluation Benchmark for Studying Lexical Ambiguity in Machine Translation. *Computational Linguistics 2024*. [link](#)
- Mathysen, J. et al. [o.a. Vandeghinste, V.] (2024). Constructing SABeD: A Spoken Academic Belgian Dutch Corpus. CLARIN2023: Selected papers.
- Norré, M. et al. (o.a. Vandeghinste, V.) (2024). Évaluation de la traduction en pictogrammes pour la communication médecin-patient par des adultes avec une déficience intellectuelle. In : *Actes du Congrès Handicap 2024 - 13ème édition*. pp. 181-186. Paris.

- Oostendorp, M. van & De Tier, V. (2024). 'Harde en zachte g'. In: *Onze Taal*, nr. 5, pp. 16-17.
- Oostendorp, M. van & De Tier, V. (2024). 'Jullie'. In: *Onze Taal*, nr. 2, pp. 16-17.
- Romein, A., et al. [o.a. Depuydt, K., Sijs, N. van der] (2024), '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'. In: *Journal of Data Mining & Digital Humanities*, <https://jdmhd.episciences.org/13242>.
- Rys, K. (2024). 6 Het adjectief. *Algemene Nederlandse Spraakkunst*.
- Segonne, V. et al. [o.a. Vandeghinste, V.] (2024). Jargon: A Suite of Language Models and Evaluation Tasks for French Specialized Domains. In: *Proceedings of LREC-COLING 2024*. Torino, Italy.
- Shterionov, D. et al. [o.a. Vandeghinste, V. & Vanroy, B.] (2024). SignON – a Co-creative Machine Translation for Sign and Spoken Languages (end-of-project results, contributions and lessons learned). In: *Proceedings of the 24th Annual Conference of the European Association for Machine Translation (EAMT)*. Sheffield, UK.
- Sijs, N. van der (2024), 'Engels in het Nederlands', iedere maandag in De Onze-Taalkalender 2025, van *Onze Taal*.
- Sijs, N. van der (2024), bewerking van het overzichtartikel 'Vlaanderen: Etymologie en betekensevolutie' van M. Gysseling t.b.v. Digitale Encyclopedie van de Vlaamse Beweging (DEVB): <https://encyclopedievlaamsebeweging.be/nl/etymologie-en-betekensevolutie-van-vlaanderen>.
- Sijs, N. van der (2024), 'Een achttiende-eeuwse fantasietaal'. In: *Verhalen in 4D*, pp. 12-15.
- Sijs, N. van der (2024), 'Taalverandering'. In: *Trouw Tijdgeest*, 4 mei, pp. 10-15.
- Sijs, N. van der (2024), 'Bully'. In: *Onze Taal* 6, 48.
- Sijs, N. van der (2024), 'Tuxedo'. In: *Onze Taal* 5, 36.
- Sijs, N. van der (2024), 'Blurb'. In: *Onze Taal* 4, 40.
- Sijs, N. van der (2024), 'Doedel'. In: *Onze Taal* 3, 58.
- Sijs, N. van der (2024), 'Ijsbergsla'. In: *Onze Taal* 2, 36.
- Sijs, N. van der (2024), 'Sedumdak'. In: *Onze Taal* 1, 36.
- Sijs, N. van der (2024), 'De opkomst van de benaming limonadewesp'. In: *Hymenovaria*: <https://www.hymenovaria.nl/category/publicaties/hymenovaria/>
- Sijs, N. van der (2024), 'Now online: The Digital Pallas'. In: CIPL newsletter: <https://ciplnet.com/newsletter/documents/now-online-the-digital-pallas/>
- Sijs, N. van der (2024), 'De twee kanten van het nieuws. Zeventiende-eeuwse kranten in de Republiek en de Spaanse Nederlanden (deel 3)'. In: *Neerlandistiek*, <https://neerlandistiek.nl/2024/10/de-twee-kanten-van-het-nieuws-zeventiende-eeuwse-kranten-in-de-republiek-en-de-spaanse-nederlanden-deel-3/>



- Sijs, N. van der (2024), 'Van "zpel" tot "sich". Zeventiende-eeuwse kranten in de Republiek en de Spaanse Nederlanden (deel 2)'. In: *Neerlandistiek*, <https://neerlandistiek.nl/2024/09/van-zpel-tot-sich-zeventiende-eeuwse-kranten-in-de-republiek-en-de-spaanse-nederlanden-deel-2/>
- Sijs, N. van der (2024), "'Morgen sal men van alles nader hooren.'" Zeventiende-eeuwse kranten in de Republiek en de Spaanse Nederlanden (deel 1)'. In: *Neerlandistiek*, <https://neerlandistiek.nl/2024/09/morgen-sal-men-van-alles-nader-hooren/>
- Sijs, N. van der (2024), 'Vragen over woordenboekgebruik'. In: *Neerlandistiek*, <https://neerlandistiek.nl/2024/12/vragenlijst-over-woordenboekgebruik/>
- Sijs, N. van der (2024), 'Etymologica: klabanus'. In: *Neerlandistiek*, <https://neerlandistiek.nl/2024/12/etymologica-klabanus/>
- Sijs, N. van der (2024), 'Etymologica: jij-bak'. In: *Neerlandistiek*, <https://neerlandistiek.nl/2024/11/etymologica-jij-bak/>
- Sijs, N. van der (2024), 'Etymologica: mozesmandje'. In: *Neerlandistiek*, <https://neerlandistiek.nl/2024/10/etymologica-mozesmandje/>
- Sijs, N. van der (2024), 'Etymologica: reuring'. In: *Neerlandistiek*, <https://neerlandistiek.nl/2024/09/etymologica-reuring/>
- Sijs, N. van der (2024), 'Etymologica: ijsco's en frisco's'. In: *Neerlandistiek*, <https://neerlandistiek.nl/2024/07/etymologica-ijscos-en-friscos/>
- Sijs, N. van der (2024), 'O etymoloog! let op u saeck'. In: *Neerlandistiek*, <https://neerlandistiek.nl/2024/07/o-etymoloog-let-op-u-saeck/>
- Sijs, N. van der (2024), 'Mode in honden- en kattennamen'. In: *Neerlandistiek*, <https://neerlandistiek.nl/2024/05/etymologica-mode-in-honden-en-kattennamen/>
- Sijs, N. van der (2024), 'Waarom maakt een komma het verschil?'. In: *Neerlandistiek*, <https://neerlandistiek.nl/2024/01/etymologica-waarom-maakt-een-komma-het-verschil/>
- Sijs, N. van der (2024), "'Dat kennen we niet"'. Het Vlaams-Russische woordenboek van Emmanuel Waegemans en Vladimir Ronin'. In: *Neerlandistiek*, <https://neerlandistiek.nl/2024/01/dat-kennen-we-niet/>
- Sijs, N. van der (2024), 'Etymologica: butterscotch'. In: *Neerlandistiek*, <https://neerlandistiek.nl/2024/04/etymologica-butterscotch/>
- Sijs, N. van der (2024), 'Etymologica: hoerenjong'. In: *Neerlandistiek*, <https://neerlandistiek.nl/2024/02/etymologica-hoerenjong/>
- Sijs, N. van der (2024), 'Etymologica: Jottem!'. In: *Neerlandistiek*, <https://neerlandistiek.nl/2024/01/etymologica-jottem/>
- Sijs, N. van der (2024), 'The Bold and the Bawdy: How Sexual Slang Became Mainstream in Dutch'. In: *The Low Countries*: <https://www.the-low-countries.com/article/the-bold-and-the-bawdy-how-sexual-slang-became-mainstream-in-dutch>.
- Sijs, N. van der (2024), 'Katenspek en de onbetrouwbaarheid van het taalgeheugen'. In: ENSIE: <https://www.ensie.nl/wetenschap/katenspek-en-de-onbetrouwbaarheid-van-het-taalgeheugen>

- Steurs, F. (2024). *Taal in Transformatie. Technologie, economie en de kracht van generatieve AI*. Schiedam: Scriptum.
- Steurs, F., Kockaert, H. & Alswlaiman, A.A. (Eds.) (2025). *Al-Ġāmi' fī 'Ilm al-Muṣṭalaḥ* = "Handbook of Terminology". Riyadh, Saudi Arabia: King Salman Global Academy for Arabic Language. Retrieved from <http://www.qu.edu.qa/qupress>
- Steurs, F. (2024). 'The Dutch language institute: Towards a comprehensive digital language infrastructure'. In: K. Ahacic & S. Kirchmeier (Eds.), *By the people, for the people. Online dictionaris, Language portals, and the role of language users* (pp. 57-75). Budapest: HUN-REN.
- Tiberius, C. (2024). *Computerlinguïstiek en fraseologie*. Leiden. [inaugural address]. [link](#)
- Tiberius, C. & Colman, L. (2024) Can ChatGPT Help Pattern Editing. In: *Woordcombinaties: A Pilot Study*. In: Štrkalj Despot, K., Ostroški Anić, A. & Brač, I. (red.) *Lexicography and Semantics Book of Abstracts of the XXI EURALEX International Congress 8–12 oktober 2024 Cavtat, Kroatië*, pp. 216-218. [link](#)
- Tiberius, C., Heylen, K., de Does, J., Vanroy, B., Vandeghinste, V. & van Doeselaar, J. (2024). LLMs and Evidence-Based Lexicography: Pilot Studies at INT. In: Krek, S. (red.) *Book of Abstracts of the Workshop Large Language Models and Lexicography*. 8 oktober 2024 Cavtat, Kroatië, pp. 44-48. [link](#)
- Tiberius, C., Ježek, E., Clementelli, A. & Colman, L. (2024). LLMs Automating Corpus Pattern Analysis: A Cross-Lingual Pilot Study for Dutch and Italian. In: Krek, Simon (ed.) *Book of Abstracts of the Workshop Large Language Models and Lexicography*. 8 oktober 2024 Cavtat, Kroatië, pp. 49-52. [link](#)
- De Tier, V. (2024). 'Gezocht voor nieuwgeborene: lokkedijzen, kindjeskak en ander klein spul'. In: *Diejalektgazette Bacht n de Kuupe*, jg. 19, nr. 1, pp. 7-9.
- De Tier, V. (2024). 'Een vogel maakt de lente niet'. In: *Diejalektgazette Bacht n de Kuupe*, jg. 19, nr. 2, pp. 7-9.
- De Tier, V. (2024). 'De Database van de Zuidelijk-Nederlandse Dialecten breidt uit met kankerbloemen en kachtels'. In: *Diejalektgazette Bacht n de Kuupe*, jg. 19, nr.3, pp. 7-9.
- De Tier, V. (2024). 'De steeperege stepert : schoonmaken in Bachten de Kupe'. In: *Diejalektgazette Bacht n de Kuupe*, jg. 19, nr. 4, pp. 7-9.
- De Tier, V. & P. De Nil (2024) 'Waar zitten die Zeeuwse broekeksters?'. In: *Zeeuws Erfgoed*, jg. 23, 1, pp. 34-35. (<https://www.erfgoedzeeland.nl/media/t11nuiur/zeeuws-erfgoed-april-2024.pdf>)
- De Tier, V. (2024) 'Bamisweer en kraaienzomers'. In: *Zeeuws Erfgoed*, jg. 23, 2, pp. 14-15. (<https://www.erfgoedzeeland.nl/media/thhj4d3d/zeeuws-erfgoed-oktober-2024.pdf>)
- De Tier, V. (2024). 'Koffiegrul en horrekoeien op de Zeeuwse kaart. De Database van de Zuidelijk-Nederlandse Dialecten breidt uit'. In: *Nehalennia*, 222, voorjaar 2024, pp. 16-18.
- De Tier, V. (2024). 'Een schort in de Zeeuwse dialecten'. In: *Nehalennia*, 223, najaar 2024, pp. 18-20.
- De Tier, V. (2024). 'Met toeters en trompen: Zeeuwse namen voor het fluitenkruid'. In: *Nehalennia*, 223, winter 2024, pp. 20-22.
- De Tier, V. & M. van Oostendorp (2024). 'Knikkernamen'. In: *Onze Taal*, nr. 1, pp. 14-15.

- De Tier, V. & Oostendorp, M. van (2024). 'Naar het strand'. In: *Onze Taal*, nr. 4, pp. 14-15.
- De Tier, V. & M. van Oostendorp (2024). 'Lekkere feestdagen'. In: *Onze Taal* nr. 6, pp. 18-19.
- De Tier, V. (2024). Elke twee weken een bijdrage aan de webrubriek Uit de streek. Op: [ivdnt.org](http://ivdnt.org)
- Vanroy, B. (2024). Fietje. An open, efficient LLM for Dutch. *arXiv preprint arXiv:2412.15450*.
- Vaičėnionienė, J. et al. [o.a. Vandeghinste, V.] (2024). Transnational Research Infrastructure: A Journey Through CLARIN Knowledge Centres. In: *8th Conference on Digital Humanities in the Nordic and Baltic Countries, Reykjavik, Island, 27 May 2024 - 31 May 2024*. Digital Humanities in the Nordic and Baltic Countries Publications.
- Vandeghinste, V. et al. (2024). Challenges with Sign Language Datasets. In: Andy Way, Lorraine Leesson and Dimitar Shterionov (eds), *Sign Language Machine Translation*. Chapter 5. Springer. pp. 117-139.
- Vandeghinste, V. et al. (2024). Language Resources for European Sign Languages. In: Andy Way Lorraine Leesson and Dimitar Shterionov (eds), *Sign Language Machine Translation*. Springer. pp. 141-170.
- Vandeghinste, V. and Kontino, T. (eds.) (2024). *CLARIN Annual Conference Proceedings 2024*. CLARIN. Barcelona.
- Vandenberghe R. & De Tier, V. (2024). 'Tuin vol wied'. In: *Onze Taal*, nr. 3, pp. 16-17.
- Vos, M. de & Meulen, M. van der (2024). Suppressed no more: prescriptivism and the evaluation of optional variability. In: N. Yáñez-Bouza, M. Rodríguez-Gil & J. Pérez-Guerra (Eds.), *New Horizons in Prescriptivism Research*, pp. 296-318.
- Vos, M. de & Vogl, U. (2024). "Wel iet wat verschelende, maar zó niet óf elck verstaat ander zeer wel". On language making in early modern mono- and multilingual grammars and textbooks in the Low Countries. In: *Belgian Journal of Linguistics* 37, pp. 37-73.
- Waszink, V. en Andree, M. (2024). *Swipen voor een kwarrel, de taal van dating*. Uitgeverij *Onze Taal*.
- Waszink, V. en Andree, M. (2024). 'Swipen voor een kwarrel, Nieuwe benamingen voor relaties van nu'. In: *Onze Taal*.
- Waszink, V. (2024). 'Geen affe verhaaltjes, maar, foto's in spreektaal, Spinvis over zijn liedteksten'. Interview met Spinvis. In: *Onze Taal*.
- Waszink, V. (2024). 'Het jaar 2024 in nieuwe woorden: van 'doemporno' tot 'knaagdierknap'. In: *De Standaard*.
- Zhou, F. & Vandeghinste, V. (2024). Prediction of Translation Techniques for the Translation Process. [arXiv:2403.14454](https://arxiv.org/abs/2403.14454).

## Lezingen en presentaties

- Bonth, Roland de, Presentatie over historisch Nederlands voor studenten Computationale Lexicografie uit Leuven [INT] (15 april 2024).
- Bonth, Roland de, Presentatie Woordenlijst.org tijdens werkvloeroverleg Taalunie-INT (12 juni 2024).

- Bonth, Roland de, Webinar Sneller spellen met Woordenlijst.org! (2 oktober 2024).
- Bonth, Roland de, Presentatie Oud nieuws en oude nieuwtjes (over Couranten Corpus, Brieven als Buit en Gekaapte brieven), voor Orde van den Prince, afdeling Rotterdam (8 oktober 2024).
- Bonth, Roland de, Presentatie Het Woordenboek der Nederlandsche Taal, voor Belgisch-Nederlandse Vereniging Brugge (17 oktober 2024).
- Bonth, Roland de, Webinar Werken met online historische woordenboeken, voor Histories – organisatie voor erfgoedvrijwilligers in Vlaanderen en Brussel (16 december 2024).
- Bouma, Gosse, Odijk, Jan & Tiberius, Carole (2024). 'Towards a Dutch Parseme Corpus' ([abstract](#), [poster](#)), poster gepresenteerd op de tweede algemene bijeenkomst van de UniDive COST-actie, 8-9 februari 2024, Napels.
- De Clercq, Orphee & Vandeghinste, Vincent. 'Begrijpelijke Taaltechnologie'. Studiedag geletterdheid: een blik op de toekomst - AI. Brugge. 13.09.2024. <https://www.west-vlaanderen.be/activiteit/studiedag-geletterdheid-een-blik-op-de-toekomst-ai>
- Colman, Lut, Woordcombinaties. Werkvloeroverleg Taalunie, 12 juni 2024.
- Colman, Lut & Tiberius, Carole 'Idioms in Woordcombinaties: how to serve both lexicography and NLP'. Linguistisches Forschungskolloquium, Leibniz-Institut für Deutsche Sprache (IDS), 3 juli 2024, Mannheim.
- Depuydt, Katrien, 'DBNL als data vanuit computationeel linguïstisch perspectief.' Feest: 25 jaar DBNL, KB, 8 november 2024.
- Depuydt, Katrien, Does, Jesse de, Haga, Thomas, Bonth, Roland de, Brouwer, Tim, Prins, Vincent & Fannee, Mathieu, 'Towards better linguistic annotation for historical Dutch' ([abstract](#), poster). Computational Linguistics in The Netherlands 34, Leiden, 30 augustus 2024.
- Eerten, Laura van en Waszink, Vivien. Frisse Taalquiz, in het kader van Letterlijk Leiden, 3 november 2024.
- Heylen, Kris. Lexical Variation and Change. 'A Distributional Semantic Approach. Gastseminarie voor promovendi, 15 mei 2024, University of Verona Doctoral School in Linguistics.
- Heylen, Kris, Jesse de Does, Mathieu Fannee, Ruud de Jong. "TermWerk: Een geïntegreerde webapplicatie voor projectgebaseerde termextractie en termbeheer in het Nederlands'. Demonstratie op de TiNT-dag, 8 november 2024, Breda.
- Tanmay Khokle, Tim Van de Cruys, Kris Heylen. 'Multilingual Definition Modeling for Neologisms'. Poster op de 34th Meeting of Computational Linguistics in The Netherlands (CLIN 34), 30 augustus, Leiden.
- Landsbergen, Frank, *Over zoeken en vinden in de digitale Algemene Nederlandse Spraakkunst*. Presentatie voor de Vertalersvakschool, Amsterdam.
- JMathysen, Jolien & Vandeghinste, Vincent. 'Een Corpusgebaseerde Woordenschatlijst Gesproken Academisch Belgisch-Nederlands'. Nederlandse Vereniging voor Toegepaste Taalwetenschap (Anela). 6-7.07.2024. Gent. <https://anela.nl/activiteiten/anelaconferentie/>.
- Niestadt, Jan en Does, Jesse de., 'BlackLab for syntax and parallel corpora' (poster en demonstratie). Clariah slotconferentie, Leiden, 23 juni 2024.

- Haga, Thomas en Bonth, Roland de, 'LAnCeLoT' (poster en demonstratie). Clariah slotconferentie, Leiden, 23 juni 2024.
- Hartog den, M., & de Vos, Machteld. Presentatie '*Ick en kan U.E. niet na-laten te adverteren de gheluckighe aenkomste van de Armade van den Koninck van Vranckrijck*': forms of address in 17<sup>th</sup>-century Dutch newspapers. HiSoN 2024, Zürich, Zwitserland. (7 juni 2024).
- Hartog den, M., Faber-van de Kraats, Iris, Opheij, E., & de Vos, Machteld, Presentatie *Aanspreekvormen in zeventiende-eeuwse Nederlandse kranten*. Grote Taaldag 2024, Utrecht. (2 februari 2024).
- Prins, Vincent, 'GaLAHaD' (poster en demonstratie). Clariah slotconferentie, Leiden, 23 juni 2024.
- Rys, Kathy, Een verkenning van het dialect van Maldegem en zijn deelgemeenten en een uitstapje over de grens. Presentatie op 27 november 2024 voor het dialectgenootschap Goedsemoeds, te Maldegem.
- Segonne, Vincent et al. [o.a. Vandeghinste, Vincent, 'Jargon: A Suite of Language Models and Evaluation Tasks for French Specialized Domains.' LREC-CoLing, Turijn.
- Sijs, Nicoline van der. 'Daar is geen woord Frans', Deventer Literair, 15 december 2024.
- Sijs, Nicoline van der. 'Het taalgebruik van Utenhove, Dathenus en Marnix van Sint-Aldegonde', Symposium over de dynamiek van de gereformeerde Liturgie, 7 december 2024.
- Sijs, Nicoline van der. 'Kranten ontsluiten. Ervaringen en resultaten', vrijwilligersmiddag 15 maart.
- Sijs, Nicoline van der. 'Ontstaan van het Standaardnederlands', 20 februari, Rotary Club Den Haag.
- Soete, Florelie & Vandeghinste, Vincent. [Optimizing Controllable Sentence Simplification for Dutch using T5 Large Language Model](#). CLIN. Leiden, 30 augustus 2024.
- Steurs, Frieda. 'Het INT, schatkamer van de Nederlandse Taal'. lezing voor Probus Lisse, 28 februari 2024.
- Steurs, Frieda. 'Taal, economie en technologie. Over het belang van het Nederlands in de internationale samenleving'. Lezing voor online congres: DutchIt! (congres vanuit Universiteit Lublin), 15 maart 2024.
- Steurs, Frieda. Opname webinar 'AI en MT: taalberoepen 2.0'. 4 april 2024.
- Steurs, Frieda. 'Taalberoepen 2.0. Onbekend is onbemind'. Lezing in het Huis van Europa, 14 april 2024.
- Steurs, Frieda. 'Terminologie pour traducteurs et interprètes: des fiches classiques à la boîte à outils terminologiques'. Congres over 'la didactique de la terminologie' Universiteit van Thessaloniki, 28 mei 2024.
- Steurs, Frieda. Lezing over Language is business. The value of language in a multilingual world. Universiteit Leiden, 5 juni 2024.
- Steurs, Frieda. 'Intelligente technologie: hoe staat het met Nederlands?'. Lezing KANTL (Gent), 21 juni 2024.
- Steurs, Frieda. 'De rol van het Instituut voor de Nederlandse Taal in het aanreiken van bronnen voor vertalers'. Lezing bij het Colloquium Neerlandicum Olomouc, 1 juni 2024.

- Steurs, Frieda. Lezing voor de Nijmeegse onderzoekers, 30 september 2024.
- Steurs, Frieda. 'Het Expertisecentrum Terminologie - INT Schatkamer van de Nederlandse Taal'.  
Lezing bij de TiNT-dag in Breda, 8 november 2024.
- Arhar Holdt, Š. & Tiberius, Carole, Challenges in data preparation and manual corpus annotation.  
Presentatie in CLARIN Café - Creating pedagogical corpora with annotation of sensitive  
content and offensive language - the CrowLL project. 4 april 2024, online.
- Tiberius, Carole, & Heylen, Kris, 'LLMs and evidence-based lexicography'. Workshop „Generative AI  
in lexicography / Generative KI in der Lexikographie“ op de ZDL-Jahrestagung 2024, Berlin-  
Brandenburgische Akademie der Wissenschaften, 16 mei 2024, Berlijn.
- Tiberius, Carole & Colman, Lut, Corpus Pattern Analysis for Dutch in Woordcombinaties. Phrasalex  
III, 10-12 juli, Innsbruck/hybrid.
- Tiberius, Carole, Čibej, J., Kallas, J., Muischnek, K. & Krek, S. (2024). 'UD syntax for the ELEXIS parallel  
sense-annotated corpus: a pilot study' - ([abstract](#), [poster](#)) poster gepresenteerd op de tweede  
algemene bijeenkomst van de UniDive COST-actie, 8-9 februari 2024, Napels.
- Tier, Veronique De, Katrien Depuydt, Jesse de Does & Koen Mertens (lezing). Op zoek naar mijn  
dialect in de database van de Zuidelijk-Nederlandse dialecten. Grote Taaldag Utrecht, 2  
februari 2024.
- Tier, Veronique De, Katrien Depuydt, Jesse de Does & Koen Mertens (poster) Zeeuwse stramienen in  
de Database van de Zuidelijk-Nederlandse dialecten (DSDD): integratie van een alfabetisch  
dialectwoordenboek in een conceptgebaseerde database. Sociolinguistic Circle, Groningen, 19  
april 2024.
- Tier, Veronique De. Ronse en 't Ronssies. Lezing bij de presentatie van een streektaal-cd. 11 mei  
2024, Ronse.
- Tier, Veronique De. Online lezing. Het Brabants in de DSDD. Brabanders en hun taal. Online. 15 mei  
2024.
- Tier, Veronique De, Katrien Depuydt en Jesse de Does. (poster) Dialect Data Integration: Paving the  
Way for New Research. 21st International Congress of Linguists (ICL). 8 to 14 September 2024,  
Poznań.
- Marta Dijkhoff & Ellen Meulens (Nationaal Taalinstituut Curaçao), Jesse de Does & Katrien Depuydt  
(Instituut voor de Nederlandse Taal) - *Een digitale lexicale infrastructuur voor het Papiaments:  
presentatie van de eerste resultaten* & Katrien Depuydt en Veronique De Tier (Instituut voor  
de Nederlandse Taal) - *Een digitale lexicale infrastructuur voor dialecten en streektalen:  
presentatie van de eerste resultaten*. Streektaalconferentie 20 september 2024. Roermond.
- Tier, Veronique De & Doreleijers, K. Workshop Hier hoor ik thuis, Workshop Taalfeest Antwerpen, 5  
oktober 2024.
- Tier, Veronique De, Katrien Depuydt en Mathieu Fannee. Developing an infrastructure for dialect  
enthusiasts: the pilot for the Bildts dialect. Lezing XXI EURALEX International Congress. 8–12  
oktober 2024.

- Tier, Veronique De, Volkse dialectuitdrukkingen. Wat vertellen ze ons? Anzegem. Bij de presentatie van Zeisels. Boek met dialectuitdrukkingen. 18 oktober 2024.
- Tier, Veronique De, Diverse lezingen over Dialecten in Vlaanderen. Waar vind ik wat in de DSDD. (12 januari, Assenede; 19 januari, Zwalm; 21 februari, Tienen; 10 juli, Izegem; 11 juli en 21 november, Biervliet; Oostakker, 23 september; Kruikebeke, 7 november; Sint-Gillis-Dendermonde, 15 november; Gent, 17 december.
- Vandeghinste, Vincent, Job van Doeselaar & Bram Vanroy. [Human Evaluation of Automated Text Simplification through Crowdsourcing](#). CLIN. Leiden, 30 augustus 2024.
- Vos de, Machteld, Presentatie & panellid (op uitnodiging) *Using Blacklab and Historical Annotation tools for studying (the use of strong verbs in) 17th-century Dutch newspaper language*. CLARIAH Conference 2024, Leiden. (13 juni 2024)
- Vos de, Machteld, Presentatie *Intertextuality and translations in 17<sup>th</sup>-century Dutch newspapers: a study into early modern newspaper editing practices*. CHINED IX “Historical News Discourse in a Comparative Perspective”, Poznań, Polen. (23 mei 2024)
- Vos de, Machteld & Van de Velde, F. Presentatie *Old news, new tidings: German influence on 17<sup>th</sup>-century Dutch newspaper language*. Internationale Tagung “Historische Mehrsprachigkeit und Sprachkontakt mit Varietäten des Deutschen: Überlieferung, Forschungsstand und Perspektiven”, Heidelberg, Duitsland. (25 april 2024)
- Vos de, Machteld, Workshop (op uitnodiging) *Spread the new(s): krantentaal vroeger en nu*. Neerlandistiekdagen 2024 ‘Festival der Zinnen’, Universiteit van Amsterdam. (5 april 2024)
- Vos de, Machteld, Presentatie *Spread the new(s): an overview of my PhD project so far*. T4D, onderzoeksgroep CLS, Radboud Universiteit Nijmegen. (20 februari 2024).
- Vos de, Machteld, Presentatie *Een bodem gelegd. Een hoopvol fundament voor de toekomst*. Op het symposium *Taal tot op de bodem* voor het afscheid van prof. dr. Nicoline van der Sijs’ van het INT, Leiden. (26 januari 2024).
- Waszink, Vivien. Diverse lezingen over inclusief taalgebruik, o.a. bij Orde van den Prince, Rijksmuseum, Ministerie voor Sociale Zaken en Werkgelegenheid en EU-vertalers Brussel.
- Waszink, Vivien. Diverse lezingen over de taal van mode en kleding, o.a. bij Bibliotheek Rotterdam.
- Waszink, Vivien. ‘Wat doe je allemaal als taalkundige?’ Lezing voor studievereniging Awater, Universiteit Utrecht, 6 maart 2024.
- Waszink, Vivien. ‘Taalkwesties Nieuw Nederlands’. Lezing op docentendag Noordhoff, 21 maart 2024.
- Waszink, Vivien en Boukje Verheij. ‘ANW en WNW’. Lezing voor studenten Universiteit Leuven, 15 april 2024.
- Waszink, Vivien. ‘Knaagdierknop, voorkauwgeneratie en waaierdenken, nieuwe woorden zijn er altijd’. Lezing over neologismen bij Ministerie voor Sociale Zaken en Werkgelegenheid, 25 juni 2024.

Waszink, Vivien en Alex Reuneker. 'Van kinderlied tot flexe beat: analyse van lexicale diversiteit en keywords in Nederlandstalige popmuziek'. Workshop zomerschool Taalunie, Universiteit Leiden, 20 augustus 2024.

Waszink, Vivien. 'Trauma's en triggers: allemaal heel toxic. De opkomst van therapietaal'. Bij: Boom Uitgevers, 30 september 2024.

Waszink, Vivien. 'Nieuwe woorden in het Nederlands', lezing in het kader van de webinarreeks 'Werken met online taalbronnen' georganiseerd door het INT in samenwerking met de Taalunie, 2 oktober 2024.

Waszink, Vivien en Karen Deschamps. 'Dat mag je óók (al niet meer) zeggen. Inclusief taalgebruik'. Lezing Taalfeest Antwerpen, 5 oktober 2024.

Zenner, Eline, Steffen Höder & Kris Heylen. 'Constructional borrowing from English.' Workshop on Anglicism research in Europe: from vocabulary to use, 57th Annual Meeting of the Societas Linguistica Europaea (SLE 57), University of Helsinki, 21–24 Augustus 2024.

## Organisatie van congressen en workshops

Eerten, Laura van. Organisatie webinars 'Nieuwe woorden in het Nederlands', Sneller spellen met [Woordenlijst.org](https://woordenlijst.org) en 'Vertaaltools voor de internationale neerlandistiek'.

Vandeghinste, Vincent. CLARIN Annual Conference: Programme Committee Chair.

## Onderwijs

Heylen, Kris en Vincent Vandeghinste, Collegereeks Computatieve lexicografie (B-KUL-FOSV1A) binnen de Master Taalkunde aan de KU Leuven: februari-mei 2024.

Heylen, Kris en Vincent Vandeghinse, Collegereeks Lexicologie en Morfologie (B-KUL- F0VG6A) binnen de Master Taalkunde aan de KU Leuven: september-december 2024.

Sijs, Noline van der (2024), 'Nederlandse leenwoorden en uitleenwoorden', gastcollege in de collegereeks Nederlands in contact voor bachelorstudenten Taal- en Letterkunde te Brussel, 22 maart.

Tiberius, Carole. MA-collegereeks Corpus Lexicography. Universiteit Leiden, februari-mei 2024.

Van pellicom, Katrien. Twee lessen in het kader van de *Europese Dag van de Talen* over het INT, de Taalunie en Nederlands in de wereld. OLVP Bornem Secundair Onderwijs, 26 september 2024.

Vandeghinste, Vincent. 'Text-to-pictograph conversion and vice versa.' Language Engineering Applications. Master of Artificial Intelligence. KU Leuven. 15 april 2024.

Vandeghinste, Vincent. 'Language Technology Infrastructures.' Language Engineering Applications. Master of Artificial Intelligence. KU Leuven. 22 april 2024.

Vos de, Machteld, Collegereeks De geschiedenis en toekomst van het Nederlands binnen de Master Leraar Nederlands aan de HAN University of Applied Sciences: september-oktober 2024.



Vos de, Machteld, Collegereeks Taalverandering in context binnen de Bachelor Nederlandse Taal en Cultuur aan de Radboud Universiteit Nijmegen (jaar 2): januari-juli 2024.

Vos de, Machteld, Bachelorscriptiebegeleiding binnen de Bachelor Taal- en Letterkunde aan de Universiteit Gent: december-juni 2024.

## In de media

Instituut voor de Nederlandse Taal. 'Alle kennis van het Nederlands in één instituut'. Artikel in de *lage landen*, mei 2024

Sijs, Nicoline van der. 'Daar is geen woord Frans bij', *De Taalstaat* NPO Radio 1, 6 januari 2024

Sijs, Nicoline van der. Hoe het plat-Amsterdams bijna kassiewijle gaat'. *Het Parool*, 16 januari 2024

Sijs, Nicoline van der. Goedemorgen Nederland over huisdiernamen, 30 mei 2024

Sijs, Nicoline van der. 'Hoe noemden onze voorouders hun huisdier?'. *Nieuwe Feiten* Radio 1 BE, 12 juni 2024

Sijs, Nicoline van der. 'Etymologie', *Radio Centraal* Babelonië, 27 juni 2024

Sijs, Nicoline van der. 'Wat heeft 'ijsbergsla' met een ijsberg te maken? In: *Nieuwe Feiten*, Radio 1 BE, 13 maart 2024

Sijs, Nicoline van der. Interview over (on)modieuze dierenamen in de *Telegraaf*, 30 mei 2024

Sijs, Nicoline van der. 'Naamval eruit, lidwoord erin'. Interview wetenschapsbijlage *de Volkskrant* 24 augustus 2024

Steurs, Frieda. Interview over het boek *Taal in transformatie*, in: *De Taalstaat* NPO Radio 1, 30 november 2024

Tier, Veronique De. Interview over de DSDD, op: *Dit is Helmond*, 24 april 2024

Waszink, Vivien. 'Vliedschaamte is de moeder van skischaamte'. Interview over woord skischaamte, in: *Leidsch Dagblad*, 17 februari 2024

Waszink, Vivien. 'Wetenschappers na discussie over taalgebruik Maurice Steijn: 'Wijf is niet eens een ontzettend grof woord''. Interview over het woord *wijf*, in: *Telegraaf*, 11 maart 2024

Waszink, Vivien. 'Tienertaaltrends, komen Kluun en Yvanka nog mee?'. Te gast in de podcast 'Help ik heb een puber' van Kluun, 12 april 2024

Waszink, Vivien. 'Iedereen heeft autisme, OCD, ADHD of een angststoornis: praten we onszelf een diagnose aan?'. Interview over therapietaal, in: *De Standaard*, 4 mei 2024

Waszink, Vivien. Radio-interview over afko's in jongeren- en studententaal, NPO radio, 3m, Wijnand in de Ochtend, 29 mei 2024

Waszink, Vivien. Diverse radio-interviews over boek *Swipen voor een kwarrel*, *De taal van dating*, o.a. bij *Nieuwe Feiten*, *Radio 538*, *Ruud de Wild* en *De Nieuws BV*

Waszink, Vivien. Diverse kranteninterviews over *Swipen voor een kwarrel*, *De taal van dating*, o.a. in *Trouw*, *Algemeen Dagblad*, *Het Nieuwsblad* en *Het Laatste Nieuws*

Waszink, Vivien. 'Lijden we massaal aan psychologische hypochondrie? 'Soms stress of angst voelen is heel normaal''. Interview over therapietaal., in: *Algemeen Dagblad*, 14 augustus 2024

Waszink, Vivien. 'Taal is een spel'. Dubbelinterview met Wouter Deprez over taal en Het Taalfeest, in: *Lage Landen*, september 2024

Waszink, Vivien. 'Skibidi, slay en sigma: waar komen nieuwe woorden vandaan?'. Interview over kinderwoord van het jaar, in: podcast NOS Jeugdjournaal, 4 december 2024

Waszink, Vivien. 'Een ode aan jeugdtaal'. Interview over kinderwoord van het jaar, in: podcast NOS De Dag, 17 december 2024

Waszink, Vivien. Radio-Interview met Gerard Ekdorf over kinderwoord van het jaar. Radio Veronica, 19 december 2024

Vandeghinste, Vincent (2024). CLARIN Annual Conference 2024: report from the CLARIAH-VL representatives. *CLARIAH-VL*. <https://clariahvl.hypotheses.org/2268>

Vandeghinste, Vincent (2024). Research Scenario: Spoken Academic Belgian Dutch (SABeD). *CLARIAH-VL*. <https://doi.org/10.58079/12rhq>

## Diversen

Maaïke Beliën

- Lid leescommissie doctoraat Arne Dhondt, *Pluricentrisme in de praktijk: Naar een methode voor de codificatie van grammaticale verschillen tussen Belgisch Nederlands en Nederlands Nederlands*, Universiteit Gent.

Roland de Bonth

- Jurylid Kruyskampprijs 2024 van de Maatschappij der Nederlandse Letterkunde.
- Auteur en juryvoorzitter van het Groot Leiderdorps Dictee (16 oktober 2024).
- Coördinator en jurylid Profielwerkstukwedstrijdprijs van Levende Talen Nederlands 2024.
- Redacteur *Trefwoord*, het digitale tijdschrift voor lexicografie en lexicologie.
- Redacteur Taalbeschouwing van het Handboek Didactiek Nederlands.
- Kolomleider literatuuronderwijs HSN 36, 22 & 23 november 2024, Erasmushogeschool Brussel.
- Auteur opdracht Nederlands voor de Taalkundeolympiade 2024 van de Universiteit Leiden.
- INT-posters voor het voortgezet onderwijs ontwikkeld en herzien.

Katrien Depuydt

- Vertegenwoordiger van de voorzitter van de executive board in de managementboard van het IMPACT Centre of Competence.
- Lid van de commissie die de werkzaamheden van het KNAW Humanities Cluster voor de periode 2018-2023 heeft geëvalueerd.

Frank Landsbergen

- Begeleiding stagiair Julia Schoppers (DAGENTA/e-ANS), samen met Nicoline van der Sijs.

- Review voor online handboek Levende Talen Nederlands.

Nicoline van der Sijs

- Jurylid Olympiade Nederlands.

Carole Tiberius

- Board of Trustees Adam Kilgarriff Prize.
- National Anchor Point LDS (European Language Data Space).
- Secretary-Treasurer EURALEX (European Association of Lexicography).
- UniDive COST-actie (CA21167): Management Committee Member voor Nederland.
- Begeleiding stagiaars Marit Janssen (PARSEME\_NL corpus); Iris Krips (Vertaalwoordenschat, TaboeLex).
- Lid oppositiecommissie doctoraar M. Fragkiadakis. Digital Tools for Sign Language Research: Towards Recognition and Comparison of Lexical Signs. Universiteit Leiden. 9 april 2024.

Vincent Vandeghinste

- Panellid 'Paving the way for a SSH Open Science Cloud for Flanders' CLARIAH-VL final event. Brussel.
- Nationaal coördinator CLARIN België, lid van National Coordinators Forum CLARIN.
- Lid van Knowledge Infrastructure Committee CLARIN.
- PhD promotor en begeleiding. Magali Norré. 'Traduction automatique du français vers les pictogrammes pour améliorer la communication entre médecins et patients avec une déficience intellectuelle en milieu hospitalier.' Universiteiten van Louvain-la-Neuve en Geneve.
- PhD jurylid. Mathieu De Coster (2024). 'Neural Sign Language Recognition and Translation: Towards Sign Language Applications Powered by Artificial Intelligence.'
- Jurylid van NWO Open competitie Humane en sociale wetenschappen.
- Redactielid Computational Linguistics in the Netherlands Journal.

Machteld de Vos

- Lid van de visitatiecommissie van de Fryske Akademy.

Vivien Waszink

- Auteur lesmethode Nieuw Nederlands, Noordhoff, methodes vmbo 3 gt, 4 havo, 5 havo, 4 vwo, 5 vwo en 6 vwo.
- Lid Taalexplorium, een samenwerkingsverband van diverse organisaties die educatieve activiteiten rond taal organiseren in o.a. bibliotheken.

## Bijlage 4: Taalmaterialen

### Downloads van taalmaterialen

Deze downloadstatistieken bevatten geen downloads door systeembeheer, webmaster en taalmaterialenmedewerkers. Dubbele downloads (van eenzelfde account) worden slechts eenmaal meegeteld. Downloads van bestelpakketten die enkel een licentie bevatten werden eveneens niet meegeteld.

<b>Materialen commercieel</b>	<b>Totaal</b>
BasiLex-corpora Commercieel	5
Corpus Gesproken Nederlands (CGN) Commercieel	4
JASMIN-spraakcorpus Commercieel	3
Referentiebestand Nederlands (RBN) Commercieel	3
BasiLex-lexicon Commercieel	2
BasiScript-corpora Commercieel	2
BasiScript-lexicon Commercieel	2
CHN N-grams Commercieel	2
Dutch Idiom Database: Native Speakers (DID-NS) Commercieel	2
Frequentielijsten corpora Commercieel	2
GiGANT-Molex Commercieel	2
Lassy Groot-corpora Commercieel	2
Referentiebestand Belgisch-Nederlands (RBBN) Commercieel	2
SoNaR Groot-corpora Commercieel	2
SoNaR Klein-corpora Commercieel	2
SumNL-samenvattingencorpora Commercieel	2
DuELME Commercieel	1
DuOMAn Subjectivity Lexicon Commercieel	1
Dutch Parallel Corpus (DPC) Commercieel	1
e-Lex Commercieel	1
Medische Termen Belgisch-Nederlands (MedTermBN) Commercieel	1
<b>TOTAAL</b>	<b>44</b>

<b>Materialen niet-commercieel</b>	<b>Totaal</b>
Corpus Gesproken Nederlands (CGN)	164
SoNaR-corpora	72
CELEX-2 Dutch	67
CGN-annotaties	54
Frequentielijsten corpora	46
CHN N-grams	36

Lassy Klein-corpus	35
GiGaNT-Molex	32
Cd-rom Middelnederlands	27
Dutch Parallel Corpus (DPC)	27
Corpus Pathologische en Normale Spraak (COPAS)	25
Spoken Academic Belgian Dutch Corpus (SABeD)	23
AI-Trainingset - Tag de Tekst voor Named Entity Recognition (NER)	22
Corpus Middelnederlands (Data)	22
SoNaR Nieuwe Media Corpus	22
DuOMAn Subjectivity Lexicon	21
Referentiebestand Nederlands (RBN)	21
BLISS Spoken Dialogue Dataset	17
INT Historische Woordenlijst	17
Lassy Groot-corpus	17
BLISS Dialogue Summaries	16
Corpus Gysseling (Data)	13
Eindhoven Corpus	13
PAROLE-lexicon	13
Wablieft-corpus	12
CombiLex	11
IFA Corpus	11
RND Woordenlijsten	11
JASMIN-spraakcorpus	10
Hotel Review Corpus in Nederlandse Gebarentaal (NGT_HoReCo)	10
Corpus Ondertitelde UvN-colleges (COUC)	9
Dutch Idiom Database: Native Speakers (DID-NS)	9
e-Lex	9
Medische Termen Belgisch-Nederlands (MedTermBN)	9
AUTONOMATA-namencorpus	8
IFA Dialogue Video corpus	8
Medische Pilot (MedPilot)	8
WAI-NOT Corpus	8
BasiLex-lexicon	7
Belgian Covid Sign Language Corpus (BeCoS Corpus)	7
DuELME	7
NAMES Corpus	7
SumNL-samenvattingencorpus	7
BasiLex-corpus	6
Dataset Synthetische Simplificatie	6
Hoger Onderwijs Terminologie in Nederland en Vlaanderen (HOTNeV)	6
Menselijke evaluatie van geautomatiseerde tekstvereenvoudiging: resultaten van crowdsourcing	6

SoNaR Character N-grams	6
Woordenboek Vlaamse Gebarentaal (Woordenboek VGT)	6
Referentiebestand Belgisch-Nederlands (RBBN)	5
AUTONOMATA-POI-corpus	4
COREA-coreferentiecorpus	4
The LiLaH Emotion Lexicon of Greek, Kurdish, Turkish, Spanish, Farsi and Chinese	4
Brieven als Buit - Gouden Standaard	3
Corpus Vlaamse Gebarentaal (Corpus VGT)	3
Dupira	3
Gold Standard Parallel Corpus of Sign and spoken Language (GoSt-ParC-Sign)	3
Hotel Review Corpus in Spanish Sign Language (LSE_HoReCo)	3
Moroccorp	3
Paco-MT Parallele Corpora	3
BasiScript-lexicon	2
Children's Oral Reading Corpus (CHOREC)	2
AUTONOMATA-transcriptietoolset	2
Hotel Review Corpus in Vlaamse Gebarentaal (VGT_HoReCo)	2
Meertalige Ondertiteldata 2BDutch	1
Afrikaans Custom Dictionary for Government Domain	1
AUTONOMATA-POI-demo	1
Autshumato English-isiZulu Parallel Corpora	1
DAESO-corpus: parallele Nederlandstalige monolinguale treebank	1
D-TUNA-corpus	1
Lwazi Afrikaans ASR Corpus	1
Lwazi isiXhosa ASR Corpus	1
Lwazi Setswana ASR Corpus	1
Lwazi Setswana Pronunciation Dictionary	1
Lwazi Siswati ASR Corpus	1
Lwazi Tshivenda Pronunciation Dictionary	1
Lwazi Xitsonga ASR Corpus	1
Lwazi Xitsonga Pronunciation Dictionary	1
MuST-Cinema-PE: post-editing in automatic subtitling	1
OMBI Nederlands-Arabisch	1
OMBI Nederlands-Deens	1
OMBI Nederlands-Indonesisch	1
Siswati Genre Classification Corpus	1
<b>TOTAAL</b>	<b>1091</b>