

/instituut voor de
Nederlandse taal/

jaarverslag

2021

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

(Woordenboek der Nederlandsche Taal)

Jaarverslag 2021

Instituut voor de Nederlandse Taal

Inhoud

Inhoud

Algemene doelstellingen INT

A. Lexicografie en Terminologie

A.1. Synchron: ANW

A.2. Synchron: Woordcombinaties

A.3. Synchron: Neologismen

A.4. Diachron: Historische woordenboeken en de GTB

A.5. Dialectologie

A.5.1. Database of the Southern Dutch Dialects (DSDD)

A.5.2. Gesproken Corpus van de Zuidelijk-Nederlandse Dialecten (GCND)

A.5.3. Atlas van het dialect in Vlaanderen

A.5.4. Overige activiteiten

A.6. Vertaalwoordenschat

A.7. Expertisecentrum Terminologie

A.7.1. Website

A.7.2. Projecten voor specifieke vakdomeinen

HOTNeV

Nederlands als wetenschapstaal

Medische terminologie

Juridische terminologie

Terminologietools

Terminologisch netwerk

A.8. Lexicon Frisicum

B. Corpora en lexica

B.1. Moderne corpora

B.2. Historische corpora

B.3. Spelling

B.3.1. Project Spelling-API

B.4. GiGaNT

B.5. Taalbank: semantisch lexicon DiaMaNT

C. Grammatica

C.1. Grammatica algemeen

C.2. Taaladvies

C.3. e-ANS

C.4. Taalportaal

D. Softwaretools (ontwikkeling, beheer en onderhoud)

D.1. INT Impactcentrum en digitization.eu

Specifieke projecten

D.2.1. CLARIAH+ Nederland

D.2.2. CLARIAH Vlaanderen

D.3. European Language Resources Coordination Initiative (ELRC)

E. Taalmaterialen

E.1. Website taalmaterialen

E.1.1. Nieuwe producten

E.1.2. Geografische verspreiding van gebruikers

E.2. CLARIN-ERIC: het INT als CLARIN-centrum

F. Onderzoek: Europese aanvragen en competitieve projecten

F.1. European Lexicographic Infrastructure (ELEXIS)

F.2. COST CA16105: European Network for Combining Language Learning with Crowdsourcing Techniques (enetCollect)

F.3. COST CA18209: European network for Web-centered linguistic datascience

F.4. H2020 - RIA 101017255 SignON: Sign Language Translation Mobile Application and Open Communications Framework

F.5. KULeuven C3: SABeD – Spoken Academic Belgian Dutch

F.6. NWO Open Competition SSH Spread the new(s)

F.7. The rise of complex verb constructions in Germanic

F.8. MacBERTh

F.9. European Language Grid

F.10. European Language Equality

G. Doelgroepenbeleid (inclusief onderwijs)

G.1. Onderzoek, onderwijs en het algemene publiek

H. Wetenschapscommunicatie en publieksbereik

H.1. Website

H.2. Nieuwsbrieven

H.3. Sociale media

H.4. Podcasts

H.5. Boek

H.6. Neerlandistiek

Bijlage 1: Raad van Toezicht en Raad van Advies

Bijlage 2: Medewerkers

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

Bijlage 4: Downloadoverzicht taalmaterialen

Algemene doelstellingen INT

Na de hervormingen van 2016 is het Instituut voor de Nederlandse Taal (INT) voluit aan de slag gegaan met alle taken die het instituut moet uitvoeren. Het meerjarenbeleidsplan (2018-2022) werd uitgevoerd, in 2021 weer met een aantal nieuwe accenten en dit in nauw overleg met de Nederlandse Taalunie (NTU). Het werkvloeroverleg (projectgericht) en het directieoverleg tussen de Taalunie en het INT werpen hun vruchten af en dit leidt tot duidelijkheid en prioritering van een aantal projecten en taken. Dit vertaalt zich ook in een heldere planning van projecten en in een begroting die deze taken duidelijk laat zien.

Bij de aanvang van 2021 werkten er 28,60 fte (37 werknemers). Eind december 2021 waren dat er 28,65 fte (35 werknemers). Er waren enkele verschuivingen door pensionering van drie werknemers, van één medewerker liep het contract af, van een andere werd de opdracht van 60 naar 100% uitgebreid en er kwamen een nieuwe webmaster (60%) en een softwareontwikkelaar (100%).

In onze werking werd aandacht besteed aan de specifieke gebieden die de Taalunie benoemt, met name: 1. Standaardtaal, 2. Nederlands, taalvariëteiten en andere talen, 3. Onderwijs Nederlands binnen het taalgebied, 4. Onderwijs Nederlands buiten het taalgebied en 5. Taal en Cultuur.

De missie van het INT staat daarbij centraal: het INT neemt een centrale positie in voor het hele Nederlandse taalgebied (Nederland, Vlaanderen, Suriname en de voormalige Antillen) op het vlak van het wetenschappelijk verantwoord ontwikkelen, bewaren en duurzaam beschikbaar stellen van taalmateriaal.

Het INT streeft ernaar om het best gesorteerde en daarmee een zeer goed, toegankelijk wetenschappelijk instituut te zijn op het gebied van de Nederlandse taal en de woordenschat. Het instituut ontwikkelt en levert data voor woordenboeken, (computationele) lexica, corpora en tools. De woordenboeken zijn online te raadplegen. Software en computerlinguïstische tools zijn open source beschikbaar.

Het instituut speelt in op de nieuwe ontwikkelingen in de geesteswetenschappen, met name op het terrein van de *digital humanities*. Om deze rol te kunnen vervullen beheert en onderhoudt het INT een digitale infrastructuur voor het Nederlands, met aandacht voor taalvariatie (terminologie, dialecten etc.). Zowel academische als niet-academische partijen kunnen gebruikmaken van deze infrastructuur. In Vlaanderen boekten we extra succes door de realisatie van Clarin.be, waarbij het INT het Clarin B-centrum werd voor België, en we werden ook erkend als Clarin K-centrum.

Er stonden in 2021 een heleboel activiteiten op het programma. Er werd vooral ingezet op het verbeteren van de digitale infrastructuur van het Nederlands, zoals in het beleidsplan werd aangekondigd. Dit bouwt voort op de inzichten van de Lorentzworkshop, waardoor een optimale

structuur voor de moderne taaldata wordt gerealiseerd. Het jaar stond ook in het teken van de visitatie. Na vijf jaar intensief werk om het INT op de rails te zetten met een breed palet aan taken, was het tijd voor een internationale audit. We schreven een zelfevaluatierapport en in augustus bezocht een internationale commissie van experts het INT. De commissie bleek opgetogen over onze werkzaamheden en realisaties. 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.

Tegelijk werden in 2021 vele van onze projecten online gelanceerd en kwam er een belangrijke update van de materialen. Zo was er de lancering van een nieuwe versie van het corpus Brieven als Buit en kwam een belangrijke uitbreiding van het Corpus Hedendaags Nederlands online. Het CHN wordt door vele onderzoekers gebruikt en de nieuwe release is een sterk uitgebreide versie van de twee eerdere versies. Op dit moment bevat het corpus al meer dan 2,5 miljoen teksten uit onder meer boeken, blogs, kranten, tijdschriften en journaaluitzendingen. Samen zijn deze teksten goed voor meer dan 1 miljard woorden. Vanaf nu zal het CHN elke maand een update krijgen.

Op het vlak van de grammatica werden drie belangrijke doelstellingen gehaald: in april werd de vernieuwde Algemene Nederlandse Spraakkunst (ANS) gelanceerd en in juli werd daar de Algemene Nederlandse Klankleer (ANK) aan toegevoegd. Ook binnen het grammaticale spectrum werd in mei het project Woordcombinaties gelanceerd, dat zich vooral richt op taalleerders van het Nederlands. Al deze projecten werden in webinars gepresenteerd door en aan verschillende experts en gebruikers.

2021 werd ook het jaar waarin de Universiteit Utrecht de leiding in drie Europese samenwerkingsverbanden overdroeg aan het INT. We werden daardoor het National Competence Centre (NCC). Dat was het geval in het ELG-project (European Language Grid) en we organiseerden daarvoor een workshop op 3 december. Eveneens namen we de leiding in het ELRC-consortium (European Language Resources Coordination) voor Nederland en organiseerden we op 11 juni in dit kader een workshop. Ten derde werden we verantwoordelijk voor het ELE-project (European Language Equality) en deden we een bevraging naar de stand van zaken van de digitale taalmaterialen en softwaretools voor het Nederlands in Nederland en Vlaanderen. De resultaten van dit onderzoek zullen in 2022 worden gepubliceerd in een nieuwe whitepaper ‘The Dutch Language in the Digital Age’.

In samenwerking met de Taalunie en de taaladviesdiensten (Onze Taal in Nederland en de Taaltelefoon in Vlaanderen) werd de website Taaladvies.net aan het INT overgedragen. De geheel vernieuwde website werd in september gelanceerd.

Ook in september werd de *Atlas van het dialect in Vlaanderen* gepresenteerd.

Ten slotte stond het jaar ook in het teken van de neologismen: in november werd het Woordenboek van Nieuwe Woorden gelanceerd en was er de presentatie van het boek *Knuffelcontact en waterwappie*. Het nieuwe woordenboek en de omgang van het INT met neologismen werden via een

webinar op 30 november verder toegelicht.

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 2022

prof. dr. Frieda Steurs, directeur/bestuurder

A. Lexicografie en Terminologie

Basistaken

A.1. Synchron: ANW

Het *Algemeen Nederlands Woordenboek* (ANW) is een corpusgebaseerd, digitaal, multimediaal woordenboek van het eigentijdse Nederlands in Nederland en Vlaanderen, in Suriname en in het Caraïbisch gebied. 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.

Evenals in 2020 werd in 2021 een groot deel van het jaar gewerkt aan het koppelen van het ANW aan GiGaNT-Molex, het grote computationele lexicon van het hedendaags Nederlands dat ook de basis vormt van de Woordenlijst Nederlandse Taal. Najaar 2020 werden in verband daarmee de updates van het ANW stopgezet. Medio 2021 waren de werkzaamheden voltooid en werd het ANW weer dagelijks geüpdatet. Nu alle artikelen van het ANW gekoppeld zijn aan de centrale database, worden de grammaticale gegevens voortaan vanuit GiGaNT-Molex aan het ANW toegevoegd, wat een aanzienlijke tijdsbesparing voor de redactie zal gaan betekenen.

Verder bewerkte de redactie van het ANW in totaal 527 artikelen voor het ANW en daarnaast ook artikelen voor het gloednieuwe *Woordenboek van Nieuwe Woorden* (WNW), dat op 17 november 2021 online ging. Gedurende het jaar werden redacteuren van het ANW ook ingezet als leden van de webredactie en incidenteel bij andere projecten.

In 2021 zijn, afgezien van de artikelen met alleen grammaticale gegevens (de zogeheten ‘kopjesartikelen’), 1309 volledig bewerkte, nieuwe artikelen online gezet. Aan het ANW zijn in 2021 niet alleen woorden toegevoegd, maar er zijn ook artikelen uit het woordenboek verwijderd. Het gaat in dit geval om 110 neologismen met een korte levensduur (de ‘eendagsvliegen’) die zijn overgeheveld naar het WNW.

A.2. Synchron: Woordcombinaties

Woordcombinaties is een onlinetaaltool in ontwikkeling die leeders van het Nederlands als vreemde taal ondersteunt bij het gebruiken van woorden in context. Het project integreert de beschrijving van collocaties, idiomen en syntactische patronen en behandelt in de eerste fase de combinatoriek van werkwoorden. Het project is corpusgebaseerd en gebruikt de *Sketch Engine* als corpusquerytool.

Woordcombinaties toont hoe woorden gebruikt worden in goede voorbeeldzinnen, welke woorden

met elkaar gecombineerd worden (collocaties of combinatiemogelijkheden) en hoe valentiepatronen en andere patronen samen met collocaties gebruikt worden voor het bouwen van zinnen. Taalleerders leren niet alleen woorden kennen, maar ook woorden gebruiken in context.

In het verslagjaar is verder gewerkt aan de voorbeeldzinnen en combinatiemogelijkheden van werkwoorden uit de schooltaalwoordenschat en het *Frequentiewoordenboek Nederlands* (Carole Tiberius en Tanneke Schoonheim, 2013). Daarnaast is gewerkt aan de patroonbeschrijving van werkwoorden en de beschrijving van idiomatische uitdrukkingen met werkwoorden.

In 2021 is een eerste release van het project beschikbaar gekomen voor het publiek. Er is aandacht besteed aan communicatie en promotie d.m.v. persberichten, een interview door de Taalunie, deelname aan congressen (Euralex 2021 en Europhras 2021) en een webinar i.s.m. de Taalunie en een docent NT2.

A.3. Synchron: Neologismen

Ieder jaar komen er veel nieuwe woorden bij; dit jaar bijvoorbeeld *alszaak*, *knalpotterreur*, *omhoogfalen*, *pandemoe* en *vaccinatievoorsteker*. Het INT onderzoekt de vorming van nieuwe woorden en houdt die woorden onder andere bij in het neologismenwoordenboek, een onlinewoordenboek van nieuwe woorden. In 2021 is een eerste versie van dit Woordenboek van Nieuwe Woorden (WNW) online gegaan. Het woordenboek is gratis beschikbaar op neologismen.ivdnt.org.

In dit woordenboek zijn nu al ruim 7000 nieuwe woorden opgenomen. Niet alleen vaak gebruikte neologismen als *appen*, *poké bowl* en *woke* staan erin, maar ook (nog) niet ingeburgerde nieuwe woorden als *besmettingsfeestje*, *beyoncéfeminisme*, *havermelkelite*, *vertiktokking* en *zoomzombie*. Het WNW wordt voortdurend aangevuld en iedere week geüpdatet. Naast het WNW kwam er dit jaar ook een bijbehorend populairwetenschappelijk boek uit over neologismen. In *Knuffelcontact en waterwappie. De kracht van nieuwe woorden* (Uitgeverij Scriptum) geven INT-taalkundigen Vivien Waszink en Veronique De Tier meer informatie over nieuwe woorden en hun achtergrond: waar, waarom en hoe zijn ze ontstaan, en wanneer noemen we een woord eigenlijk nieuw?

Ook worden alle woorden in het WNW digitaal gekoppeld aan de bijbehorende lemmata in het spellingbestand Molex en daarmee ook aan de bijbehorende ingangen in de Woordenlijst Nederlandse Taal.

Ten slotte is er ook dit jaar iedere week een neologisme behandeld in de webrubriek 'Nieuw woord van de week'. Het neologisme van de week wordt niet alleen wekelijks op de INT-website gepubliceerd, maar ook in het *Leidsch Dagblad*. Er zijn meer (populair)wetenschappelijke activiteiten ontplooid: het INT verzorgde een rubriek over neologismen in de scheurkalender van Onze Taal, er werden lezingen en webinars over nieuwe woorden verzorgd en er werden diverse interviews in

kranten en op de radio gegeven over onder andere het neologismenwoordenboek en het populairwetenschappelijke boek.

A.4. Diachroon: Historische woordenboeken en de GTB

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 de [historische woordenboekenportal](#) GTB: het Oudnederlands Woordenboek (ONW), het Vroegmiddelnederlands Woordenboek (VMNW), het Middelnederlandsch Woordenboek (MNW) en het Woordenboek der Nederlandsche Taal (WNT). De data van deze woordenboeken zijn een bron voor GiGaNT en DiaMaNT (zie B.4 en B.5).

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. Met hetzelfde doel werd op het [YouTube-kanaal](#) van het INT een [webinar](#) geplaatst over doeltreffend zoeken in de online historische woordenboeken; ook werd er een korte, instructieve [animatievideo](#) gemaakt om de betekenis van een oud woord op te kunnen zoeken. 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 2021 dit kader zijn uitgevoerd (zie B.4), leveren correcties op die ook voor de online woordenboeken relevant zijn. Deze zullen uitgevoerd worden zodra een nieuwe, voor 2022 geplande TEI-P5-versie van de woordenboekbestanden beschikbaar is.

Tot slot is gedurende twee stages 95% van de definities in het Oudnederlands Woordenboek (ONW) in het Engels en Duits vertaald om de internationalisering van de medioneerlandistiek te stimuleren. De overige 5% zal het INT zelf aanvullen.

Specifieke projecten

A.5. Dialectologie

A.5.1. Database of the Southern Dutch Dialects (DSDD)

Begin januari 2021 bevatte de DSDD 19.000 concepten: de hele Algemene Woordenschat van de drie onderliggende dialectwoordenboeken (Woordenboek van de Vlaamse, Brabantse en Limburgse dialecten) en een deel van de vaktaal- en de landbouwwoordenschat. In de loop van 2021 werden de resterende concepten van de landbouwwoordenschat en de vaktalen toegevoegd en gekoppeld en gebeurde er datacuratie. Eind 2021 bevatte de DSDD ongeveer 25.000 concepten. Dit aantal zal nog licht stijgen, omdat nog een klein deel van de vaktalen niet is ingevoerd (smid, loodgieter en timmerman).

In 2021 werd het Woordenboek der Zeeuwse dialecten geanalyseerd en werden de bestanden van het woordenboek geïmporteerd in lex'it. In samenwerking met de diverse partners wordt de DSDD in 2022 uitgebreid met Zeeuwse dialectdata.

A.5.2. Gesproken Corpus van de Zuidelijk-Nederlandse Dialecten (GCND)

Het INT is partner in het project Gesproken Corpus van de Zuidelijk-Nederlandse Dialecten, samen met de Universiteit Gent (hoofdaanvrager), Variaties vzw (Koepelvereniging voor dialecten en oraal erfgoed in Vlaanderen), Meertens Instituut en Rijksuniversiteit Groningen. Het beoogt de ontsluiting van een collectie van dialectopnames uit 768 plaatsen in België, Frankrijk en het zuiden van Nederland, opgenomen tussen 1963 en 1976. Het INT zal hier een substantiële infrastructurele rol in spelen. De opnames worden volgens een nieuw ontwikkeld transcriptieprotocol getranscribeerd. Eind 2021 maakten de Universiteit Gent en Variaties vzw 190 transcripties klaar. Het INT zal de audio, de transcripties en de annotaties op termijn vrij online beschikbaar en doorzoekbaar maken en duurzaam bewaren. In de loop van 2021 zijn hiervoor een aantal voorbereidende overlegvergaderingen geweest.

A.5.3. Atlas van het dialect in Vlaanderen

In september verscheen de *Atlas van het dialect in Vlaanderen*. Het project is een initiatief van Uitgeverij Lannoo i.s.m. Johan De Caluwe, Anne-Sophie Ghyselen, Roxane Vandenberghe (UGent) en Veronique De Tier (INT). Het boek telt 289 bladzijden, vijftig hoofdstukken waarvan er dertig over de dialecten gaan (Wat is dialect, kenmerken van dialectgroepen, enz.) en twintig over dialectwoorden rond bepaalde thema's (kleding, ziekte, flora, fauna, ...). Iedere auteur nam ongeveer een kwart van de bijdragen voor zijn of haar rekening. Het boek verscheen eind september met een boekpresentatie in De Minard in Gent, met veel pers aandacht (zie daar). N.a.v. de presentatie hebben de vier auteurs ook

een enquête over kennis en gebruik van dialect verspreid die op veel respons kon rekenen (iets meer dan 12.000 deelnemers). Ondertussen wordt een derde druk voorbereid die in april 2022 verschijnt. Er verschenen enkele positieve recensies in tijdschriften en kranten en het boek is eind 2021 genomineerd voor de *LOT-populariseringsprijs* die begin februari 2022 wordt toegekend en voor de LIA's, een prijs die eind april uitgereikt wordt door *De Taalsector*.

A.5.4. Overige activiteiten

In 2021 verscheen op de website tweewekelijks 'Uit de streek', een rubriek over dialectwoorden uit de DSDD en de woordenbank.

Het INT zetelt in de werkgroep contactvariëteiten, een werkgroep van de Nederlandse Taalunie en de Fryske Akademy die onderzoekt hoe dialectgemeenschappen (met als pilot het Bildts) eigen informatie kunnen samenbrengen en delen. De bedoeling is dat er een infrastructuur komt die niet alleen door het Bildts maar ook door andere dialectgemeenschappen gebruikt kan worden. Er wordt overlegd of het INT hierin een rol kan spelen, o.a. op het vlak van hosting en aanbieden van infrastructuur. In de loop van 2021 waren er enkele voorbereidende vergaderingen.

Veronique De Tier is vanuit het INT betrokken bij de Stichting Nederlandse Dialecten, die jaarlijks een streektaalconferentie organiseert. Dit jaar vond die plaats in Oldenzaal op 1 oktober met als thema 'Streektaal op de werkvloer'.

Nicoline van der Sijs ontving samen met Roeland van Hout en Henk van den Heuvel subsidie uit het Stimuleringsfonds Nedersaksisch in Gelderland ten behoeve van het e-WGD Veluwe, oftewel het elektronische Woordenboek van de Gelderse dialecten, de delen die gaan over de Veluwe.

A.6. Vertaalwoordenschat

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

Inmiddels staan het Nederlands-Nieuwgrieks / Nieuwgrieks-Nederlands, het Nederlands-Portugees / Portugees-Nederlands, het Nederlands-Estisch en, sinds het voorjaar van 2021, het Nederlands-Fins / Fins-Nederlands online. De woordenboeken zijn gratis raadpleegbaar op vertaalwoordenschat.ivdnt.org; er is ook een app Vertaalwoordenschat.

Om correcties en inhoudelijke updates in de toekomst op een gebruiksvriendelijke manier te kunnen realiseren is in 2021 gewerkt aan een nieuwe redactieomgeving voor de Vertaalwoordenschat. De werkzaamheden hieraan zullen begin 2022 worden afgerond, waarna correcties in de bestanden kunnen worden verwerkt.

Verder is de applicatie uitgebreid zodat ook het flexiotype van trefwoorden getoond kan worden. Er is echter nog een handmatige correctie van de inputdata nodig voordat deze informatie ook online beschikbaar komt.

Daarnaast is er in 2021 verder gewerkt aan de koppeling van de bestanden aan GiGaNT-Molex, zodat alle informatie hieruit op termijn ook gebruikt kan worden ten behoeve van de Vertaalwoordenschat. Tot slot zijn de bestanden Arabisch-Nederlands / Nederlands-Arabisch gereedgemaakt voor online release. De bestanden zijn nu nog alleen intern toegankelijk en zullen in overleg op een latere datum officieel worden gereleased.

A.7. Expertisecentrum Terminologie

Veldondersteuning en positionering in een terminologisch netwerk zijn traditioneel de twee kerntaken waar het Expertisecentrum Nederlandstalige Terminologie zich op richt. In 2021 heeft dit tot de volgende resultaten geleid.

A.7.1. Website

De visuele band tussen genoemde ondersteuning en positionering, en de ENT-webpagina werd verder versterkt door toevoeging van de nieuwe rubriek van Europese terminologiegerelateerde instellingen. Deze rubriek biedt een overzicht in de vorm van een interactieve landkaart met routeplanner. De onderliggende dataverzameling kwam tot stand op basis van drie selectiecriteria: een bredere terminologische oriëntatie inzake onderzoek en/of domeinen, de toegankelijkheid op niet-commerciële basis en/of de terbeschikkingstelling van geproduceerde inzichten en data, en een website waaruit voldoende activiteit blijkt.

Tot de nieuwe onderdelen van de ENT-webpagina behoort tevens de webrubriek Terminologieprojecten. Hier zijn de exporteerbare bestanden verzameld die resulteren uit recent terminologisch onderzoek op het INT, inclusief die uit lopende projecten. Tot deze laatste groep behoren het HOTNeV-project (Hoger Onderwijs Terminologie in Nederland en Vlaanderen) en Nederlands als wetenschapstaal. Eveneens zijn hier twee databases opgenomen uit de medische sfeer, de Medische Pilot (MedPilot) en de Medische Termen Belgisch-Nederlands. De Medische Pilot beschrijft bij wijze van experiment een klein deel van de medische woordenschat op verschillende niveaus, van wetenschappelijk tot toegankelijk voor laaggeletterden. Medische Termen Belgisch-

Nederlands bestaat uit een lijst met termen waarvan niet mag worden aangenomen dat ze zowel in België als in Nederland worden gebruikt en begrepen.

Op de ENT-webpagina bood verder de agenda met evenementen rond terminologie opnieuw een overzicht van initiatieven in 2021 in Nederland, Vlaanderen en Europa.

Naast de communicatie door informatievoorziening via de website zijn geïnteresseerden in terminologie tevens actief benaderd met nieuwsbrieven terminologie. In totaal verschenen in 2021 vier nieuwsbrieven met een vaste rubriekenreeks over voorbije evenementen, nieuwe ontwikkelingen en publicaties, tools en bronnen, een uitgelichte term, berichten van NL-Term en een agenda. Zoals gebruikelijk blijven ze beschikbaar in het archief op de terminologiepagina.

A.7.2. Projecten voor specifieke vakdomeinen

Veldondersteuning werd in 2021 eveneens gerealiseerd met projecten die door hun thematiek en online raadpleegbare materiaalverzamelingen een toegepaste waarde hebben.

HOTNeV

HOTNeV, in eerste instantie opgezet als een pilot, betreft een tweetalige termenbank (Nederlands en Brits-Engels) over hogeronderwijsterminologie en wordt sinds 2017 voortgezet. Het doel van het HOTNeV-project is om te streven naar meer duidelijkheid en eenduidigheid in het Nederlandstalige termgebruik en om in plaats van ad-hocvertalingen uit en naar het Engels passende termequivalenten aan te reiken na termanalyse. Het termonderzoek impliceert tracing, interpretatie en vergelijking van online beschikbare documenten van onderwijsinstellingen en -instanties, waarna de beschrijvingen van de geanalyseerde termen worden ingevoerd in de HOTNeV-termenbank. Deze is in 2021 toegankelijk gemaakt via de ENT-webpagina en wordt gaandeweg uitgebreid. De termbeschrijving in 2021 resulteerde mede uit de begeleiding van vier stages. In het kader van dit project werden ook de stageplannen van twee studenten begeleid van wie de stage in 2022 zal plaatsvinden. Vier vergaderingen van de Termraad Nederlands boden verder een platform voor discussie over HOTNeV-termen en hun concepten.

Nederlands als wetenschapstaal

Uit onderzoek naar onderwijs en woordenschat is gebleken dat voor de opstap naar het eerste bachelorjaar een betere begeleiding wenselijk is 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. Het pilootproject heeft betrekking op zowel alfa-, bèta- als gamma-richtingen. Er is gekozen voor een bredere benadering die de selectie en beschrijving van termen geschikt maakt voor een ruimere studentenpopulatie zodat bijvoorbeeld scheikundige en wiskundige termen niet louter voor scheikundigen of wiskundigen in opleiding geschikt zijn maar ook voor verpleegkundigen respectievelijk economiestudenten. Het project dient dus een bredere onderwijsdoelstelling die het INT zal steunen door de resulterende termenlijsten te hosten. De bestanden voor de exacte vakken scheikunde, wiskunde en natuurkunde zijn in 2020 aan het INT opgeleverd. Aan de Nederlandstalige termen en definities zijn tevens Engelse termequivalenten gekoppeld. De bestandsstructuren ondergingen op het INT eerst een uniformeringsproces en vervolgens vond tekstinhoudelijke correctie plaats. Deze bestanden zijn op de INT-website al downloadbaar voor de gebruiker in voorlopige pdf-versie. Daarnaast vond in 2021 een conversie van deze data plaats van Word- naar TBX-formaat met het oog op online publicatie. Daarvoor werd tevens een acceptatietestomgeving (ato) ontwikkeld die door haar generiek ontwerp ook voor andere terminologieprojecten inzetbaar zal zijn. In 2022 gaat de definitieve webinterface met zoekapplicatie online. Ook het natuurkundebestand zal dan de nodige datastructurering en correctie ondergaan voor online publicatie.

Medische terminologie

Op het vlak van de medische vaktaal ging de aandacht in 2021 uit naar het databestand van het *Pinkhof Geneeskundig Woordenboek*. Met de Stichting beheer Database Pinkhof Geneeskundig Woordenboek besprak het INT verder de gegevensoverdracht en werden juridische voorwaarden vastgelegd. Oplevering van de medische termbank vond plaats op 23 april 2021 en de definitieve samenwerkingsovereenkomst tussen genoemde Stichting en het INT werd medio december ondertekend. Ook is een raad van advies samengesteld. Wat de bestandsinhoud betreft, vond in 2021 overleg plaats over bestandscategorieën en werden de data ter voorbereiding van de publicatie naar TBX-formaat geconverteerd. De gegevens werden ingevoerd in een applicatietestomgeving die ook voor de bestanden voor het project Nederlands als wetenschapstaal wordt ingezet. Het woordenboek is nu online beschikbaar. In 2022 wordt na de nodige tests de definitieve webinterface ontwikkeld.

Juridische terminologie

Het *Juridisch Woordenboek* van M.C. Oosterveld-Egaz en J.B. Vuyk-Bosdriesz gericht op de rechtspraktijk, handel en bedrijfscommunicatie, behandelt de terminologie van het Nederlandse rechtssysteem wat betreft de uitgangstermen, en de rechtstermen van het Spaanse rechtssysteem als doeltaal. De demoapplicatie die is ontwikkeld bij de Fryske Akademy om de inhoud van het woordenboek te tonen is in 2021 verbeterd. Het is nu mogelijk een overzicht te krijgen van de

bewerkte termen. Verder is onderzoek gedaan naar vereenvoudiging van de structuur van de data.

Terminologietools

Voor de termextractor TermTreffer vonden in 2021 overleg en onderzoek plaats naar de mogelijkheden om van dit taaltechnologische hulpmiddel een nieuwe versie te ontwikkelen. Aan deze herimplementatie wordt verder gewerkt in 2022. In 2021 zijn op aanvraag meermaals accounts uitgegeven voor gebruik van deze tool in het kader van colleges en opdrachten aan universiteiten.

Terminologisch netwerk

Positionering in een netwerk kreeg onder meer vorm door deelname aan of bijwonen van congressen en symposia (Pre-TSS Day TermNet, 4th International Conference on Terminology – Institute of the Lithuanian Language, TiNT 2021). Verschillende online lezingen vonden plaats in Europese context, met name in Budapest over het ENT bij de Hungarian Academy of Sciences, en in Wrocław, Brno, Olomouc en Triëst voor studenten Nederlands. Voor de Taalsector Vlaanderen werd een online lezing gehouden over de aanbodmogelijkheden van het INT met speciale aandacht voor terminologie. Eveneens werd een cursus Terminologie en ICT met gebruik van de TermTreffer gegeven aan de KU Leuven. Het ENT nam verder deel aan terminologieoverleg tijdens de vergaderingen die de Termraad Nederlands belegde voor Nederlandstalige (ver)taaldiensten van EU-instellingen en van Nederlandse, Belgische en Vlaamse overheden, en aan stagevoorlichting van de Termraad Academy. Voor hogeronderwijsinstellingen in Nederland werden vier stagiairs begeleid, waarvan een in samenwerking met het Departement Nederlands van het directoraat-generaal voor vertaling van de EU.

A.8. Lexicon Frisicum

Het Lexicon Frisicum is 148 jaar geleden geschreven door Joast Hiddes Halbertsma, in het Latijn, en was een eerste aanzet tot een woordenboek van het Fries. Het woordenboek beschrijft 6400 woorden, wat slechts een klein deel is van de totale woordenschat. Halbertsma overleed namelijk voor hij het kon afmaken. Lexicograaf en taalkundige Anne Dykstra vertaalde het Lexicon naar het Fries, samen met zijn collega Bouke Slofstra en verschillende vrijwilligers. Het INT, dat via het historisch woordenboekportaal ook het Wurdboek fan de Fryske taal / Woordenboek der Friese taal (WFT) online heeft gezet, is ook verantwoordelijk voor het online zetten van dit unieke woordenboek. In 2021 heeft het woordenboek een data-update gekregen, en zijn er aanpassingen gedaan aan de userinterface.

B. Corpora en lexica

Basistaken

B.1. Moderne corpora

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

Het interne corpus bevat inmiddels ruim 2,3 miljard tokens; het extern gepubliceerde deel bevat ruim 2.500.000 documenten en meer dan 1 miljard tokens, meer dan het dubbele van de vorige externe versie.

In november 2021 is een nieuwe versie van de zoekapplicatie voor het corpus gepubliceerd op chn.ivdnt.org. Hiertoe is de applicatie geoptimaliseerd, en is de complete automatische workflow van bronbestanden naar geïndexeerd corpus omgezet naar een continue integratie, zodat maandelijkse updates mogelijk zijn.

Op het gebied van de data zijn diverse bronnen toegevoegd, waaronder met name forumdata en de in toegankelijk Nederlands opgestelde Startkrant, waarmee het materiaal diverser is geworden, en is een systematische verbetering van de corpusmetadata uitgevoerd. Het Europese CLARIN-project ParlaMint, in het kader waarvan materiaal van het Belgisch federaal parlement is aangeworven en geconverteerd naar TEI, is afgerond.

B.2. Historische corpora

In 2019 is flink geïnvesteerd in frontend en backend voor historische corpora. Daarbij is gekeken naar zoveel mogelijk uniformering, met de mogelijkheid om efficiënt een aanpassing van de userinterface en de zoekmogelijkheden te maken indien een specifiek historisch corpus dat nodig heeft. In 2020 is de BlackLab-gebaseerde applicatie voor de online corpora verder ingrijpend verbeterd. In de nieuwe versie, voortbouwend op het in het kader van CLARIAH-core verrichte werk aan de OpenSoNaR-

applicatie, is flexibeler configuratie van de zoekmogelijkheden gerealiseerd. Hiermee kan een snellere onlinegang van een nieuw corpus gerealiseerd worden. Er is ook gewerkt aan de gebruikersvriendelijkheid van het zoeken op met name features in de taalkundige verrijking.

In januari 2021 is de update van het oorspronkelijke Brieven als Buit-corpus online gekomen, de daaropvolgende maand gevolgd door de aanvullingen daarop, Brieven als Buit 2. Met de publicatie van die aanvullingen worden transcripties, foto's en metadata van 1386 nieuwe 17e- en 18e-eeuwse brieven beschikbaar gesteld voor onderzoek. Het corpus is nog niet taalkundig verrijkt.

Ook het Corpus Juridisch Nederlands zag in 2021 het licht. Oorspronkelijk maakte dit corpus deel uit van het Corpus Hedendaags Nederlands, maar omdat de meer dan 5000 juridische teksten afkomstig zijn uit de periode 1814-1989 was een eigen historisch corpus meer op zijn plaats.

In 2021 is volop gewerkt aan het Corpus Oudnederlands, dat begin 2022 online zal komen. Voor dit corpus is een betere genrebeschrijving uitgewerkt en toegepast. Er is ruim aandacht besteed aan een gebruikersvriendelijke helpfunctie. Het Corpus Oudnederlands zal doorzoekbaar zijn met de TDN-tagset (zie verderop).

In 2021 is gestart met de voorbereiding voor een corpus van 17e-eeuwse kranten van de Koninklijke Bibliotheek die door de vrijwilligers van Nicoline van der Sijs zijn overgetikt en gecorrigeerd. Daartoe is in de op het INT ontwikkelde bewerkingsomgeving Lex'it gewerkt aan het corrigeren van metadata, waaronder plaatsnamen. Deze werkzaamheden worden voortgezet, waarna het corpus in de loop van 2022 online zal komen.

Daarnaast is mede in de context van CLARIAH+ gewerkt aan de Tagset Diachroon Nederlands, die toepasbaar moet zijn op alle taalfasen van het Nederlands, en die ter beoordeling is voorgelegd aan een keur van met name historisch taalkundigen in Vlaanderen en Nederland. De bedoeling is om van de tagset een breed gedragen standaard te maken voor de verrijking van tekstmateriaal, om zo diachroon onderzoek beter te kunnen faciliteren.

In mei 2021 is versie 1.0 van deze tagset online gezet. Op Neerlandistiek verscheen vervolgens een bijdrage over het nut en het belang ervan. Het is de bedoeling dat de tagset ook toegepast zal worden in het centraal computationeel lexicon GiGaNT, en uiteindelijk in de online historische woordenboekapplicatie. Voor de uitwerking van die tagset is onder andere een mapping gemaakt van de taalkundige verrijking van bestaande verrijkte corpora, waaronder het Corpus Gysseling, het Brieven als Buit-corpus, het Corpus Van Reenen Mulder en het Corpus Oudnederlands.

Tot slot zijn de werkzaamheden gestart om evaluatie- en trainingsdata te maken van taalkundig verrijkt diachroon corpusmateriaal, verrijkt volgens de TDN-kerntagset. Dit werk wordt uitgevoerd met behulp van de Cobalt-tool (zie D 2.1). Het aldus verkregen trainingsmateriaal zal automatisch taalkundige annotatie van historische teksten volgens de richtlijnen van de tagset TDN mogelijk maken. Deze werkzaamheden zullen in 2022 worden voortgezet.

B.3. Spelling

Ook in 2021 is de onlineversie van de Woordenlijst Nederlandse Taal weer uitgebreid. Er werden niet alleen nieuwe trefwoorden toegevoegd, maar ook de woordvormen, glossen en zinvolle reeksen werden aangevuld. Kwaliteit staat nog steeds voorop. We willen uiteraard dat taalgebruikers maximaal antwoord krijgen bij opzoekingen en dat ze ook bijkomende informatie kunnen vinden, als ze dat zouden wensen. Gebruikers met onbeantwoorde (spelling)vragen werden uiteraard zo snel mogelijk geholpen, hetzij met een duidelijk antwoord, hetzij met een doorverwijzing naar een gepastere dienst of persoon.

De Commissie Spelling ziet toe op de nieuwe en gewijzigde informatie in het spellingbestand en beantwoordt eventuele vragen daarover. De omstandigheden hebben er in 2021 voor gezorgd dat er meer (maar kortere) vergadermomenten waren, tien in totaal.

B.3.1. Project Spelling-API

Op grond van diverse ingrijpende verzoeken tot verbetering en uitbreiding van de functionaliteit van woordenlijst.org (o.a. zoeken met woordsoort en uitspraak, verbeterde zoeksuggesties) is onderzocht hoe het best op deze behoeften kan worden ingespeeld. Op grond hiervan is ervoor gekozen een nieuwe applicatie met meer mogelijkheden te ontwikkelen, waarmee we taalkundig geïnteresseerd publiek ook de mogelijkheid kunnen bieden veel flexibeler in het materiaal te zoeken. Door de software te ontwikkelen op een manier die parallel is aan andere recentelijk op het INT ontwikkelde zoeksystemen, maken we het onderhoud door ons ook eenvoudiger.

In 2021 zijn de vereisten uitgewerkt, is onderzoek verricht naar verbeterde zoeksuggesties, en is een eerste interne prototype-versie van de applicatie gebouwd. Om de gewenste functionaliteit te kunnen realiseren wordt uitspraakinformatie toegevoegd en wordt de werkwoordskenmerk-informatie uitgebreid. Er is gewerkt aan de principes voor het toevoegen van uitspraakinformatie, aan het inrichten van het bewerkingstraject voor deze informatie, en aan een eerste invulling van een deel van de data. Verder is de BAN-database van de werkgroep Buitenlandse Aardrijkskundige Namen in het Nederlands geïntegreerd in het moderne lexicon, met behoud van de verwijzingen naar de website van de BAN.

B.4. GiGaNT

GiGaNT (Groot Geïntegreerd Lexicon van de Nederlandse Taal) is het computationele woordvormenlexicon dat de door het INT beschreven woordenschat bevat van het Nederlands van de 6e eeuw tot nu. Het vormt de centrale database van het INT. Werkzaamheden aan deze centrale data-

infrastructuur zijn modulair opgezet en worden dienovereenkomstig uitgevoerd. GiGaNT bestaat uit twee grote componenten: GiGaNT-Molex, dat de kern is van de formele beschrijving van de hedendaagse woordenschat, en GiGaNT-Hilex, de historische lexiconcomponent, die onderdeel is van de lexicografische infrastructuur voor het historisch Nederlands. Beide modules gebruiken het modern Nederlands lemma + woordsoort als ingang.

GiGaNT-Hilex is gebaseerd op de historische woordenboeken van het Nederlands. Het lexicon bevat nu de data gebaseerd op het WNT en het MNW. De twee subcomponenten zijn via een GiGaNT-lemma onderling gekoppeld. De koppeling is gebaseerd op etymologische gronden. Omdat het lexicon gebaseerd is op deze gedrukte woordenboeken, is de lemma-indeling van de woordenboeken niet een-op-een te vertalen naar een GiGaNT-lemma. Daarnaast is ook op woordvormniveau de op het woordenboeklemma uitkomende eenvoudige analyse van een woordvorm niet altijd correct, bijvoorbeeld omdat clitische vormen regelmatig voorkomen in historisch Nederlands. In de huidige bewerkingsfase van Hilex worden met name op deze twee punten datacorrecties uitgevoerd. Tot nog toe heeft dat bijvoorbeeld geresulteerd in ruim 22.000 nieuwe lemmata in Hilex, en bijna 38.000 woordvormtypes die een meervoudige analyse nodig hadden. De woordsoorttoekenning in Hilex is tot nog toe gebaseerd op de informatie uit de woordenboeken, die mede bron van onderzoek zijn geweest voor de ontwikkeling van de Tagset Diachron Nederlands. Het is de bedoeling dat het lexicon op termijn verrijkt wordt met de TDN-tagset en volgens de TDN-tagsetprincipes. Sinds september 2020 wordt de lexiconservice, waarmee GiGaNT Hilex toegankelijk wordt gemaakt voor gebruik in bijvoorbeeld corpusapplicaties als Nederlab, Delpher of het Corpus Middelnederlands, maandelijks geüpdatet.

GiGaNT-Molex is de moderne lexiconcomponent. Het bevat lemmata met hun complete paradigma, in de correcte spelling, en is sinds 2015 de bron voor woordenlijst.org (zie het onderdeel spelling). Vanuit het spellingproject worden lemmata aan het lexicon toegevoegd, maar het doel is dat alle producten die hedendaags Nederlandse woordenschat beschrijven, gekoppeld zijn aan de moderne lexiconcomponent. Nadat in 2020 voor het ANW, het neologismenproject en het project woordcombinaties de technische inbedding van het lexicon in de workflow van deze projecten is gerealiseerd, is nu ook de inhoudelijke koppeling een feit. Dit heeft geleid tot ruim 4700 nieuwe lemmata vanuit deze projecten aan het centrale lexicon. Het totale aantal koppelingen tussen Molex en de lemmata van projecten is ruim 97000. En deze aantallen groeien bij iedere toevoeging van nieuwe lemmata in deze met Molex gekoppelde bronnen. De koppeling van RBN met Molex, die dit jaar is afgerond, heeft geleid tot ruim 1700 nieuwe lemmata voor Molex en ruim 62.000 koppelingen tussen de betekenissen in het RBN en het corresponderende lemma in Molex. De toevoeging en koppeling van de exoniemen vanuit de BAN-lijst (zie B.3.3.1) heeft geresulteerd in ruim 2400 nieuwe lemmata en ruim 3200 koppelingen. De werkzaamheden om Molex met de vertaalwoordenschat te koppelen worden in 2022 afgerond.

B.5. Taalbank: semantisch lexicon DiaMaNT

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

In 2021 is als voorbereiding op corpusgebaseerde toepassing en uitbreiding gewerkt aan het verwerven van een betrouwbaar gedateerd corpus Historisch Nederlands. Een probleem met beschikbare historische teksten is dat nauwkeurige datering van de tekst niet altijd beschikbaar is. Een datering van historische tekst verwijst vaak naar de publicatiegegevens van de teksteditie in plaats van naar de datering van de tekstgetuige. Er zijn aanzienlijke inspanningen geleverd om de taalkundig relevante metadata aan te vullen en zo een betrouwbaar corpus te verkrijgen.

Verder is gewerkt aan een visualisatietool om de mogelijkheden te onderzoeken om met behulp van contextgevoelige distributionele modelleringstechnieken inzicht te krijgen in diachrone betekenisprofielen. Er is contact gelegd met het MacBerth-project, met het doel gezamenlijk aan evaluatie van deze technieken te werken.

Het verkennende werk aan de koppeling van het ONW aan de Thesaurus of Old English is afgerond met een presentatie op ICEHL 2021 op 7 juni 2021 en publicatie in de *Amsterdamer Beiträge zur älteren Germanistik* (inclusief de dataset).

C. Grammatica

Basistaken

C.1. Grammatica algemeen

In 2021 zijn grote stappen gezet in de ontwikkeling van de verschillende grammaticaproducten op het INT. De vernieuwde e-ANS is online gegaan (zie hieronder), de site van Taaladvies.net is volledig vernieuwd en er is een eerste, interne versie ontwikkeld van het geïntegreerde grammaticaportaal, een site die als startpagina zal gaan dienen voor de verschillende grammaticaproducten van het INT.

C.2. Taaladvies

In de zomer van 2021 is de vernieuwde site van Taaladvies.net succesvol online gegaan. Deze Wordpress-site is ontwikkeld door Frank Landsbergen op het INT in nauwe samenwerking met de taaladviseurs van Taaladvies.net. De site, nu in beheer van webmaster Job van Doeselaar, is niet alleen gemoderniseerd en beter toegankelijk gemaakt, maar ook is de werkomgeving voor de auteurs volledig vernieuwd en verbeterd. De site draait voortaan op de servers van het INT.

C.3. e-ANS

In 2021 konden we de eerste vruchten plukken van de nieuwe situatie, waarin het INT verantwoordelijk is voor de verdere herziening van de e-ANS. Dankzij de structurele financiering door de Taalunie kan het INT deze nieuwe taak waarmaken. We moeten hierbij opmerken dat het in de jaarrekening 2021 lijkt alsof de Taalunie minder budget beschikbaar heeft gesteld dan voorzien in de begroting, maar dit is niet zo. Het INT had in de begroting het bedrag toegekend voor de ANS verkeerdelijk tweemaal opgenomen.

In april van het jaar is de nieuwe ANS-website gelanceerd, inclusief de eerste herziene hoofdstukken, tijdens een feestelijk, drukbezocht webinar. In juli volgde de lancering van het geheel nieuwe, uitgebreide hoofdstuk Klankleer van het Nederlands, ook door middel van een webinar. Een update met hoofdstukken uit het eerdere ANS-herzieningsproject is verschoven naar volgend jaar.

Op inhoudelijk vlak zijn grote stappen gezet. Ook in het afgelopen jaar konden, mede dankzij een financiële versnellingsimpuls, extra contracten worden afgesloten voor de herziening van nieuwe hoofdstukken. Maaike Beliën houdt als coördinator van de inhoudelijke herziening contact met de externe auteurs, leest mee en bewaakt de toegankelijkheid en interne consistentie van de hoofdstukken. Zo kwamen de conceptversies tot stand van de hoofdstukken over het telwoord, tussenwerpsel,

woordvorming, de werkwoordelijke constituent en modaliteit.

Als resultaat van een pilot voor de ontwikkeling van een didactische laag voor de ANS is in het voorjaar een onderwijspagina toegevoegd, gepresenteerd en geëvalueerd. Aan de hand van deze evaluatie is besloten om de opzet van de pagina's op enkele punten te wijzigen. In het najaar van 2021 is verder gewerkt aan een onderwijspagina voor het nieuwe hoofdstuk over negatie.

C.4. Taalportaal

De conversietool voor de Syntax of Dutch (SoD) van Hans Broekhuis, waarmee de Wordbestanden automatisch kunnen worden omgezet naar Taalportaal-xml, is volledig vernieuwd met het oog op de omzetting van het nieuwe deel van de SoD, *Coordination*. De geplande update van de Taalportaalapplicatie is door corona en andere oorzaken verschoven naar een later tijdstip.

D. Softwaretools (ontwikkeling, beheer en onderhoud)

Basistaken

D.1. INT Impactcentrum en digitization.eu

Het IMPACT Centre of Competence 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 sinds 2019 ook CLARIN Knowledge (K) Centre. In 2020 heeft het centrum een plek gekregen bij de Fundación General van de Universidad de Alicante (FGUA). In 2021 is Sally Chambers aangetreden als nieuwe directeur. Het INT is voorzitter van de executive board en is vertegenwoordigd in het kernteam dat verantwoordelijk is voor het runnen van het centrum. Het levert expertise op het gebied van historisch Nederlands en digitalisering.

Specifieke projecten

D.2.1. CLARIAH+ Nederland

Het CLARIAH (Common Lab for Research in the Arts and Humanities) CORE-project (2015-2018) was erop gericht een gemeenschappelijke infrastructuur tot stand te brengen voor data-intensief wetenschappelijk onderzoek in de geesteswetenschappen. Vanaf begin 2019 tot en met 2023 loopt het vervolgproject CLARIAH-PLUS, waarin het accent nog meer gericht is op het concreet ondersteunen van de onderzoeker door middel van het tot stand brengen van (virtuele) onderzoeksomgevingen. Het INT houdt zich onder andere bezig met een verbetering van de infrastructuur voor historisch Nederlands, uitbreiding op de corpuszoekmachine BlackLab naar parallelle corpora en treebanks, hulpmiddelen voor het aanbrenen van persistente gebruikersannotaties in corpuszoekresultaten, een gebruikersvriendelijker digitalisatieworkflow en curatie van dialectwoordenboekdata.

In het kader van de infrastructuur voor historisch Nederlands is voortgewerkt aan het platform GALAHAD voor betere en gebruikersvriendelijke taalkundige verrijking van historisch Nederlands, dat inmiddels een eerste publieke testrelease nadert. De nieuwe versie van de annotatietool COBALT is klaar voor intern gebruik, en kan, na begin 2022 op deze wijze uitvoerig getest te zijn, ook publiek ter beschikking gesteld worden. Wat betreft de zoekmachine BlackLab is ervoor gekozen om in eerste instantie in te zetten op een integratie met het gedistribueerde platform Apache Solr, waardoor de

schaalbaarheid van de applicatie verbeterd kan worden.

Binnen de infrastructuur voor digitalisatie en conversie is gewerkt aan een workflow-builder, waarmee de gebruiker een met behulp van de OCR-D-toolset gebaseerd digitalisatietraject kan opbouwen.

Voor de use-case 17^e-eeuwse kranten nadert de curatie van de reeds gedigitaliseerde set haar voltooiing, en staat het traject voor digitalisatie van de aanvullende data in de steigers.

D.2.2. CLARIAH Vlaanderen

Het INT was betrokken bij de indiening van CLARIAH-VL: Advancing the open humanities service infrastructure, als 3^e partij bij de Vlaamse FWO/EWI-call met betrekking tot International Research Infrastructures. Dit project werd gehonoreerd en het INT voorziet de nodige infrastructuur voor het opzetten van het Digital Text Analysis Dashboard & Pipeline, dat grotendeels overeenkomt met het CLARIN-luik in deze aanvraag. Er werd in die hoedanigheid regelmatig overlegd met onderzoekers van de UGent, UAntwerpen en KU Leuven.

D.3. European Language Resources Coordination Initiative (ELRC)

ELRC is een doorlopend initiatief dat als doel heeft tekstdata te verzamelen in alle EU-lidstaten, IJsland en Noorwegen, die gebruikt kunnen worden om eTranslation (de automatische vertaaldienst van de Europese Commissie) verder te ontwikkelen. De kwaliteit van een automatische vertaling hangt onvermijdelijk samen met de kwaliteit en kwantiteit van de taalbronnen die worden gebruikt om het systeem te ‘trainen’. Grote hoeveelheden taaldata zijn dan ook nodig om de kwaliteit van de Nederlandse vertalingen te verbeteren. Dit alles is van groot belang om taalbarrières in Europa te slechten en de nationale talen, in dit geval de Nederlandse taal, te behouden in de digitale informatiemaatschappij.

Het INT leidt inmiddels het ELRC-initiatief. Op 11 juni is vanuit het INT de derde ELRC-workshop in Nederland georganiseerd. Doel van de workshop was om het belang van het verzamelen van tekstdata t.b.v. het trainen van eTranslation verder te benadrukken en om de bekendheid en het gebruik van eTranslation binnen Nederland te promoten. De workshop vond online plaats en genoot een grote belangstelling.

E. Taalmaterialen

E.1. Website taalmaterialen

In 2020 is een groot aantal productpagina's van de website voor Taalmaterialen aangepast om zo een grotere uniformiteit te bereiken. Verder is een aantal scripts ontwikkeld voor het automatiseren van veelvoorkomende handelingen en de documentatie daarvan verbeterd.

Het totale aantal downloads voor zowel commerciële als niet-commerciële producten is afgenomen in vergelijking met het voorgaande jaar:

- Niet-commerciële downloads 930 (1412 in 2020)
- Commerciële downloads 27 (44 in 2020)

Het is niet duidelijk waar dit aan ligt. Een mogelijke verklaring is dat er ten gevolge van de lancering van de nieuwe website en de vereenvoudigde methode om te kunnen downloaden er in 2020 een kortdurende toename van belangstelling is geweest.

E.1.1. Nieuwe producten

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

- Een belangrijke update van het Corpus Hedendaags Nederlands (CHN). Het Corpus Hedendaags Nederlands (CHN) is een tekstverzameling met meer dan 2,5 miljoen teksten uit boeken, blogs, kranten, tijdschriften en journaaluitzendingen uit Nederland, Vlaanderen, Suriname en de Nederlandse Antillen. Samen zijn deze teksten goed voor bijna 1 miljard woorden. Vanwege auteursrechten is dat corpus alleen toegankelijk met een CLARIN-account.
- Hoger Onderwijs Terminologie in Nederland en Vlaanderen (HOTNeV). De aanzet voor dit project vormde de sterke toename van onderwijstermen, gegenereerd door het onderwijsbeleid van de EU en geïmplementeerd door het Tuning Project. HOTNeV beoogt een tweeledig doel. Waar tot op heden Nederlandstalige equivalenten voor de Engelstalige terminologie vooral ad hoc werden gecreëerd, vestigt dit project de aandacht op de noodzaak om gecoördineerd te voorzien in termen die zijn geaccordeerd door partijen in de Nederlandstalige onderwijswereld. Ook wil het de haalbaarheid van dit streven laten zien.
- BLISS Dialogue Summaries. Deze dataset bestaat uit Nederlandse spraakopnames van deelnemers die spreken met het BLISS-dialoogsysteem (v1) over alledaagse bezigheden en activiteiten waar ze plezier aan beleven. De data bevat 55 opnames met een gemiddelde duur van 2 minuten en 34 seconden.
- WAI-NOT Corpus. Het WAI-NOT-corpus bestaat uit 874 krantenartikels, afkomstig uit de

WAI-NOT-krant. De artikels zijn opgesteld in eenvoudig te lezen Nederlands en zijn afkomstig uit de periode 2009-2021. Het corpus bevat ongeveer 75.000 woorden.

- Een update van RND Woordenlijsten. De Reeks Nederlandse Dialectatlassen is een serie boeken, uitgegeven tussen 1925 en 1976, met daarin de fonetische transcripties van een vast aantal zinnen. Daarbij is de uitspraak vastgelegd van zegslieden uit een groot aantal dorpen en steden in Nederland en Vlaanderen.
- Algemene Nederlandse Spraakkunst (e-ANS). De ANS wil een zo volledig mogelijke beschrijving geven van de grammaticale aspecten van het hedendaagse Standaardnederlands (in zijn geografische en stilistische verscheidenheid), die bedoeld is voor een breed publiek van grammaticaal geïnteresseerden en dus niet uitsluitend – en ook niet in de eerste plaats – voor gespecialiseerde taalkundigen.
- Woordcombinaties. Woordcombinaties is een online naslagwerk in ontwikkeling dat gevorderde leerders en gebruikers van het Nederlands ondersteunt bij het gebruiken van woorden in context. De tool toont hoe woorden gebruikt worden in voorbeeldzinnen, welke woorden typisch en/of vaak met elkaar gecombineerd worden en hoe (valentie)patronen samen met collocaties gebruikt worden voor het bouwen van zinnen.
- Nederlands als Wetenschapstaal: Scheikunde en Wiskunde. In het proefproject komen twee aspecten aan bod: enerzijds het ontwikkelen van corpusmateriaal van gesproken academisch Nederlands, dit om het academische taalgebruik in kaart te brengen en taalondersteuning voor beginnende studenten te ontwikkelen; anderzijds het ontwikkelen van terminologielijsten voor een aantal kernvakken uit de BA1-programma's in Nederland en Vlaanderen. Wat betreft de terminologielijsten richt het pilootproject zich op de alfa-, bèta- en gamma-richtingen op basis van schriftelijk cursusmateriaal uit Nederland en Vlaanderen. Dit project kwam tot stand in overleg met en met steun van de Nederlandse Taalunie.
- Een update van het Lassy Klein-corpus. Het Lassy Klein-corpus is een corpus van ongeveer 1 miljoen woorden met manueel geverifieerde syntactische annotaties. Lemma's en POS-tags werden automatisch toegevoegd aan het corpus m.b.v. Tadpole (nu: Frog). De syntactische dependentiestructuren werden toegevoegd m.b.v. Alpino. De lemma's, POS-tags en syntactische boomstructuren werden geverifieerd en gecorrigeerd.
- Medische Pilot (MedPilot). Medische Pilot is een bij wijze van experiment ingerichte database waarin een klein deel van de medische woordenschat beschreven wordt op verschillende niveaus, van wetenschappelijk tot toegankelijk voor laaggeletterden, en waarin ook verschillen tussen Vlaamse en Nederlandse termen worden getoond.
- Brieven als Buit-2. Brieven als Buit-2 (BAB2) is een spin-off van het onderzoeksprogramma Brieven als Buit / Letters as Loot. Op diverse punten verschilt BAB2 van het oorspronkelijke BAB-corpus. Ten eerste omvat het een extra brievencollectie uit de Prize Papers (National

Archives, Kew, UK). Ten tweede is een groter deel (26%) van de 1386 brieven in BAB2 zakelijk of deels zakelijk van aard, waar dat bij BAB voor slechts 10% gold.

E.1.2. 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 (155), daarna België (70) en verder Duitsland (19), Zwitserland (5), Verenigd Koninkrijk (5), en lager.

De e-mailadressen geven niet een volledig beeld van waar de gebruikers zitten, omdat veel een ambigue extensie hebben als .com of .org. Daarom hebben we gekeken naar de geolocatie van de ip-adressen die zijn geregistreerd bij de downloads. Dat leverde het volgende overzicht op: Nederland (195), België (68), Zweden (11), Verenigde Staten (8), Duitsland (8), Canada (6), Spanje (4) en een groot aantal lagere scores.

E.2. CLARIN-ERIC: het INT als CLARIN-centrum

Het INT was als derde partij betrokken bij de indiening van CLARIN-VL bij de Vlaamse FWO/EWI-call met betrekking tot International Research Infrastructures. Dit project werd goedgekeurd, en heeft inmiddels geleid tot de oprichting van CLARIN-België en het volwaardig Belgisch lidmaatschap van CLARIN. Het INT is tot nog toe het enige CLARIN-B Centre (technisch centrum) voor België. Het INT levert met Vincent Vandeghinste de nationale coördinator voor CLARIN België, en maakt deel uit van het National Coordinators Forum. Daarnaast levert het INT met Jesse de Does ook een vertegenwoordiger voor de Standing Committee of CLARIN Technical Centres (SCCTC). Het INT host ook de [website van CLARIN-BE](#).

In het kader van *user involvement* werd CLARIN door het INT voorgesteld aan studenten van de KU Leuven bij het vak Computationale Lexicografie voor de Master Taalkunde en bij het vak Language Engineering Applications, een onderdeel van de Master Artificial Intelligence.

In de zomer van 2021 kreeg het INT de officiële goedkeuring van de Europese onderzoeksinfrastructuur CLARIN om een zogenaamd CLARIN Knowledge Centre te worden: [K-Dutch](#).

F. Onderzoek: Europese aanvragen en competitieve projecten

F.1. European Lexicographic Infrastructure (ELEXIS)

Het INT is partner in ELEXIS, een Europees Horizon 2020-project dat in februari 2018 van start is gegaan en een looptijd heeft van 4 jaar. Vanwege de coronacrisis is de looptijd van het project met 6 maanden verlengd tot en met juli 2022. ELEXIS is een samenwerkingsverband tussen 17 partners uit Europa en Israël en wordt geleid door het Jožef Stefan Instituut (Slovenië).

Het doel van het project is om een infrastructuur voor e-lexicografie op te zetten. ELEXIS streeft ernaar de lexicografische inspanningen binnen Europa zoveel mogelijk te harmoniseren door best practices te formuleren, conversietools te ontwikkelen en, belangrijker nog, door de bestaande lexicografische bronnen aan elkaar te koppelen, zodat ze kunnen worden gebruikt om nieuwe data, technologieën, producten en diensten te ontwikkelen. Tevens zal de ELEXIS-infrastructuur door training en educatie bijdragen aan het verkleinen van verschillen in expertise tussen lexicografen in Europa.

Het INT leidt het werkpakket ‘Lexicographic data and workflow’. Binnen dit werkpakket is in 2021 verder gewerkt aan het ontwikkelen van een common vocabulary voor het ELEXIS-datamodel. Ook is verder gewerkt aan Elexifier, een cloudgebaseerde conversietool voor het converteren van legacy XML- en pdf-woordenboeken naar een uniform formaat conform het ELEXIS-datamodel (zie o.a. D1.4 ELEXIS Conversion Tools). Deze tool wordt binnen ELEXIS ingezet om de lexicografische bronnen van partners en observers te converteren en te integreren in de ELEXIS-infrastructuur. Tot slot is in het kader van dit werkpakket de survey ‘Lexicographic Practices in Europe: A Survey of User Needs for Observers’ afgerond en zijn de resultaten geanalyseerd.

Daarnaast werkt het INT mee aan andere werkpakketten, met name de werkpakketten ‘Interoperability and Linked (Open) Data’ (o.a. ontwikkelen van datasets voor het (monolingual en crosslingual) linken van woordenboeken), ‘Lexicographic data for NLP’ (o.a. ontwikkelen van een parallel sense-annotated dataset in 10 Europese talen), ‘NLP for lexicography’ (o.a. verder ontwikkelen van Lexonomy) en ‘Training and Education’ (ontwikkelen van trainingsmateriaal voor het ELEXIS-curriculum dat via DARIAH-CAMPUS beschikbaar zal worden gesteld).

Door de coronacrisis zijn er in 2021 geen gastonderzoekers met een ELEXIS trans-national access-beurs ontvangen op het INT.

F.2. COST CA16105: European Network for Combining Language Learning with Crowdsourcing Techniques (enetCollect)

De COST-actie *enetCollect* ging in 2021 het laatste jaar in. Deze COST-actie heeft vijf werkgroepen die alle aspecten van het combineren van crowdsourcing met het leren van (een) taal dekken. Frieda Steurs is vicevoorzitter van werkgroep 3 ‘User-oriented design strategies for a competitive solution’. Carole Tiberius is lid van het Management Committee. Verder is Mathieu Fannee bij deze actie betrokken.

Er werden tijdens de pandemie veel webinars georganiseerd, en het managementteam vergaderde online.

Het INT werkte wel verder met een van de andere partners in het netwerk, Ben Vaske en Anne-Lies Schrijvers van oefenen.nl, aan implementaties om computerondersteund taalleren aantrekkelijker te maken. Verder werden de laatste rapporten aangeleverd om het project te finaliseren.

F.3. COST CA18209: European network for Web-centered linguistic data science

Het INT neemt deel aan de NexusLinguarum COST-actie. Het thema van deze actie is ‘linguistic data science’, een deelgebied binnen het opkomende wetenschappelijk dataonderzoek. Taalkundige data vormen een specifiek geval en zijn tot nu toe nog grotendeels onontgonnen in een bigdatacontext. Het hoofddoel van NexusLinguarum is om taalkundigen, computerwetenschappers, terminologen en andere belanghebbenden in één netwerk bij elkaar te brengen om zo samenwerking en kennisdeling op het gebied van ‘linguistic data science’ te bevorderen. De actie is eind oktober 2019 van start gegaan en heeft een looptijd van 4 jaar.

Vanwege de coronacrisis vonden de meeste evenementen van deze COST-actie in 2021 virtueel of in hybride vorm plaats. Van 8-12 februari werd de EUROLAN Training school on ‘Linked Data for Linguistics’ georganiseerd. Daarnaast was er de jaarlijkse plenaire bijeenkomst (in Skopje en virtueel) van 28-30 september.

F.4. H2020 - RIA 101017255 SignON: Sign Language Translation Mobile Application and Open Communications Framework

Het INT is consortiumpartner in het SignON-project, een Horizon 2020-project gefinancierd door de Europese Commissie. Het hoofddoel van dit project is het opzetten van automatische vertaalservices

tussen gebarentalen en zogenaamde *gesproken* talen. 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 Nederlandse Taalunie, Radboud Universiteit Nijmegen, Tilburg University, en als derde partij Beeld en Geluid. Het project wordt geleid door Dublin City University. Het INT voorziet in dit project de data-infrastructuur en verzamelt bestaande gebarentaalcorpora, zoals het corpus Vlaamse Gebarentaal en het corpus Nederlandse Gebarentaal, en maakt die beschikbaar voor andere consortiumpartners. Hiervoor werd een sftp-server opgesteld met ruime opslagmogelijkheden, aangezien het hier om grote hoeveelheden videodata gaat. Er werd ook een krachtige server met gpu klaargemaakt waarop de toepassingen die binnen dit project ontwikkeld worden kunnen draaien. Er wordt ook gewerkt aan een Belgian.news.covid-corpus: alle persconferenties van het COVID-overlegcomité (met gebarentaaltolken) in België worden verzameld, en er werd een prototype gebouwd van een spraakherkenner voor het Belgisch Nederlands, zodat dit corpus aangevuld kan worden met automatische transcripties.

F.5. KULeuven C3: SABeD – Spoken Academic Belgian Dutch

Het SaBeD-project loopt aan de KU Leuven, waar in 2021 videodata verzameld werd. Dit project wordt door het INT nauwgezet opgevolgd via Vincent Vandeghinste, die bij de KU Leuven copromotor is van dit project. Eerste tests met de spraakherkenner voor het Belgisch Nederlands (cf. SignON project) doen vermoeden dat die in de transcriptiefase van dit project zal leiden tot een aanzienlijke versnelling ten opzichte van volledig manuele transcriptie.

F.6. NWO Open Competition SSH Spread the new(s)

Het project ‘Spread the new(s). Understanding standardization of Dutch through 17th-century newspapers’, dat wordt uitgevoerd met subsidie van NWO Open Competition SSH, maakte in 2021 goede vorderingen: promovenda Machteld de Vos, die gedeeltelijk op het INT en gedeeltelijk bij de Radboud Universiteit werkt, heeft in 2021 de normatieve werken uit de 16e en 17e eeuw in kaart gebracht. In 2022 zal zij de gegevens uit deze werken toetsen aan het daadwerkelijke taalgebruik binnen een corpus dat onder andere bestaat uit de 17e-eeuwse kranten die door vrijwilligers zijn getranscribeerd. Dit corpus wordt door het INT voor alle onderzoekers beschikbaar gesteld.

F.7. The rise of complex verb constructions in Germanic

Het onderzoeksproject ‘The rise of complex verb constructions in Germanic’, uitgevoerd door Evie

Coussé, Gerlof Bouma en Nicoline van der Sijs en gefinancierd door de Vetenskapsrådet (Swedish Research Council), is in 2021 voltooid met de oplevering van het zogenoemde EDGeS Corpus, een parallel historisch bijbelcorpus met 36 bijbels in vier talen (Engels, Nederlands, Duits, Zweeds) van de 14e tot de 21e eeuw, gelanceerd op 28 mei 2021. Achter log-in is het hele corpus beschikbaar op opus.nlpl.eu, terwijl de 30 bijbels waarop geen copyright rust, kunnen worden gedownload op spraakbanken.gu.se. Het corpus zal ook beschikbaar worden gesteld via de zoekapplicatie van het INT met extra Nederlandse en Zuid-Afrikaanse bijbelvertalingen en bijbelvertalingen in Nederlandse dialecten.

F.8. MacBERTh

In 2021 heeft het INT samengewerkt met het project MacBERTh - Bidirectional Encoder Representations from Transformers of historical English and Dutch met als hoofdaanvrager Lauren Fonteyn binnen de call PDI-SSH digitale infrastructuur, dat in dat jaar is gestart.

F.9. European Language Grid

In de loop van 2021 nam het INT de taken van de Universiteit Utrecht over m.b.t. de European Language Grid (ELG). Het INT werd National Competence Centre (NCC) en Vincent Vandeghinste (INT) werd NCC-lead voor ELG voor Nederland. Op 3 december werd samen met de Nederlandse Organisatie voor Taal- en Spraaktechnologie (NOTaS) een workshop georganiseerd waarin de ELG aan de hand van een aantal lezingen geïntroduceerd werd.

F.10. European Language Equality

In 2021 startte het project European Language Equality (ELE). Het project loopt parallel aan ELG (zie hierboven). Ook hier is het INT het national competence centre (ncc) voor Nederland en zorgt het samen met de Universiteit Antwerpen (ncc voor België) voor de gegevens omtrent het Nederlands. Er werd een onderzoek ingesteld naar de stand van de digitale taalmaterialen voor het Nederlands, de database werd up-to-date gemaakt en we startten met de voorbereiding van een verslag over de situatie van de digitale ondersteuning van het Nederlands. Hier wordt in 2022 tijdens de Metaforumconferentie verslag van uitgebracht.

G. Doelgroepenbeleid (inclusief onderwijs)

G.1. Onderzoek, onderwijs en het algemene publiek

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

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

Op de website van het INT is een apart menu-item voor onderwijs ingericht, onderverdeeld in informatie voor docenten, leerlingen en studenten. Het INT heeft webinars georganiseerd over de lancering van de nieuwe e-ANS en over het *Woordenboek van Nieuwe Woorden*. Tijdens de Week van het Nederlands presenteerde het INT een webinar over Woordcombinaties, een nieuwe tool die vooral voor leerders van Nederlands als tweede en vreemde taal van groot nut is. Verder heeft er verschillende malen overleg plaatsgevonden met de Nederlandse Taalunie over gerichte activiteiten die inspelen op de noden van docenten Nederlands in het buitenland (Nederlands als Vreemde Taal, NVT).

Ook het algemene publiek wordt niet uit het oog verloren. Zo was het INT in 2021 betrokken bij de denktank rondom een nieuw Letterenfestival in Leiden. Op onze website en socialemediakanalen verschenen veel actuele bijdragen en het instituut stuurde nieuwsbrieven en podcasts de wereld in. Zie voor dit alles het volgende hoofdstuk. Daarnaast hielden medewerkers voordrachten, waren zij te horen in radioprogramma's en droegen zij bij aan publicaties voor een algemeen publiek dat belangstelling heeft voor taal in het algemeen en Nederlands in het bijzonder. Zie verder de bijlagen.

H. Wetenschapscommunicatie en publieksbereik

H.1. Website

De website is het gezicht van het instituut naar buiten en daarmee het belangrijkste communicatiemiddel. Bijna alle communicatie-uitingen zijn erop gericht om bezoekers naar de website te trekken. In 2021 telde ivdnt.org een half miljoen paginaweergaven. Minder dan in 2020, maar de cijfers zijn vermoedelijk vertekend doordat in 2021 veel meer ‘onzuivere’ bezoeken, van bots e.d., zijn weggefilterd. De toegenomen gemiddelde tijd op de pagina lijkt dat te bevestigen.

De nieuwe woorden van de week worden het vaakst bezocht; daarna de historische woordenboeken. Hoewel het algemene bezoek gedaald lijkt te zijn, waren de volgende delen van de site juist beduidend meer in trek dan in 2020:

- historische woordenboeken
- hedendaags Nederlands
- spelling en grammatica
- jaaroverzicht nieuwe woorden
- terminologie
- podcast

Een ‘hit’ was de doorverwijzing vanaf de website [Joop](#) in een artikel over *blank* en *wit* in het WNT.

De populairwetenschappelijke rubrieken Nieuw woord van de week en Terug in de Taal scoren elke week goed en worden veel gedeeld. Ze hebben gezelschap gekregen van de rubriek Uit de streek, over dialectwoorden. Ook nieuw op de site waren de columns van Ewoud Sanders (WoordHoek, een voortzetting van zijn column in *NRC Handelsblad*) en Ludo Permentier (Woorden weten alles). Elke aflevering trekt een paar duizend lezers, met als uitschieter Sanders’ column ‘Jodenkoek’ (ruim 5000) en van Permentier ‘Over mannetjes en vrouwtjes’ (ruim 3500).

Aan de nog nieuwe website (2020) is sinds het aantreden van de nieuwe webmaster in maart 2021 op technisch vlak veel verbeterd. Heel wat handelingen zijn geautomatiseerd en vereenvoudigd, waardoor bijdragers aan de site geen html meer hoeven te bewerken. Diverse plugins maken het werk van de webredactie sneller en simpeler en voorkomen dat bezoekers op fouten stuiten, bijvoorbeeld op de voorpagina en in de Termenlijst. Ook de publicatielijst is ‘strakgetrokken’.

Tot slot is de site beter toegankelijk, zowel qua leesbaarheid op grote en kleine schermen als qua snelheid. De laadtijd van pagina’s zal nog verder verbeterd worden.

H.2. Nieuwsbrieven

De algemene nieuwsbrief is om geïnteresseerden te blijven binden aan en te informeren over het instituut. De nieuwsbrief terminologie is een aparte thematische nieuwsbrief, waarop abonnees zich los van de algemene nieuwsbrief kunnen inschrijven. In 2021 zijn er zes algemene nieuwsbrieven verschenen en vier nieuwsbrieven over terminologie. Daarnaast zijn er nog diverse aparte mailings verzonden.

De algemene nieuwsbrief telde eind 2021 4.380 abonnees, van wie per editie bijna de helft van de ontvangers de nieuwsbrieven opent; van de 3.239 geadresseerden van de nieuwsbrief terminologie doet een derde dat.

H.3. Sociale media

Het instituut gebruikt met name Twitter als socialemediakanaal voor het delen van nieuws, evenementen en kennis over taal. Het aantal volgers op Twitter blijft al jaren gestaag groeien en bedroeg eind 2021 zo'n 5.500. Daarnaast is het INT aanwezig op Facebook en LinkedIn. Via het eigen YouTube-kanaal waren diverse webinars, soms met honderden kijkers, live te volgen en achteraf terug te zien. In 2021 debuteerde het INT bovendien op Instagram, waar de huisstijl er goed uit springt.

H.4. Podcasts

Na het succes van de eerste podcast 'Waar komt pindakaas vandaan?' in 2020 volgde in 2021 een tweede reeks van zeven afleveringen waarin onder meer eten, drinken, kleding en uitdrukkingen voor de dood werden besproken. Daarnaast lanceerden Onze Taal en het INT in mei een nieuwe podcast met langere afleveringen, 'Over taal gesproken'. Daarin spreken Laura van Eerten en Raymond Noë met een deskundige over het Nederlands, bijvoorbeeld over grammatica, nieuwe woorden of het oudste Nederlands.

Elke aflevering van 'Over taal gesproken' is tot eind 2021 twee- tot zesduizend keer gedownload. Het publiek komt voor 76% uit Nederland en voor 11% uit België. De eerste aflevering, over het Nederlands in Nederland en België, scoort het hoogst, maar is ook al het langst beschikbaar. Het aantal downloads van de tweede serie 'Waar komt pindakaas vandaan?' is heel constant: tussen twee- en drieduizend per aflevering. De eerste serie uit 2020 scoort inmiddels tussen vijf- en dertienduizend downloads per aflevering. Het publiek is voornamelijk Nederlands (87%, tegen 4% uit België).

H.5. Boek

In november verscheen bij uitgeverij Scriptum *Knuffelcontact en waterwappie. De kracht van nieuwe woorden* van Vivien Waszink en Veronique De Tier, over nieuwe woorden en hoe ze ontstaan. Het kreeg alom veel aandacht en het verkocht meteen goed. Een uitgebreid overzicht van publicaties en media-optredens is te vinden in bijlage 3.

H.6. Neerlandistiek

Het INT is sinds 2020 samen met het Meertens Instituut verantwoordelijk voor de (technische) ondersteuning van het elektronische tijdschrift voor de Nederlandse taalkunde, letterkunde en taalbeheersing *Neerlandistiek*, en een medewerker van INT heeft zitting in de vierkoppige redactie. Daarnaast faciliteert INT de maandelijks verschijnende nieuwsbrief *Neerlandistiek* voor de klas. In april werd bekend dat *Neerlandistiek* een subsidie van tienduizend euro krijgt uit het Pilotfonds Wetenschapscommunicatie door Wetenschappers Gewaardeerd! van de KNAW.

Bijlage 1: Raad van Toezicht en Raad van Advies

Raad van Toezicht

Drs. Paul Rüpp, voorzitter

Prof. mr. Jan Cerfontaine

Drs. Gertine van der Vliet

Raad van Advies

Prof. dr. Hans Bennis (voorzitter)

Prof. dr. Willy Vandeweghe (vicevoorzitter)

Prof. dr. Paula Fikkert

Prof. dr. Dirk Geeraerts

Prof. dr. Jack Hoeksema

Prof. dr. Veronique Hoste

Drs. Jan Jaap Knol

Prof. dr. Reinhild Vandekerckhove

Lic. Wim Vanseveren

Prof. dr. Gert Oostindie (afscheid in november 2021 wegens emeritaat)

Bijlage 2: Medewerkers

Prof. dr. Frieda Steurs – directeur/bestuurder

Dr. Maaike Beliën – onderzoeker/taalkundige

Marjolijn van Bennekom – taalkundig assistent

Dr. Bob Boelhouwer – computerlinguïst

Dr. Roland de Bonth – onderzoeker/taalkundige

Tim Brouwer – softwareontwikkelaar

Lic. Lut Colman – onderzoeker/taalkundige

Lic. Griet Depoorter – onderzoeker/taalkundige

Lic. Katrien Depuydt – onderzoeker/taalkundige

Lic. Veronique De Tier –

onderzoeker/taalkundige

Dr. Jesse de Does – computerlinguïst

Job van Doeselaar – webmaster (sinds 1 maart

2021)

Laura van Eerten MA – communicatieadviseur

Drs. Mathieu Fannee – systeemontwikkelaar

Lic. Dirk Geirnaert – onderzoeker/taalkundige

(pensioen in oktober 2021)

Drs. Thomas Haga – taalkundig assistent

Dr. Kris Heylen – onderzoeker/taalkundige

Ruud de Jong – systeemontwikkelaar (sinds 1

februari 2021)

Dr. Dirk Kinable – onderzoeker/taalkundige

Drs. Marco van der Laan – systeembeheerder

Dr. Frank Landsbergen – computerlinguïst

Koen Mertens – systeemontwikkelaar

Drs. Jan Niestadt – systeemontwikkelaar

Henk van der Pol – systeemontwikkelaar

Wil de Ruyter – taalkundig assistent (pensioen in

juni 2021)

Dr. Kathy Rys – onderzoeker/taalkundige

Prof. dr. Nicoline van der Sijs –

onderzoeker/taalkundige

Rob van Strien – programmeur/ontwikkelaar

(pensioen in april 2021)

Paulette Tacx – managementassistent

Drs. Rob Tempelaars – onderzoeker/taalkundige

Dr. Carole Tiberius – computerlinguïst

Dr. Vincent Vandeghinste –

onderzoeker/taalkundige en CLARIN-

coördinator

Lic. Katrien Van pellicom – projectleider

Spelling

Drs. Boukje Verheij – onderzoeker/taalkundige

Rayvano van Vliet – systeembeheerder

Machteld de Vos, MPhil – onderzoeker in

opleiding

Drs. Vivien Waszink – onderzoeker/taalkundige

Karin van Weerlee – managementassistent

Laetitia de Winter – administratief medewerker

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

Publicaties

- Beelen, Hans, Nicoline van der Sijs (2021). Topinamboer. In: *Onze Taal* 10, 30.
- Beelen, Hans, Nicoline van der Sijs (2021). Elften en finten. *Onze Taal* 5, 30.
- Beelen, Hans, Nicoline van der Sijs (2021). Groenlands. *Onze Taal* 4, 30.
- Beelen, Hans, Nicoline van der Sijs (2021). In cohort. *Onze Taal* 2/3, 30.
- Beelen, Hans, Nicoline van der Sijs (2021). Krambamboeli. *Onze Taal* 1, 30.
- Beelen, Hans, Nicoline van der Sijs (2021). Otaksa en ginkgo. *Onze Taal* 6, 26.
- Beelen, Hans, Nicoline van der Sijs (2021). De reuzenkoeskoes. In: *Onze Taal* 7/8, 30.
- Beelen, Hans, Nicoline van der Sijs (2021). De vier seizoenen. In: *Onze Taal* 9, 30.
- Beelen, Hans, Nicoline van der Sijs (2021). Koper. In: *Onze Taal* 12, 30.
- Beelen, Hans, Nicoline van der Sijs (2021). Tarp en palliatief. In: *Onze Taal* 11, 30.
- Beliën, Maaïke (2021). 17 De adpositieconstituent. *Algemene Nederlandse Spraakkunst*.
- Beliën, Maaïke (2021). 9 De adpositie (het voorzetsel, achterzetsel, omzetsel en meer). *Algemene Nederlandse Spraakkunst*.
- Bonth, Roland de (2021). Achter het achtervoegsel (1): wasserette. Op: www.neerlandistiek.nl
- Bonth, Roland de (2021). Nieuwe versie corpus Brieven als Buit online. Op: www.neerlandistiek.nl
- Bonth, Roland de (2021). Een schitterend lexicon: DiaMaNT. Op: www.neerlandistiek.nl
- Bonth, Roland de (2021). Praten over historischeletterkundefragmenten. Op: www.neerlandistiek.nl
- Bonth, Roland de (2021). Brieven als Buit-2 staat online. Op: www.neerlandistiek.nl
- Bonth, Roland de (2021). Achter het achtervoegsel (3): kitchenette. Op: www.neerlandistiek.nl
- Bonth, Roland de (2021). De populairste voornamen uit de 17^e en 18^e eeuw. Op: www.neerlandistiek.nl
- Bonth, Roland de (2021). Vragen over Vestdijk – een Vestdijkquiz. Op: www.neerlandistiek.nl
- Bonth, Roland de (2021). Een andere methode. Op: www.neerlandistiek.nl
- Bonth, Roland de (2021). *Verba van Vestdijk*. Op: ivdnt.org
- Bonth, Roland de (2021). *Verba van Vestdijk, een gelegenheidswoordenboekje*. Op: www.neerlandistiek.nl
- Bonth, Roland de (2021). Uitslag Vragen over Vestdijk – een Vestdijkquiz. Op: www.neerlandistiek.nl
- Bonth, Roland de (2021). De populairste voornamen uit de middeleeuwen. Op: www.neerlandistiek.nl
- Bonth, Roland de (2021). *Nederlandsch hertaelt junior*. Op: www.neerlandistiek.nl
- Bonth, Roland de (2021). *Nederlandsch hertaelt junior – De uitslag*. Op: www.neerlandistiek.nl

- Bonth, Roland de (2021). Achter het achtervoegsel (5): Opelette. Op: www.neerlandistiek.nl
- Bonth, Roland de (2021). Met betrekking tot de E-ANS. Op: www.neerlandistiek.nl
- Bonth, Roland de (2021). “Met de Y van Ysco”. Op: www.neerlandistiek.nl
- Bonth, Roland de (2021). Jozef en zijn jas. Op: www.neerlandistiek.nl
- Bonth, Roland de (2021). Teksten verrijken met tags van TDN. Op: www.neerlandistiek.nl
- Bonth, Roland de (2021). Etymologica: bedenkers van woorden. Op: www.neerlandistiek.nl
- Bonth, Roland de (2021). “Geen huisgezin meer zonder muziek”. Achter het achtervoegsel (7). Op: www.neerlandistiek.nl
- Bonth, Roland de (2021). Etymologica zomervakantiequiz. Op: : www.neerlandistiek.nl
- Bonth, Roland de (2021). Antwoorden en uitslag van de Etymologica zomervakantiequiz. Op: www.neerlandistiek.nl
- Bonth, Roland de (2021). “De u doet int uyt spreken de lippen voor uyt puilen”. Op: www.neerlandistiek.nl
- Bonth, Roland de (2021). Willem Frederik Hermans en het WNT. Op: www.neerlandistiek.nl
- Bonth, Roland de (2021). Een moordette in de kitchenette. Achter het achtervoegsel (9). Op: www.neerlandistiek.nl
- Bonth, Roland de (2021). Etymologica: het hora van de waarheid. Op: www.neerlandistiek.nl
- Bonth, Roland de (2021). Woorden van W.F. Hermans. Op: ivdnt.org
- Bonth, Roland de (2021). Op reis in Delpher. Op: www.neerlandistiek.nl
- Bonth, Roland de (2021). Neerlandistiek ín de klas. Op: www.neerlandistiek.nl
- Bonth, Roland de (2021). Corpus Hedendaags Nederlands online. Op: www.neerlandistiek.nl
- Bonth, Roland de (2021). Mobylette. Achter het achtervoegsel (11). Op: www.neerlandistiek.nl
- Bonth, Roland de (2021). Etymologica: Bilderdijks bedenkzels. Op: www.neerlandistiek.nl
- Bonth, Roland de (2021). “Mijn woordspelen is leeren”. Op: www.neerlandistiek.nl
- Bonth, Roland de (2021). Van kaper naar corpus. Onderzoek naar historisch taalgebruik met *Brieven als Buit*. In: *de lage landen*, nummer 3, 38-43.
- Bonth, Roland de (2021). Privateers Show How the Dutch Used to Write. Op: the-low-countries.com
- Bonth, Roland de (2021). Elke vier – vanaf oktober elke twee – weken een bijdrage aan de webrubriek Terug in de Taal. Op: ivdnt.org.
- Booij, Geert, N. Jansen, Kathy Rys (2021). 1 De klankleer van het Nederlands. *Algemene Nederlandse Spraakkunst*.
- Chambers, Sally, Wout Dillen, Tom Gheldof, Wouter Ryckbosch, Vincent Vandeghinste, Christophe Verbruggen, Els Lefever (2021). Interdisciplinary Resilience: critical humanities infrastructure in the times of a global pandemic: a Flemish case study. *DH Benelux 2021*. Online / Leiden, Netherlands.
- Colleman, Timothy, Johan De Caluwe, Walter Haeseryn, Ronny Boogaart, Frank Landsbergen, Johan

- Van Hoorde (2021). Over de Algemene Nederlandse Spraakkunst. *Algemene Nederlandse Spraakkunst*.
- Colman, Lut, Carole Tiberius (2021). Woordcombinaties / Word Combinations: all-in-one resource for Dutch words in context. In Jean-Pierre Colson (ed.), *Book of Abstracts Europhras 2021*. Louvain-la-Neuve, 6-9 September, 2021. 40-41.
- Colman, Lut, Carole Tiberius (2021). Woorden combineren met Woordcombinaties. In *Neerlandistiek*. 21 July 2021.
- De Caluwe, Johan, Veronique De Tier, A.S. Ghyselen, R. Vandenberghe (2021). *Atlas van het Dialect in Vlaanderen*. Lannoo.
- Depuydt, Katrien, Jesse de Does (2021). Linking the Dictionary of Old Dutch to A Thesaurus of Old English. In: *Amsterdamer Beiträge zur älteren Germanistik*, 81(3-4), 493-513. doi: <https://doi.org/10.1163/18756719-12340240>
- De Tier, Veronique (2021). Het immateriële landschap in Zeeland. In: *Zeeuws Erfgoed*, jg. 20, 2, pp. 4-7.
- De Tier, Veronique (2021). Ingveoonse kenmerken in West-Vlaanderen (1). In: *Diejalektgazette Bachtn de Kuupe*, jg. 16, nr.1, pp. 9-10.
- De Tier, Veronique (2021). Ingveoonse kenmerken in West-Vlaanderen (2). In: *Diejalektgazette Bachtn de Kuupe*, jg. 16, nr.2, pp. 9-10.
- De Tier, Veronique (2021). Leuntje en Merien: Zeeuwse regionale voornamen. In: *Zeeuws Erfgoed*, jg. 20, 4, 9-11.
- De Tier, Veronique (2021). Over flipflops, teensletsen en zeesletsen. In: *Het grote taalmuseumvakantieboek*, pp. 98-99.
- De Tier, Veronique (2021). Van Zeeuwse puierrek en padderdoelekes. In: *Zeeuws Erfgoed*, jg. 20, 1, pp. 24-25.
- De Tier, Veronique (2021). West-Vlaanderen boven in de grote dialecttest. In: *Diejalektgazette Bachtn de Kuupe*, jg. 16, nr.4, pp. 8-12.
- De Tier, Veronique (2021). Zeeuws-Vlaamse dialectkaarten op de computer. In: *Nehalennia*, 211, pp. 21-24.
- De Tier, Veronique (2021). Zwoele avonden in West-Vlaanderen. In: *Diejalektgazette Bachtn de Kuupe*, jg. 16, nr. 3, p. 7.
- De Tier, Veronique, Piet van Sterkenburg (2021). De klaproos in de Zeeuwse dialecten. In: *Nehalennia*, 213, pp. 28-31.
- De Tier, Veronique, S. Chambers, L. Hellebaut, Jacques Van Keymeulen, R. Vandenberghe, Katrien Depuydt, Jesse de Does, K. Mertens, Tanneke Schoonheim (2021). The Dictionary portal of the Southern Dutch Dialects. In Z. Gavriilidou, L. Mitits, & S. Kiosses (Eds.), *EURALEX 19, Congress of the European Association for Lexicography, Abstracts*, pp. 87-90.

- De Tier, Veronique, Marjo van Koppen, Miet Ooms, Jos Swanenberg, Gunter De Vogelaer (2021). *Wie zegt wat waar? Regionale taal in Nederland en Vlaanderen*. Genootschap Onze Taal/Stichting Nederlandse Dialecten.
- Van Eerten, Laura (2021). Professor Weijnens reis naar Leiden. *Neerlandistiek*, 13 juli.
- Erjavec Tomaž, Maciej Ogrodniczuk, Petya Osenova, Andrej Pančur, Nikola Ljubešič, Tommaso Agnoloni, Starkaður Barkarson, María Calzada Pérez, Çağrı Çöltekin, Matthew Coole, Roberts Dargis, Luciana D. de Macedo, Jesse de Does, Katrien Depuydt, Sascha Diwersy, Dorte Haltrup Hansen, Matyáš Kopp, Tomas Krilavičius, Giancarlo Luxardo, Maarten Marx, Vaidas Morkevičius, Costanza Navarretta, Paul Rayson, Orsolya Ring, Michał Rudolf, Kiril Simov, Steinþór Steingrímsson, István Üveges, Ruben van Heusden, Giulia Venturi (2021). ParlaMint: comparable corpora of European parliamentary data. In: Proceedings of CLARIN annual conference 2021, 27-29 September, 2021.
- Geirnaert, Dirk (2021). Marcus Gheeraerts, Source of Inspiration for Tapestry-Designers. Sixteenth-Century Fable-Illustrations Used in Seventeenth-Century Tapestries. In: Van de Haar, A. & Schulte Nordholt, A. (eds.), *Figurations animalières à travers les textes et l'image en Europe, du Moyen Âge à nos jours. Essais en hommage à Paul J. Smith*, pp. 193-209. Brill.
- Haga, Thomas, Katrien Depuydt, Jesse de Does, m.m.v. Roland de Bonth, Dirk Geirnaert (2021). *Tagset voor Diachroon corpusmateriaal van het Nederlands (TDN)*. INT Working Papers 1 (Instituut voor de Nederlandse Taal)
- Hellebaut, L., A. Breitbarth, M. Farasyn, A.S. Ghyselen, Veronique De Tier, Jacques Van Keymeulen (2021). Stemmen uit het Verleden: een schat aan informatie voor heemkundigen en andere erfgoedactoren. In: *Histories Magazine* Histories vzw.
- Joby, C., Noline van der Sijs (2021). Japanse aanpassingen van leenwoorden. In: *Onze Taal* 6, 10.
- Joby, C., Noline van der Sijs (2021). Soja en kokku: taaluitwisselingen tussen Nederland en Japan. In: *Onze Taal* 6, 7-10.
- Kosem, Iztok, Michal Cukr, Miloš Jakubíček, Jelena Kallas, Simon Krek, Carole Tiberius (eds.) (2021). *Electronic lexicography in the 21st century. Proceedings of the eLex 2021 conference*. 5–7 July 2021, virtual. Brno: Lexical Computing CZ, s.r.o.
- Kosem, Iztok, Sanni Nimb, Carole Tiberius, Bob Boelhouwer & Simon Krek (2021). License to use: ELEXIS survey on licensing lexicographic data and software. In: Zoe Gavriilidou, Maria Mitsiaki, Asimakis Fliatouras (eds.), *Proceedings of XIX EURALEX Congress: Lexicography for Inclusion*, Vol. II, Democritus University of Thrace. 705-712.
- Landsbergen, Frank (2021). *Het dikke alfabetboek*. Tielt.
- Martelli, Federico, Roberto Navigli, Simon Krek, Jelena Kallas, Polona Gantar, Svetla Koeva, Sanni Nimb, Bolette Sandford Pedersen, Sussi Olsen, Margit Langemets, Kristina Koppel, Tiiu Üksik, Kaja Dobrovoljc, Rafael-J. Ureña-Ruiz, José-Luis Sancho-Sánchez, Veronika Lipp, Tamás

- Váradi, András Györffy, Simon László, Valeria Quochi, Monica Monachini, Francesca Frontini, Carole Tiberius, Rob Tempelaars, Rute Costa, Ana Salgado, Jaka Čibej & Tina Munda (2021). Designing the ELEXIS Parallel Sense-Annotated Dataset in 10 European Languages. In: *Electronic lexicography in the 21st century. Proceedings of the eLex 2021 conference*. 377-395.
- Norré, Magali, Vincent Vandeghinste, Pierrette Bouillon, Thomas François (2021). Extending a Text-to-Pictograph System to French and to Arasaac. *International Conference on Recent Advances in Natural Language Processing (RANLP)*
- Rys, Kathy (2021). De adpositie. *De ANS in het onderwijs*.
- Rys, Kathy, U. Balesar, Johan De Caluwe, W. Heeringa, Frans Hinksens, S.D. Misier, J.S. Rutten (2021). *De Staat van het Nederlands 2021. Over de taalkeuzes van Surinamers, Vlamingen en Nederlanders in het dagelijks leven*. Den Haag & Amsterdam: Taalunie & Meertens Instituut.
- Saggion, Horacio, Dimitar Shterionov, Gorka Labaka, Tim Van de Cruys, Vincent Vandeghinste and Josep Blat (2021). SignON: Bridging the gap between Sign and Spoken Languages. *XXXVII International Conference of the Spanish Society for Natural Language Processing*.
- Sijs, Nicoline van der (2021). *Taalwetten maken en vinden: het ontstaan van het Standaardnederlands*. Uitgeverij Sterck & De Vreese.
- Sijs, Nicoline van der, Lauren Fonteyn, Marten van der Meulen (red.) (2021). *Wat gebeurt er in het Nederlands?! Over taal, frequentie en variatie*. Sterck & De Vreese.
- Sijs, Nicoline van der (2021). 15 eeuwen Nederlands: Historische taalkunde in het curriculum! In: *34ste Conferentie Onderwijs Nederlands (HSN34)*, 102-106.
- Sijs, Nicoline van der (2021). Christoffel Plantijn: bouwmeester van een taalpaleis met vele kamers. In: *Noord & Zuid* 4, 7, 26-29.
- Sijs, Nicoline van der (2021). The Digital Pallas, an eighteenth-century multilingual Russian dictionary with 60.000 language data. *CIPL Newsletter March 2021*.
- Sijs, Nicoline van der (2021). Etymologica: moe(de)rneuker. *Neerlandistiek*, 17 mei.
- Sijs, Nicoline van der (2021). Etymologica: vogels in het wild en op tafel. *Neerlandistiek*, 28 juni.
- Sijs, Nicoline van der (2021). Etymologica: Hoe basaal zijn Nederlandse uitleenwoorden? *Neerlandistiek*, 26 juli.
- Sijs, Nicoline van der (2021). Etymologica: hotemetoot en hittepetit. *Neerlandistiek*, 11 oktober.
- Sijs, Nicoline van der (2021). Hoeveel dialecten werden er gesproken in negentiende-eeuws Amsterdam? *Taal & Tongval* 73 (1), 45-79.
- Sijs, Nicoline van der (2021). Hunnen zijn een volk en hennen leggen eieren. Dooddoeners bij taalfouten. In: *Onze Taal* 9, 22-23.
- Sijs, Nicoline van der (2021). Leve Karel van het Reve. *Neerlandistiek*, 19 mei.
- Sijs, Nicoline van der, Margit Rem (2021). Medeklinkerwisseling tussen enkel- en meervoudsvormen: een combinatie van spellingsbeginselen. *Taal & Tongval* 73 (2), 137-166.

- Sijs, Nicoline van der (2021). Ode aan de punt. In: *Drs. P en de punt* (pp. 30-59).
- Sijs, Nicoline van der (2021), Review of Michael T. Putnam & B. Richard Page (editors): *The Cambridge Handbook of Germanic Linguistics*, Cambridge University Press, 2020. In: *The Canadian Journal of Netherlandic Studies / Revue canadienne d'études néerlandaises* (CJNS/RCÉN) 41 (1), 81-86.
- Steurs, Frieda (2021). Centre of Expertise for Dutch Terminology. A digital platform for professional language. *Academic Journal of Modern Philology* (accepted).
- Steurs, Frieda (2021). Terminology at the Dutch Language Institute: towards a centre of expertise for professional language. *Terminologija* (accepted).
- Steurs, Frieda, K. Tryczyńska (2021). 'Europe, its member countries and Labour Law: Challenges in multilingual terminology and translation'. In: *Sendeban, Revista de Traducción y la Interpretación*. University of Granada. (accepted)
- Steurs, Frieda, Kris Heylen, Vincent Vandeghinste (2021). Hoe automatische vertaling de genderbias van AI verraaft. In: Van der Sijs, N., Fonteyn, L. & Van der Meulen, M. (Eds.). *Wat gebeurt er in het Nederlands? Over taal, frequentie en variatie*. Sterck & De Vreese.
- Steurs, Frieda, K. Tryczyńska (2021). On the untranslatability of legal texts: Can technology help in harmonizing the inequivalences in the different legal concepts and terms? The case of labour law. *Revista de Llengua i Dret - Journal of Language and Law* (accepted).
- Tempelaars, Rob, Vivien Waszink, Boukje Verheij, rubriek 'Neologisme van de week' / 'Nieuw woord van de week' op de website van het INT. Ook wekelijks gepubliceerd in het *Leidsch Dagblad*.
- Tiberius, Carole, Jan Odijk (2021). *Deliverable D3.2.10 Task 3 ELRC Workshop Report for the Netherlands*.
- Tiberius, Carole, Simon Krek, Katrien Depuydt, Polona Gantar, Jelena Kallas, Iztok Kosem, Michael Rundell (2021). *Towards the ELEXIS data model: defining a common vocabulary for lexicographic resources*. In: *Electronic lexicography in the 21st century. Proceedings of the eLex 2021 conference*. 56-77.
- Vincent Vandeghinste, Els Lefever, Walter Daelemans, Tim Van de Cruys, Sally Chambers (2021). CLARIN Flanders: new prospects. *CLARIN Annual Conference abstracts*.
- Vandeghinste, Vincent, E. Peters (2021). Jeugdjournaal versus Duidelijke Taal. In: Nicoline van der Sijs, Lauren Fonteyn, Marten van der Meulen (eds.), *Wat gebeurt er in het Nederlands?* Sterck & De Vreese.
- Vandeghinste, Vincent, Adéline Muller, Thomas François, Orphée Declercq (2021). Easy Language in Belgium. In: Camilla Lindholm and Ulla Vanhatalo (eds.): *Handbook of Easy Languages in Europe*. Frank & Timme, Berlin. pp. 53-90.
- Vandeghinste, Vincent, A. Muller, T. François, T., O. Declercq (2021). Easy Language in Europe -

Belgium. In: C. Lindholm & U. Vanhatalo (eds.), *Easy Language in Europe*. Frank & Timme.

Waszink, Vivien (2021). De kracht van nieuwe woorden. Op: *Taalvoutjes*.

Waszink, Vivien, Laura van Eerten (2021): Alcoflexen, thuisjuf, zonnehuid. Nieuwe woorden van 2020. In: *Onze Taal*.

Waszink, Vivien, Laura van Eerten (2021). Bijdragen scheurkalender Onze Taal over nieuwe woorden.

Waszink, Vivien, Veronique De Tier (2021). *Knuffelcontact en waterwappie. De kracht van nieuwe woorden*. Scriptum/Instituut voor de Nederlandse Taal.

Zviel-Girshin, Rina, Tanara Zingano Kuhn, Ana R. Luís, Kristina Koppel, Branislava Šandrih Todorović, Špela Arhar Holdt, Carole Tiberius, Iztok Kosem (2021). Developing pedagogically appropriate language corpora through crowdsourcing and gamification. In: Naouel Zoghliami, Cédric Bruderemann, Cédric Sarré, Muriel Grosbois, Linda Bradley, Sylvie Thouésny (eds.), *CALL and professionalisation: short papers from EUROCALL 2021*. 312-317.

Lezingen en presentaties

Bonth, Roland de, ‘Woorden uit de verleden tijd’. Kasteel Amerongen, 1 november 2021.

Bonth, Roland de, Uitreiking Kruyskampprijs, Laureatenmiddag van de Maatschappij der Nederlandsche Letterkunde, 30 oktober 2021.

Bulté, Bram, Vincent Vandeghinste, Leen Sevens, Ineke Schuurman, Frank Van Eynde (2021). Pictograph translation technologies as a tool for facilitating communication and integration in migration settings. Presented at *31st Annual conference on Computational Linguistics in the Netherlands (CLIN 31)*. 9 July. <https://www.clin31.ugent.be/programme/>

Chambers, Sally, Wout Dillen, Tom Gheldof, Wouter Ryckbosch, Vincent Vandeghinste, Christophe Verbruggen, Els Lefever (2021). Interdisciplinary Resilience: critical humanities infrastructure in the times of a global pandemic: a Flemish case study. Presented at *DH Benelux 2021*. 2-4 June. <https://2021.dhbenelux.org/>

Depuydt, Katrien, Jesse de Does, Linking the Dictionary of Old Dutch to A Thesaurus of Old English, ICEHL 2021, 7 juni 2021

Heylen, Kris, Vincent Vandeghinste (2021). A White Paper on the Future of Academic Lexicography. Keynote at *eLex2021 Conference*. 5 July. <https://elex.link/elex2021/conference-programme/>

Hubers, Ferdy, Catia Cucchiari, Nicoline van der Sijs (2021). ‘Does emigration affect idiomatic knowledge in the native language?’, postersession bij de *27th LIPP-Symposium ‘Language and Migration’*, 16 november 2021; wint de posterprijs.

Krauwier, Steven (moderator), Franciska de Jong, Hans Overbeek, Ted van der Togt, Vincent Vandeghinste (2021). Language data creation, management and sharing: existing practices and

- challenges - Panel session. *The third European Language Resource Coordination (ELRC) workshop in the Netherlands*. 11 June. <https://lr-coordination.eu/index.php/thenetherlands3rd>
- Norré, Magali, Vincent Vandeghinste, Pierrette Bouillon, Thomas François (2021). Experiments for the adaptation of Text2Picto to French. Presented at *31st Annual conference on Computational Linguistics in the Netherlands (CLIN 31)*. 9 July. <https://www.clin31.ugent.be/programme/>
- Norré, Magali, Vincent Vandeghinste, Pierrette Bouillon and Thomas François (2021). Extending a Text-to-Pictograph System to French and to Arasaac. Presentation at *International Conference Recent Advances*
- Rigouts, Ayla (moderator), Vincent Vandeghinste, Lore Vandevoorde, Nathalie De Sutter (2021). Language data creation, management and sharing: existing practices and challenges - Panel session. *The third European Language Resource Coordination (ELRC) workshop in Belgium*. 8 July. <https://lr-coordination.eu/belgium3rd>
- Shterionov, Dimitar, Vincent Vandeghinste, Horacio Saggion, Josep Blat, Mathieu De Coster, Joni Dambre, Henk Van den Heuvel, Irene Murtagh, Lorraine Leeson, Ineke Schuurman (2021). The SignON project: a Sign Language Translation Framework. Presented at *31st Annual conference on Computational Linguistics in the Netherlands (CLIN 31)*. 9 July. <https://www.clin31.ugent.be/programme/>
- Sijs, Nicoline van der (2021), 'Is a bible corpus suitable for research into language change?', Webinar release of the EDGeS Diachronic Bible Corpus, 28 mei 2021.
- Sijs, Nicoline van der (2021), '17th-century newspapers; History and future', Golden Agents colloquium 7 mei 2021.
- Sijs, Nicoline van der (2021), '17^e-eeuwse kranten. Ervaringen, lessen, mogelijkheden', Webinar CLARIAH 8 april 2021.
- Sijs, Nicoline van der (2021), 'Digitalisering vanuit het perspectief van de onderzoeker', Radboud Academy cursus Duurzaam Digitaal Erfgoed, 19 maart 2021.
- Sijs, Nicoline van der (2021), 'Crowdsourcing 17de-eeuwse kranten: ervaringen en lessen', lezing voor de Collegagroep Digitalisering van Periodieken georganiseerd door de Vlaamse Erfgoedbibliotheken, 16 februari 2021.
- Sijs, Nicoline van der (2021), 'De reis van Nederlandse woorden door de wereld', gastcollege voor masterstudenten neerlandistiek van de Universiteit Keulen, 2 februari 2021.
- Sijs, Nicoline van der (2021), 'Historische taalkunde in het curriculum', HSN Conferentie 34, 28 januari 2021.
- Steurs, Frieda, Dirk Kinable. Het ENT: een actief en attractief platform voor terminologie. Brugge, 15 oktober 2021, TiNT-dag NL-Term.
- Tiberius, Carole. ELRC, Language Technology and AI for Dutch. Presentatie op de derde ELRC-workshop in Nederland. 11 juni 2021, online.

- Tiberius, Carole. Towards the ELEXIS data model: defining a common vocabulary for lexicographic resources. Presentatie op eLex 2021. 7 juli 2021, online.
- Tiberius, Carole. Elexifier: a cloud-based dictionary conversion tool. Demo op eLex 2021. 7 juli 2021, online.
- Tiberius, Carole. Skema: A New Tool for Corpus-driven Lexicography. Demo op EURALEX 2021. 8 september 2021, online.
- Vandeghinste, Vincent, Elke Peters (2021). Vocabulary Demands for Different Types of Dutch Easy Language. 31st Annual conference on Computational Linguistics in the Netherlands (CLIN 31). 9 juli 2021.
- Vandeghinste, Vincent, Els Lefever, Walter Daelemans, Tim Van de Cruys, Sally Chambers (2021). Common Language Resources Infrastructure in Flanders: new prospects. 31st Annual conference on Computational Linguistics in the Netherlands (CLIN 31). 9 juli 2021.
- Vandeghinste, Vincent (2021). Easy Language in Belgium - prospects for automatic simplification. ELRC Technical Workshop: Simplify Language – Capture Audience. 24 september 2021.
- Vandeghinste, Vincent, Els Lefever, Walter Daelemans, Tim Van de Cruys, Sally Chambers (2021). CLARIN Flanders: new prospects. CLARIN Annual Conference Poster presentation. 27-29 september 2021.
- Vandeghinste, Vincent (2021). Demonstratie van de ELG, met focus op het Nederlands. Presentatie op de European Language Grid Workshop. 3 december 2021.
- Vos, Machteld de. *De standaard uitstippelen: hoe praat je over een standaardtaal in een tijd dat die nog in ontwikkeling is?* Presentatie op de Grote Taaldag. Utrecht (online), 29 januari 2021.
- Vos, Machteld de, Marten van der Meulen. *Suppressed no more: prescriptivism and the appreciation of optional variability*. Presentatie op de 6th Prescriptivism Conference. Vigo (online), 24 september 2021.
- Vos, Machteld de. *'Geeft gyse maer den duym, sy neemt de heele hand.'* *Argumenten voor de grenzen van variatie in de 17^e eeuw*. Presentatie op de boekpresentatie van *Wat gebeurt er in het Nederlands?! Over taal, frequentie en variatie* (red.: Nicoline van der Sijs, Lauren Fonteyn en Marten van der Meulen). Nijmegen, 8 oktober 2021.
- Waszink, Vivien. *Hoe maak je een woordenboek van modern Nederlands?* Webinar, 31 maart 2021.
- Waszink, Vivien. *Word senses, sense relations and definitions*. College Corpus Lexicography, Universiteit Leiden, 21 april 2021.
- Waszink, Vivien. *Interview over nieuwe woorden*. Podcast *Over Taal gesproken*, INT/Onze Taal.
- Waszink, Vivien. *Anderhalvemetersamenleving, alszaak en vaccinatievoordringer, Nieuwe woorden zijn er altijd*. Webinar KTV Kennisnet, 8 juli 2021.
- Waszink, Vivien. *Beyoncéfeminisme, die meisje en ombudsvrouw. Taal voor vrouwen?* Lezing Haags Vrouwen Netwerk (HVN), 18 oktober 2021.

Waszink, Vivien. *Eenzaamheidsvirus, keukentafelgesprek en wlz, Taal in de zorg*. Lezing Actiz Dutch Design Week, 21 oktober 2021.

Waszink, Vivien en Veronique De Tier. *Knuffelcontact en waterwappie, Nieuw boek en nieuw woordenboek!* Webinar, 30 november 2021.

Congressen & workshops

Baisa, Vit, Carole Tiberius, Elisabetta Ježek, Lut Colman, Constanza Marini, Emma Romani. Skema: A New Tool for Corpus-driven Lexicography. EURALEX 2020-2021, 7-9 september 2021.

Colman, Lut & Carole Tiberius: *Woordcombinaties / Word Combinations: all-in-one resource for Dutch words in context*. EUROPHRAS 2021, 6-9 september 2021.

De Tier, Veronique, Jesse de Does, Katrien Depuydt e.a. 'The Database of the Southern Dutch Dialects: New opportunities for digital lexicographical research'. Poster. Sociolinguistic circle. Online. 19 mei 2021.

De Tier, Veronique, Jesse de Does, Katrien Depuydt e.a. 'Interdisciplinary synergies: opening up new perspectives for data-driven research of the Southern Dutch Dialects'. Presented at DH Benelux. Online. 2 juni 2021.

De Tier, Veronique, Jesse de Does, Katrien Depuydt e.a. 'The Dictionary portal of the Southern Dutch Dialects'. Poster presentation at Euralex XIX. Lexicography for inclusion. 7 september 2021.

De Tier, Veronique, Jesse de Does, Katrien Depuydt e.a. 'The Dictionary portal of the Southern Dutch Dialects'. Software demonstration at Euralex XIX. Lexicography for inclusion. 8 september 2021.

De Tier, Veronique, Jacques Van Keymeulen 'Onomasiological and semasiological dialect dictionaries for the southern Dutch dialects (Flemish, Brabantian and Limburgian)'. Kolloquium Großlandschaftliche Dialektwörterbücher zwischen Linguistik und Landeskunde. Münster, 29 oktober 2021.

De Tier, Veronique, Jesse de Does, Katrien Depuydt e.a. 'De droom van Weijnen. Het DSDD-project voor de dialectlexicografie'. Presentatie op Taal en Tongvalcolloquium. Big data: Perspectieven voor onderzoek naar taalvariatie en taalverandering /Perspectives for research on language variation and change. 26 november 2021.

Lismont, E., Machteld de Vos. Organisatie roundtable *Methods and approaches in determining normative success*. 6th Prescriptivism Conference. Vigo (online), 23 september 2021.

Tiberius, Carole. Organisatie derde ELRC-workshop in Nederland. 11 juni 2021, online.

Tiberius, Carole. Deelname aan projectbijeenkomst: ELEXIS Project Management Board. 17 juni 2021, online.

Tiberius, Carole. Medeorganisatie en deelname aan eLex 2021. 5-7 juli 2021, online.

Tiberius, Carole. Deelname aan EURALEX 2021. 7-9 september 2021, online.

Tiberius, Carole. Deelname aan EUROPHRAS 2021, 6-9 september 2021, online.

Onderwijs

Bonth, Roland de. 'Woorden uit de verleden tijd'. Instituut voor de Nederlandse Taal, 30 maart 2021.

Bonth, Roland de. 'Woorden uit de verleden tijd'. Proteus, Leidse Vereniging van Renaissancisten, 8 april 2021.

Bonth, Roland de. 'Woorden uit de verleden tijd'. Vertalersplatform, 18 mei 2021.

Bonth, Roland de. Tweemaal gast(werk)college aan studenten Nederlands van de Universiteit Leiden over het zoeken in de GTB, 29 en 30 september 2021.

Bonth, Roland de. Presentatie webinar Woordcombinaties, 5 oktober 2021.

Colman, Lut. Webinar Woordcombinaties. Week van het Nederlands 2021, 2-9 oktober.

Heylen, Kris, Carole Tiberius, Vincent Vandeghinste (2021). Computatieve Lexicografie. Master Taalkunde en Master Taal- en Letterkunde, KU Leuven.

Tempelaars, Rob. College Woordbetekenissen, betekenisrelaties en definities in de MA-collegereeks Computatieve Lexicografie. KU Leuven, 29 maart 2021.

Tiberius, Carole. MA-collegereeks Corpus Lexicography. Universiteit Leiden, februari-mei 2021.

Tiberius, Carole. College over Dictionary Writing Systems in de MA-collegereeks Computatieve Lexicografie. KU Leuven, 29 maart 2021.

Tiberius, Carole. Crowdsourcing and gamification: a workshop with a purpose. College voor Language and Technology: a successful combination, workshop georganiseerd door de Europese Commissie. Utrecht, 12 november 2021.

Van de Cruys, Tim, Marie-Francine Moens, Frank Van Eynde, Hugo Van hamme, Vincent Vandeghinste, Astrid van Wieringen. Language Engineering Applications. KU Leuven, 2021.

Vandeghinste, Vincent, Tim Van de Cruys. A Linguist's Guide to Natural Language Processing with Neural Networks. LOT Summer School 2021.

Vandeghinste, Vincent. Machine Translation and Post-Editing. Postgraduate in Specialised Translation. KU Leuven, 2021.

Vandeghinste, Vincent. Treebank Querying with GrE TEL. In het vak Computerlinguïstiek. KU Leuven, 2021.

In de media

Colman, Lut. Woordcombinaties: veelbelovende taalhelp voor leerders van het Nederlands. Interview door de Nederlandse Taalunie, 21 juni 2021.

Instituut voor de Nederlandse Taal. Wekelijkse rubriek 'Neologisme van de week', in *Leidsch Dagblad*, van 1 januari 2021 t/m 31 december 2021.

Tempelaars, Rob. Interview voor Simone Paauw. 'Krijg de ziekte, teringlijer!' In: *Medisch Contact*, 30 september 2021.

Tempelaars, Rob. Interview over de vraag waarom criminelen vaak bijnamen hebben. EditieNL, RTL 4, 17 december 2021.

Waszink, Vivien. Straattaalwoord van de week: neef. In: *de Volkskrant*, 13 januari 2021.

Waszink, Vivien. Interview over jongerentaal. De Madammen, Radio 2 België, 20 januari 2021.

Waszink, Vivien. Interview over oude en nieuwe woorden. Editie.nl, RTL 4, 3 februari 2021.

Waszink, Vivien. Interview over de populairste koosnaampjes. Gerard Ekdome, Radio 10, 10 februari 2021.

Waszink, Vivien. Interview Persco, kowa en lawa: afko's (afkortingen) populair onder studenten. De Nieuws BV, BNN/VARA, NPO 1 radio en tv, 13 april 2021.

Waszink, Vivien. Interview voor artikel *Racisme is lastig te bestrijden als niemand het woord mag uitspreken* over de definitie van het woord 'racisme'. In: *de Volkskrant*, 22 mei 2021.

Waszink, Vivien. Interview voor artikel *Ewa niffo?! Praatsto in knap wurdje 'straattaal'?* In: Fries tijdschrift *LinKk*, september 2021.

Waszink, Vivien. Interview over *Knuffelcontact en Waterwappie: de kracht van nieuwe woorden*. Groot Nieuws Radio, 17 november 2021.

Waszink, Vivien. Interview over *Knuffelcontact en Waterwappie: de kracht van nieuwe woorden*. De Nieuws BV (BNN/VARA), NPO 1 radio en tv, 19 november 2021.

Waszink, Vivien. Interview over *Knuffelcontact en Waterwappie: de kracht van nieuwe woorden*. Gerard Ekdome, Radio 10, 25 november 2021.

Waszink, Vivien. Interview over *Knuffelcontact en Waterwappie: de kracht van nieuwe woorden* en de Woord van het Jaar-verkiezing. BNR Nieuwsradio, 7 december 2021.

Waszink, Vivien. Interview over *Knuffelcontact en Waterwappie: de kracht van nieuwe woorden*. Haandrikman (Omroep Max), 21 december 2021.

Waszink, Vivien. Interview over *Knuffelcontact en Waterwappie: de kracht van nieuwe woorden*. Lunchroom, Radio Noord-Holland, 21 december 2021

Diversen

Bonth, Roland de

- Coördinator en jurylid Profielwerkstukwedstrijdprijs van Levende Talen Nederlands (LTN) in samenwerking met de Internationale Vereniging voor Neerlandistiek (IVN)
- Jurylid Kruyskampprijs
- Redacteur *Trefwoord*

Colman, Lut

- PC-member EURALEX
- PC-member eLex

De Tier, Veronique

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

Depuydt, Katrien en Jesse de Does

- Maken deel uit van de werkgroep die verantwoordelijk is voor TEI Lex-0, een basiscodering in TEI voor lexicografische bronnen

Eerten, Laura van

- Programmapartner eerste Neerlandistiekdagen (i.s.m. Roland de Bonth)
- Adviesraad/programmapartner nieuw festival Letterlijk Leiden

Sijs, Nicoline van der

- Bekleder van de Koning Willem-Alexander leerstoel voor Lage Landen-studies te Luik
- Subsidie door Pilotfonds Wetenschapscommunicatie door Wetenschappers Gewaardeerd! van de KNAW voor de redactie van Neerlandistiek
- Redacteur Internationale Neerlandistiek
- Redacteur van Neerlandistiek
- Hoofdredacteur, met Peter-Arno Coppen en Marc van Oostendorp, van de nieuwsbrief Neerlandistiek voor de klas
- Lid beoordelingscommissie van het Onderwijsfonds van de Maatschappij der Nederlandse Letterkunde
- Lid van de GLAD (Global Anglicism Database) Network: www.gladnetwork.org
- Lid van het bestuur van de ‘Stichting Jiddische Lexicografie Amsterdam’
- Member of the steering committee for the DHBenelux conference
- Vaste medewerker van het tijdschrift Onze Taal
- Redacteur van *Trefwoord*, tijdschrift op het gebied van lexicografie en lexicologie

Tempelaars, Rob

- Redacteur en redactiesecretaris van *Trefwoord*, tijdschrift op het gebied van lexicografie en lexicologie
- Betrokken bij een door diverse partijen opgezet project over geuren, dat uiteindelijk moet leiden tot het boek *NeusWijzer* (te verschijnen in het voorjaar van 2023)

Tiberius, Carole

- Board of Trustees Adam Kilgarriff Prize
- Technical National Anchor Point ELRC (European Language Resource Coordination)

Van pellicom, Katrien

- Lid van de werkgroep ‘Persoonsnamen’ van de ‘commissie anderstalige namen’ van de Taalunie

Waszink, Vivien

- Redactielid *Neerlandia*

Begeleiding thesis, stage, vrijwilligers

Dirk Kinable:

- Nina Zwiers, Universiteit Utrecht
- Anouk de Wit, ITV Hogeschool voor Tolken en Vertalen
- Rick van Mechelen, Universiteit Leiden
- Ryanne de Winter, Universiteit Leiden

Nicoline van der Sijs, promovendi:

- Machteld de Vos. ‘Spread the new(s). Understanding standardization of Dutch through 17th-century newspapers’
- Marten van der Meulen. ‘Change your language! Effects of prescriptivism on the grammar of Dutch in the 20th century’(RU)
- Emmie Hoebens. ‘Sociolinguistic research into the language and culture of the Mennonite settlement of ejido Salamanca, Bacalar Municipality, Quintana Roo State, Mexico

Nicoline van der Sijs, stages:

- Corinne Hendriks (etymologiebank)
- Martijn Knapen (Digitale Pallas)
- Joanna Kruyt (etymologie)
- Katja van Rooten (etymologie)
- Iris Szigalyj (etymologie)

Vincent Vandeghinste:

- Tess Dejaeghere (2021). Benchmarking Named Entity Recognizers for Dutch. Master in Linguistics: text data analysis. Universiteit Antwerpen
- Jannis Raemaekers (2021). Automated Open Dutch Wordnet Link Creation using embeddings. Master in Artificial Intelligence. Universiteit Leuven
- Joni Kruijsbergen (2021-). Wat is tekst? Filteren van tekst vs. niet-tekst in krantenberichten. Master in Taalkunde. KU Leuven
- Nick Vanackere (2021-). Creation of a comparable corpus of Newspaper Dutch with Easy-to-read Dutch. Master in Artificial Intelligence. Universiteit Leuven.

Bijlage 4: Downloadoverzicht taalmaterialen

Overzicht downloads commercieel

In totaal is er 27 keer een product afgenomen voor commerciële toepassingen.

SoNaR	8
CGN	6
BasiLex Lexicon	2
BasiScripts Corpus	2
Frequentielijsten Corpora	2
BasiLex Corpus	1
combilex	1
DPC	1
E-Lex	1
Lassy Klein	1
OMBI Ned-Arabisch	1
RBN	1

Overzicht downloads niet-commercieel

Er zijn in totaal 930 producten gedownload voor niet-commercieel gebruik.

Corpus Gesproken Nederlands (CGN)	120
SoNaR-corpus	89
Eindhoven-corpus	83
CGN-annotaties	42
DuOMAn Subjectivity Lexicon	39
Dutch Parallel Corpus (DPC)	30
CHN N-grams	30
Cd-rom Middelnederlands	28
Frequentielijsten corpora	28
SoNaR Nieuwe Media Corpus	27
Referentiebestand Nederlands (RBN)	27
Lassy Klein-corpus	24
Corpus Ondertitelde UVN-Colleges (COUC)	23
Wablieft-corpus	22
Corpus Middelnederlands	21
GiGaNT-Molex	19
Hoger Onderwijs Terminologie in Nederland en Vlaanderen (HOTNeV)	18
Dutch Idiom Database: Native Speakers (DID-NS)	17
e-Lex	15
Medische Termen Belgisch-Nederlands (MedTermBN)	15
PAROLE-lexicon	15
Medische Pilot (MedPilot)	13
Corpus Gysseling	12
INT Historische Woordenlijst	11
Basilex Lexicon	10
WAI-NOT Corpus	10
Boarnsterhim Corpus (BHC)	10
Brieven als Buit - Gouden Standaard	10
NAMES Corpus	9
RND Woordenlijsten	8
COREA-coreferentiecorpus	8
BLISS Dialogue Summaries	7

Corpus Pathologische en Normale Spraak (COPAS)	7
IFA Dialoog-videocorpus	6
BLISS Spoken Dialogue Dataset	6
Afrikaans Custom Dictionary for Government Domain	5
DuELME	5
DAESO-corpus: parallele Nederlandstalige monolinguale treebank	5
BasiLex Corpus	4
INT IMPACT NE-lexicon	4
SumNL-samenvattingencorpus	4
AUTONOMATA-namencorpus	4
Referentiebestand Belgisch-Nederlands (RBBN)	4
D-TUNA-corpus	4
AUTONOMATA-POI-corpus	3
CombiLex	3
IFA-corpus	3
OMBI Nederlands-Indonesisch	3
BasiScript Corpus	2
Dupira	2
Moroccorp	2
BasiScript Lexicon	1
Jasmin	1
OMBI Nederlands-Deens	1
Paco-MT Parallele Corpora	1
Lwazi Tshivenda Pronunciation Dictionary	1
AUTONOMATA-POI-demo	1
Lwazi Sesotho ASR Corpus	1
Lwazi isiZulu ASR Corpus	1
Lwazi isiXhosa ASR Corpus	1
Lwazi Afrikaans Pronunciation Dictionary	1
Tshivenda Custom Dictionary for Government Domain	1
Tshivenda Genre Classification Corpus	1
Afrikaans Genre Classification Corpus	1
Lwazi Afrikaans ASR Corpus	1