

/instituut voor de
Nederlandse taal/

jaarverslag

2020

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

(Woordenboek der Nederlandsche Taal)

Jaarverslag 2020

Instituut voor de Nederlandse Taal

Inhoud

Algemene doelstellingen INT

A. Lexicografie en Terminologie

Basistaken

- A.1. Synchron: ANW
- A.2. Synchron: Woordcombinaties
 - A.2.1. Redactionele werkzaamheden
 - A.2.2. Corpuswerkzaamheden, data-extractie en lexicografische tools
 - A.2.3. Webapplicatie
- A.3. Synchron: Neologismen
- A.4. Diachron: Historische woordenboeken en de GTB
- A.5. Etymologiebank en Uitleenwoordenbank

Specifieke projecten

- A.6. Dialectologie
 - A.6.1. Database of the Southern Dutch Dialects (DSDD)
 - A.6.2. Gesproken Corpus van de Zuidelijk-Nederlandse Dialecten (GCND)
 - A.6.3. Atlas van het dialect in Vlaanderen
 - A.6.4. Elektronische Woordenbank van de Nederlandse Dialecten (eWND) en regionale woordenboeken
 - A.6.5. Overige activiteiten
- A.7. Vertaalwoordenschat
- A.8. Expertisecentrum Terminologie
 - A.8.1. Website
 - A.8.2. Projecten voor specifieke vakdomeinen (HOTNeV)
 - A.8.3. Medische terminologie
 - A.8.4. Juridische terminologie
 - A.8.5. Terminologietools
 - A.8.6. Terminologisch netwerk

A.9. De digitale Pallas

A.10. Lexicon Frisicum

B. Corpora en lexica

Basistaken

B.1. Moderne corpora

B.2. Historische corpora

B.3. Spelling

B.4. GiGaNT

B.5. Taalbank: semantisch lexicon DiaMaNT

C. Grammatica

Basistaken

C.1. Grammatica algemeen

Specifieke projecten

C.2. e-ANS

C.3. ANK

C.4. Taalportaal Zuid-Afrika

D. Softwaretools (ontwikkeling, beheer en onderhoud)

Basistaken

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.1.3. Downloads
- E.2. CLARIN-ERIC; het INT als CLARIN-centrum

F. Onderzoek: Europese aanvragen en competitieve projecten

- F.1. European Lexicographic Infrastructure (ELEXIS)
- F.2. enetCollect (COST)
- F.3. COST CA18209: European network for Web-centered linguistic data science
- 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

G. Doelgroepenbeleid

- G.1. Onderzoek, onderwijs en het algemene publiek

H. Wetenschapscommunicatie

- H.1. Sociale media
- H.2. Nieuw woord van de week; Terug in de Taal; Uitgelichte term; Gelegenheidswoordenboekjes; Woordbaak
- H.3. Populairwetenschappelijke uitgaven en activiteiten: boeken, webinars, podcasts
- H.4. Website
- H.5. Nieuwsbrieven, mailings, persberichten
- 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

Bijlage 5: Rapport externe communicatie

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 2020 weer met een aantal nieuwe accenten en dit in nauw overleg met de Nederlandse Taalunie (NTU). Het werkvloeroverleg (projectgericht) en het directieoverleg tussen de NTU 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.

Het jaar 2020 moest een speciaal jaar worden, met name het “Matthias de Vriesjaar”. Op 9 november 2020 was het precies tweehonderd jaar geleden dat de Nederlandse taal- en letterkundige Matthias de Vries werd geboren, de aartsvader van de lexicografie. Er waren tal van evenementen gepland die Leiden als Dictionopolis en het INT als instituut in de kijker moesten zetten. Door het plotse opduiken van het coronavirus moest bij de meeste plannen omgeschakeld worden naar een onlinegebeuren, wat vlot verlopen is. Er was een Matthias de Vrieswandeling, een quiz, een Dicté (in de spelling-De Vries en Te Winkel), een symposium en er verschenen twee boeken *Geen woord teveel. De geschiedenis van INL naar INT* en *Matthias de Vries aan het woord*. Daarnaast werd er ook een dialectmiddag gehouden met de presentatie van de DSDD, de Database van de Zuidelijk-Nederlandse Dialecten. Ten slotte werden we creatiever door de pandemie en gingen we aan de slag met nieuwe communicatievormen, zoals podcasts en webinars.

Bij de aanvang van 2020 werkten er 28,8 fte bij het INT (36 werknemers), eind december 2020 waren dat er 28,75 fte (37 werknemers). Er waren enkele verschuivingen in het personeelsbestand door een overlijden en enkele medewerkers die aan einde van hun contract waren, en door aanwerving van drie taalkundigen (respectievelijk gespecialiseerd in historische taalkunde, dialectologie en grammatica) en een systeemontwikkelaar.

In onze werking werd aandacht besteed aan de specifieke gebieden die de NTU 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.

Zo hebben we in 2020 meer aandacht besteed aan de ondersteuning van taaldocenten Nederlands in het buitenland, dit door hen te betrekken bij de verdere ontwikkeling van het project Woordcombinaties. Daarnaast blijft de verdere uitbouw van het Expertisecentrum Terminologie onze aandacht opeisen en willen we verder resultaten boeken in enkele grote terminologieprojecten, met name in het medische en juridische domein (taalvariatie en het beschikbaar stellen van vaktaal voor het grote publiek in een inclusieve samenleving). Meer en meer komen ook het academisch Nederlands, moeilijke woorden en Nederlands als wetenschapstaal centraal te staan in onze toepassingen voor het onderwijs.

We vielen ook in de prijzen: de Nederlandse Dataprijs is toegekend aan de 17e-eeuwse krantenset die door honderden vrijwilligers is getranscribeerd binnen een samenwerkingsproject tussen het Instituut voor de Nederlandse Taal (INT), de Koninklijke Bibliotheek (KB) en het Meertens Instituut onder leiding van Nicoline van der Sijs.

In 2020 werd ook werk gemaakt van het uitvoeren van de conclusies van de Lorentzworkshop, door het centraal stellen van de infrastructuur (GiGaNT) en het daaruit afleiden van alle hedendaagse taalproducten. Dat betekent een nieuwe workflow met meer teamoverleg over het hedendaags Nederlands (ANW, neologismen, vaktaal, woordcombinaties).

Door corona werd er vooral thuis gewerkt, er was geen reistijd en de meeste congressen werden ofwel uitgesteld, ofwel online gehouden. Dat resulteerde in meer output, meer onderzoek en meer dataverwerking.

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 Nederlandse 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. We werken dan ook hard aan het meer zichtbaar maken van al onze taalmaterialen. Om dat te realiseren is er in 2020 een totaal nieuwe website gebouwd, met uitgebreide zoekfuncties, gericht op een divers doelpubliek en ook in een responsief design. In 2020 werd tevens resultaat geboekt in twee belangrijke dossiers in Vlaanderen: CLARIAH VL-II werd goedgekeurd en ook het CLARIN-lidmaatschap verkreeg de nodige steun.

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.

Het jaar 2020 was voor het ANW een atypisch jaar, dat overschaduwd werd door het plotselinge overlijden van hoofdredacteur Tanneke Schoonheim. Uiteraard hadden wij ook te maken met coronamaatregelen, maar die hebben de werkzaamheden dankzij een goede thuiswerkomgeving niet nadelig beïnvloed. Het jaar werd verder vooral bepaald door het ondernemen van enkele zeer ingrijpende werkzaamheden. Een groot deel van het jaar werd 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*. Met ingang van 17 september 2020 werden in verband daarmee de dagelijkse updates van het ANW stopgezet; in de loop van 2021 wordt het ANW weer geüpdatet. Het doel van deze operatie is niet alleen dat alle lemmata van het ANW in GiGaNT-Molex worden opgenomen, maar ook een verstrekkende en rigoureuze wijziging van de workflow: voortaan worden de grammatische gegevens vanuit GiGaNT-Molex aan het ANW toegevoegd, wat een aanzienlijke tijdbesparing voor de redactie zal gaan betekenen. Een andere grootscheepse operatie was het beoordelen van ruim 23.000 mogelijke neologismen die in de loop der tijd verzamelend waren en zo nodig aan de woordfamilies van de ANW-artikelen en/of de database van neologismen toegevoegd werden.

Gedurende het jaar werden redacteuren van het ANW ook ingezet bij andere projecten, te weten bij Woordcombinaties, Neologismen, ELEXIS, enetCollect en als leden van de webredactie. Daarnaast bewerkten zij voor het ANW 286 artikelen. In 2020 werden in totaal 215 nieuwe artikelen online gezet, waaronder zeer omvangrijke woorden als *beleid*, *bewering*, *klimaat*, *percentage*, *salaris* en *stroom*, actuele woorden en woordgroepen als *anderhalvemetersamenleving*, *corona*, *COVID-19*, *pfas*, *social distancing* en *sociale onthouding*, en een flink aantal werkwoorden die beginnen met het voorvoegsel *be-* en *ont-*. Ondanks dat de output achter de schermen geringer was dan in andere jaren, steeg het aantal ingangen in het ANW online in 2020 toch exponentieel. De reden daarvoor is dat vaste verbindingen en spreekwoorden voortaan als apart lemma behandeld worden, met een verwijzing

daarnaar bij het grondwoord. Op 17 september 2020, de datum van de laatste update, telde het ANW 84.585 trefwoorden, 42.971 betekenissen, 35.055 synoniemen, 251.339 voorbeeldzinnen, 130.683 combinaties, 9676 vaste verbindingen en 362 spreekwoorden. Verder zijn 4753 afbeeldingen, 729 videofilmjes en 178 geluiden opgenomen.

In 2020 was Laura van Oeveren (Universiteit Leiden) als onderdeel van haar master ‘Translation in Theory and Practice’ bij het ANW werkzaam als stagiaire. Onder begeleiding van projectleider Rob Tempelaars hield zij zich bezig met het in het Engels vertalen van relevante ANW-semagrammen en diverse Nederlandse termen uit het arbeidsrecht.

A.2. Synchroon: 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 (www.sketchengine.co.uk). Woordcombinaties toont hoe woorden gebruikt worden in goede voorbeeldzinnen, welke woorden met elkaar gecombineerd worden (collocaties of combinatiemogelijkheden) en hoe valentiepatronen en andere patronen samen met collocaties gebruikt worden voor het bouwen van zinnen. Taalleeders leren niet alleen woorden kennen, maar ook woorden gebruiken in context.

A.2.1. Redactionele werkzaamheden

De pilot van Woordcombinaties is door een groep docenten van de doelgroep geëvalueerd en positief ontvangen. De redactie heeft de constructieve feedback verwerkt in de nieuwe applicatie die in de loop van 2021 voor de doelgroep beschikbaar komt. De bewerking van de voorbeeldzinnen en combinatiemogelijkheden is uitgebreid met de werkwoorden uit de *Basislijst Schooltaalwoorden* van het iTTA (Instituut voor Taalonderzoek en Taalonderwijs Anderstaligen) van de Universiteit van Amsterdam. De basislijst bevat algemene schooltaalwoorden en vakgerelateerde woorden die anderstaligen nodig hebben om de lessen in het secundair onderwijs goed te kunnen volgen. Schooltaalwoorden die ook in het *Frequentiewoordenboek* voorkomen zijn met voorrang bewerkt. Daarnaast werden ook andere werkwoorden uit het frequentiewoordenboek bewerkt. In het verslagjaar lag de focus op de bewerking van de voorbeeldzinnen en combinatiemogelijkheden. Voor de patroonbeschrijving zijn proefartikelen gemaakt. In 2021 zal er tijd gereserveerd worden voor de redactionele bewerking van de patronen.

A.2.2. Corpuswerkzaamheden, data-extractie en lexicografische tools

In 2020 kwam de geparseerde versie van het *Woordcombinatie*-corpus beschikbaar. Dit corpus verbetert de workflow voor de redactionele bewerking van samengestelde scheidbare werkwoorden aanzienlijk. De combinatie-editor is in 2020 geoptimaliseerd.

A.2.3. Webapplicatie

In het eerste kwartaal van 2020 kwam de demoversie van *Woordcombinaties* online. In 2021 komt een eerste release beschikbaar voor het publiek.

A.3. Synchroon: Neologismen

Ieder jaar komen er veel nieuwe woorden bij; dit jaar bijvoorbeeld *anderhalvemetersamenleving*, *kot couture*, *fleet* en *gravinfluencer*. Het INT onderzoekt de vorming van nieuwe woorden en houdt die woorden onder andere bij in het neologismenwoordenboek, een onlinewoordenboek van nieuwe woorden. Ook in 2020 is er verder gewerkt aan dit woordenboek. Een testversie van dit woordenboek is intern beschikbaar; hierin zijn al ongeveer 6000 woorden opgenomen. In de loop van 2021 komt een eerste versie van het woordenboek beschikbaar voor iedereen. Dit neologismenwoordenboek wordt vervolgens voortdurend aangevuld en geüpdatet.

Verder is er in 2020 een op het INT bijgehouden lijst met ruim 23.000 neologismen doorgewerkt en is er voor elk nieuw woord op die lijst bepaald of het woord in aanmerking komt voor opname in het Neologismenwoordenboek, het *Algemeen Nederlands Woordenboek* (ANW) of in beide woordenboeken. Ook worden alle woorden in het ANW op dit moment digitaal gekoppeld aan de bijbehorende lemmata in het spellingbestand Molex en daarmee ook aan de bijbehorende ingangen in de *Woordenlijst Nederlandse Taal*. Hetzelfde wordt gedaan met alle neologismen – niet alleen de neologismen die in het ANW staan, maar ook de neologismen die (voorlopig) alleen in het Neologismenwoordenboek komen.

Ten slotte is er ook dit jaar iedere week een neologisme behandeld in de rubriek “Nieuw woord van de week”. Nieuw sinds 2020 is dat het neologisme van de week niet alleen wekelijks op de INT-website wordt 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 werd een artikel gepubliceerd over nieuwe coronawoorden; er verscheen een artikel over het neologismenproject in het Amerikaanse tijdschrift *Dictionaries*.

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](#): 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 tijdens de *Week van het Nederlands* een webinar verzorgd over doeltreffend zoeken in de online historische woordenboeken; ook werd in een bijdrage op [neerlandistiek.nl](#) ten behoeve van het onderwijs informatie gegeven over het *Woordenboek der Nederlandsche Taal*, over de onlinebeschikbaarheid ervan en over twee lesbrieven die bij het INT ontwikkeld zijn. Uit de loggegevens én uit de reacties en vragen die we regelmatig zowel van specialistische als niet-specialistische gebruikers ontvangen, blijkt inderdaad dat de webapplicatie ondertussen een veel gebruikt werkinstrument geworden is.

In 2020 is ten behoeve van het historisch lexicon binnen GiGaNT intensief gewerkt aan het taalkundig verrijken van de woordvormen die voorkomen in de citaten van het MNW en WNT. Bij deze werkzaamheden werden ruim 2.500 woordenboekcitaten gevonden waarbij nog datacorrectie nodig is. Deze correcties zullen uitgevoerd worden zodra een nieuwe, voor 2021 geplande TEI-P5-versie van de woordenboekbestanden beschikbaar is.

Twee stagiairs hebben dit jaar in het kader van hun opleiding werk verricht voor het *Oudnederlands Woordenboek* (ONW). Sune Gregersen, in oktober van dit verslagjaar gepromoveerd aan de Universiteit van Amsterdam (UvA) op modale werkwoorden in het Oud- en Middelnederlands, voegde aan diverse ONW-lemmata Oudengelse en Middelnederlandsse cognaten (woorden met een gelijke etymologische oorsprong) toe. Alan Davis, MA-student bij het Leiden University Centre for Linguistics (LUCL), vertaalde voor zijn stage Translation Studies 700 definities van het ONW in het Engels en Duits; hij hoopt dit werk in 2021 voort te kunnen zetten. Dergelijk vertaalwerk kan een interessante opmaat zijn tot het vervullen van een al lang bestaande wens binnen de medioneerlandistiek om de internationalisering ervan te stimuleren door het aanbieden van Engelstalige tools.

Tot slot is er ook gewerkt aan een aantal verbeteringen van de woordenboekapplicatie, ten behoeve van een update die begin 2021 zal plaatsvinden.

A.5. Etymologiebank en Uitleenwoordenbank

In 2020 zijn de Etymologiebank en de Uitleenwoordenbank van het Meertens Instituut overgegaan naar het INT. Beide banken zijn bedoeld als hulpmiddel voor etymologen en historisch taalkundig onderzoekers, maar ook voor geïnteresseerden. Aan de Uitleenwoordenbank wordt vooralsnog niet (meer) gewerkt, maar de Etymologiebank is nog volop in bewerking. Het doel van die bank is om alle belangrijke etymologische publicaties van het Nederlands op woordniveau aan te bieden op één centraal punt. Momenteel bevat de Etymologiebank meer dan zestig woordenboeken, met 293.512 ingangen, waarvan 114.894 unieke ingangen. Stagiairs en vrijwilligers voegen nog steeds nieuwe ingangen en nieuwe woordenboeken toe aan de Etymologiebank.

Specifieke projecten

A.6. Dialectologie

Vanaf 1 juni 2020 stelt het INT een Vlaamse dialectoloog, Veronique De Tier, tewerk die het instituut versterkt op het gebied van dialectologie en variatielinguïstiek.

A.6.1. Database of the Southern Dutch Dialects (DSDD)

In 2020 is het pilootproject *Database van de Zuidelijk-Nederlandse Dialecten/Database of the Southern Dutch Dialects* (DSDD) afgerond. Dit Herculesproject met als voornaamste deliverables een gecureerde database, waarin een overkoepelende laag concepten zorgt voor gezamenlijke doorzoekbaarheid, een portaalapplicatie voor een breed publiek, en een API voor onderzoekers, gebeurde in samenwerking met de Universiteit Gent en de Radboud Universiteit Nijmegen. In 2020 ontwikkelde het INT de portaalapplicatie met karteermogelijkheden voor het brede publiek. Het doel was op het einde van het pilootproject minstens 1.500 concepten online te hebben. Op 30 september is de website met zoekfuncties en kaartapplicatie (dsdd.ivdnt.org) gepresenteerd aan subsidiërende partijen, collega's en het brede publiek op een onlinedialectmiddag tijdens de [Week van het Nederlands](#) met iets meer dan 10.000 concepten¹. De database wordt nu gehost, onderhouden en verder aangevuld door het INT.

Eind december 2020 was de hele Algemene Woordenschat van de drie onderliggende dialectwoordenboeken (*Woordenboek van de Vlaamse, Brabantse en Limburgse dialecten*) doorzoekbaar en een deel van de vaktaal- en de landbouwwoordenschat (in totaal ongeveer 19.000 concepten). In de loop van 2021 worden de resterende concepten (landbouwwoordenschat en vaktalen)

¹ Het evenement is nog te bekijken via <https://vimeo.com/461766071>.

toegevoegd. In samenwerking met de diverse partners zal het in het kader van DSDD ontwikkelde dialectportaal ook uitgebreid worden met andere dialectdata, zoals het Zeeuws, het Gelders en het Overijssels.

A.6.2. Gesproken Corpus van de Zuidelijk-Nederlandse Dialecten (GCND)

In 2020 werd de FWO-aanvraag Gesproken Corpus van de Zuidelijk-Nederlandse Dialecten ingediend door de Universiteit Gent (hoofdaanvrager), Variaties vzw (Koepelvereniging voor dialecten en oraal erfgoed in Vlaanderen), Meertens Instituut, Rijksuniversiteit Groningen en INT. Het project is goedgekeurd en 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 project ging van start op 1 juni 2020. Het INT zal hier een substantiële infrastructurele rol in spelen. De opnames worden volgens een nieuw ontwikkeld transcriptieprotocol getranscribeerd. Het INT zal de audio, de transcripties en de annotaties op termijn vrij online beschikbaar en doorzoekbaar maken en duurzaam bewaren.

A.6.3. Atlas van het dialect in Vlaanderen

In de tweede helft van 2020 is het project *Atlas van de dialecten in Vlaanderen* in samenwerking met de Universiteit Gent (o.l.v. Johan De Caluwe) en Uitgeverij Lannoo van start gegaan. Het boek verschijnt in het najaar van 2021. Het INT zorgt voor een kwart van het tekstmateriaal.

A.6.4. Elektronische Woordenbank van de Nederlandse Dialecten (eWND) en regionale woordenboeken

In 2020 is de elektronische Woordenbank van de Nederlandse Dialecten (eWND) van het Meertens Instituut overgegaan naar het INT. De eWND bevat meer dan zeventig lokale woordenboeken van 51 verschillende Nederlandse dialecten, die vaak speciaal ten behoeve van eWND zijn gedigitaliseerd. De woordenboeken dateren van eind 18e eeuw tot de 21e eeuw. Er worden nog steeds woordenboeken aan toegevoegd. Alle dialectingangen worden door vrijwilligers of stagiairs verrijkt met vernederlandsing, zodat gebruikers en onderzoekers alle dialectwoorden bij elkaar kunnen vinden zonder dat ze hoeven te weten hoe het in het dialect wordt gespeld.

In 2020 heeft INT ook samengewerkt met de Radboud Universiteit aan de ontsluiting van regionale woordenboeken. Dat heeft geresulteerd in de lancering van het e-WALD, het elektronische Woordenboek van de Achterhoekse en Liemerse Dialecten, op 11 december 2020.

A.6.5 Overige activiteiten

Eind 2020 verscheen een eerste ‘Uit de Streek’, een rubriek over dialectwoorden uit de DSDD en de woordenbank die vanaf 2021 tweewekelijks wordt gepubliceerd op de website.

In het kader van CLARIAH⁺ (zie aldaar) neemt het INT deel aan een deelproject ten behoeve van de integratie van dialectwoordenboeken in de CLARIAH-infrastructuur.

A.7. Vertaalwoordenschat

In september 2017 heeft het INT het onlineplatform, 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.

In voorgaande jaren zijn het Nederlands-Nieuwgrieks / Nieuwgrieks-Nederlands, het Nederlands-Portugees / Portugees-Nederlands, het Nederlands-Estisch ontsloten. Eind 2020 is het Nederlands-Fins/Fins-Nederlands toegevoegd. Dit is nu nog alleen intern toegankelijk. In 2021 zal dit officieel gereleased worden.

Daarnaast is er in 2020 gewerkt aan de omspelling van glossen en voorbeelden naar de huidige spellingregels. Eerst is geïnventariseerd hoeveel unieke glossen en voorbeeldzinnen er in de reeds ontsloten OMBI-bestanden voorkwamen. De unieke glossen zijn vervolgens alle 77.662 handmatig gecontroleerd en indien nodig gecorrigeerd. Voor de 238.533 unieke voorbeeldzinnen is een deels automatische en deels handmatige controle gedaan. Alleen die voorbeeldzinnen waarvan de afzonderlijke woorden niet in GiGaNT-Molex voorkwamen, zijn handmatig gecontroleerd. Dit waren ruim 19.000 voorbeeldzinnen. Bij de spellingcorrectie viel op dat een inhoudelijke update op termijn ook wenselijk is. De resultaten van de spellingcorrectie zullen voor de verschillende taalparen in de nieuwe update zichtbaar zijn.

Tot slot is er in 2020 voorzichtig begonnen met de voorbereidingen om afhankelijk van de mogelijkheden Nederlands-Arabisch/Arabisch-Nederlands in 2021 te ontsluiten.

A.8. Expertisecentrum Terminologie

De ontwikkeling van het ENT als een gewaardeerd en erkend expertisecentrum voor terminologie heeft in 2020 verder haar beslag gekregen. Kerntaken in dat verband zijn veldondersteuning en positionering in een terminologisch netwerk die het ENT op verschillende wijzen heeft gerealiseerd.

A.8.1. Website

Ondersteuning door informatieaanbod en voorzieningen kwam tot stand via de intussen uitgebouwde communicatiekanalen. Zo is de terminologiepagina van de INT-website geconsolideerd en deels verder ontwikkeld. Nieuw toegevoegd is de rubriek met de bibliografie over terminologie. Deze maakt boeken en artikelen over dit vakgebied doorzoekbaar in 11 deelrubrieken en voorziet in de mogelijkheid om het materiaal chronologisch te sorteren en te downloaden.

Eveneens hebben de werkzaamheden voor een nieuwe rubriek over Europese terminologiegerelateerde instellingen hun beslag gekregen. Ze resulteerden in een dataverzameling die nog verder is uitgebreid en een update onderging. Daarbij zijn vooral drie selectiecriteria gehanteerd: 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. Deze data zijn gevisualiseerd en raadpleegbaar in de vorm van een landkaart met routeplanner. Publicatie op de website wordt in de loop van 2021 verwacht.

Voor de al bestaande voorziening van online beschikbare Nederlandstalige termenlijsten, doorzoekbaar volgens de internationaal erkende onderwerpenclassificatie van de Library of Congress, is verder materiaal verzameld of toegevoegd.

De agenda met *Evenementen rond terminologie* die een overzicht levert van initiatieven in Nederland, Vlaanderen en Europa, werd eveneens verder aangevuld.

Naast voorzieningen op de website is ook het communicatiekanaal van de nieuwsbrieven terminologie ieder kwartaal actief benut. Deze vier nieuwsbrieven bieden informatie in een vaste rubriekenreeks over voorbije evenementen, nieuwe ontwikkelingen en publicaties, tools en bronnen, een uitgelichte term, berichten van NL-Term en een agenda. Ze blijven beschikbaar in het archief op de terminologiepagina.

A.8.2. Projecten voor specifieke vakdomeinen (HOTNeV)

Ook projecten met toegepaste waarde die online raadpleegbare materiaalverzamelingen moeten opleveren, vormden in 2020 deel van de veldondersteuning.

Het *HOTNeV-project* (Hoger Onderwijs Terminologie in Nederland en Vlaanderen) heeft als doel om Engelstalige EU-terminologie te voorzien van Nederlandstalige equivalenten die binnen de

Nederlandstalige onderwijswereld zijn geaccordeerd in plaats van ad hoc gecreëerd. De basis vormt een databestand aangeleverd vanuit de RUGroningen. De beschrijving van onderwijstermen is in 2020 voortgezet met onder meer vier stages. Verder werd op drie vergaderingen van de Termraad Nederlands over een aantal HOTNeV-termen en concepten gediscussieerd. De resultaten van de werkzaamheden zijn opgeslagen in een HOTNeV-databank met terminologiebeheertool QTerm. In het kader van dit project werden ook de stageplannen van studenten begeleid voor drie nieuwe stages die begin 2021 van start gaan. Via een enquête werd bij leden van de werk- en stuurgroep van het vroegere HOTNeV-pilootproject getoetst of er mogelijkheden zijn voor een corpus van teksten over hoger onderwijs bij de respectievelijke onderwijsinstellingen. De databank met HOTNeV-termen zal begin 2021 op de daartoe bestemde rubriek op ENT-webpagina worden gezet.

Het INT ondersteunde tevens het *Proefproject Nederlands als wetenschapstaal - van corpora naar terminologielijsten*. In samenwerking met de Stichting Nederlands / Vlaams Platform Taalbeleid Hoger Onderwijs, en ook met KU Leuven en UGent, worden taalhulpmiddelen ontwikkeld voor studenten om de overstap van het middelbaar naar het hoger onderwijs makkelijker te maken. Van de beoogde vakgebieden uit de alfa-, bèta- en gammastudierichtingen zijn voor de exacte vakken al effectief resultaten geboekt. Cursusmateriaal BA1 uit Nederland en Vlaanderen is door het INT gedigitaliseerd en termen werden geëxtraheerd. Er is een uitgebreide scheikundige termenlijst met bijbehorende definities en Engelse termvertalingen verrijkt, gevalideerd en afgerond. Met de redactionele begeleiding hiervan is het ENT intussen begonnen en ook voor een volgend onderdeel wiskunde vond reeds termextractie plaats.

A.8.3. Medische terminologie

De medische terminologie en een verbrede aandacht voor zorgterminologie bleven een centraal aandachtspunt in 2020. Doelstelling is terminologisch bij te dragen aan een platform voor medische taal. Daarmee werd een onderzoeksproces gecontinueerd dat in 2019 startte. Meer specifiek is in 2019 een licentie aangevraagd voor de internationale medische terminologiestandaard SNOMED/Nictiz en werden de gegevens ervan in een Lex'it-database ingevoerd om de structuur van de ontologie te verkennen. Tevens vond overleg plaats voor een geschikte terminologische tool en met Nictiz werden besprekingen gevoerd over de mogelijkheden van de SNOMED-data. Na deze voorbereidingen werd 2020 het jaar om de terminologische beschrijving en dataverwerking in de praktijk te testen.

Dit is uitgevoerd, met als resultaat een databank van 1800 medische termen die a) de SNOMED-code en wetenschappelijke classificatie omvatten, b) de woorden uit het ANW (algemene standaardtaal) en de taalvariatie in Vlaanderen en Nederland. De databank werd gemaakt in Lex'it.

In samenhang daarmee bereidt het INT de publicatie van het *Pinkhof Geneeskundig woordenboek* voor, en in het kader daarvan werden extra termen voor het woordenboek bewerkt, die ook in Lex'it werden opgenomen.

De resultaten van deze medische pilot zullen vanuit Lex'it worden geëxporteerd naar een formaat dat leesbaar en raadpleegbaar is voor elke gebruiker en eind februari 2021 online staan via de rubriek "Taalmaterialen". Er zal een inleiding en documentatie worden bijgevoegd.

Op het vlak van de medische vaktaal zijn nog nieuwe termen toegevoegd aan het databestand van *Pinkhof Geneeskundig Woordenboek*. Met de Stichting beheer Database Pinkhof Geneeskundig woordenboek besprak het INT de overdracht van de databank en werden juridische voorwaarden vastgelegd. Voor de concrete exploitatie vond binnen het INT computationeel-linguïstisch overleg plaats inzake selectie van informatiecategorieën uit dit bestand.

Ook de COVID-lijst behoort hierbij; het INT heeft als taak de actuele woordenschat te monitoren en de evolutie in de vaktaal, zeker bij een crisis in de gezondheidszorg zoals de pandemie te volgen. Deze pandemie heeft ervoor gezorgd dat er uitzonderlijk veel nieuwe termen gecreëerd werden.

A.8.4. Juridische terminologie

In het kader van het *project Juridische Terminologie* kreeg de beoogde uitbreiding van het *Juridisch Woordenboek* van M.C. Oosterveld-Egaz en J.B. Vuyk-Bosdriesz haar beslag. Dit woordenboek, 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. Tevens werd overleg gevoerd met de Stichting Rechtstaal over de overdracht. Bedoeling is om het woordenboek in 2021 online beschikbaar te maken.

A.8.5. Terminologietools

Naast informatie en voorzieningen biedt het ENT in het kader van haar hoofdtaak ook producten en diensten aan. Wat betreft de producten vond advies plaats voor onderhoud en functionele verbetering van de termextractor TermTreffer. Computationeel-linguïstische werkzaamheden vanuit het instituut resulteerden in de ontwikkeling van een actueel referentiebestand voor deze applicatie dat na implementatie door IT het huidige verouderde en kleine referentiebestand zal vervangen. Meer specifiek is een corpus samengesteld door randomselectie uit het SoNaR-corpus. In verband met auteursrechten is gekozen voor individuele zinnen van voldoende lengte. Eveneens voerde IT een versterking door van de beveiliging van de TermTrefferwebsite en vonden daarvoor bij ENT tests plaats voor de werking van de applicatie. Traditioneel werd nog ondersteuning geboden bij gebruikersvragen.

A.8.6. Terminologisch netwerk

Positionering in een netwerk kreeg onder meer vorm door deelname aan of bijwonen van congressen en symposia (Termos 2020, EFNIL, TiNT 2020, Verbinden & Innoveren met SNOMED). Het ENT nam ook 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. Verder vond in 2020 externe dienstverlening plaats. Voor hogeronderwijsinstellingen in binnen- en buitenland werden vier stagiaires begeleid, waarvan twee in samenwerking met het Departement Nederlands van het directoraat-generaal voor vertaling van de EU. Advies over de beschrijving van COVID-19-terminologie werd verstrekt in het kader van online bijeenkomsten van Nictiz. Voor de KULeuven-campus Antwerpen werd een gastcollege over conceptanalyse en termbeschrijving gegeven.

A.9. De digitale Pallas

In 2020 is het project ‘De digitale Pallas’ gestart. Dit houdt in de vervaardiging van een geannoteerde en geüpdatete digitale editie van het Russische vierdelige *Vergelijkende woordenboek van alle talen en dialecten* (Сравнительный словарь всех языков и наречий, по азбучному порядку расположенный) door Peter Simon Pallas (1790-1791). De Duitse onderzoeker Pallas stelde op initiatief van de Russische tsarina Catharina de Grote een programma en vragenlijst samen om materiaal van de talen van Rusland en de hele wereld te verzamelen ten behoeve van taalvergelijkend onderzoek. De vragenlijst werd niet alleen verstuurd naar de administrateurs van de Russische provincies, maar ook naar Europa, China, Brazilië en Noord-Amerika. Op basis van het verzamelde materiaal publiceerde Pallas een vergelijkend woordenboek in twee delen (1787-1789), dat in 1790-1791 opnieuw werd uitgegeven in vergrote en gewijzigde vorm. De tweede druk bevat 62.000 woorden voor 297 concepten in 328 talen. Het materiaal is nog steeds erg belangrijk voor de vergelijkende taalkunde en de historische taalkunde, en bevat ook data van bedreigde talen. Via een aangepaste versie van de databewerkingsapplicatie Lex’it zijn in 2020 door vrijwilligers en stagiaires alle woorden uit het vierdelige woordenboek overgetypt. In 2021 wordt het materiaal verder ontsloten met onder andere transcripties van de cyrillische woorden.

A.10. Lexicon Frisicum

In februari werd onlinevertaling van het Lexicon Frisicum gepresenteerd in het provinciehuis in Leeuwarden. Het lexicon 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 6.400 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. Werkzaamheden aan deze applicatie kaderden mede in de ontwikkeling van een nieuwe generieke woordenboekapplicatie, waarmee op termijn alle woordenboekdata online zal worden gezet. Het woordenboek is te vinden op <http://lexiconfrisicum.ivdnt.org/>.

B. Corpora en lexica

Basistaken

B.1. Moderne corpora

De moderne lexiconbouw en de beschrijving van het hedendaags Nederlands worden 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, dataprocesing en taalkundige verrijking, en anderzijds uit het beschikbaar stellen van het materiaal in een corpusapplicatie voor intern en extern gebruik.

Een belangrijk resultaat van 2020 is de verwerving van de Vlaamse kranten van het Mediahuis, die dagelijks binnenkomen en verwerkt worden. Daarnaast is het automatisch binnenhalen en verwerken van drie fora, respectievelijk uit Suriname, Vlaanderen en Nederland, in productie gegaan. Afgezien van de reguliere updates van het materiaal uit Vlaanderen, Nederland en Suriname en de Antillen zijn er een aantal corpora voor eenvoudig Nederlands aan het CHN toegevoegd, namelijk Basilex en het WAI NOT-corpus. Ten slotte is in de context van het Europese CLARIN-project ParlaMint, in december gewerkt aan het converteren naar ParlaMint-TEI van het Belgisch federaal parlement. Omdat TEI ook de standaard is van het CHN kunnen de parlementaire data komend jaar ook aan het CHN worden toegevoegd. Meer informatie over het ParlaMintproject is te vinden op de [CLARIN-website](#).

Omdat het corpus richting de twee miljard woorden gaat, is voor de geplande onlinegang in de lente van 2021 gewerkt aan het corpus backend (BlackLab). De vele verbeteringen die gerealiseerd zijn van het corpus frontend ten behoeve van de historische corpora komen ook het CHN online ten goede.

B.2. Historische corpora

In 2019 is flink geïnvesteerd in corpus frontend en backend voor historische corpora. Daarbij is gekeken naar zoveel mogelijk uniformering, met de mogelijkheid om efficiënt een aanpassing van de user interface en de zoekmogelijkheden te maken indien een specifiek historisch corpus dat nodig heeft. In 2020 is de BlackLab-gebaseerde applicatie voor de onlinecorpora 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. 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. 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, en het Brieven als Buit-corpus.

Op 1 september zijn het Corpus Middelnederlands en het Corpus Gysseling in de nieuwe applicatie online gegaan. Voor beide corpora is een betere genrebeschrijving uitgewerkt en toegepast. Er is ruim aandacht besteed aan een gebruikersvriendelijke help. Het Corpus Gysseling is ook doorzoekbaar met de TDN-tagset. Het Corpus Middelnederlands is nog niet taalkundig verrijkt, maar maakt gebruik van GiGaNT-Hilex via de lexicon service voor een betere doorzoekbaarheid. Het is de bedoeling dat dit corpus op termijn ook taalkundig verrijkt zal worden. Hiervoor zal gebruikgemaakt worden van het werk dat hiervoor in het kader van CLARIAH plus wordt gedaan.

Tot slot is de update van het Brieven als Buitcorpus verder afgerond, en het werk aan de aanvullingen op Brieven als Buit grotendeels voltooid, die respectievelijk in januari 2021 en februari 2021 online komen.

B.3. Spelling

In 2020 werden wederom nieuwe trefwoorden toegevoegd aan de nog steeds groeiende onlineversie van de *Woordenlijst Nederlandse Taal*. Aangezien de nadruk vanaf dit jaar ook niet meer ligt op kwantiteit maar wel op kwaliteit, werd vooral gekeken naar trefwoorden met een potentieel spellingprobleem én zijn er ook extra woordvormen en andere woordinformatie toegevoegd. Daarnaast werd een begin gemaakt met het aanvullen van allerlei reeksen (o.a. Europese mannelijke en vrouwelijke inwonernamen, chemische elementen enz.) en het consistent maken van bestaande trefwoorden en woordvormen. De wijzigingen en toevoegingen aan het spellingbestand werden beoordeeld door de Commissie spelling, die ook mede de probleemgevallen beoordeelde. Dit gebeurde afgelopen jaar in twee onlinevergaderingen.

Er is in 2020 verder gewerkt aan de hostingovername van *woordenlijst.org*. E.e.a. gaf aanleiding tot het definiëren van een plan voor een nieuwe spelling-API die eind 2020 aan de Taalunie en de spellingcommissie is voorgelegd.

Ook in 2020 werden door het INT spellingvragen (en algemenere taalvragen) van gebruikers zo snel mogelijk beantwoord of doorgestuurd naar de juiste persoon (of dienst).

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

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 ca. 24.000 nieuwe lemmata in Hilex, en ruim 35.000 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 Diachroon 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 lexicon service, waarmee GiGaNT Hilex toegankelijk wordt gemaakt voor gebruik in bijvoorbeeld corpusapplicaties als Nederlab, of het Corpus Middelnederlands, maandelijks geüpdatet. GiGaNT-Molex is de moderne lexiconcomponent. Het bevat lemmata met hun compleet paradigma, in de correcte spelling, en is sedert 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. In 2020 is voor het ANW, het neologismenproject en het project

Woordcombinaties technisch de inbedding van het lexicon in de workflow van deze projecten gerealiseerd. De inhoudelijke koppeling met Woordcombinaties is een feit, en voor het volledige ANW is de koppeling van de volledige artikelen ook gerealiseerd. Ten behoeve van het project Woordcombinaties is aan de ruim 13.000 werkwoorden in Molex informatie toegevoegd over het hulpwerkwoordgebruik in de o.t.t. en v.t.t., en er is per werkwoord aangegeven of het transitief, intransitief, reflexief of onpersoonlijk gebruikt wordt. De koppeling met het ANW heeft ruim 1.600 nieuwe lemmata voor Molex opgeleverd, waarvan de bewerking eind 2020 is gestart. Om de koppeling met de vertaalwoordenschat van het INT te realiseren, is gestart met de koppeling van Molex aan het Referentiebestand Nederlands. De koppeling met lemmata die al in Molex stonden, is gerealiseerd. In een volgende fase moet Molex verrijkt worden met lemmata die het RBN heeft, maar Molex nog niet.

Beide GiGaNT-modules zijn opgeslagen in een PostgreSQL-database en maken gebruik van de op het INT ontwikkelde bewerkingsomgeving Lex'it. Afgezien van werkzaamheden aan de API en waar nodig aan de tool is er in 2020 gezorgd dat er gewerkt kan worden op een geschiktere database-infrastructuur die de groei en het intensieve gebruik van de bewerkingsomgeving op de databases aankan.

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 de afgelopen jaren is gewerkt aan het opzetten van een infrastructuur voor de lexiconbouw op basis van woordenboek- en corpusdata. De in het lexicon opgenomen data zijn nu merendeels gebaseerd op de onomasiologische informatie (synoniemen) in de definities van het MNW en het WNT. Het resultaat is, mede in het kader van CLARIAH, gepubliceerd als linked open data en in een eerste versie van een applicatie voor het bredere publiek.

In 2020 is het datamodel van de linked open data-versie van het lexicon in het kader van het werk aan de OntoLex module voor FRequency, Attestaties en Corpus data (FRAC)² verder uitgewerkt en afgestemd met het OntoLex datamodel (Resultaat gepubliceerd in Chiarcos et. al. 2020).

Er is onderzoek gedaan naar de mogelijkheden om de data te verrijken door middel van de koppeling van een externe thesaurus of ontologie. Hierbij is met name gekeken naar de Thesaurus of Old English³ en de Historical Thesaurus of English⁴. Uiteindelijk is er bij wijze van proef een deelset van

² <https://github.com/ontolex/frequency-attestation-corpus-information>

³ <https://oldenglishthesaurus.arts.gla.ac.uk/>

het *Oudnederlands Woordenboek* gekoppeld aan de TOE. Deze deelset bestond uit de familietermen uit het ONW en een random selectie. Dit heeft tot inzicht geleid over de mate waarin een voor een specifieke taalfase van het Engels ontwikkelde ontologie bruikbaar is voor een indeling van Oudnederlandse woordvelden. Op grond van deze handmatige koppeling is ook onderzocht in hoeverre dit werk te ondersteunen is met op computationele technieken gebaseerde definitie-similariteiten, zowel gebaseerd op string matching als op distributionele technieken. Het resultaat is gepresenteerd op de EASE-workshop van 17 oktober 2020.

Tot slot is er een begin gemaakt met het onderzoek naar de bruikbaarheid van contextgevoelige vectoriële woordrepresentaties voor de ondersteuning van lexiconbouw- en toepassing. Als eerste verkenning is onderzocht of op BERT-modellen gebaseerde inductie van betekenisgroepen voor het modern Nederlands vergelijkbare resultaten oplevert met wat reeds voor het Engels gedaan is⁵. Dit blijkt het geval te zijn. Dit onderzoek is uitgevoerd met de moderne Nederlands SemCor-data; de volgende stap is het samenstellen van een historisch semantisch testcorpus op basis van de DiaMaNT-data.

⁴ <https://ht.ac.uk/>

⁵ Vgl Amrami en Goldbert 2019, <https://arxiv.org/abs/1905.12598>

C. Grammatica

Basistaken

C.1. Grammatica algemeen

Vanaf 2020 is *grammatica* een van de structurele basistaken van het INT geworden. Dit betekent dat het INT zorgt voor de ontwikkeling, het beheer en de beschikbaarstelling van diverse digitale grammaticaproducten, waaronder het Taalportaal en de e-ANS, en in een later stadium ook de taaladviezen van taaladvies.net.

Hiervoor is in 2020 een plan van aanpak geschreven, waarin drie kerndoelen worden geschetst: het opzetten van een centraal grammaticaportaal – een website die als *hub* en startpagina dient voor de grammaticaproducten, de voortdurende inhoudelijke actualisering van de producten, en het beter ontsluiten en onder de aandacht brengen van de inhoud.

Voor Taalportaal heeft het INT in het begin van 2020 de website geactualiseerd, en in samenwerking met vIvA in Potchefstroom, Zuid-Afrika, veel nieuw Afrikaans materiaal gepubliceerd. Dit is in mei feestelijk gepresenteerd via een webinar, met Frank Landsbergen van het INT als eregast.

C.2. e-ANS

Na de afronding van het herzieningsproject van de ANS eind 2019, is de ANS formeel ondergebracht bij het INT. Vanaf 2020 is het INT verantwoordelijk voor de verdere herziening van de ANS. Hiervoor is aan het begin van het jaar Maaïke Beliën in dienst gekomen. Zij regelt en begeleidt de verdere inhoudelijke herziening door externe redacteurs, en zal ook zelf meeschrijven aan de herziening. Aan het eind van het jaar zijn de eerste contracten afgesloten voor de herziening van twee hoofdstukken in 2021. De lancering van de vernieuwde website stond gepland voor 2020, maar is mede door corona en (andere) ziekte opgeschoven naar voorjaar 2021.

Op technisch gebied is de nieuwe ANS-website verder ontwikkeld, met links naar relevante taaladviezen en een betere responsive opmaak.

C.3. ANK

In 2020 is het onderdeel Klankleer (ANK – Algemene Nederlandse Klankleer) van de ANS herschreven in een project onder leiding van Geert Booij. Dit project is in november 2020 succesvol afgerond, en het vernieuwde deel zal in de loop van 2021 worden gepubliceerd. De opzet en succesvolle uitwerking ervan vormen een blauwdruk voor de toekomstige herzieningen van de resterende hoofdstukken.

C.4. Taalportaal Zuid-Afrika

Voor het project *Taalportaal Zuid-Afrika* is aan het begin van 2020 een in 2019 op het INT aangepaste Taalportaal-java-applicatie naar Sadilar in Zuid-Afrika gestuurd. De bedoeling is dat Sadilar deze applicatie in gebruik zal nemen voor zijn eigen lokale Taalportaal, waarna de zoekresultaten van dit portaal en het Nederlandse Taalportaal kunnen worden gekoppeld.

D. Softwaretools (ontwikkeling, beheer en onderhoud)

Basistaken

D.1. INT Impactcentrum en digitization.eu

Het INT is voorzitter van het IMPACT Centre of Competence (www.digitisation.eu). Dit is een non-profitorganisatie bestaande uit publieke en commerciële organisaties met als doel de digitalisering van historisch materiaal “beter, sneller, en goedkoper” te maken. Het centrum voorziet in data, tools, services en expertise op het gebied van document imaging, taaltechnologie en het verwerken van historisch tekstmateriaal. Het IMPACT Centre of Competence is sedert 2019 ook CLARIN Knowledge Centre. In 2020 heeft het centre een plek gekregen bij de Fundación General van de Universidad de Alicante (FGUA).

In november 2020 is een herwerkte COST-action aanvraag ingediend: *DIALOGUE* (DIGITISAtion knowLedge fOr diGital hUMANities and cultural hERitage), met als doel tot een Europees netwerk voor het delen van kennis op het gebied van digitalisatie te komen. Het INT is een van de aanvragers van dit project.

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 aanbrengen van persistente gebruikersannotaties in corpuszoekresultaten, een gebruikersvriendelijker digitalisatieworkflow en curatie van dialectwoordenboekdata.

In het kader van de infrastructuur voor historisch Nederlands is voortgewerkt aan het platform GALAHAD voor betere en gebruikersvriendelijke taalkundige verrijking van historisch Nederlands, waarbij goede voortgang geboekt is met zowel de automatische component als met de werkomgeving voor handmatige correctie, en met het uitwerken van de Tagset Diachroon Nederlands die als doel heeft de diverse coderingen van taalkundige verrijking in de beschikbare historische corpora op een

flexibele manier te harmoniseren. Het voorstel is in twee besprekingsronden positief ontvangen door de externe gebruikerscommissies en zal begin 2021 gefinaliseerd worden. Voorts is via mapping van de tagsets van ONW, Corpus Gysseling en CGN naar de TDN gewerkt aan de totstandkoming van trainings- en evaluatiemateriaal. Eerste evaluaties en foutanalyses van onder andere de tagger-lemmatizer PIE (<https://github.com/emanjavacas/pie>) zijn uitgevoerd.

Het werk aan de infrastructuur voor digitalisatie en conversie is, mede door capaciteitsissues bij de afdeling IT, doorgeschoven naar 2021. Het uitgevoerde werk aan de CLARIAH use-cases draagt wel bij aan de voorbereiding van deze taak.

Aan de zoekmachine Blacklab zijn verbeteringen doorgevoerd die noodzakelijk waren voor de interactie met het GALAHAD-platform, met name de performance van het opvragen van frequentielijsten drastisch is verbeterd. Deze verbetering komt ook de doorzoekbaarheid van grotere corpora ten goede, zoals het CHN.

Als voorbereiding voor het opnemen van materiaal buiten de DSDD-regio in de CLARIAH-infrastructuur is vooronderzoek gedaan en is er overlegd met de diverse partners. De DSDD-kaartenapplicatie is in dit verband gedeeld met de partners uit Nijmegen.

D.2.2. CLARIAH Vlaanderen

Binnen *CLARIAH-Vlaanderen* werd samengewerkt aan een vervolgaanvraag voor de oproep naar Internationale Research Infrastructuur van het FWO en Departement Economie, Wetenschap en Innovatie (EWI) van de Vlaamse regering. *CLARIAH-VL: Advancing the Open Humanities Service Infrastructure* werd aangevraagd, waarbij er gelet werd op een goede integratie tussen de DARIAH-componenten en de CLARIN-componenten van de aanvraag. Deze aanvraag werd inmiddels goedgekeurd en wordt gefinancierd met een subsidie van meer dan 3,6 miljoen euro.

Binnen het huidige project *CLARIAH-VL: Open Humanities Service Infrastructure*, dat loopt van 2019 tot en met januari 2021, en waarbij het INT de niet-gefinancierde partner is die de CLARIN kant van het voorstel vertegenwoordigt, werd in de loop van 2020 een pilootproject opgestart samen met het Vlaamse Supercomputer Centrum (VSC). Binnen dit pilootproject, dat verder loopt in het vervolgproject *CLARIAH-VL: Advancing the Open Humanities Service Infrastructure*, worden door VSC trainingssessies georganiseerd die expliciet gericht zijn op gebruikers uit de *humanities*, waaronder ook gebruikers uit de computerlinguïstiek. Er werd afgesproken om samen met UAntwerpen een state-of-the-art neuraal taalmodel te trainen op een Nederlandstalig corpus van meer dan 10 miljard woorden. Deze data zijn door het INT klaargezet op de VSC high performance computing cluster. Hier wordt verder aan gewerkt in 2021.

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

Het INT is betrokken bij het ELRC-initiatief. Dit is een doorlopend initiatief dat als doel heeft tekstdata te verzamelen in alle EU-lidstaten, IJsland en Noorwegen, die gebruikt kunnen worden om CEF eTranslation verder te ontwikkelen. CEF eTranslation is een automatische vertaaldienst die door de Europese Commissie ter beschikking wordt gesteld om meertalige communicatie tussen openbare diensten, ministeries en burgers mogelijk te maken. 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.

In 2020 stonden de ELRC-werkzaamheden grotendeels on hold. Er zijn geen nieuwe taalbronnen verzameld voor het ELRC-SHARE repository. Wel waren er twee online Language Resource Board-bijeenkomsten. De workshop om ELRC en het gebruik van CEF eTranslation te promoten binnen de openbare diensten in Nederland zal plaatsvinden in 2021.

E. Taalmaterialen

E.1. Website taalmaterialen

In 2019 hebben we de website voor Taalmaterialen vernieuwd. Daar waren een aantal redenen voor. Ten eerste was de oude website gebaseerd op verouderde technieken (Joomla) die niet meer voldoende ondersteund werden. Een tweede belangrijke reden is dat de oude website geïntegreerd was in de website van het INT. Dat gaf nog wel eens verwarring voor de gebruikers.

Het totaal aantal downloads is iets toegenomen in vergelijking met het voorgaande jaar:

- Niet-commerciële downloads 1.412 (1.169 in 2019).
- Commerciële downloads 44 (41 in 2019).

E.1.1. Nieuwe producten

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

- Names (commercieel en niet-commercieel): Een corpus van Nederlandse voor- en achternamen zoals gevonden in 19de-eeuwse geboorte-, huwelijks- en overlijdensakten. De naamvarianten zijn gekoppeld aan een standaardvorm.
- Diachroon seMantisch lexicon van de Nederlandse Taal – DiaMaNT (niet-commercieel): een computationeel semantisch lexicon dat diachrone semantische informatie biedt. Het is gebouwd als laag bovenop [GiGaNT](#), door betekenisinformatie toe te voegen. Het lexicon legt relaties tussen woordvormen en betekenseenheden (concepten), en plaatst deze in de tijd.
- BLISS Spoken Dialogue Dataset (niet-commercieel): Nederlandse spraakopnames van deelnemers die spreken met het BLISS-dialogosysteem (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.
- Dutch Idiom Database: Native Speakers DID-NS (commercieel en niet-commercieel): Een database met beoordelingen van 390 moedertaalsprekers van 374 Nederlandse uitdrukkingen. In een onlineonderzoek hebben de deelnemers de idiomatische uitdrukkingen beoordeeld op een aantal aspecten: frequentie, gebruik, bekendheid, voorstelbaarheid en transparantie. Ook is onderzocht of ze de juiste betekenis van de uitdrukkingen kenden.
- Corpus Ondertitelde UVN-Colleges COUC (niet-commercieel): Het corpus bevat de ondertiteling van 57 colleges van de Universiteit van Nederland.
- Federated Search Lexica (online service): Een belangrijk onderdeel van het [CLARIAH](#)-project is het aanbieden van de mogelijkheid om verschillende dataverzamelingen gelijktijdig te doorzoeken: "Federated Content Search". Het INT heeft een aantal lexica geconverteerd naar RDF en een interface ontwikkeld om die in combinatie te doorzoeken.

- Greedy Extraction of Trees for Empirical Linguistics – GrETEL (online service): GrETEL is een gebruiksvriendelijke interface voor het doorzoeken van syntactisch geannoteerde corpora oftewel treebanks.

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 (254), daarna België (104) en verder Duitsland (31), Frankrijk (7), Italië (7), 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 ook gekeken naar de geo-locatie van de IP-adressen die zijn geregistreerd bij de downloads. Dat leverde het volgende overzicht op: Nederland (836), België (334), Duitsland (41), UK (27), US (25), Zweden (19), Canada (13), Frankrijk (12), Spanje en Zuid-Afrika (10), Indonesië (6) en een groot aantal lagere scores.

E.1.3. Downloads

In totaal is er 41 keer een product afgenomen voor commerciële toepassingen. CGN Commercieel is het vaakst afgenomen (14), gevolgd door CGN Annotaties Commercieel (3) en Gigant-Molex Commercieel (3).

Er zijn in totaal 1.412 producten gedownload voor niet-commercieel gebruik. Het Corpus Gesproken Nederlands (CGN) is het vaakst gedownload (224), gevolgd door het SoNaR-corpus (147) en de Frequentielijsten Corpora (105). Zie bijlage 4 voor een gedetailleerd overzicht.

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

Om Vlaanderen toegang te bezorgen tot de CLARIN-infrastructuur, wat slechts mogelijk is als België lid wordt van, en lidgeld betaald aan CLARIN-ERIC, werd op aanraden van het FWO en EWI, een aanvraag tot politieke steun vanuit Vlaanderen ingediend binnen de oproep Internationale Research Infrastructuur. Deze aanvraag, met als coördinator Prof. Dr. Els Lefever (UGent. LT3), en als consortium partners UGent-IDLab, KULeuven-CCL, KULeuven-ESAT-PSI, KULeuven-LIIR en UAntwerpen-CLIPS, werd inmiddels goedgekeurd en de procedures ter oprichting van CLARIN-België en het lidmaatschap van België bij CLARIN-ERIC zijn opgestart, en moeten nog alleen een aantal formele fases doorlopen.

Er werden in 2020 ook voorbereidingen getroffen om van het INT een CLARIN-Knowledge Centre te maken dat expertise aanbiedt aan Europese (CLARIN) onderzoekers op het vlak van het Nederlands, zowel qua tools en resources als qua inhoudelijke, taalkundige kennis.

F. Onderzoek: Europese aanvragen en competitieve projecten

F.1. European Lexicographic Infrastructure (ELEXIS)

Het INT is partner in ELEXIS (<https://elex.is>), 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.

Begin maart stond de reviewmeeting van ELEXIS gepland. Door de coronacrisis werd deze in eerste instantie gecancelled. Uiteindelijk heeft de review in juni online plaatsgevonden. Het project heeft een positieve review gekregen.

Begin 2020 zijn een aantal deliverables opgeleverd waar het INT aan mee heeft gewerkt, namelijk

- [Deliverable 1.2: Best practices for lexicography – intermediate report](#)
- [Deliverable 6.2: Recommendations on legal and IPR issues for lexicography](#)
- [Deliverable 5.2: Guidelines for Producing ELEXIS tutorials and instruction manuals](#)

Het INT leidt het werkpakket “Lexicographic data and workflow”. Binnen dit werkpakket is in 2020 verder gewerkt aan het ontwikkelen van een common vocabulary voor het ELEXIS datamodel. In oktober werd een onlineworkshop over dit onderwerp gehouden met de titel ‘Digital Lexicographic Data Models’.

Verder zijn in de loop van het jaar de lexicografische bronnen van de verschillende partners en observers binnen ELEXIS verzameld en is begonnen met de conversie hiervan, zodat ze geïntegreerd kunnen worden in de ELEXIS infrastructuur. Dit alles in samenwerking met de andere partners.

Om een nog beter inzicht te krijgen in de lexicografische workflows, metadata en dataformaten die worden gebruikt in lexicografische projecten binnen Europa is in de zomer van 2020 de survey die eerder onder de ELEXIS-partners is gehouden uitgebreid naar de [observers](#). De resultaten van deze nieuwe survey zullen in 2021 worden gepubliceerd.

Binnen de andere werkpakketten heeft het INT bijgedragen aan verschillende werkzaamheden. Zo was het INT betrokken bij de shared task over het linken van woordenboeken (gepresenteerd op de

GlobalLex workshop op [LREC 2020](#) in mei 2020), en heeft het INT de [MWE Lex-2020-workshop](#) op COLING 2020 in december 2020 meegeorganiseerd.

Door de coronacrisis zijn er in 2020 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](#) beleefde in 2020 het vierde jaar. 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 en Tanneke Schoonheim zijn lid van het Management Committee. Verder is Mathieu Fannee bij deze actie betrokken.

In 2020 had de General Assembly moeten plaatsvinden in Napels eind maart, maar door corona ging dat niet door. Het COST-agentschap besloot de actie, die normaal gezien zou eindigen in maart 2021, te verlengen tot september 2021. Er werden tijdens de pandemie veel webinars georganiseerd, en het managementteam vergaderde online.

Het project 'Crowdsourcing corpus cleaning for language learning' zou worden gepresenteerd op EURALEX (8 tot 12 september 2020 in Alexandroupolis (Griekenland)), maar ook dat is uitgesteld tot 2021. 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.

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 de opkomende 'data science'. Taalkundige data vormen een specifiek geval en zijn tot nu toe nog grotendeels onontgonnen in een big data-context.

Het hoofddoel van NexusLinguarum is om taalkundigen, computerwetenschappers, terminologen en andere belanghebbenden in één netwerk bij elkaar te brengen om zo samenwerking en kennisdeling op het gebied van 'linguistic data science' te bevorderen. De actie is eind oktober 2019 van start gegaan en heeft een looptijd van 4 jaar.

In 2020 werden er twee plenaire netwerkbijeenkomsten gehouden, één in januari in Praag en één in oktober in Lissabon (en online). Daarnaast waren er maandelijks onlinemeetings van de verschillende werkgroepen. Carole Tiberius en Vincent Vandeghinste volgen met name de werkzaamheden in werkgroep 1 ‘Linked data-based language resources’ en werkgroep 2 ‘Linked data-aware NLP services’.

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

Er werd gewerkt aan een projectaanvraag met als doel het bouwen van mobiele applicaties waarbij het mogelijk wordt om automatisch te vertalen tussen gebarentalen en gesproken talen (inclusief tekst). De aanvraag werd inmiddels gehonoreerd met een subsidie van meer dan 5,6 miljoen euro, en het project zal lopen van 1 januari 2021 tot en met 31 december 2024 en wordt geleid vanuit Dublin City University (Prof. Dr. Andy Way) en Universiteit Tilburg (Prof. Dr. Dimitar Shterianov). De talen waarop het project zich onder meer zal focussen zijn Vlaamse Gebarentaal, Nederlandse Gebarentaal en het Nederlands, en het consortium bevat dan ook heel wat partners uit Nederland en Vlaanderen, zoals de Nederlandse Taalunie, VRT, UGent, KULeuven, Vlaams Gebarentaalcentrum, Radbouduniversiteit Nijmegen, Universiteit Tilburg. De rol van het INT bestaat uit dataverzameling en -ontsluiting en het voorzien van de infrastructuur om de mobiel applicatie aan te sturen.

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

Binnen de KULeuven werd een aanvraag goedgekeurd voor het bouwen van een corpus Gesproken Academisch Nederlands in België. Deze aanvraag werd ingediend met als promotor Elke Peters (Taal en Onderwijs) en als copromotoren Patrick Wambacq (ESAT PSI), Vincent Vandeghinste (Computationele en Formele Taalkunde), Serge Verlinden en Lieve Dewachter (beide Instituut voor Levende Talen). Het INT zal in dit project instaan voor de opslag en de ontsluiting van het corpus, via de CLARIN-infrastructuur.

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

In 2020 is een promotietraject gestart binnen het project ‘Spread the new(s). Understanding standardization of Dutch through 17th-century newspapers’, waarvoor subsidie is verkregen van NWO Open Competition SSH. De promovenda werkt gedeeltelijk op INT en gedeeltelijk bij de Radboud Universiteit; het INT neemt de technische voorzieningen voor dit project op zich, waaronder de ontsluiting van de 17e-eeuwse kranten, en de educatieve module die is voorzien binnen de kennisbenutting.

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

Het INT is in 2020 partner geworden in het project ‘The rise of complex verb constructions in Germanic’, dat wordt uitgevoerd door Evie Coussé, Gerlof Bouma en Nicoline van der Sijs en dat is gefinancierd door de Vetenskapsrådet (Swedish Research Council). Het INT is medeverantwoordelijk voor de samenstelling van een diachroon meertalig bijbelcorpus, lopend van de 14e eeuw tot de 21e met Nederlandse, Duitse, Engelse en Zweedse Bijbels. Met de hulp van vrijwilligers zijn vijf oude Nederlandse en Duitse Bijbels (Nieuwe Testament) gedigitaliseerd.

F.8. MacBERTh

In 2020 is het project MacBERTh - Bidirectional Encoder Representations from Transformers of historical English and Dutch met als hoofdaanvrager Lauren Fonteyn goedgekeurd binnen de call PDI-SSH digitale infrastructuur. Het INT is als medeaanvrager betrokken bij dit project dat in 2021 zal starten.

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. Zo was het INT in 2020 betrokken bij de invulling van het programma voor de eerste Neerlandistiekdagen op 6 en 7 maart 2020, geïnitieerd door de Raad voor de Neerlandistiek. Aan de KU Leuven is het vak Computationale 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.

Daarnaast verschuift het INT zijn werkerrein, gezien de brede taakomschrijving, nadrukkelijk naar docenten en studenten. In dat verband is het Instituut aanwezig op en profileert het zich op beurzen, conferenties (HSN-conferentie) en festivals (Week van het Nederlands).

Op de in 2020 vernieuwde website van het INT is een nieuwe categorie voor onderwijs ingericht, onderverdeeld in informatie voor docenten, leerlingen en studenten. Ook is er in de Week van het Nederlands een webinar georganiseerd om te leren zoeken in de Historische Woordenboeken. Verder is er overleg met de Nederlandse Taalunie en het IVN 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. Op onze website verschijnen wekelijks populairwetenschappelijke rubrieken over nieuwe en historische woorden, we zijn actief op verschillende sociale media en tien keer per jaar worden twee soorten nieuwsbrieven verstuurd aan geïnteresseerden (zie H.5). Daarnaast houden medewerkers voordrachten, zijn zij te horen in radioprogramma's en dragen zij bij aan publicaties voor een algemeen publiek dat belangstelling heeft voor taal in het algemeen en Nederlands in het bijzonder.

H. Wetenschapscommunicatie

H.1. 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 met dit jaar een toename van 474. Naast Twitter heeft het INT nog een Facebookpagina, een LinkedIn-pagina en een YouTube-kanaal. Sinds de uitbraak van het coronavirus heeft het werken met video een vlucht genomen. In 2021 zetten we deze ontwikkeling voort in het maken van nieuwe video's. Daarnaast voegen we Instagram aan onze sociale media toe.

H.2. Nieuw woord van de week; Terug in de Taal; Uitgelichte term; Gelegenheidswoordenboekjes; Woordbaak

De aandachttrekkers op de website zijn de populairwetenschappelijke rubrieken Nieuw woord van de week, Terug in de Taal en de Gelegenheidswoordenboekjes. Ze worden goed gelezen, vaak gedeeld op de sociale media en overgenomen in taalnieuwsbrieven van de Taalunie en *Onze Taal*. Deze rubrieken blijven in 2021 verzorgd, vernieuwd en uitgebreid worden.

Bezoekers van de website zijn ook zeer geïnteresseerd in etymologie: de rubriek Woordbaak (herkomst van woorden) wordt sinds twee jaar niet meer actief bijgehouden maar blijft desondanks de meest bezochte rubriek op de website. Vooral de artikelen over *doni en barki* en *pappa en mamma* zijn in 2020 vaak bekeken. De terminologierubriek Uitgelichte term wordt minder goed gelezen, maar deze rubriek verschijnt dan ook maar vier keer per jaar en is daarnaast in zijn geheel te lezen in de terminologienieuwsbrief.

In 2020 zijn er geen nieuwe (online)gelegenheidswoordenboekjes verschenen. De rubriek is daarom minder onder de aandacht gebracht en ook minder vaak bezocht dan het voorgaande jaar. Van de eerder gepubliceerde woordenboekjes is *Het Groene Woordenboekje* (2.787 views) het vaakst bekeken. Deze lijst is eind 2020 nog geüpdatet en aangevuld.

In 2021 worden de webrubrieken uitgebreid met twee wekelijkse columns van Ludo Permentier en Ewoud Sanders, en Uit de streek: een nieuwe rubriek over dialectwoorden.

H.3. Populairwetenschappelijke uitgaven en activiteiten: boeken, webinars, podcasts

Op 9 november 2020 was het precies tweehonderd jaar geleden dat de Nederlandse taal- en letterkundige Matthias de Vries werd geboren. Daarom vierde het INT in 2020 het Matthias de Vriesjaar en organiseerde het hele jaar door verschillende activiteiten, zoals een stadswandeling, een Matthias de Vriesquiz, een onlinelezing over het werk en leven van De Vries, een kranslegging op begraafplaats Groenesteeg, een Groot Dicté der Nederlandsche Taal in de spelling-De Vries en Te Winkel, het symposium ‘Op je woorden letten’ (5 november) en een bakkebaardverkiezing. In november verschenen ter ere van het gedenkjaar bij uitgeverij Scriptum twee boeken: *Geen woord te veel. Geschiedenis van INL tot INT* van Piet van Sterkenburg en Frieda Steurs, en *Matthias de Vries aan het woord. Uitgave ter gelegenheid van het Matthias de Vriesjaar 2020* van Roland de Bonth, Dirk Geirnaert en Rob Tempelaars. Het laatste boek werd als nieuwjaarsgeschenk verstuurd naar onze relaties.

Als gevolg van de uitbraak van het coronavirus zijn in 2020 verschillende evenementen uitgesteld of veranderd van opzet, maar uiteindelijk kon alles in aangepaste vorm doorgaan. Het bood ook nieuwe mogelijkheden, en we ontwikkelden nieuwe vormen van communicatie door te werken met video, audio, livestreams, webinars en evenementen deels op locatie en deels online.

Zo maakten we een podcast. Op 5 oktober, tijdens de Week van het Nederlands, kwam ‘Waar komt pindakaas vandaan?’ online. Een podcastserie van 12 korte afleveringen waarin de herkomst van alledaagse maar opmerkelijke woorden besproken wordt. De podcast is in totaal al 31.345 keer gedownload.

Tijdens de Week van het Nederlands 2020 organiseerden we een middag rondom de lancering van de Database van de Zuidelijk-Nederlandse Dialecten (DSDD): *Horen, zien en proeven van dialect*. Verder hielden we in die week nog een webinar over zoeken in de Historische Woordenboeken en een tutorial over OpenSoNaR.

H.4. Website

De website is in 2020 volledig vernieuwd. Het Joomla!-CMS is vervangen door WordPress en de gehele site is opnieuw opgebouwd, ingericht en herschreven. De indeling van de website is overzichtelijker en compacter en het (hamburger)menu is duidelijker en platter van structuur. Ook is de website nu responsive, dat wil zeggen: geschikt voor gebruik op mobiel en tablet. Er zijn nieuwe foto's van alle medewerkers gemaakt en iedereen heeft een eigen profielpagina. Veel teksten zijn herschreven en weergegeven in een vast format, er is een verklarende lijst met taalkundige begrippen toegevoegd en de toegankelijkheid voor mensen met een visuele beperking is verbeterd. Ten slotte is

er een overzicht van alle webpagina's die vertaald zijn in het Engels. De component *Taalmaterialen* is een aparte, zelfstandige website geworden om de gebruikers ervan beter te kunnen bedienen.

Om de vernieuwde website te onderhouden en het plaatsen van nieuwe berichten en teksten beter te stroomlijnen is er een webredactie gevormd (Laura van Eerten (eindredactie), Roland de Bonth, Vivien Waszink, Boukje Verheij, Rob van Strien (webmaster)).

De website heeft in 2020 646.216 bezoekers getrokken, opnieuw flink veel meer dan in het voorgaande jaar (576.575). De naamsbekendheid van het INT groeit en de activiteiten nemen toe, en dat is goed te zien aan de bezoekersaantallen. Opvallend is dat het aantal bezoeken uit België fors is gestegen. Het gaat om een toename van zo'n 36% ten opzichte van 2019.

H.5. Nieuwsbrieven, mailings, persberichten

We versturen de INT-nieuwsbrief en de Nieuwsbrief Terminologie om geïnteresseerden te blijven binden aan en te informeren over het instituut. In 2020 zijn er zes INT-nieuwsbrieven verschenen en vier terminologienieuwsbrieven. Daarnaast zijn er nog diverse aparte mailings verzonden.

Het abonneebestand van de twee nieuwsbrieven is in 2020 gegroeid: de INT-nieuwsbrief kreeg er 362 abonnees bij en de terminologienieuwsbrief 368. De cijfers voor het aantal opens en kliks voor de algemene nieuwsbrief zijn al jaren relatief hoog in vergelijking met nieuwsbrieven van andere organisaties (gemiddeld 38% opens en 16,3% CTO volgens de Nationale E-mail Benchmark van 2018). De Nieuwsbrief Terminologie zit 10% onder dat gemiddelde.

Gerichte mailings over een specifiek onderwerp (uitnodigingen, nieuwjaarswens) naar een selecte groep uit het adressenbestand worden gemiddeld gezien het vaakst geopend. Persberichten worden relatief het minst vaak geopend, maar zijn wel effectief.

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.

Bijlage 1: Raad van Toezicht en Raad van Advies

Raad van Toezicht

Drs. Paul Rüpp, voorzitter

Drs. Gertine van der Vliet

Prof. mr. Jan Cerfontaine

Raad van Advies

Prof. dr. Hans Bennis (voorzitter)

Prof. dr. Willy Vandeweghe (vicevoorzitter)

Prof. dr. Dirk Geeraerts

Prof. dr. Veronique Hoste

Prof. dr. Reinhild Vandekerckhove

Prof. dr. Jack Hoeksema

Prof. dr. Paula Fikkert

Drs. Jan Jaap Knol

Lic. Wim Vanseveren

Prof. dr. Gert Oostindie

Bijlage 2: Medewerkers

Prof. dr. Frieda Steurs - directeur/bestuurder

Marjolijn van Bennekom – taalkundig assistent

Dr. Bob Boelhouwer – computerlinguïst

Dr. Roland de Bonth – onderzoeker/taalkundige

Tim Brouwer - softwareontwikkelaar

Lic. Lut Colman – onderzoeker/taalkundige

Iulianna Ciudin MA – onderzoeker/taalkundige

Lic. Griet Depoorter – CLARIN-coördinator

Lic. Katrien Depuydt – onderzoeker/taalkundige

Dr. Jesse de Does – computerlinguïst

Laura van Eerten MA – communicatieadviseur

Drs. Mathieu Fannee – programmeur/ontwikkelaar

Lic. Dirk Geirnaert – onderzoeker/taalkundige

Drs. Thomas Haga – taalkundig assistent

Dr. Kris Heylen – onderzoeker/taalkundige

Dr. Dirk Kinable – onderzoeker/taalkundige

Drs. Marco van der Laan – systeemontwikkelaar/programmeur

Dr. Frank Landsbergen – computerlinguïst

Koen Mertens – programmeur/ontwikkelaar

Drs. Jan Niestadt – systeemontwikkelaar/programmeur

Henk van der Pol – systeemontwikkelaar

Wil de Ruyter – taalkundig assistent

Dr. Kathy Rys – onderzoeker/taalkundige

Dr. Tanneke Schoonheim – onderzoeker/taalkundige († 25 augustus 2020)

Prof. dr. Nicoline van der Sijs – onderzoeker/taalkundige (sinds 1 februari 2020)

Rob van Strien – programmeur/ontwikkelaar

Paulette Tacx – managementassistent

Drs. Rob Tempelaars – onderzoeker/taalkundige
Dr. Carole Tiberius – computerlinguïst
Lic. Veronique De Tier – onderzoeker/taalkundige (sinds 1 juni 2020)
Lic. Katrien Van pellicom – projectleider Spelling
Dr. Vincent Vandeghinste – onderzoeker/taalkundige
Drs. Boukje Verheij – onderzoeker/taalkundige
Rayvano van Vliet – systeembeheerder
Machteld de Vos, MPhil – onderzoeker in opleiding
Drs. Vivien Waszink – onderzoeker/taalkundige
Karin van Weerlee – managementassistent
Laetitia de Winter – administratief medewerker

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

Publicaties

- Baisa Vit, Carole Tiberius, Elisabetta Ježek, Lut Colman, Costanza Marini & Emma Romani (2020). Skema: A New Tool for Corpus-driven Lexicography. In: *Proceedings of the XIX EURALEX International Congress: Lexicography for Inclusion*.
- Beelen, Hans en Nicoline van der Sijs (2020). Complotdenken. In: *Onze Taal* 12, 30.
- Beelen, Hans en Nicoline van der Sijs (2020). Digitale tekstuitgave van William Sewel, *Nederduytsche Spraakkonst, Waarin de Gronden der Hollandsche Taale naauwkeuriglyk opgedólven, en zelfs voor geringe Verstanden, zo ten aanzien der Spellinge als bewoordinge, duydelyk aangewezen zyn*, Amsterdam: Assuerus Lansvelt/ Erfg. van J. Lescaillje, 1708.
- Beelen, Hans en Nicoline van der Sijs (2020). Digitale tekstuitgave van J.G.A. Kirchhoff en J.A.W. Schröder, *Holländisch-deutsches Handwörterbuch besonders für Unstudirte, Oldenburg: Schulzesche Buchhandlung*, 1810.
- Beelen, Hans en Nicoline van der Sijs (2020). Digitale tekstuitgave van J.G.A. Kirchhoff en J.A.W. Schröder, *Holländisch-deutsches Handwörterbuch besonders für Unstudirte, Oldenburg: Schulzesche Buchhandlung*, 1810.
- Beelen, Hans en Nicoline van der Sijs (2020). Dinosauriërs. In: *Onze Taal* 6, 24. (2020).
- Beelen, Hans en Nicoline van der Sijs (2020). Feestvarken. In: *Onze Taal* 1, 30.
- Beelen, Hans en Nicoline van der Sijs (2020). Huidhonger. In: *Onze Taal* 7/8, 30.
- Beelen, Hans en Nicoline van der Sijs (2020). Kinine. In: *Onze Taal* 4, 30.
- Beelen, Hans en Nicoline van der Sijs (2020). Kwakzalvers. In: *Onze Taal* 5, 30.
- Beelen, Hans en Nicoline van der Sijs (2020). Mondkapje. In: *Onze Taal* 10, 30.
- Beelen, Hans en Nicoline van der Sijs (2020). Sprinkhaan. In: *Onze Taal* 9, 30.
- Beelen, Hans en Nicoline van der Sijs (2020). Stikstof. In: *Onze Taal* 2/3, 30.
- Beelen, Hans en Nicoline van der Sijs (2020). Taser. In: *Onze Taal* 11, 30.
- Beelen, Hans en Nicoline van der Sijs (2020). Wie is toch die berisper van een Nederlandse spelling? In: *Neerlandistiek* 26/10/2020: <https://www.neerlandistiek.nl/2020/10/wie-is-toch-die-berisper-van-een-nederlandse-spelling/#more-123118>.
- Bonth, Roland de (2020). ‘Sterrix & Kruisix’. Op: www.neerlandistiek.nl
- Bonth, Roland de (2020). ‘O schrijverke, schrijverke’. Op: www.neerlandistiek.nl
- Bonth, Roland de (2020). ‘Multatuli-triviaquiz’. Op: www.neerlandistiek.nl
- Bonth, Roland de (2020). “‘De daad bij de term voegen’”. Op: www.neerlandistiek.nl
- Bonth, Roland de (2020). ‘Uitslag Multatuli-triviaquiz’. Op: www.neerlandistiek.nl

Bonth, Roland de (2020). 'De vlag dekt de lading niet'. Op: www.neerlandistiek.nl

Bonth, Roland de (2020). 'De naam van de vaders. Een paasquiz'. Op: www.neerlandistiek.nl

Bonth, Roland de (2020). 'Vitamine N(amen)'. Op: www.neerlandistiek.nl

Bonth, Roland de (2020). 'Uitslag van De naam van de vaders. Een paasquiz'. Op:
www.neerlandistiek.nl

Bonth, Roland de (2020). 'Laat ze maar lallen'. Op: www.neerlandistiek.nl

Bonth, Roland de (2020). 'Wa zedde nou?'. Op: www.neerlandistiek.nl

Bonth, Roland de (2020). 'Vermakelijk bio- en bibliografisch toetsenbord: zoeken op internet'. Op:
www.neerlandistiek.nl

Bonth, Roland de (2020). 'Matthias de Vriesquiz'. Op: www.neerlandistiek.nl

Bonth, Roland de (2020). 'Was Balthazar Huydecoper een letterkundig initiator? Een promotie in 1962'. Op: www.neerlandistiek.nl

Bonth, Roland de (2020). 'Uitslag Matthias de Vriesquiz'. Op: www.neerlandistiek.nl

Bonth, Roland de (2020). 'Gratis drie meter woordenboek voor elke leerling'. Op:
www.neerlandistiek.nl

Bonth, Roland de (2020). 'Letterzetter? Puzzelen met schrijvers'. Op www.neerlandistiek.nl

Bonth, Roland de (2020). 'Letterzetter? Puzzelen met schrijvers – uitslag & antwoorden'. Op:
www.neerlandistiek.nl

Bonth, Roland de (2020). 'Groot Dicté der Nederlandsche Taal'. Op: www.neerlandistiek.nl

Bonth, Roland de (2020). 'Een noot bij Voskuils Afgang (Het Bureau, deel 6)'. Op:
www.neerlandistiek.nl

Bonth, Roland de (2020). 'Word woordenschatgraver!'. Op: www.neerlandistiek.nl

Bonth, Roland de (2020). '(Nog even) gratis toegang to wetenschappelijke publicaties voor leraren'.
Op: www.neerlandistiek.nl

Bonth, Roland de (2020). 'Wè neukt dè! De Database van de Zuidelijk-Nederlandse Dialecten (DSDD)'. Op: www.neerlandistiek.nl

Bonth, Roland de (2020). 'Waarde en dierbare lezer,'. Op: www.neerlandistiek.nl

Bonth, Roland de (2020). 'Willem Elsschotquiz'. Op: www.neerlandistiek.nl

Bonth, Roland de (2020). 'Uitslag en antwoorden van de Willem Elsschotkwis'. Op:
www.neerlandistiek.nl

Bonth, Roland de (2020). 'Gerard Outhof. "Een schrander letterkenner"'. Op: Schrijverskabinet.nl

Bonth, Roland de (2020). 'Hendrik Graauwhart. Een voorbeeldelijk en godvruchtig dichter'. Op:
Schrijverskabinet.nl

Bonth, Roland de (2020). 'Simon Doekes. Wijnkoper en zangbergstichter'. Op: Schrijverskabinet.nl

Bonth, Roland de (2020). 'Hendrik van Bracht. Een Dordtenaar in het Dichterkunstpaleis'. Op:
Schrijverskabinet.nl

- Bonth, Roland de, Dirk Geirnaert & Rob Tempelaars (2020), *Matthias de Vries aan het woord. Uitgave ter gelegenheid van het Matthias de Vriesjaar 2020*. Instituut voor de Nederlandse Taal/Scriptum, Schiedam.
- Bonth, Roland de (2020), 'Vondel voor het voortgezet onderwijs'. In: *Voltreffer! Populariteit en popularisering van de historische Nederlandse letterkunde. Bijdragen ter gelegenheid van het emeritaat van Lia van Gemert* (Amsterdam, 2020), 74-78.
- Bonth, Roland de i.s.m. Quest, 'Weet jij hoe je moest spellen in 1870?'. Op: tests.quest.nl
- Bouma, Gerlof, Evie Coussé, Trude Dijkstra, Nicoline van der Sijs (2020). The EDGeS Diachronic Bible Corpus. In: *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*, 5232–5239: <http://www.lrec-conf.org/proceedings/lrec2020/pdf/2020.lrec-1.644.pdf>.
- Chiarcos, Christian, Max Ionov, Jesse de Does, Katrien Depuydt, Anas Fahad Khan, Sander Stolk, Thierry Declerck, en John Philip McCrae, "Modelling Frequency and Attestations for OntoLex-Lemon." In: *Proceedings of the 2020 Globalex Workshop on Linked Lexicography*. Marseille: European Language Resources Association, 2020, p. 1-9.
- Creten, Silke, Peter Dekker, Vincent Vandeghinste (2020). Linguistic Enrichment of Historical Dutch using Deep Learning. In: *Computational Linguistics in the Netherlands Journal*, vol. 10. p. 57-72. Op: <https://clinjournal.org/clin/article/view/104>
- Depuydt, Katrien, Jesse de Does (2020). "DiaMaNT, een diachroon semantisch lexicon van de Nederlandse taal." In: *Dixit. Tijdschrift over taal- en spraaktechnologie* jg. 17 (2), p. 27-30.
- De Tier, Veronique (2020). Het Vlaams, Brabants en Limburgs in kaart gebracht. Op: vrttaal.net
- De Tier, Veronique (2020). Bachten de Kupe op de dialectkaart. In: *Diejalektgazette van Bachtin de Kuupe*, jg 15, nr.4, 9-10
- De Tier, Veronique (2020). De Zeeuws-Vlaamse dialecten staan op de kaart. In: *Erfgoed Zeeland*, nr. 4, december, 18-20.
- De Tier, Veronique (2020). Rubriek 'Uit de streek' op ivdnt.org.
- Doreleijers, Kristel en Nicoline van der Sijs (2020). Studenten Nederlands in het buitenland: taalambassadeurs in de wereld. In: *Vaktaal* 2/3, 34-35.
- Doreleijers, Kristel, Mathilde Jansen en Nicoline van der Sijs (2020). Taal onderzoek je zó. In: *Levende Talen Magazine* 4: 38-39.
- Eerten, Laura van, Vivien Waszink (2020). Bijdragen *Taalkalender* over neologismen, Onze Taal.
- Eerten, Laura van, Caspar Stalenhoef (2020). Waar komt pindakaas vandaan?, 12-delige podcastserie.
- Fannee, Mathieu (2020). "[Het Hoge Pad, een middeleeuws pad herontdekt](#)". In: *De Hekkenluiter* 34, Historisch Genootschap Warmelda, 12-21.
- Geirnaert, Dirk (2020). Diverse bijdragen in de webrubriek [Terug in de taal](#) op de website van het INT.

- Geirnaert, Dirk, Roland de Bonth en Rob Tempelaars (2020). *Matthias de Vries aan het woord. Uitgave ter gelegenheid van het Matthias de Vriesjaar 2020*. Instituut voor de Nederlandse Taal/Scriptum, Schiedam.
- Geirnaert Dirk (2020). A Gift for Hanno: the Fictitious Booklist of Eduard de Dene. In: Pouey-Mounou, Anne-Pascale and Smith, Paul J. (eds.), *Early Modern Catalogues of Imaginary Books. A Scholarly Anthology*. Leiden-Boston: Brill Leiden, 2020, 112-134.
- Geirnaert, Dirk, Paul J. Smith (2020): Wieringa – Doedijns – Anna Folie – Van Lennep: Dutch Versions of Rabelais’s Library of Saint-Victor. In: Pouey-Mounou, Anne-Pascale and Smith, Paul J. (eds.), *Early Modern Catalogues of Imaginary Books. A Scholarly Anthology*. Leiden-Boston: Brill Leiden, 2020, 320-348.
- Id-Youss, L., Alsulaiman, A., Steurs, F. (2020). New Forms of Translation: The Need for New Professional Profiles. In: S. Shiyab (Eds.), *Translators and Interpreters’ Training: Arab and Middle Eastern Perspectives*. Amsterdam / Philadelphia: John Benjamins Publishing Company.
- Joubert, Sidney en Nicoline van der Sijs (2020). [Antilliaans-Nederlandse woorden en hun herkomst](#). In: *Trefwoord*, november 2020.
- Meulen, Marten van der & Nicoline van der Sijs (2020). Aan het Nederlands valt best wat te knutselen. In: Jubileumnummer van *Nederlandse Taalkunde / Dutch Linguistics*, 387-396.
- Sijs, Nicoline van der (2020). “[Een schaap, dat geen wol had, zag paarden](#)”. Een aanstekelijk boekje over historische taalkunde (bespreking van Yannick Fritschy. De stam van het woord. Over taalevolutie en de eerste taal ter wereld). In: *Neerlandistiek*.
- Sijs, Nicoline van der (2020). “Spiegheling” en “daet” van Simon Stevin. In: *Onze Taal* 10, 28.
- Sijs, Nicoline van der (2020). 180.000 antwoorden op 15.000 vragen. In: *E-data & Research*, februari, p. 6.
- Sijs, Nicoline van der (2020). [Appelflappen and Clog Wogs. On Flemish and Dutch Emigrants in the Anglophone World](#). In: *The Low Countries*.
- Sijs, Nicoline van der (2020). Digitale tekstuitgave van 16 satirische kranten: Haegsche Middel-Weecksche Courant, july 1661; Haeghsche Mercurius, of Verscheide Post-tydingen, july 1661. Extraordinaris Parysche Courant, augusti 1661; Nieuwe Haegse Middel-Weeckse Courant, 1661; Extraordinaire Amsterdamsche Maendaegsche Courant, 1678; De courant tot gemeene best, 1684; Extraordinaire courant in t troubeljaar, 1684; Extraordinaire Courant in twistmaand, 1684; Extraordinaris courant in twyffel-maent, 1684; Hora novissima, 1684; Korte aanmerkingen over de extraordinaire courant in twistmaand, 1684; Mercurius Gallo-Belgicus, 1684; De Darmstadse Thee Courant, 1687; De Oost-Indische Thee-Post, 1687; Amsterdamse post-tijding, 1689; Nieuwe Ypersz Maendaegse postyding, 1691.

- Sijs, Noline van der (2020). Digitale tekstitgave van 374 exemplaren van de *Nieuwe Tijdinghen* van Abraham Verhoeven, gepubliceerd tussen 1605 en 1628 en als scan beschikbaar bij de Erfgoedbibliotheek Hendrik Conscience.
- Sijs, Noline van der (2020). En chaussettes blanches et sandales. Langue et culture des émigrés néerlandais et flamands en francophonie. In: *Septentrion* 2 (26), 134-139.
- Sijs, Noline van der (2020). Englische Lehnwörter in der niederländischen Sprache: ein anschwellender Strom (Übersetzung von Roland Duhamel). In: *Englisch in europäischen Sprachen*, Hg. von Barbara Kaltz, Gerhard Meiser, Horst Haider Munske, FAU University Press, 125-166 (<https://doi.org/10.25593/978-3-96147-319-9>).
- Sijs, Noline van der (2020). In hoeverre houden geëmigreerde Nederlanders en Vlamingen in de eenentwintigste eeuw vast aan de Nederlandse taal en cultuur?. In: *Internationale Neerlandistiek* 58 (1): 5-21.
- Sijs, Noline van der (2020). Paol, son, loisig. De vele soorten Amsterdams in de negentiende eeuw. In: *Onze Taal* 6, 8-9.
- Sina Ahmadi; John P. McCrae; Sanni Nimb; Thomas Troelsgård; Sussi Olsen; Bolette S. Pedersen; Thierry Declercq; Tanja Wissik; Monica Monachini; Andrea Bellandi; Fahad Khan; Irene Pisani; Simon Krek; Veronika Lipp; Tamás Váradi; László Simon; András Gyórfy; Carole Tiberius; Tanneke Schoonheim; Yifat Ben Moshe; Maya Rudich; Raya Abu Ahmad; Dorielle Lonke; Kira Kovalenko; Margit Langemets; Jelena Kallas; Oksana Dereza; Theodorus Fransen; David Cillessen; David Lindemann; Mikel Alonso; Ana Salgado; José Luis Sancho; Rafael-J. Ureña-Ruiz; Kiril Simov; Petya Osenova; Zara Kancheva; Ivaylo Radev; Ranka Stanković; Cvetana Krstev; Biljana Lazić; Aleksandra Marković; Andrej Perdih; Dejan Gabrovšek (2020). A Multilingual Evaluation Dataset for Monolingual Word Sense Alignment. In *Proceedings of the 12th Language Resource and Evaluation Conference (LREC 2020)*, European Language Resources Association, Marseille, France, 3232-3242.
- Steurs, F. (2020). Van INL naar INT: een unieke schatkamer van alle taalmaterialen voor het Nederlands. In: *Leuvense Bijdragen*, 249-263. [doi: 10.2143/LB.102.0.3288696](https://doi.org/10.2143/LB.102.0.3288696) [Open Access](#)
- Steurs, F. (2020). Nederlands een grote taal? Een kwestie van meten. In: *Acta Universitatis Wratislaviensis. Neerlandica Wratislaviensia*.
- Steurs, F. (2020). CLARIN-CLARIAH-DARIAH: Hoe staat het met de digitale infrastructuur voor de Nederlandse taal? In: *Neerlandia : Algemeen-Nederlands Tijdschrift*.
- Steurs, F., van Sterkenburg, P. (2020). *Geen woord te veel: Geschiedenis van INL tot INT*. Scriptum, Schiedam. ISBN: 9789463190558.
- Steurs, F. (2020). Language is business. The challenges of a single digital market in a multilingual society. In: *Language and Economy. Language Industries in a multilingual Europe*, Chapt. 3,

- (39-57). (EFNIL). Budapest: Research Institute for Linguistics Hungarian Academy of Science. ISBN: 9789639074842.
- Steurs, F. (2020). Rekenen op taal. Hoeveel kost taal en wat is de impact op onze economie? In: W. Coudenys (Eds.), *Taal en de wereld*, Chapt. 1, (11-31). Universitaire Pers Leuven, Leuven. ISBN: 9789462702134.
- Steurs, F. (2020). Computer en taal. Worden vertalers op lange termijn overbodig? In: *Romaneske*, 69-76.
- Tempelaars, Rob, Vivien Waszink en Boukje Verheij, rubriek 'Neologisme van de week' / 'Nieuw woord van de week' op de website van het INT. <http://www.ivdnt.org/> Met ingang van 7 maart 2020 ook wekelijks gepubliceerd in het *Leidsch Dagblad*.
- Vonk, Frank, Jan Noordegraaf, Roland de Bonth & Gijsbert Rutten (2020). 'In memoriam: Frans Wilhelm (1945-2020). Op: www.neerlandistiek.nl
- Vonk, Frank, Jan Noordegraaf, Roland de Bonth & Gijsbert Rutten (2020). 'Frans Wilhelm (1945-2020)'. In: *Language & History*, 63:3, 237-239
- Vos, Machteld de (2020). De weerbarstige grammatica's van het 17e-eeuwse Standaardnederlands. Op: www.neerlandistiek.nl
- Vos, Machteld de (2020). 'Dat onse tale geenzins lijden en kan'. Op: [Filter. Tijdschrift over vertalen](http://www.filter.nl).
- Vos, Machteld de (2020). Van de Gezangen. Op: [Those Who Wander](http://www.thosewhowander.nl).
- Vos, Machteld de (2020). Neerlands trots. Op: [Those Who Wander](http://www.thosewhowander.nl).
- Waszink, Vivien. Dushi, bae en schattebout. In: *Onze Taal* 2/3, 2020.
- Waszink, Vivien. Hoestschaamte, #blijfinuwkot en coronahuffer. Nieuwe woorden in tijden van corona. Op: [ivdnt.org](http://www.ivdnt.org), 24 april 2020.
- Waszink, Vivien. Neologisms in an Online Portal: The Dutch Neologismenwoordenboek (NW). In: *Dictionaries*, Volume 41, Issue 1, 2020, 27-44.
- Waszink, Vivien. Directeur (m/v/x) gezocht. Over het nut van genderneutraal taalgebruik. In: *Tijdschrift voor Ontwikkeling in Organisaties (TvOO)*, nummer 3, september 2020, 100-104.

Lezingen en presentaties

- Bonth, Roland de. [Het Groot Dicté der Nederlandsche Taal](http://www.hetgrootdictie.nl). Koninklijke Academie voor Nederlandse Taal en Letteren. Gent, 16 september 2020.
- De Tier, Veronique, Katrien Depuydt en Jesse de Does. *De DSDD-database en applicatie*. Lancering Database van de Zuidelijk-Nederlandse dialecten. Gent, 30 september 2020.
- De Tier, Veronique, Roxane Vandenberghe en Lien Hellebaut. *De DSDD-database: demonstratie*. Lancering Database van de Zuidelijk-Nederlandse dialecten. Gent, 30 september 2020.

- Depuydt, Katrien, Jesse de Does. *Linking the dictionary of Old Dutch to the Thesaurus of Old English. A First Exploration*. Lezing in het kader van DiaMaNT op de Second Evoke Workshop in Leiden, 17 oktober 2020.
- Doreleijers, Kristel, Mathilde Jansen & Nicoline van der Sijs (2020). *Taal onderzoek je zo!* De eerste resultaten van de LOT-populariseringsprijs 2018, lezing op de Grote Taaldag, 31 januari.
- Sijs, Nicoline van der (2020). *De standaardisering van het Nederlands in de 16/17e eeuw*, als special guest bij college neerlandici aan de Universiteit van Gent, 7 december.
- Sijs, Nicoline van der (2020). Lancering van het e-WALD, elektronische Woordenboek van de Achterhoekse en Liemerse Dialecten, 11 december.
- Sijs, Nicoline van der (2020). *Work-in-progress*, Oudgermanistendag, 31 oktober.
- Sijs, Nicoline van der (2020). [*Stevin als taalkundige*](#), Symposium In de geest van Stevin van de werkgroep Geschiedenis van de Nederlandse Vereniging van Wiskundeleraren, 10 oktober.
- Sijs, Nicoline van der (2020). *Use case historische kranten. Voorbewerkingen & onderzoeksmogelijkheden*, Werkpakket 6 “Tekst” van CLARIAH-PLUS, 12 mei.
- Sijs, Nicoline van der (2020). *Casus DH. Hoeveel dialecten werden er in de 19e eeuw in Amsterdam gesproken?*, lezing op de Neerlandistiekdagen, 7 maart 2020.
- Steurs, F. (2020). *De schatkamer van het Nederlands*. Invited by Orde van den Prince, online.
- Steurs, F. (2020). *Simon Stevin in Leiden. Het Nederlands als wetenschapstaal*. Invited by NL-Term. Presented at the TiNT-dag, online, 23 Oct 2020-23 Oct 2020.
- Steurs, F. (2020). *The Dutch Language is coronaproof!* Invited by EFNIL. Presented at the Efnil annual conference, online, 30 Sep 2020-01 Oct 2020. (professional oriented)
- Steurs, F. (2020). *Legal translation: a terminological challenge*. Invited by TermNet. Presented at the Terminology Summer School, online, 01 Jul 2020-04 Jul 2020.
- Tempelaars, Rob. Toespraak ter nagedachtenis aan Tanneke Schoonheim (1965-2020), namens de collega's en oud-collega's van het Instituut voor Nederlandse Lexicologie (INL) en het Instituut voor de Nederlandse Taal (INT) een korte in de Hooglandse kerk te Leiden. 30 oktober 2020.
- Tiberius, Carole. *ELEXIS WPI Lexicographic data and workflow*. Presentatie op ELEXIS review meeting. juni 2020, online.
- Vos, Machteld de. *Laying the foundation: annotating linguistic prescriptivism in the formative period of Standard Dutch*. Presentatie voor Languages in Transition Stages (LiTS, RU onderzoeksgroep). Nijmegen, 28 augustus 2020.
- Vos, Machteld de. *Setting the norm: annotating 17th-century linguistic prescriptivism for the Dutch Normative Databases (NODE)*. Presentatie op de 16^e Sprachwissenschaftliche Tagung für Promotionsstudierende (STaPs). Wenen (online), 25 september 2020.

- Vos, Machteld de. *Setting the norm: norm variability in the formative period of Standard Dutch*. Presentatie in de workshop ‘Standardization and usage’, op de 20th International Conference of the Department of Linguistics (University of Bucharest). Bucharest (online), 20 november 2020.
- Vos, Machteld de. *‘Grondlyke ende alghemene reghelen int licht brenghen’: NODE & de zoektocht naar usage items*. Presentatie in de workshop Taalnormen en taalgebruik. Online, 18 december 2020.
- Waszink, Vivien. *Zieke zinnen zullen zegevieren, De taal van nederhop*. Lezing Neerlandistiekdagen, 7 maart 2020.

Onderwijs

- Bonth, Roland de, webinar ‘Woorden uit de verleden tijd’. Week van het Nederlands 2020, 7 oktober 2020.
- Heylen, Kris, Vincent Vandeghinste, Carole Tiberius. MA-college Computatieve Lexicografie. KU Leuven, februari-mei 2020.
- Kinable, Dirk. MA-college Concepts: systems, diagrams and definitions. KULeuven – campus Antwerpen, 19 oktober 2020.
- Odiijk, Jan, Jesse de Does, Katrien Depuydt, Kris Heylen, Jan Niestadt, Carole Tiberius, en Vincent Vandeghinste. Online tutorial over Open SoNaR. 9 oktober 2020.
- Steurs, F. (2020). Language Technology: what's next? Presented at the Abralin ao Vivo, online, 23 Jun 2020-30 Jun 2020. (URL)
- Steurs, F. (2020) Legal Translation: a terminological challenge. Presented at TermNet, online, 4 July 2020. Tiberius, Carole. MA-collegereeks Corpus Lexicography. Universiteit Leiden, februari-mei 2020.
- Tiberius, Carole College in MA-collegereeks Computatieve Lexicografie. KU Leuven, 30 maart 2020.
- Waszink, Vivien. Workshop jongerentaal. ROC Mondriaan Den Haag, 12 februari 2020.
- Waszink, Vivien. Webinar over neologismen KTV Kennisnet, 12 mei 2020.

In de media

- Bonth, Roland de. Item over het Matthias de Vriesjaar in [De Taalstaat](#) (NPO Radio 1), 25 januari 2020 (6'07''-7'38'')
- De Tier, Veronique. [Interview over de Database van de Zuidelijke-Nederlandse Dialecten](#). Radio 2 (België), 30 september 2020.

- Instituut voor de Nederlandse Taal. Wekelijkse rubriek ‘Neologisme van de week’, in *Leidsch Dagblad*, van 1 januari 2020 t/m 31 december 2020.
- Instituut voor de Nederlandse Taal. [INT roept 2020 uit tot Matthias de Vriesjaar](#). Op [vrtaal.net](#), 22 januari 2020.
- Instituut voor de Nederlandse Taal. 2020: het jaar van Matthias de Vries. In *Trouw*, 23 januari 2020.
- Instituut voor de Nederlandse Taal. [‘Lexicon Frisicum’ toegankelijk gemaakt voor de lezer van vandaag](#). Omrop Fryslân, 26 februari 2020.
- Instituut voor de Nederlandse Taal. [Utrechtse mysteries: waarom is het DE Neude en niet HET Neude?](#) Op [indebuurt.nl/utrecht](#), 5 april 2020.
- Instituut voor de Nederlandse Taal. [Is het nu eindelijk afgelopen met ‘papadag’?](#) Op [oneworld.nl](#), 1 mei 2020.
- Instituut voor de Nederlandse Taal. [Onderzoekers lanceren het grootste dialectenwoordenboek van de Nederlandse Taal](#) in *Het Belang van Limburg*, 25 september 2020.
- Instituut voor de Nederlandse Taal. [Hoe is het met jouw kennis van de dialecten gesteld? Ontdek het in deze quiz](#). VRT NWS, 30 september 2020.
- Instituut voor de Nederlandse Taal. [Nieuw digitaal woordenboek maakt je wegwijs in lokale woordenschat. Hier een vlinder, daar een pepel](#). *Het Nieuwsblad*, 30 september 2020.
- Instituut voor de Nederlandse Taal. [UGent en Instituut voor de Nederlandse Taal lanceren online dialectwoordenboek](#), *Het Laatste Nieuws*, 30 september 2020.
- Instituut voor de Nederlandse Taal. [Duurzame taal: van vaag tot vijandig](#) (*Het Groene Woordenboekje*). In *Reformatorisch Dagblad*, 20 oktober 2020.
- Sijs, Noline van der (2020). Gast bij De Nachtoen, EO, https://www.npostart.nl/de-nachtoen/22-12-2020/VPWON_1323333, 21/12/2020.
- Sijs, Noline van der (2020). SBS Dutch, radio Australië, [Taalverloedering bestaat niet](#), 17/10/2020.
- Sijs, Noline van der (2020). De Taalstaat, Taalboekenprijs voor “15 eeuwen Nederlandse taal” van Noline van der Sijs, 10/10/2020, <https://www.nporadio1.nl/de-taalstaat/onderwerpen/66129-2020-10-10-taalboekenprijs-voor-15-eeuwen-nederlandse-taal-van-noline-van-der-sijs>.
- Sijs, Noline van der (2020). Wat is het verschil tussen besluiten en beslissen?, Saturday Mike Fever, RN7, 19/9/2020, <https://soundcloud.com/rn7smf/wat-is-het-verschil-tussen-beslissen-en-besluiten>.
- Sijs, Noline van der (2020). Origineel Amsterdams: waar hoor je nog godskolere tegenwoordig?, 14/15/16/9-2020, <https://www.at5.nl/artikelen/204388/origineel-amsterdams-waar-hoor-je-nog-godskolere-tegenwoordig>.
- Steurs, Frieda. [Interview over het Matthias de Vriesjaar](#) in *De Taalstaat*, 8 augustus (11.00) 2020.

- Tempelaars, Rob. Laura Obdeijn, ‘Nederlanders schelden graag met ziektes. Eindigt *corona* ook als scheldwoord?’, in: *Het Parool*, 22 juni 2020. <https://www.parool.nl/nederland/nederlanders-schelden-graag-met-ziektes-eindigt-corona-ook-als-scheldwoord~b08a7e34/>
- Tempelaars, Rob. Over het gebruik en de betekenis van het woord *appartement*. Radio 538, 9 augustus 2020. <https://www.538.nl/gemist>
- Vos, Machteld de. Interview ‘Waarom verdween ‘aalmoes’ uit de Bijbel?’. [Faces of Science, NEMO Kennislink](#), 11 november 2020.
- Waszink, Vivien. Drie keer een interview over jongerentaal in de rubriek *Woord van de Week*, de *Volkscrant*.
- Waszink, Vivien. Diverse bijdragen voor een taalrubriek in *Het Klokhuis*, NTR televisie.
- Waszink, Vivien. Interview voor een artikel over hiphoptaal. Jongens van de straat, in: *Groene Amsterdammer*, 13 mei 2020.
- Waszink, Vivien. Interview over coronawoorden. Hoestdab en kuchsot: woorden volgen de realiteit. In: *Leidsch Dagblad*, 4 april 2020.
- Waszink, Vivien. Interview over nieuwe woorden in 2020. Alcoflexen in veganuary. In: *Leidsch Dagblad*, 23 december 2020.
- Waszink, Vivien. Interview over nieuwe coronawoorden. Taaljaar 2020: Quarantainekapsel, huidhonger, zorgapplaus en coronakilo's. Taalkundigen Vivien Waszink, Wim Daniëls en Ton den Boon over hun favoriete coronawoorden. In: *HP/De Tijd*, 28 november 2020.
- Waszink, Vivien. Interview over o.a. de betekenis van het woord *leuk*. De Nieuws BV, NPO Radio 1, 11 augustus 2020.
- Waszink, Vivien. 'Is het woord *moorkop* racistisch of niet?' In: *Algemeen Dagblad*, 8 februari 2020.
- Waszink, Vivien. Interview over coronawoorden bij Wout 2day, NPO Radio 2, 29 april 2020.
- Waszink, Vivien. Podcast *Woordwaarde* over coronawoorden BNR Nieuwsradio, 4 december 2020.
- Waszink, Vivien. Twee interviews over coronawoorden. *RTL 4 Editie.nl*, april 2020.

Congressen & workshops

- Creten, Silke, Peter Dekker, Vincent Vandeghinste. Computational Linguistics in the Netherlands. Utrecht. Mondelinge presentatie Linguistic Enrichment of Historical Dutch using Deep Learning.
- Eerten, Laura van. Medeorganisatie *Horen, zien en proeven van dialect*, lancering DSDD, 30 september 2020.
- Eerten, Laura van. Medeorganisatie *Groot Dicté der Nederlandsche Taal*, 16 september 2020.

Eerten, Laura van. Medeoraganisatie en online host *Op je woorden letten*, symposium ter ere van het Matthias de Vriesjaar, 5 november 2020.

Kinable, Dirk. Deelname aan het symposium *Verbinden & Innoveren met SNOMED* (13 februari, Zeist).

Kinable, Dirk. Deelname aan de bijeenkomsten van de Termraad Nederlands (6 maart Den Haag; 28 mei en 25 september, online).

Kinable, Dirk. Deelname aan de Nictiz-bijeenkomsten over patiëntvriendelijke COVID-19-terminologie (26 mei en 23 juni, online).

Kinable, Dirk. Deelname aan de bijeenkomst van de Termraad Academy (15 oktober, online).

Kinable, Dirk. Deelname aan het congres: TiNT-dag (23 oktober, online).

Kinable, Dirk. Deelname aan het symposium *Op je woorden letten* (5 november, online).

Steurs, F. (2020). Taal is Business! Over de impact van taalkundige toepassingen in onze samenleving. Presented at the Onderwijs- en Inspiratiedag AUHL: Nederlands in het hoger onderwijs, PXL Congress Hasselt, 23 Jan 2020-23 Jan 2020.

Tiberius, Carole. Deelname aan projectbijeenkomst: *COST CA18209 eerste plenaire bijeenkomst*. 27 en 28 januari 2020, Praag.

Tiberius, Carole. Deelname aan projectbijeenkomst: *ELEXIS Project Management Board* . 10 maart 2020, online.

Tiberius, Carole. Deelname aan projectbijeenkomst: *ELRC LRB meeting*. 29 april 2020, online.

Tiberius, Carole. Deelname aan projectbijeenkomst *ELEXIS Review Meeting*. 12 juni 2020, online.

Tiberius, Carole. Deelname aan workshop: *7th Workshop on Linked Data in Linguistics: Building tools and infrastructure*. 22 en 23 juni 2020, online.

Tiberius, Carole. Deelname aan projectbijeenkomst: *ELEXIS Project Management Board* . 30 september 2020, online.

Tiberius, Carole. Deelname aan workshop: *ELEXIS Digital Lexicographic Data Models*, 6 oktober 2020, online.

Tiberius, Carole. Deelname aan projectbijeenkomst: *COST CA18209 tweede plenaire bijeenkomst*. 26 en 27 oktober 2020, online.

Tiberius, Carole. Deelname aan workshop: *2nd Globalex Workshop on Lexicography and Neology (GWLN2)*, 4 november 2020, online.

Tiberius, Carole. Deelname aan training: *Training Time Management: GRIP*, 25 november 2020, online.

Tiberius, Carole. Deelname aan projectbijeenkomst: *ELRC LRB meeting*. 27 november 2020, online.

Vandeghinste, Vincent, Koen Mertens. Computational Linguistics in the Netherlands. Utrecht. Posterpresentatie GrETEL @ INT: Querying Very Large Treebanks by Example

Vandeghinste, Vincent. Neerlandistiekdagen 2020. Leiden. Tools en data voor het Nederlands.
Vos, Machteld de. Organisatie workshop Taalnormen en taalgebruik, in samenwerking met Eline Lismont van de Vrije Universiteit Brussel. 18 december 2020.

Prijzen

- Nederlandse Dataprijs voor 17e-eeuwse kranten van projectleider Nicoline van der Sijs.
- *15 eeuwen Nederlandse taal* door Nicoline van der Sijs krijgt de Onze Taal/ANV taalboekenprijs 2020.
- Het werk aan TEI Lex-0 is in 2020 beloond met de [2020 Rahtz Prize for TEI Ingenuity](#).

Diversen

Bonth, Roland de

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

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. Sedert 2020 zijn de huidige guidelines hier te vinden: <https://dariah-eric.github.io/lexicalresources/pages/TEILex0/TEILex0.html>.

Eerten, Laura van

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

Sijs, Nicoline van der

- Vaste medewerker van het tijdschrift *Onze Taal*
- Lid Stichting beheer database *Pinkhof Geneeskundig woordenboek*
- Redacteur *Internationale Neerlandistiek*
- Redacteur van *Neerlandistiek*
- Redacteur van het tijdschrift *Trefwoord*
- 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 Pool van deskundigen van de Digitale Bibliotheek voor de Nederlandse Letteren (DBNL)
- Voorzitter van het bestuur van de Kiliaanstichting ter bevordering van het etymologisch onderzoek in Nederland en België
- Promotor van Hercules-project Digital Dialects: an integrated lexicological infrastructure for the Southern Dutch Dialects Dictionary of the Southern Dutch Dialects (DSDD), woordvoerder Jacques van Keymeulen, Gent
- Lid advisory board ERC Consolidator project Mediate (Middlebrow Enlightenment: Disseminating Ideas, Authors, and Texts in Europe, 1665 - 1820) van Alicia Montoya
- Lid van de GLAD (Global Anglicism Database) Network: www.gladnetwork.org
- Lid van het bestuur van de ‘Stichting Jiddische Lexicografie Amsterdam’
- Curator namens de Universal Esperanto Association van de bijzondere leerstoel Esperanto en Interlinguïstiek bij de Universiteit van Amsterdam
- Member of the steering committee for the DHBenelux conference
- Voorzitter van de Stichting Vrijwilligersnetwerk Nederlandse Taal (SVNT)
- Lid van de Fryske Akademy
- Lid van de Commissie voor Taal- en Letterkunde van de Maatschappij der Nederlandse Letterkunde
- Lid van de Maatschappij der Nederlandse Letterkunde
- Redacteur van Trefwoord, tijdschrift op het gebied van lexicografie en lexicologie
- Begeleiding stagiairs: Hanna Ruiz (etymologie), Rein Brobbel (Leids), Cato Verschaik (leenwoorden), Djoeke Leguijt (Digitale Pallas), Melle Groen (Digitale Pallas).

Tiberius, Carole

- Board of Trustees Adam Kilgarriff Prize

Tempelaars, Rob

- Redacteur en redactiesecretaris van het tijdschrift *Trefwoord*
- Namens het INT betrokken bij een door diverse partijen op te zetten project over geuren, dat uiteindelijk ook tot een publicatie moet leiden

Van pellicom, Katrien

- Lid van de werkgroep 'Persoonsnamen' van de 'commissie anderstalige namen' van de Taalunie

Waszink, Vivien

- Redactielid *Neerlandia*

Bijlage 4: Downloadoverzicht taalmaterialen

Overzicht downloads commercieel

In totaal is er 41 keer een product afgenomen voor commerciële toepassingen.

CGN Commercieel	14
CGN Annotaties Commercieel	3
Gigant-Molex Commercieel	3
Lassy Klein Commercieel	3
Frequentielijsten Corpora Commercieel	2
Medische Termen Belgisch-Nederlands (MedTermBN) Commercieel	2
SONAR Groot Commercieel	2
SONAR Klein Commercieel	2
Autonomata Namencorpus Commercieel	1
Basilex Lexicon Commercieel	1
Basiscript Lexicon Commercieel	1
Combilex Commercieel	1
DPC Commercieel	1
Duelme Commercieel	1
Duoman Commercieel	1
E-lex Commercieel	1
Jasmin Commercieel	1
Names Commercieel	1
RBBN Commercieel	1
RBN Commercieel	1
SumNL Commercieel	1

Overzicht downloads niet-commercieel

Er zijn in totaal 1.412 producten gedownload voor niet-commercieel gebruik.

Corpus Gesproken Nederlands (CGN)	224
SoNaR-corpus	147
Frequentielijsten corpora	105
DuOMAn Subjectivity Lexicon	75
CGN-annotaties	73
Lassy Klein-corpus	62
Medische Termen Belgisch-Nederlands (MedTermBN)	54
Cd-rom Middelnederlands	47
Dutch Parallel Corpus (DPC)	46
SoNaR Nieuwe Media Corpus	45
Referentiebestand Nederlands (RBN)	45
Dutch Idiom Database: Native Speakers (DID-NS)	32
e-Lex	32
CHN N-grams	31
GiGaNT-Molex	29
CombiLex	25
Corpus Gysseling	24
Wablieft-corpus	21
PAROLE-lexicon	19
Eindhoven-corpus	18
INT Historische Woordenlijst	17
Corpus Middelnederlands	16
Moroccorp	16
Boarnsterhim Corpus (BHC)	15
Referentiebestand Belgisch-Nederlands (RBBN)	13

Brieven als Buit - Gouden Standaard	12
DuELME	11
AUTONOMATA-namencorpus	11
SumNL-samenvattingencorpus	11
Corpus Pathologische en Normale Spraak (COPAS)	10
RND Woordenlijsten	8
COREA-coreferentiecorpus	8
BLISS Spoken Dialogue Dataset	7
Afrikaans Custom Dictionary for Government Domain	6
Basilex Corpus	6
Basilex Lexicon	6
NAMES Corpus	6
IFA-corpus	6
INT IMPACT NE-lexicon	6
Basiscript Corpus	4
Jasmin	4
OMBI Arabisch-Nederlands	4
IFA Dialoog-videocorpus	4
RBN-klein	4
Lwazi Afrikaans Pronunciation Dictionary	3
Lwazi English Pronunciation Dictionary	3
Dupira	3
D-TUNA-corpus	3
OMBI Nederlands-Deens	3
OMBI Nederlands-Indonesisch	3
Lwazi Sesotho Pronunciation Dictionary	2
OMBI Nederlands-Arabisch	2
Setswana Genre Classification Corpus	2

BasiScript Lexicon	1
Paco-MT Parallele Corpora	1
Tshivenda Custom Dictionary for Government Domain	1
Xitsonga Custom Dictionary for Government Domain	1
Xitsonga Genre Classification Corpus	1
Siswati Custom Dictionary for Government Domain	1
Lwazi isiNdebele Pronunciation Dictionary	1
Afrikaans Genre Classification Corpus	1
Autshumato Afrikaans-English Translation Memory	1
Autshumato English-isiZulu Parallel Corpora	1
Autshumato English-Sesotho sa Leboa Parallel Corpora	1
DAESO-corpus: parallele Nederlandstalige monolinguale treebank	1
isiNdebele Custom Dictionary for Government Domain	1
isiXhosa Genre Classification Corpus	1
Lwazi Afrikaans ASR Corpus	1
Lwazi Xitsonga Pronunciation Dictionary	1
Lwazi isiXhosa Pronunciation Dictionary	1
Lwazi isiZulu ASR Corpus	1
Lwazi isiZulu Pronunciation Dictionary	1
Lwazi Sepedi Pronunciation Dictionary	1
Lwazi Setswana Pronunciation Dictionary	1
Lwazi Siswati Pronunciation Dictionary	1
Lwazi Tshivenda Pronunciation Dictionary	1

Bijlage 5: Rapport Externe Communicatie

In dit rapport worden de in 2020 ingezette communicatiemiddelen en de belangrijkste cijfers gepresenteerd.

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. Voor onderstaande analyse is gebruikgemaakt van Google Analytics (periode 1 januari 2020 - 1 januari 2021).

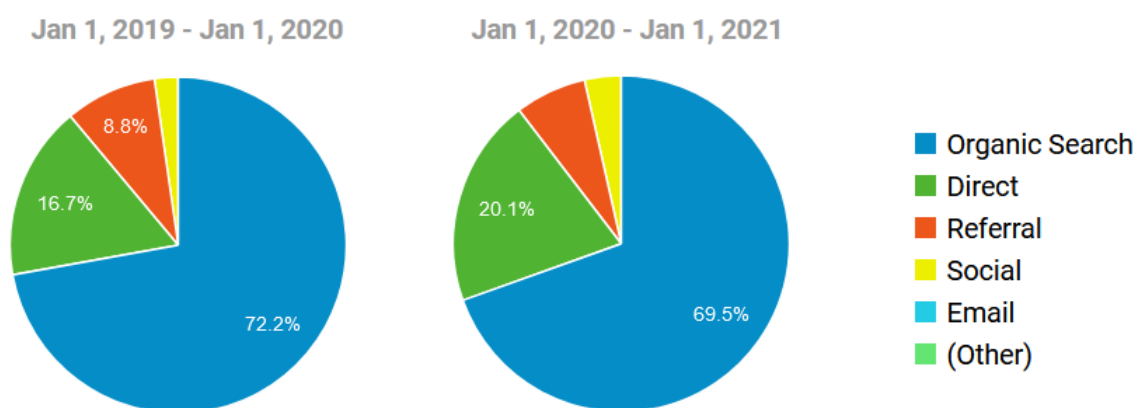
Cijfers

Algemeen	2019	2020
Totaal aantal paginaweergaven	576.575	646.216
Bezoeken uit Nederland	203.701	228.285
Bezoeken uit België	54.045	73.366
Gemiddelde tijd op pagina	00.01.57	00.01.51
Nieuwe bezoekers	76%	75%
Terugkerende bezoekers	24%	25%

Paginaweergaven rubrieken	2019	2020
Woordbaak	124.447	124.760
Neologismen (neo v/d week + neo v/d week archief + neologismen algemeen)	77.837	87.357
Historische woordenboeken (algemeen + ONW, VMNW, MNW, WNT, WFT)	51.212	52.964
Index/home	36.542	42.615
Nieuws	13.792	17.573
<i>Algemeen Nederlands Woordenboek</i>	10.855	10.366
Taalmaterialen	7.143	9.356
Gelegenheidswoordenboekjes	11.485	8.322
Terug in de taal	5.886	8.223
Terminologie	2.678	6.762
Spelling & grammatica	3.668	4.680
Podcast (nieuw in 2020)	-	3.775
Themapagina Historisch Nederlands	1.110	3.635

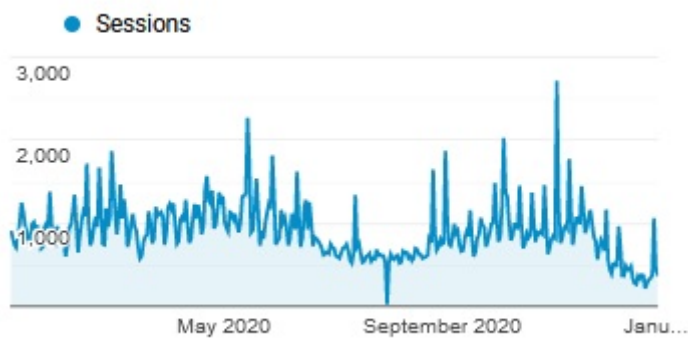
Etymologie	5.240	3.625
Jaaroverzicht in neologismen	3.124	3.587
Contact/medewerkers	1.347	3.297
Matthias de Vriesjaar 2020 (nieuw in 2020)	-	2.006
Themapagina Hedendaags Nederlands	1.197	1.686
CLARIN	1.587	1.601
Themapagina Terminologie	574	607
Uitgelichte term	169	475

Bronnen/verwijzingen



Het grootste deel van de bezoekers (70%) komt op de website terecht via een zoekopdracht in Google. 20% komt direct via de URL op de website, iets meer dan het voorgaande jaar. 7% van de bezoekers komt via een link op een andere website (referral) en 3,5% via een link op social media. Van de doorverwijzingen komt het merendeel (44%) op ivdnt.org terecht via het *Algemeen Nederlands Woordenboek* en de *Historische Woordenboeken* (gtb.ivdnt.org). Een klein deel komt via Wikipedia, (6%) nrc.nl (2,2%) en neerlandistiek.nl (1,7%). Van de doorverwijzingen via social media komen de meeste bezoekers via Facebook (51%) en Twitter (44%). Een veel kleiner percentage komt via LinkedIn (2,3%) en YouTube (0,4%).

Verloop bezoekersaantallen



Het beeld van het verloop van de bezoekersaantallen door het jaar heen is vergelijkbaar met voorgaande jaren: relatief stabiel met hier en daar wat uitschieters en een dip in de zomermaanden. De pieken zijn gerelateerd aan actuele neologismen (14 mei: *kloofporno*, 3 september *kipmobiel*, 5 november *hagelslagbeleid*) en opvallende gebeurtenissen (27 augustus: In memoriam Tanneke Schoonheim, 6 oktober: lancering podcast).

De website heeft in 2020 opnieuw flink veel meer bezoekers getrokken dan het voorgaande jaar. De naamsbekendheid van het INT groeit en de activiteiten nemen toe, en dat is goed te zien aan de bezoekersaantallen op de website. Opvallend is dat het aantal bezoeken uit België fors is gestegen. Het gaat om een toename van zo'n 36% ten opzichte van 2019.

De populairwetenschappelijke rubriek *Woordbaak* (etymologie) wordt sinds twee jaar niet meer actief bijgehouden en aangevuld, maar het blijft de meest bezochte rubriek op de website. Vooral de artikelen over *doni en barki* en *pappa en mamma* zijn in 2020 vaak geraadpleegd. Na *Woordbaak* volgt *Neologisme van de week* (nieuwe woorden), eind 2020 hernoemd tot *Nieuw woord van de week*. *Terug in de taal* (historische woorden) is vaker bezocht dan in 2019. De rubriek *Uitgelichte term* is het minst vaak bekeken. Een oorzaak kan zijn dat de rubriek maar heel sporadisch verschijnt, namelijk 4x per jaar. In de terminologienieuwsbrief is die rubriek in zijn geheel te lezen, doorklikken naar de website wordt dus niet gestimuleerd. Om meer aandacht voor *Uitgelichte term* te genereren kan het helpen om er bijvoorbeeld een maandelijkse rubriek van te maken, en een link in de nieuwsbrief te plaatsen in plaats van het hele artikel.

In 2021 worden de webrubrieken uitgebreid met twee wekelijkse columns van Ludo Permentier en Ewoud Sanders, en *Uit de streek*: een rubriek over dialectwoorden (afgewisseld met *Terug in de taal*). Er is opnieuw een verdubbeling in het aantal bezoeken van webpagina's over terminologie. De informatie over terminologie is dan ook sterk uitgebreid, het onderwerp wordt steeds meer geassocieerd met het INT en is met ingang van de themapagina's ook direct vindbaar op de voorpagina.

In 2020 zijn er geen nieuwe (online) gelegenheidswoordenboekjes verschenen. De rubriek is daarom minder onder de aandacht gebracht en ook minder vaak bezocht dan een jaar eerder. Van de eerder gepubliceerde woordenboekjes is *Het Groene Woordenboekje* (2.787 views) het vaakst bekeken. Dat is te verklaren doordat deze lijst eind 2020 is geüpdatet en aangevuld.

Van alle themapagina's is de themapagina Historisch Nederlands het best bezocht. Dit overzicht is ook het meest uitgebreid. De themapagina van Hedendaags Nederlands wordt in 2021 nog verder uitgebreid, vergelijkbaar met Historisch Nederlands.

De meeste zoektermen die op de website gebruikt worden hebben te maken met taalmaterialen, zoals *sonar*, *Afrikaans*, *cgn*, *diamant* en *basilex*. In de top tien zoekwoorden komen ook *podcast* en *frequentie* voor.

Vernieuwing website

De website is in 2020 volledig vernieuwd. Het Joomla!-CMS is vervangen door WordPress en de gehele site is opnieuw opgebouwd, ingericht en herschreven. De indeling van de website is overzichtelijker en compacter en het (hamburger)menu is duidelijker en platter van structuur. Ook is de website nu responsive, dat wil zeggen: geschikt voor gebruik op mobiel en tablet (website past zich aan het formaat van het apparaat aan). Om het instituut meer 'gezicht' te geven zijn er nieuwe foto's van alle medewerkers gemaakt, en iedereen heeft nu een eigen profielpagina. Veel teksten zijn herschreven en weergegeven in een vast format, er is een verklarende lijst met taalkundige begrippen toegevoegd en de toegankelijkheid voor mensen met een visuele beperking is verbeterd. Ten slotte is er een overzicht van alle webpagina's die vertaald zijn in het Engels.

Om de vernieuwde website te onderhouden en het plaatsen van nieuwe berichten en teksten beter te stroomlijnen is er een webredactie gevormd (Laura van Eerten (eindredactie), Roland de Bonth, Vivien Waszink, Boukje Verheij, Rob van Strien (webmaster)). De component Taalmaterialen is een aparte, zelfstandige website geworden om de gebruikers ervan beter te kunnen bedienen.

2. Sociale Media

Het instituut gebruikt met name Twitter als socialemediakanaal voor het delen van nieuws, evenementen en kennis over taal. Via Twitter worden links naar de website en taalweetjes/taalnieuws gedeeld, en contacten met volgers en partnerorganisaties onderhouden. Sinds 2019 is er een Facebookpagina waar net als op Twitter de webrubrieken en het nieuws geplaatst worden. Het INT heeft daarnaast nog een pagina op LinkedIn, daar is het minste activiteit. Ook is er sinds 2020 een YouTube-kanaal waar een aantal video's en opnames van (online)evenementen te bekijken zijn.

Cijfers

	Januari 2020	Januari 2021
Volgers op Twitter	4.492	4.966
Volgers op Facebook (ivdnt)	255	563

Twitter blijft al jaren groeien, met dit jaar een toename van 474 volgers. Het aantal volgers van de Facebookpagina is dan wel verdubbeld, maar nog steeds niet heel groot. Het is goed dat we op dit kanaal te vinden zijn maar er is meer uit te halen. Door bijvoorbeeld posts te plaatsen, beter gericht op het Facebookpubliek en meer interactie aan te gaan. Dat blijft vooralsnog liggen. 2021 biedt perspectief: met de samenstelling van de webredactie en het vergemakkelijken van doorposten van websiteberichten op social media via WordPress kan de activiteit op o.a. Facebook en LinkedIn aanzienlijk worden vergroot. Beheer van de socialmediakanalen wordt over de teamleden van de redactie verspreid. Dat is nodig om alle kanalen zo goed mogelijk te kunnen blijven onderhouden, want ze zullen voorlopig alleen maar uitgebreid worden. Begin 2021 is er alweer een nieuw kanaal bij: Instagram.

3. Nieuwsbrieven, mailings & persberichten

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 2020 zijn er zes algemene nieuwsbrieven verschenen en vier nieuwsbrieven over terminologie. Daarnaast zijn er nog diverse aparte mailings verzonden. Voor het versturen van nieuwsbrieven, mailings (ook intern) en persberichten maken we gebruik van de software MailChimp.

Cijfers

Nieuwsbrief algemeen	2019	2020
Aantal inschrijvingen	3.813	4.175
Gemiddeld aantal opens	48%	45%
Gemiddeld aantal kliks (CTO)*	27%	31%
Nieuwsbrief terminologie		
Aantal inschrijvingen	2.676	3.044
Gemiddeld aantal opens	28%	27%
Gemiddeld aantal kliks (CTO)	14%	11%

* CTO = click-to-open ratio, een percentage dat wordt bepaald door het aantal ontvangers dat de mail opent en doorklikt (clicks e-mail gedeeld door geopende e-mails)

- De februari-editie van de algemene nieuwsbrief is het vaakst geopend (51%)
- Best gelezen artikelen: *In memoriam Tanneke Schoonheim* (419 kliks), *Woordbaak* over de herkomst van *bakkebaarden* (317 kliks), *Terug in de taal over lepelzucht* (317 kliks), gelegenheidspublicatie *Woordenschat van de coronacrisis* (308 kliks), startpagina *Matthias de Vriesjaar* (203 kliks)
- Meest geopende nieuwsbrief terminologie is de maarteditie (28%)
- Links waar het vaakst op geklikt is: *Coronawoordenboek* van Ton de Boon (188 kliks), *Snomed-symposium* (29 kliks) en vaktechnisch glossarium *gastropedia* (28 kliks)

Mailings & persberichten	opens
22-1 Persbericht: 2020 is het Matthias de Vriesjaar	28%
28-5 Wie was Matthias de Vries? Doe de quiz!	39%
29-7 Uitnodiging 9 augustus Groenesteeg	69%
10-9 Groot Dicté der Nederlandsche Taal	36%
23-9 Week van het Nederlands 2020	37%
5-10 Persbericht: Nieuwe taalpodcast ‘Waar komt pindakaas vandaan?’	27%
7-10 Verhuizing taalmaterialen	51%
28-10 Uitnodiging symposium 5 november: Op je woorden letten	60%
18-11 Persbericht: Nederlandse Dataprijs gaat naar 17e-eeuwse krantenset	25%
10-12 Nieuwjaarswens 2021	54%

De cijfers van de mailings en persberichten zijn lastig met elkaar te vergelijken omdat ze naar verschillende (doel)groepen binnen het adressenbestand worden verstuurd.

Het abonneebestand van de twee nieuwsbrieven is in 2020 gegroeid: de algemene nieuwsbrief kreeg er 362 abonnees bij en de terminologienieuwsbrief 368. De cijfers voor het aantal opens en kliks voor de algemene nieuwsbrief zijn al jaren relatief hoog in vergelijking met nieuwsbrieven van andere organisaties (gemiddeld 38% opens en 16,3% CTO volgens de Nationale E-mail Benchmark van 2018). Dat laat zien dat lezers geïnteresseerd zijn, zich betrokken voelen en optimaal kennis nemen van de inhoud. De nieuwsbrief terminologie zit 10% onder het gemiddelde percentage opens van de Benchmark. Daarnaast wordt er in de terminologienieuwsbrief minder vaak op links geklikt dan in de algemene nieuwsbrief. Dat verschil zit waarschijnlijk in de opzet: in de algemene nieuwsbrief staan vaak korte berichten gevolgd door een ‘lees meer’ die naar de website verwijst. In de terminologienieuwsbrief staan meestal volledige berichten, met links naar extra informatie. In 2020 is er bijzonder weinig doorgeklikt, maar 11%. Misschien loont het om nog eens kritisch naar de opzet van de terminologienieuwsbrief te kijken, en het meer aan te passen aan de opzet van de algemene nieuwsbrief.

Aan de cijfers van de verstuurde mailings naast de reguliere nieuwsbrieven is te zien dat gerichte mailings over een specifiek onderwerp (uitnodigingen, nieuwjaarswens) naar een geselecteerde groep uit het adressenbestand gemiddeld gezien het vaakst geopend worden. Het percentage opens van persberichten is het laagst.

4. Populairwetenschappelijke uitgaven en activiteiten

Matthias de Vriesjaar

Op 9 november 2020 was het precies tweehonderd jaar geleden dat de Nederlandse taal- en letterkundige Matthias de Vries werd geboren. Daarom vierde het INT in 2020 het Matthias de Vriesjaar en organiseerde het hele jaar door verschillende activiteiten:

- Stadswandeling op 7 maart tijdens de eerste Neerlandistiekdagen, geïnitieerd door de Raad voor de Neerlandistiek (Het INT was ook inhoudelijk betrokken bij de invulling van het programma van de Neerlandistiekdagen)
- Matthias de Vriesquiz op neerlandistiek.nl
- onlinelezing over het werk en leven van De Vries
- kranslegging op begraafplaats Groenesteeg
- Groot Dicté der Nederlandsche Taal in de spelling-De Vries en Te Winkel
- Symposium ‘Op je woorden letten’ (5 november)
- Boek *Geen woord te veel* van Piet van Sterkenburg en Frieda Steurs
- Bakkebaardenverkiezing
- Nieuwjaarsgeschenk *Matthias de Vries aan het woord*

Als gevolg van de uitbraak van het coronavirus zijn verschillende evenementen uitgesteld of veranderd van opzet, maar uiteindelijk kon alles in aangepaste vorm doorgaan. Het bood ook nieuwe mogelijkheden, en we ontwikkelden nieuwe vormen van communicatie door te werken met video, livestreams, webinars en evenementen deels op locatie en deels online.

Podcast

Ook maakten we onze eerste podcast. Op 5 oktober, tijdens de Week van het Nederlands, kwam ‘Waar komt pindakaas vandaan?’ online. Een podcastserie van 12 korte afleveringen waarin de herkomst van alledaagse maar opmerkelijke woorden besproken wordt. De serie is geïnspireerd op het boek *Waar komt pindakaas vandaan?* (2013) dat het INT schreef in samenwerking met *Onze Taal*. Voor het ontwikkelen van de podcast werd een freelance podcastmaker aangetrokken. De podcast is in totaal al 31.345 keer gedownload. De afleveringen ‘Waar komt pindakaas vandaan?’ (3.597) en ‘Waar komt podcast vandaan?’ (3.265) zijn het vaakst beluisterd. 50% van de luisteraars vindt de podcast op

Spotify, 11% op Google Podcasts en 10% op Apple Podcasts. Maar 3% luistert de podcast via ivdnt.org.

Week van het Nederlands

Tijdens de Week van het Nederlands 2020 organiseerden we een middag rondom de lancering van de Database van de Zuidelijk-Nederlandse Dialecten (DSDD): *Horen, zien en proeven van dialect*. Het aantal bezoekers op locatie was i.v.m. de coronamaatregelen beperkt, daarom was er voorzien in een gratis te volgen livestream. De dialectmiddag werd omlijst met muziek van Wannes Cappelle, ambassadeur van de Week van het Nederlands. Verder hielden we in de Week nog een webinar over zoeken in de Historische Woordenboeken en een tutorial over OpenSoNaR.

Publicaties

In november verschenen bij uitgeverij Scriptum twee INT-boeken: *Geen woord te veel. Geschiedenis van INL tot INT* van Piet van Sterkenburg en Frieda Steurs, en *Matthias de Vries aan het woord. Uitgave ter gelegenheid van het Matthias de Vriesjaar 2020* van Roland de Bonth, Dirk Geirnaert en Rob Tempelaars. Het laatste boek werd als nieuwjaarsgeschenk verstuurd naar onze relaties.

Een uitgebreid overzicht van publicaties en media-optredens is te vinden in bijlage 3 van het jaarverslag 2020.