

/instituut voor de  
Nederlandse taal/

# jaarverslag

2019

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

*(Woordenboek der Nederlandsche Taal)*

# Inhoud

## Inleiding

1. Algemeen Nederlands Woordenboek
2. Neologismen
3. Woordcombinaties
4. Historische woordenboeken
5. DSDD (Database of the Southern Dutch Dialects)
6. Vertaalwoordenschat
7. Expertisecentrum Nederlandstalige Terminologie
8. Taalbank Nederlands
9. Spelling
10. Taalmaterialen
11. Taalportaal
12. Algemene Nederlandse Spraakkunst
13. Onderwijs
14. Externe communicatie & wetenschapscommunicatie
15. COST CA16105: European Network for Combining Language Learning with Crowdsourcing Techniques
16. COST CA18209: European network for Web-centered linguistic data science
17. European Infrastructure for Lexicography (ELEXIS)
18. European Language Resources Coordination Initiative (ELRC)
19. European Federation of National Institutes of Language (EFNIL)
20. NIAS-Lorentz-workshop “The Future of Academic Lexicography”
21. CLARIAH en CLARIAH-PLUS
22. CLARIN ERIC
23. Digitization.eu
24. Infrastructuur
25. IT-afdeling: projecten en organisatie

Bijlage 1: Raad van Toezicht en Raad van Advies

Bijlage 2: Medewerkers

Bijlage 3: Publicaties, lezingen, media etc.

Bijlage 4: Verslag Taalbank Nederlands

Bijlage 5: Rapport Externe Communicatie

# Inleiding

## Missie en visie van het instituut

### *Missie*

- Het bevorderen van kennis en gebruik van de Nederlandse taal door het uitvoeren van toepassingsgericht wetenschappelijk onderzoek.
- Het stimuleren en coördineren van de wetenschappelijke beschrijving van de Nederlandse woordenschat en grammatica in al zijn facetten door de eeuwen heen.
- De productie, koppeling en ontsluiting van bronnenmateriaal voor het Nederlands in de vorm van historische en eigentijdse corpora, woordenboeken, lexicale digitale databases, grammatica's en de daarbij behorende technologische hulpmiddelen.

### *Visie*

Het Instituut voor de Nederlandse Taal wil een breed toegankelijk wetenschappelijk instituut zijn op het gebied van het Nederlands. Daarnaast wil het instituut een centrale positie innemen in het hele Nederlandse taalgebied (Nederland, Vlaanderen, Suriname en de Antillen) op het vlak van het wetenschappelijk verantwoord ontwikkelen, bewaren en duurzaam beschikbaar stellen van taalmateriaal.

## **Wat hebben we gerealiseerd in 2019?**

Er waren vier vergaderingen van de Raad van Toezicht en een vergadering van de Raad van Advies (zie bijlage 1 voor de samenstelling).

De Raad van Toezicht kwam in 2019 bijeen op 27 februari, 29 mei, 11 september en op 11 december. Belangrijke thema's die aan bod kwamen waren personeel, inrichting financiële dienst, begroting en beleidsplan.

De Raad van Advies kwam in 2019 bijeen op 12 februari en op 11 september. Thema's die werden besproken waren: Grammatica - Taaladvies.net, Taalportaal, e-ANS, Antilliaans en Surinaams Nederlands, CLARIN Vlaanderen, Website INT: nieuwe knoppen Historisch, Hedendaags, Terminologie, Algemene Nederlandse Klankleer (ANK), de Lorentz-workshop en de wetenschapspopularisering.

In 2019, vier jaar na de oprichting van het nieuwe INT, werd op volle kracht gewerkt aan een heleboel projecten en vielen een aantal initiatieven op de juiste plaats. Het INT is na de reorganisatie duidelijk geëvolueerd tot een sterke partner voor vele doelgroepen, niet alleen de onderzoekers, maar ook docenten, bedrijven en het algemene publiek leerden het INT en de dienstverlening beter kennen. We versterkten ons imago door het opstarten van het Expertisecentrum Nederlandstalige Terminologie en

gingen tegelijk ook dieper in op de uitdagingen om een academisch woordenboek van het hedendaags Nederlands uit te bouwen met de huidige mogelijkheden van big data en AI. We worden meer en meer bekend, en vele partijen willen met ons samenwerken. Dat is positief, maar brengt meteen een grote uitdaging met zich mee: hoe kunnen we al deze interessante projecten de nodige aandacht geven en in onze planning een plaats geven? De limieten van ons potentieel worden bepaald door de budgetten, en groei is nauwelijks mogelijk.

Het personeelsbestand omvatte op 1 januari 2019 27,5 fte. Enkele collega's namen afscheid van het INT, anderen kwamen de ploeg versterken. Op 31 december waren er 27,7 fte in dienst. Het gaat hier om 33 werknemers, waarvan er een aantal deeltijds werken. Het personeelsbestand bestaat uit 16 taalkundigen, 6 computerlinguïsten, 5 systeemontwikkelaars, 2 systeembeheerders en 4 werknemers in de algemene dienst.

Het jaar begon al meteen met een gelukkig toeval: in het kader van het Rembrandtjaar deed men beroep op onze expertise van het historisch Nederlands. Dit thema liep als een rode draad door het werkjaar met een aantal mooie en geslaagde publieksevenementen: de taal van Rembrandt voor de Rembrandt Tutorials werd door onze historisch taalkundigen gereconstrueerd. Daarna volgde een Rembrandtwoordenboekje, een tentoonstelling in samenwerking met het Taalmuseum in Leiden, de verkiezing van het Rembrandtwoord van het jaar en de publieke bekendmaking in de Pieterskerk. We zetten verder in op een algemene bekendmaking van het instituut, zowel in Vlaanderen als in Nederland. Het INT vervult een duidelijke maatschappelijke functie en staat ten dienste van alle taalgebruikers. Er was in 2019 meer aandacht voor bekendmaking naar het bredere publiek en dienstverlening voor docenten en studenten, deelname aan de conferentie van Het Schoolvak Nederlands (HSN), bijdragen aan Neerlandistiek in de klas, en weer meer stagiair(es) die bij het INT aan de slag gingen. Verder leverden medewerkers van het INT ook bijdragen aan onderwijs bij de universiteiten van Leiden en Leuven.

Het overleg met de Taalunie, met name het "werkvloeroverleg" en het "directieoverleg" werd voortgezet, en hierdoor is het mogelijk de samenwerking te versterken. Daarbij heeft het INT zijn takenpakket kunnen uitbouwen, met name het werk aan de ANS (Algemene Nederlandse Spraakkunst) en de ANK (Algemene Nederlandse Klankleer). Websites zoals die van Taaladvies.net zullen door het INT worden gehost. Het INT functioneert zo meer en meer als dé centrale plaats voor alle taalmaterialen voor het Nederlands.

In 2019 was er vier keer werkvloeroverleg, met name op 19 maart, 18 juni, 10 september en 2 december. Projecten die werden besproken waren: spelling, terminologie, onderwijs Nederlands in het

buitenland (samenwerking met IVN), ANS en Taalportaal, Taaladvies.net, vertaalwoordenboeken en de corpora rond eenvoudige taal. Er was ook aandacht voor de medische terminologie en de medische pilot werd gepresenteerd.

Het directieoverleg richt zich op de bespreking van het beleidsplan, het prioriteren van bepaalde projecten en de financiële zaken: subsidie en geormerkt projectgeld. Ook de begroting komt hier aan bod. Aan het directieoverleg nemen naast de algemeen secretaris van de NTU en de directeur van het INT ook de financiële verantwoordelijken deel. In 2019 waren er overlegmomenten op 18 maart en 24 oktober.

De grote woordenboekprojecten werden voortgezet. Voor de hedendaagse woordenschat werd er hard gewerkt op drie vlakken: de uitbouw van het ANW, de neologismen met de voorbereiding van een nieuwe editor *Neoloog* en het project *Woordcombinaties*, dat nu een pilot klaar heeft.

Zo mogelijk nog belangrijker was het organiseren van de Lorentz-workshop: *The future of academic dictionaries* (4-8 november 2019). Deze intensieve brainstorming resulteerde in tal van aanbevelingen voor het optimaliseren van het werk aan de hedendaagse woordenboeken. De aanbevelingen zullen in 2020 leiden tot aanpassingen van de workflow en het inzetten van nieuwe technologische ontwikkelingen in NLP en AI.

Het derde vertaalwoordenboek, deel van een hele collectie vertaalwoordenboeken uitgebracht door de Commissie Lexicografische Vertaalvoorzieningen (CLVV), is online gezet: het betreft de collectie Estisch/Nederlands.

In 2019 werd veel aandacht besteed aan het luik vaktaal, met name door de lancering van het Expertisecentrum Nederlandstalige terminologie (ENT). Dit expertisecentrum werd sterk gepromoot en zichtbaar gemaakt via een speciale pagina op de website. Het ENT is lid van de Termraad (een interinstitutionele vergadering van vertaaldiensten van de Europese Instellingen en de Belgische, Vlaamse en Nederlandse overheid) en is betrokken bij de activiteiten van NL-TERM.

Op het vlak van deelnamen aan wetenschappelijke projecten scoorden we uitstekend: CLARIAH-PLUS-deelname betekent voor het INT een sterke wetenschappelijke activiteit op het vlak van digitale taalmaterialen en het ontwikkelen van de nodige softwaretools. Dit leidde in 2019 tot een verbeterde infrastructuur voor het historisch Nederlands.

Op Europees vlak loopt het ELEXIS-project heel vlot, waarbij het INT verantwoordelijk is voor het werkpakket “Lexicographic data and workflow”. De inzichten die hier verworven worden, kunnen we

meteen toepassen in onze eigen datacollecties. In het kader van het COST-netwerk *enetCollect* over *computer assisted language learning and crowd sourcing techniques* werkt het INT aan projecten zoals Taalradar. In 2019 ging ook een nieuwe COST-actie van start: *European network for Web-centered linguistic data science*. Ook hier maakt het INT deel uit van het managementcomité en zal zo meer kennis kunnen opdoen over de wijze waarop taalkundig materiaal verwerkt kan worden.

Op het vlak van de interne organisatie van het INT werd de financiële dienst opnieuw ingericht en een nieuwe software gekozen voor de planning. Dat is erg belangrijk aangezien het INT meer en meer taken krijgt die goed op elkaar moeten zijn afgestemd. Ook werd de dienst systeembeheer sterker gemaakt met de aanwerving van een extra personeelslid, wat resulteert in een planmatige aanpak van de verbetering van de digitale infrastructuur. De IT-afdeling had het bijzonder druk en realiseerde een aantal zeer mooie taaltoepassingen. Voor heel het instituut werd de interne workflow grondig aangepakt; tegelijk is er veel aandacht geweest voor teambuilding en meer intern overleg.

Ten slotte is er de nodige aandacht voor wetenschapscommunicatie. Het instituut werkte in 2019 ook weer actief mee aan de Week van het Nederlands, het DRONGO talenfestival, de conferentie van Het Schoolvak Nederlands en tal van andere activiteiten. Een nieuw initiatief is het publiceren van een boekje over bepaald aspect van onze werkzaamheden; in 2019 was dat het boekje *Klimaatschieter, dwangzoen en kastelen in Spanje* waarin de historische woordenschat centraal stond. Vanaf nu zal er elk jaar een publicatie worden verzorgd als nieuwjaarsgeschenk voor onze relaties.

A handwritten signature in black ink, appearing to read 'F. Steurs', enclosed within a large, horizontal, teardrop-shaped oval stroke.

Frieda Steurs  
Februari 2020

# 1. Algemeen Nederlands Woordenboek

## Taak en werkwijze

Het *Algemeen Nederlands Woordenboek* (ANW) is een corpusgebaseerd, digitaal woordenboek dat het eigentijdse Nederlands in Nederland, Vlaanderen, Suriname en het Caribisch gebied zo uitgebreid mogelijk beschrijft. Deze beschrijving strekt zich uit over de hele algemene woordenschat en wordt aangevuld met actuele neologismen. In 2019 werd vooral gewerkt aan de beschrijving van substantieven en adjectieven die behoren tot de kern van de Nederlandse woordenschat. Daarnaast zijn gangbare woorden van andere woordsoorten, woorden uit Suriname en het Caribisch gebied en tal van woorden met een beperkt profiel (d.w.z. met alleen informatie over grammaticale gegevens, de woordvorming, de uitspraak e.d.) bewerkt en deels ook online gezet.

## Databewerking

In totaal bewerkte de redactie in het verslagjaar 584 woordenboekartikelen en werden er – naast ruim 2000 trefwoorden met een beperkt profiel – 615 volledig bewerkte artikelen aan het ANW toegevoegd. De focus lag vooral op woorden uit de domeinen ‘verkeer en vervoer’ en ‘ziekten en aandoeningen’.

Gedurende drie maanden liepen achtereenvolgens Nicole van der Meijden en Marike van Dam stage bij het ANW. Zij hielden zich vooral bezig met het in het Engels vertalen van de informatie in de semagrammen van uiteenlopende categorieën substantieven.

Door Tanneke Schoonheim en Wil de Ruyter is verder gewerkt aan de ontbrekende fonetische schrijfwijzen aan het veld ‘Uitspraak’ in het ANW. De transcripties worden nu nog niet direct getoond in het ANW online, maar worden achter de schermen bewerkt in de bewerkingsomgeving Lex’it. De werkzaamheden zijn nagenoeg gereed en de gegevens kunnen op korte termijn aan het ANW toegevoegd worden.

Net als in voorgaande jaren werd in samenwerking met betrokken collega’s gewerkt aan een volledige koppeling van het ANW aan de centrale spellingdatabase, zodat het in de toekomst mogelijk wordt automatisch benodigde gegevens uit de spellingdatabase te halen.

Aan het eind van het jaar telde het ANW 260.704 woorden die behandeld zijn in 77.800 trefwoorden. De trefwoorden bevatten in totaal 43.249 betekenissen, 31.437 synoniemen, 246.108 voorbeeldzinnen, 125.529 combinaties, 8897 vaste verbindingen en 314 spreekwoorden. Verder zijn 4725 afbeeldingen, 727 videofilmmpjes en 178 geluiden opgenomen.

## Medische pilot

In 2019 werden zoals gezegd tal van woorden uit het domein ‘ziekten en aandoeningen’ aan het ANW toegevoegd. Directe aanleiding was daarvoor het opzetten van een pilot waarin de medische taal op

een unieke wijze wordt benaderd, met als noviteit bijvoorbeeld definities die ook voor laaggeletterden toegankelijk zijn. Boukje Verheij en Rob Tempelaars bewerkten in totaal 152 woorden en verbindingen uit Nederland en Vlaanderen in een speciaal voor het project naar hun wensen ingerichte werkomgeving in Lex'it. Gekozen werd in eerste instantie voor de bewerking van woorden uit de categorieën 'diabetes' (suikerziekte), 'hartproblemen', 'overgevoeligheid voor geluiden' en 'schildklierproblemen'. Bij diverse gelegenheden heeft Boukje Verheij de medische pilot gepresenteerd.

### **Technische ondersteuning**

De redactionele werkzaamheden worden ondersteund door een computerlinguïst en een ontwikkelaar/programmeur. Op technisch gebied werden zowel voor als achter de schermen de nodige verbeteringen gerealiseerd. Als voorbeelden noemen we: a) een andere opzet van de woordfamilie, noodzakelijk om de centrale spellingdatabase succesvol aan het ANW te kunnen koppelen; b) het toevoegen van de zoekmogelijkheid om uitsluitend in de woordfamilie te zoeken met een joker; c) uniformering van de volgorde van afleidingen in de woordfamilies van alle artikelen.

## **2. Neologismen**

In 2019 is er een testversie van het neologismenproject, een woordenboek met alleen nieuwe woorden, intern online gegaan en in 2020 komt een eerste versie van dit neologismenwoordenboek op internet beschikbaar voor iedereen. De testversie van het woordenboek bevat al ruim 6000 woordenboeklemma's. De onlinegang van dit woordenboek in 2020 zal worden afgestemd met de verschijning van een populairwetenschappelijk boek over neologismen bij Uitgeverij Scriptum.

In het *Neologismenwoordenboek* (NW) worden nieuwe woorden opgenomen, te weten nieuwe samenstellingen (bv. *vliegschaamte*), nieuwe afleidingen (*ontspullen*), nieuwe leenwoorden (*app*, *selfie*) en bestaande woorden in een nieuwe betekenis (*slim* in bv. *slimme meter*). Ook in het *Algemeen Nederlands Woordenboek* (ANW) worden neologismen opgenomen, maar alleen die woorden die ingeburgerd zijn in de Nederlandse taal en al langere tijd gebruikt worden. In het NW is ook ruimte voor nieuwe woorden die weliswaar geen blijvertjes zijn, maar die om een andere reden interessant zijn om te beschrijven, bijvoorbeeld omdat ze passen in een morfologische reeks of in een bepaald betekenthema of omdat ze typerend zijn voor een bepaalde tijd.

Het ontwerp van het NW verschilt van dat van het ANW: er is gewerkt met een tabbladstructuur en de verschillende taalkundige informatiecategorieën zijn op een andere wijze geordend en benoemd.



Het neologismenproject werd dit jaar op twee internationale congressen gepresenteerd, te weten op de *Globalex workshop on Lexicography and Neologisms* op de conferentie van the *Dictionary Society of North America* (DSNA) in Bloomington, Indiana en bij de workshop *Neologismen: korpuslinguistische Ermittlung und lexikographische Bearbeitung* op het Leibniz-Institut für Deutsche Sprache in Mannheim. Op de workshop in Mannheim werd niet alleen het neologismenwoordenboek geïntroduceerd, maar werd ook het programma Neoloog gepresenteerd. Neoloog is een op het INT ontworpen programma waarin krantenteksten geladen worden waarin vervolgens semi-automatisch neologismen kunnen worden opgespoord.

Afgelopen jaar is er ook een begin gemaakt met het systematisch doorwerken van de Neologismenlijst, een lijst van ongeveer 20.000 nieuwe woorden die op het INT wordt bijgehouden. Van alle woorden die op deze lijst staan, wordt bekeken of zij voor bewerking in aanmerking komen voor het ANW en/of het NW. Ook dit jaar weer behandelden Rob Tempelaars en Vivien Waszink elke dinsdag op [ivdnt.org](http://ivdnt.org) een nieuw woord in de webrubriek Neologisme van de week.

### **3. Woordcombinaties**

*Woordcombinaties* wordt een online taaltool die leerders van het Nederlands als vreemde taal en taalgebruikers in het algemeen ondersteunt bij het gebruiken van woorden in context. In een pilot met werkwoorden wordt bekeken in hoeverre een collocatie- en idioomwoordenboek en een patroonwoordenboek geïntegreerd kunnen worden.

Het project is corpusgebaseerd en gebruikt [Sketch Engine](#) als corpusquerytool. Voor de beschrijving van patronen en hun betekenis wordt de Corpus Pattern Analysis (CPA) gebruikt van [Patrick Hanks](#). *Woordcombinaties* toont hoe woorden gebruikt worden in goede voorbeeldzinnen, welke woorden met elkaar gecombineerd worden en hoe (valentie)patronen samen met collocaten gebruikt worden voor het bouwen van zinnen. Taalleerders en taalgebruikers leren met de tool niet alleen woorden kennen, maar ook woorden gebruiken.

#### **Redactionele werkzaamheden**

In de eerste helft van 2019 zijn alle voorbeeldzinnen en combinatiemogelijkheden voor de werkwoorden uit de pilotlemmalijst bewerkt. Daarnaast is begonnen met voorbeeldzinnen en combinatiemogelijkheden voor werkwoorden uit de *Basislijst Schooltaalwoorden vmbo en ISK* van het Instituut voor Taalonderzoek en Taalonderwijs Anderstaligen van de Universiteit van Amsterdam. In de tweede helft van 2019 is de patrooneditor getest met een aantal proefartikelen. Het annoteren en beschrijven van semantisch gemotiveerde patronen is een complexe en arbeidsintensieve taak. De proefartikelen kunnen een basisset opleveren op basis waarvan een methodiek voor patroonbeschrijving verder uitgewerkt kan worden.

Patronen die een aparte aanpak vereisen, zijn conversatieformules. Ze zijn gelinkt aan taalhandelingen en taalsituaties. Er zijn voorbereidende werkzaamheden uitgevoerd voor de bewerking van dit type patronen.

### **Corpuswerkzaamheden en data-extractie**

Voor de redactionele bewerking zijn we in 2019 uitgegaan van een getagd pilotcorpus. Omdat naast de woordsoortinformatie ook de structuur van de zinnen belangrijk is, met name voor het bewerken van patronen, is in 2019 een geparseerde versie van het pilotcorpus gemaakt dat in de SketchEngine gebruikt kan worden. Hiervoor is de Alpino-parser gebruikt. Aangezien de SketchEngine al ondersteuning bood voor corpora in CONLL-UD-formaat, is in eerste instantie de Alpino-output geconverteerd naar dit formaat. Deze omzetting naar de universele UD-labels gaat echter ten koste van de fijnmazigheid van de Alpinolabels. Om dit te vermijden werd ook nog een aanpak geïmplementeerd waarbij de Alpino-output omgezet wordt naar CONLL-formaat, zonder de mapping naar Universal Dependencies, dus met behoud van de originele Alpinolabels. Er wordt nog overwogen om deze twee bronnen samen te voegen tot een geheel, waarbij zowel de fijnmazige Alpinolabels als de universele UD labels beschikbaar zijn.

Daarnaast is in 2019 een begin gemaakt met automatische data-extractie voor de verschillende onderdelen van het project, d.w.z. voorbeeldzinnen, combinatiemogelijkheden en patronen.

### **Ontwikkeling lexicografische tools**

Binnen het ELEXIS-project is in 2019 verder gewerkt aan de ontwikkeling van de Skema patrooneditor voor de bewerking van de patronen in *Woordcombinaties*.

Binnen het INT zijn ook voorbereidingen getroffen voor het creëren van een werkomgeving voor het inventariseren van conversatieformules, taalhandelingen en situaties.

### **Webapplicatie**

De werkzaamheden voor het online zetten van een demoapplicatie voor taalleerders zijn gestart. In het eerste kwartaal van 2020 komt de demoversie online.

## **4. Historische woordenboeken**

In 2019 is de Helpfunctie van de webapplicatie van de historische woordenboeken geactualiseerd, herzien en online gezet. Verder zijn in het XML-bestand van het *Supplement op het Middelnederlandsch Handwoordenboek* van J. J. van der Voort van der Kleij enkele bewerkingen ter optimalisering van de data uitgevoerd, teneinde dit supplement te zijner tijd op de meest efficiënte manier aan de onlineapplicatie van de historische woordenboeken toe te kunnen voegen.

Het afgelopen jaar hebben uiteenlopende publicaties onze historische woordenboeken ruimer onder de aandacht gebracht. Net als in voorgaande jaren werd in de webrubriek *Terug in de Taal* een twintigtal bijdragen rond historische woorden en begrippen gepubliceerd. Op basis van deze en eerder verschenen stukken in deze rubriek is het taalboek *Klimaatschieter, dwangzoen en kastelen in Spanje* samengesteld, dat eind 2019 in de boekhandel lag en als nieuwjaarsgeschenk onder de relaties van het INT is verspreid.

Verder heeft het INT ook het Rembrandtjaar 2019 gebruikt om met enkele producten naar buiten te treden waarbij vooral het *Woordenboek der Nederlandsche Taal* (WNT) een centrale rol speelde. Zo werd het digitale gelegenheidswoordenboekje *Woorden voor Rembrandt* samengesteld, dat in 63 woordenboeklemma's (evenveel lemmata als Rembrandt levensjaren telde) een honderdtal 17<sup>e</sup>-eeuwse schilderstermen behandelde. Dit woordenboekje werd ook integraal opgenomen in *De taal van Rembrandt. De kunst van het overtuigen*, een publicatie van het Leidse Taalmuseum. Ook bij het hertalen naar het 17<sup>e</sup>-eeuwse Nederlands van de zogeheten *Rembrandt Tutorials*, een reeks door ING en het Amsterdamse Rijksmuseum geïnitieerde schilderslessen, is intensief en met vrucht gebruik gemaakt van de digitale versie van het WNT. Over de werkwijze en de hertaalprincipes die daarbij gehanteerd werden, is verslag gedaan op [Neerlandistiek.nl](http://Neerlandistiek.nl).

Het afgelopen jaar is een extra inspanning gedaan om aandacht voor de webapplicatie van de historische woordenboeken te genereren bij het onderwijs. Voor het voortgezet onderwijs werden twee lesplannen ontwikkeld, te vinden op de website bij de landingspagina Historisch Nederlands. Ook zijn aan eerste- en tweedejaarsstudenten Nederlands van de Universiteit Leiden werkcolleges gegeven over de opzet en de onderzoeksmogelijkheden van de historische woordenboeken online.

## **5. DSDD (Database of the Southern Dutch Dialects)**

Het DSDD-project behelst de samenvoeging en standaardisering van drie grote dialectlexicografische databases, waarvan de samenstelling tientallen jaren heeft gekost: het *Woordenboek van de Vlaamse Dialecten* (WVD, 1972 -), het *Woordenboek van de Brabantse Dialecten* (WBD, 1961-2005) en het *Woordenboek van de Limburgse Dialecten* (WLD, 1961-2008). De woordenboeken werden 40 à 50 jaar geleden parallel opgezet, om een uiteindelijke samenvoeging mogelijk te maken. Ondanks deze parallelle opzet lopen de woordenboeken zowel methodologisch als in de keuze van de behandelde concepten uiteen. Ook in technisch opzicht (bestandsformaten en logische structuur) is sprake van heterogeniteit.

De geïntegreerde database heeft tot doel vernieuwend onderzoek mogelijk te maken, vooral op het gebied van kwantitatieve lexicologie en dialectgeografische analyse. Het project, dat loopt van 1/1/2017 tot 30/4/2020, zal resulteren in een geharmoniseerde dataset, een API ten behoeve van onderzoekers en een portaalapplicatie waarin het resultaat van het project voor een breed publiek beschikbaar zal zijn.

In 2019 is verder gewerkt aan het vervolledigen van de database-inhoud: Concepten zijn ingevoerd en gekoppeld en definities zijn toegevoegd; correcties op lemma- en trefwoordinhoud zijn uitgevoerd. Er is een eerste versie van de portaalapplicatie ontwikkeld, die volgens planning in de eerste helft van 2020 opgeleverd wordt. Daarmee is het pilotproject dan afgerond.

## 6. Vertaalwoordenschat

De *Vertaalwoordenschat* is een applicatie voor tweetalige woordenboeken met Nederlands als bron- of doeltaal, ontwikkeld door het Instituut voor de Nederlandse Taal. Rond de eeuwwisseling zijn er verschillende tweetalige bestanden ontwikkeld voor talen die voor de Nederlandstalige wel relevant zijn, maar op de commerciële markt niet spontaan aan bod kwamen, zoals Nederlands - Deens, Nederlands - Nieuwgrieks en Nederlands - Arabisch. Deze bestanden zijn veelal gemaakt in opdracht van een speciale *Commissie Lexicografische Vertaalvoorzieningen*, ingesteld door de toenmalige ministers van Onderwijs van Nederland en Vlaanderen. In de meeste gevallen beschikt de Taalunie over het volledige auteursrecht op deze bestanden en zijn met diverse uitgevers afspraken gemaakt over de papieren publicatie ervan. In enkele gevallen deelt de Taalunie het auteursrecht met andere partijen of beschikt ze enkel over een uitgavelicentie. Enkele talenparen zijn nu niet meer in druk omdat uitgevers er geen commerciële mogelijkheden meer in zien. Het gevolg is dat deze bestanden niet meer beschikbaar zijn voor gebruikers en vertalingen tussen diverse talenparen niet meer worden ondersteund.

Daarom is door het INT in september 2017 het onlineplatform de *Vertaalwoordenschat* gelanceerd, waarmee deze tweetalige bestanden ter beschikking worden gesteld aan gebruikers. Op dit moment zijn Nederlands – Nieuwgrieks / Nieuwgrieks – Nederlands, Nederlands - Portugees / Portugees - Nederlands en sinds de zomer van 2019 Nederlands-Estisch via dit platform ontsloten. In de toekomst zal het platform verder worden uitgebreid met andere talen.

In 2019 is ook begonnen met het bijwerken van de spelling van de Nederlandse lemma's van de bestanden. De bronbestanden van de woordenboeken zijn immers veelal 10 tot 20 jaar oud.

Momenteel wordt de spelling handmatig per taalpaar bijgewerkt. Door de Nederlandse lemma's te koppelen aan de spellingdatabase, hoeft dat in de toekomst niet meer handmatig te gebeuren.

## **7. Expertisecentrum Nederlandstalige Terminologie**

De uitbouw van het Expertisecentrum Nederlandstalige Terminologie (ENT) heeft in 2019 zijn verdere beslag gekregen in de vorm van veldondersteunende activiteiten en de positionering in een netwerk van relaties.

### **Website**

Tussen de veldondersteuning en de website van het instituut bestaat een duidelijke, ook gevisualiseerde band. De voorzieningen op de ENT-pagina zijn er verder uitgebreid of geactualiseerd. Onder meer werd een introductierubriek toegevoegd met algemene informatie over terminologie. De gebruiker wordt er op weg geholpen met verwijzingen naar inleidende literatuur, richtlijnen voor terminologiebeschrijving en een praktische termtest. Tot de webwerkzaamheden van 2019 behoort ook de opname van een geactualiseerde rubriek voor taaltechnologie op basis van de vroegere NedTerm-pagina van het Taaluniversum. De nieuwe rubriek vergemakkelijkt oriëntatie op het vlak van hulpmiddelen die het terminologiewerk ondersteunen door termextractie, termbeheer, tekstalignering enz. Actualisering vond eveneens plaats voor de onderdelen online beschikbare termenlijsten en de opleidingen en cursussen, hoofdzakelijk in Nederland en Vlaanderen. De al bestaande agenda met terminologie-evenementen binnen Europa werd gecontinueerd. De data voor de rubriek terminologiegerelateerde instellingen zijn reeds verzameld en zullen in de loop van 2020 worden verwerkt in de vorm van een interactieve landkaart.

Raadpleegbaar op de terminologiepagina zijn ook de nieuwsbrieven Terminologie. In 2019 verschenen er opnieuw vier. Ze boden, afhankelijk van het nieuwsaanbod, volgens een vaste rubrieksindeling informatie over voorbije evenementen, nieuwe ontwikkelingen en publicaties, tools en bronnen, berichten van NL-Term en een agenda. Als smaakmaker is een rubriek 'uitgelichte term' toegevoegd die eveneens apart op de webpagina wordt gearchiveerd.

### **Terminologietools**

Vanuit het veld blijkt geregeld belangstelling voor de terminologietools TermTreffer en TermBeheerder. Overeenkomstig beleidsdocument *ENT: visie en missie* was veldondersteuning op dit vlak dan ook aan de orde. INT en ENT hebben activiteiten ontplooid om deze terminologietools voor termextractie, respectievelijk terminologiebeheer te onderhouden en verder te ontwikkelen. Dit hield

concreet voorstellen in voor verbeteringen en opvolging door ENT, en onderzoek en implementatie door de IT-afdeling en het externe ICT-bedrijf OGD. Tot de verbeteringen behoren onder meer het bijwerken van verouderde software, de mogelijkheid van inladen van geëxporteerde TBX-bestanden uit TermTreffer en de koppeling aan de CLARIN-login die voor TermBeheerder al is gerealiseerd. Na verdere technische updates en de nodige tests, voor TermTreffer bovendien na invoering van recentere referentiecorpora, worden zo mogelijk de verbeterde versies in 2020 ter beschikking gesteld van het publiek.

### **Projecten met toegepaste waarde**

Ook projecten met toegepaste waarde die moeten resulteren in online raadpleegbare materiaalverzamelingen, maakten in 2019 deel uit van de veldondersteuning.

In het kader van het project *Hoger Onderwijs Terminologie in Nederland en Vlaanderen (HOTNeV)* vond samenwerking plaats met de KULeuven - Campus Antwerpen. Werkopdrachten werden samengesteld voor 25 studenten en het ENT zorgde als ondersteuning bij de termbeschrijving voor begeleidend onderwijs en een introductie in corpuscompilatie en zoektools. Tevens werd een corpus van teksten met betrekking tot hogeronderwijsterminologie gevormd. Ook is begonnen met een concept voor een projectaanvraag in het kader van een EU Call for Proposals voor 2020. Het doel daarbij is de beschikbare Nederlands-Engelse database van de Rijksuniversiteit Groningen te harmoniseren met gebruikmaking van computationeel-linguïstische methoden, deze publiek ter beschikking te stellen en vertalers van Nederlandse en Vlaamse universiteiten in staat te stellen bij te dragen. Tevens werden enige HOTNeV-termen vanuit het ENT lexicografisch bewerkt voor het *Algemeen Nederlands Woordenboek*.

Een tweede project waarvoor in 2019 onderzoek is verricht, betreft het platform medische taal. Deze voorziening beoogt de koppeling van medische woordenschat voor verschillende gebruikersgroepen: zorgprofessionals, paramedici en perimedici, en groepen die behoefte hebben aan laagdrempeligere medische communicatie (leken, minder taalvaardige zorgvragers, anderstalige professionelen en studenten in de zorgsector met behoefte aan Nederlandstalige zorgterminologie). Daarvoor dienen vooral drie taalregisters te worden gekoppeld: een vaktalig register, een vereenvoudigd register voor communicatie met patiënten en een register met informele termen. Als structuur voor het platform is gekozen voor het internationale medische indelingssysteem: SNOMED CT. Daarvoor is een licentie aangevraagd, zijn gegevens en hun structuur bestudeerd en vond een basisverwerking van terminologiegegevens plaats in Lex'it. Tevens werd overlegd over het gebruik van de Ontoterminology Editor met de ontwikkelaar aan de Université Savoie Mont-Blanc en over de mogelijkheid van een visualisering van de ontologie die verder gaat dan de huidige SNOMED-weergave. Ook werd vergaderd met de Taalunie en NICTIZ, en volgde een verkennend werkbezoek

bij NICTIZ. Daarnaast werd deelgenomen aan vergaderingen die NICTIZ belegde over patiëntvriendelijke termen.

Per 1 september 2019 is verder het project Juridische Terminologie volgens afspraak overgenomen van de Stichting Rechtstaal. Een van de eerste acties betrof het maken van afspraken over het beheer van de data in de juridische database (voorlopig nog bij de Fryske Akademy) en het inventariseren van benodigde werkzaamheden aan de (nog vrij rudimentaire) gebruikersapplicatie.

De juridische database bevatte bij overname 2642 termen, in verschillende stadia van bewerking. In 2019 zijn beleid en afspraken van voor de overname door het INT gerespecteerd. Dit heeft in de periode van september tot en met december geresulteerd in de definitieve bewerking van ca. 80 juridische termen, voornamelijk op het gebied van het burgerlijk procesrecht. Dit is conform de afspraken die bij de overname van het project in september 2019 zijn gemaakt.

### **Terminologisch netwerk**

Verder is gestreefd naar positionering door deelname aan uiteenlopende initiatieven. Het INT bood medewerking aan het programma en de organisatie van het jaarlijkse TiNT-symposium van NL-term. Met de Nederlandstalige (ver)taaldiensten van EU-instellingen, vertegenwoordigers van Nederlandstalige (ver)taaldiensten van Nederlandse, Belgische en Vlaamse overheden en de Nederlandse Taalunie was het ENT ook betrokken bij de heractivering van de Termraad Nederlands. Doel van de samenwerking is terminologische afstemming en overleg tussen de deelnemende instanties. Contact en samenwerking kwamen ook tot stand door bijdragen aan het Nordterm-congres (11-14 juni) in Kopenhagen, de vierde International Summer School in Translation Technology (2-6 september) in Antwerpen, en de Workshop Terminology in Language Technology (29 november) in Brussel. Informatie over ENT-werkzaamheden en terminologische toelichting zijn ook verstrekt aan bezoekers in het kader van hun werkzaamheden voor ELEXIS, EMLex, de universiteit van Wrocław en het Croatian Standard Language Department. Ten slotte is vanuit het ENT is ook deelgenomen aan de door ECQA gecertificeerde opleiding terminologiebeheer van de internationale terminologievereniging TermNet.

## 8. Taalbank Nederlands

De Taalbank Nederlands omvat corpora en computationele lexica van zowel modern als historisch Nederlands. Het lexicon is bedoeld om ingezet te worden voor onderzoek in de vorm van diverse producten die niet alleen in principes van verrijking en structuur op elkaar aansluiten, maar bovendien in toenemende mate inhoudelijk aan elkaar gekoppeld zijn. De zichtbaarheid van de resultaten is in 2019 vergroot door het realiseren van de themapagina's historisch en modern Nederlands. Hieronder volgt een stand van zaken (zie bijlage 4 voor een uitgebreid verslag).

### GiGaNT

GiGaNT (Groot Geïntegreerd Lexicon van de Nederlandse Taal) is het computationele woordvormenlexicon dat de door het INT beschreven woordenschat moet gaan bevatten van het Nederlands vanaf de 6<sup>e</sup> eeuw tot nu en de centrale database vormt van het Nederlands van het INT. In GiGaNT zit de formele beschrijving van de woorden. De semantische informatie zit in DiaMaNT, en in de diverse woordenboeken van het INT.

Werkzaamheden aan deze centrale data-infrastructuur zijn modulair opgezet. De twee grote componenten zijn GiGaNT-Molex, waarin de hedendaagse woordenschat wordt beschreven, en GiGaNT-Hilex, de historische lexiconcomponent. Voor de historische lexiconcomponent moeten de vier historische woordenboeken van het Nederlands (ONW, VMNW, MNW en WNT) de kern vormen, aangevuld met nieuw materiaal. De moderne lexiconcomponent bevat het hedendaags taalmateriaal. Lexiconinformatie wordt ter beschikking gesteld via de lexicon service en het doel is ook om downloadbare datasets te distribueren.

Afgelopen jaar is gewerkt aan systematische verfijningen van de verrijking in de MNW- en WNT-module van GiGaNT- Hilex. Er is verder gewerkt aan de koppeling van de historische en moderne component en de uitbreiding van de moderne lexiconcomponent. Daarnaast zijn belangrijke stappen gezet in het realiseren van de centrale rol van het moderne lexicon in de beschrijving van de woordenschat in de diverse producten van het INT. De structuur en de inhoud van GiGaNT-Molex is uitgebreid om de koppeling met het combinatiewoordenboek mogelijk te maken, en er is verder gewerkt aan de technische koppeling van ANW, Neoloog en Molex. De API op Molex is hiervoor aangepast, en er zijn aanpassingen gedaan in de verschillende werkomgevingen. Wat betreft de werkomgeving voor Molex, deze heeft technische aanpassingen gehad om de werkzaamheden ten behoeve van spelling efficiënter te laten verlopen en om de performance te verhogen.



## **DiaMaNT**

In dit project wordt gewerkt aan de ontwikkeling van een Diachroon semantisch lexicon van het Nederlands (DiaMaNT). Het diachroon semantisch lexicon heeft als doel een hulpmiddel te bieden bij tekstontsluiting en bij het onderzoek naar begrippen door de eeuwen heen. Het lexicon is het derde onderdeel van de lexicografische infrastructuur voor historisch Nederlands die in ontwikkeling is en als kern de vier historische woordenboeken van het Nederlands bevat, ONW, VMNW, MNW en WNT. Het eerste onderdeel van deze infrastructuur is het historische woordenboekenportaal, waar de gebruiker de betekenis van woorden kan opzoeken. Het tweede onderdeel is het morfosyntactische lexicon GiGaNT – ook functionerend als centrale database – dat informatie bevat over mogelijke variatie in spelling en vorm van de historische Nederlandse taal, waarmee het zoeken in historische tekst vergemakkelijkt wordt. De derde component is het DiaMaNT-lexicon. Het vormt een laag bovenop het GiGaNT-lexicon en is bedoeld om te hulp te bieden bij het omgaan met historische semantische variatie.

Het doel van het lexicon is om relaties te leggen tussen woordvormen en betekenseenheden (concepten), en deze in de tijd te plaatsen en zo diachrone onomasiologie, d.i. de veranderende uitdrukking/verbalisatie van een concept, en semasiologie, d.i. de verschuiving van betekenis(nuance) van woorden in de tijd, systematisch vast te leggen op een zodanige wijze dat de informatie voor mens en computer bruikbaar is.

Het datamodel van het uiteindelijke lexicon ligt vast. De bewerking van de data tot een gestructureerd semantisch lexicon gebeurt in stappen, waarbij geprobeerd wordt om tussentijdse resultaten ook ter beschikking te stellen. Het ontwikkelen van intelligenter strategieën voor lexiconbouw is een continue taak.

In 2019 is gewerkt aan een verbetering van het datamodel in RDF. Het werk daaraan is in internationaal verband uitgevoerd en resultaten van die werkzaamheden hebben hun neerslag gehad in de OntoLex Lemon module voor lexicografische producten, verschenen in het najaar van 2019 (OntoLex Lemon Lexicography Module; <https://www.w3.org/2019/09/lexicog/>). Omdat DiaMaNT attestaties met woordvormen koppelt aan lemmata en betekenissen, en attestaties geen onderdeel van de vermelde module vormden, is in 2019 het werk aan het modelleren ervan in RDF uitgevoerd binnen een nieuw te definiëren OntoLex Lemon module voor frequentie, attestaties en corpus data (FRAC; <https://github.com/acoli-repo/ontolex>).

De correctie van de automatische synoniemdetectie is in 2019 afgerond. Eind 2019 is de DiaMaNT-applicatie versie 1 gereleased. De gebruikersapplicatie laat toe om te zoeken naar historische

synoniemen van moderne concepten en kan daarvan een visualisatie zien door de tijd heen. De applicatie is te vinden op <http://diamant.ivdnt.org/diamant-ui/>.

DiaMaNT is verder ingezet in het CLARIAH-project ATHENA.

### **Modern corpusmateriaal: Corpus Hedendaags Nederlands (CHN)**

Het CHN bevat al het hedendaags Nederlands corpusmateriaal waarover het INT beschikt en wordt door diverse projecten in het INT gebruikt. Het corpus wordt continu uitgebreid. Daarbij wordt beoogd om materiaal binnen te halen van Nederland, Vlaanderen, Suriname en de Antillen. De acquisitie wordt mede bepaald door de projecten die van het corpus gebruikmaken. Een belangrijke functie van het corpus is het monitoren van de Nederlandse taal. Daarvoor proberen we data te verwerven die systematisch en continu aangeleverd worden. Kranten spelen daarbij een belangrijke rol, maar ook sociale media en parlementair materiaal leveren een continue stroom aan taaldata op. Voor het project spelling was er behoefte aan meer materiaal uit Suriname en de Antillen. Voor de Antillen hebben we *ABC Online Media* en de *Achterkant van Curaçao* aan het corpus kunnen toevoegen en voor het Surinaams het forum van *culturu.com*.

Voor het combinatiewoordenboek is er behoefte aan materiaal dat de gesproken taal benadert en ook behoefte aan eenvoudiger Nederlands. Er is daarom gewerkt aan het materiaal van Open Subtitles. Er zijn verder onderhandelingen gestart met de VRT. Er is ook een eerste aanzet gedaan tot het benaderen van educatieve uitgeverijen, als aanvulling op het materiaal uit *Basilex*.

Parlementair materiaal is een welkome aanvulling op het krantenmateriaal dat we gebruiken als monitorcorpus. We krijgen het materiaal van het tweede kamer uit Nederland systematisch binnen via de ftp-server. Inmiddels is ook een eerste verkenning gedaan van de data van het Vlaams parlement en het federaal parlement van België.

Het monitorcorpus van kranten willen we graag uitbreiden met meer Vlaamse en Nederlandse kranten, en het liefst zodanig dat we het materiaal kunnen distribueren als dataset aan geïnteresseerde gebruikers. De onderhandelingen met het Vlaams Mediahuis verlopen voorspoedig. Tot slot zijn de recente aan de Taalmaterialencollectie toegevoegde corpora *Morrocorp* en *Wablieft* geconverteerd en aan het CHN toegevoegd.

Er zijn in 2019 ook tests gedaan met het nieuwe corpus frontend dat in het kader van CLARIAH is gerealiseerd. En er is een investering gedaan in een veel krachtiger corpusserver. In 2020 moet het CHN online komen.

## **Historisch corpusmateriaal**

In 2019 zijn de data van de volgende corpora onderhanden genomen en van een update voorzien: *Brieven als buit*, *Corpus Gysseling*, *Corpus Oudnederlands*, *Eindhoven Corpus* en *Corpus Middelnederlands*. Dat betreft zaken als datacorrecties, aanpassingen in de taalkundige verrijking, of verdere curatie van de data, in het geval van het *Corpus Oudnederlands*. De aanvulling op het *Brieven als buit*-corpus is ontvangen en ook verwerkt.

Na het online gaan van Open SoNaR+ in de nieuwe corpusapplicatie, is gewerkt aan een versie van de bovengenoemde corpora in het nieuwe frontend. Dat betekende dat het corpus frontend zo aangepast moest worden om het makkelijker configureerbaar te maken voor corpora met andere metadata of andere taalkundige verrijking. Ook aan het BlackLab backend zijn aanpassingen gedaan. Omdat het *corpus Middelnederlands* en de *Brieven als Buit aanvullingen* geen taalkundige verrijking hebben, is de lexiconservice als functionaliteit aan het frontend toegevoegd. De corpora worden in 2020 aan een laatste test onderworpen en komen dan online.

## **9. Spelling**

### **Woordenlijst**

Sinds 1995 is de Woordenlijst Nederlandse Taal bij het INT ondergebracht. Het gedrukte boekje werd in 2015 voor het laatst geactualiseerd, terwijl de onlineversie sindsdien regelmatig geüpdatet wordt. Ook in 2019 was dit het geval: voor het vervolgtraject werden ongeveer 3500 nieuwe trefwoorden opgeleverd. Naast deze nieuwe trefwoorden werden extra woordvormen en andere informatie aangevuld en vonden ook de nodige correcties van het reeds aanwezige taal materiaal plaats. Correcties van eerder opgeleverd materiaal wordt bijgehouden in een erratalijst. In 2019 heeft de bewerkingsomgeving een technische update gekregen waardoor erratarapportages automatisch gegenereerd kunnen worden. Andere technische verbeteringen aan de bewerkingsomgeving zijn: het implementeren van functionaliteiten om zonder tussenkomst van een computerlinguïst consistentiechecks te doen; en het toevoegen van een functionaliteit om op elk gewenst moment proefopleveringen te genereren.

In 2019 werd ook de lijst met Surinaams-Nederlandse trefwoorden, die in 2018 door het INT was samengesteld en beoordeeld door deskundigen uit Suriname, aan de onlinewoordenlijst toegevoegd. Een soortgelijk traject is in 2019 uitgevoerd voor het Antilliaans-Nederlands, als voorbereiding op de online toevoeging van dat materiaal.

In 2019 is het besluit genomen, in overleg met de Taalunie, om de hosting van woordenlijst.org bij het INT onder te brengen. Eind 2019 is hiervoor het technisch overleg gestart. De uitvoering moet in de loop van 2020 zijn beslag krijgen.

### **Commissie Spelling**

Alle toevoegingen aan en wijzigingen van het onlinespellingbestand werden ook in 2019 voorgelegd aan en goedgekeurd door de Commissie Spelling. In 2019 werden 6 vergaderingen van deze Commissie georganiseerd (incl. een extra vergadering met Helen Chang over de Surinaams-Nederlandse trefwoorden). Door het INT werden voor deze vergaderingen o.a. volgende taken uitgevoerd: samenstellen en aanleveren van woordenlijsten; voorbereiden van inhoudelijke kwesties; verwerken van opmerkingen, aanvullingen en correcties in het spellingbestand.

### **Keurmerk Spelling**

Eind 2018 werd beslist dat het INT zou stoppen met de reguliere keurmerkcontroles. Voor de zeer omvangrijke doch voor het INT erg interessante woordverzamelingen van OpenTaal (ruim 230.000 trefwoorden) en TaalTik (ruim 915.000 trefwoorden) worden in 2020 (met een eerste aanzet in 2019) aangepaste controletrajecten uitgewerkt, gezien de grootte en complexiteit van deze materialen. De spellingcontrole van (substantiële delen van) deze trefwoordenverzamelingen kan namelijk voor beide partijen een win-winsituatie betekenen.

### **Spelspiek & spellingvragen**

In het kader van Spelspiek zijn tijdens de eerste maanden van 2019 nog een twintigtal vragen beantwoord. Vanaf begin april 2019 is de service opgehouden te bestaan. Daarnaast kwamen ook via e-mail spellingvragen (en bij uitbreiding taalvragen) binnen, die zo snel mogelijk werden beantwoord of doorgestuurd aan de juiste persoon of dienst.

Zoals ook vermeld in het meerjarenbeleidsplan, is in 2019 nagedacht over het 'inhouse' ontwikkelen van een Spelspiek 2.0, aangezien het potentieel van deze interactieve toepassing volgens ons onvoldoende benut werd. We zouden hiervoor dan gebruik kunnen maken van eigen, dynamische data. Prioriteit wordt voorlopig echter gegeven aan het overnemen van de hosting van woordenlijst.org.

## **10. Taalmaterialen**

In 2018 is een nieuwe website in gebruik genomen en is een nieuw beleid in gevoerd voor de distributie en licensering van de taalmaterialen. Ook voor commercieel gebruik waren de producten gratis. Dat had in 2018 nog geen grote gevolgen voor de hoeveelheid afgesloten licenties voor commercieel gebruik. Aan het begin van 2019 zijn we daarom begonnen om dat aspect meer te

benadrukken. Op alle productpagina's van commerciële (versies van) producten werd nadrukkelijk vermeld dat het gebruik daarvan gratis was. Dat heeft wellicht bijgedragen aan het grotere aantal afgesloten licenties voor commercieel gebruik. In 2019 waren dat er 41, terwijl in 2018 het aantal slechts 9 bedroeg. Wel moet worden opgemerkt dat één commercieel bedrijf dit jaar tegelijkertijd voor tien producten licenties heeft afgesloten.

Behalve dat het aantal commerciële licenties toenam, was er ook een beduidende groei bij de niet-commerciële downloads. In 2019 zijn er 1169 geregistreerd, tegen 582 in 2018.

Verder is de systematiek voor verwijzing naar producten aangepast. Voorheen werd gebruikers van de taalmaterialen aanbevolen om in publicaties te verwijzen naar een referentiepapier dat door de makers van het product was gepubliceerd. Die referentiepapier zou dan weer een verwijzing naar de locatie van het product bevatten. Deze systematiek heeft als nadeel dat de verwijzingen in de referentiepapieren naar de productpagina's verouderen zoals bijvoorbeeld de verwijzingen naar de TST-Centrale. Verder is niet voor alle producten een referentiepapier beschikbaar. Daarom is besloten om voor alle producten een aparte PID (Permanent Identifier) aan te maken en die met een verwijzingsvoorschrift aan de productpagina's toe te voegen. Voor de PID's wordt de Handle service ('<http://handle.net>') gebruikt omdat het INT die al langere tijd gebruikt en ook omdat het binnen CLARIN een standaard is.

Omdat het mogelijk moet zijn te verwijzen naar oudere versies van producten is de taalmaterialenwebsite uitgebreid met z.g. 'Tombstone'-pagina's. Die bevatten een beschrijving van oude (versies van) producten en een verwijzing naar de actuele versies.

## **Nieuwe producten**

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

- CHN N-grams. De verzameling van n-grammen met lengten van één, twee en drie letters uit het Corpus Hedendaags Nederlands. Dit product is samengesteld op verzoek van een particulier bedrijf. Omdat het ook interessant kan zijn voor andere partijen is het opgenomen in de Taalmaterialencatalogus.
- Moroccorp. Dit is een corpus van communicatie via internet-chat tussen Marokkaans-Nederlandse taalgebruikers, bestaande uit tien miljoen woorden. De data hebben inmiddels een aantal updates gehad.
- RND Woordenlijsten. Dit zijn de fonetische transcripties van dialectwoorden verzameld in Nederland en België. Oorspronkelijk gepubliceerd in de "Reeks Nederlandse Dialectatlassen".
- GiGaNT-Molex. Dit lexicon bevat Nederlands materiaal uit Nederland, Vlaanderen, de Antillen en Suriname, afkomstig uit hedendaags corpusmateriaal van het Instituut voor de Nederlandse Taal.

### **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 (244), daarna België (75) en verder Duitsland (26), Canada (9), UK (7), Hongarije (4), Zweden, Italië en UK ieder 2.

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 (702), België (209), Duitsland (52), US (30), Bulgarije (20), UK (19), Canada (12), Zweden (11), China, Hongarije, Italië (6), Tsjechië, Frankrijk, Oostenrijk (5), Spanje (4), Ierland, Rusland (3) en een groot aantal lagere scores.

### **Downloads commercieel**

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

### **Downloads niet-commercieel**

Er zijn in totaal 1169 producten gedownload voor niet-commercieel gebruik.

### **Servicedesk**

De Servicedesk heeft dit jaar ongeveer 200 berichten ontvangen.

## **11. Taalportaal**

Voor het *Taalportaal* zijn in 2019 twee reguliere updates uitgevoerd, waarmee er opnieuw enkele honderden artikelen aan het grammaticaportaal zijn toegevoegd, de meeste voor het Afrikaans. Daarnaast is er onderhoud verricht aan de auteursomgeving en de achterkant van de webapplicatie. In het voorjaar van 2019 is een grote onderhoudsklus afgerond voor de literatuurdatabase. Omdat de auteurs in Zuid-Afrika (VivA) en in Nederland verschillende databases gebruiken, is een protocol ontwikkeld waarbij beide partijen in hun eigen database blijven werken, maar er bij de update van *taalportaal.org* een *merge* van beide databases plaatsvindt. Daarnaast is eind 2019 begonnen met de conversie van het laatste deel van de Syntax of Dutch, *Coördinatie* naar Taalportaal-XML. De werkzaamheden hiervoor worden naar verwachting in de loop van 2020 afgerond.

In 2019 is een visiedocument opgesteld voor de toekomst van Taalportaal en andere grote grammaticaonderdelen binnen het INT (ANS en Taaladvies.net). De kern van dit document was dat het INT vanaf 2020 niet langer alleen een ad hoc, faciliterende rol zal spelen bij grammaticaprojecten, maar zich structureel en inhoudelijk met het onderwerp grammatica zal bezighouden. De werkzaamheden hiervoor beginnen in de loop van 2020.

Voor het project *Taalportaal-Zuid-Afrika* is in 2019 de huidige Java-applicatie van Taalportaal door een extern ingehuurde programmeur omgebouwd naar een generieke applicatie. Deze applicatie is aan het begin van 2020 aangeboden aan Sadilar, de Zuid-Afrikaanse partner in het project, die met deze applicatie een eigen portaal zal bouwen voor twee Afrikaanse talen. Het was de bedoeling om in 2019 de zoekresultaten van de twee portalen aan elkaar te koppelen, maar dit is uitgesteld tot 2020 vanwege vertraging bij Sadilar.

## **12. Algemene Nederlandse Spraakkunst**

In 2019 is het project *Herziening eANS* afgerond. Het INT verzorgde hierin WP4, *Ergonomie en digitale omgeving*. Binnen het INT is de demosite die in 2018 is opgeleverd uitgebouwd tot een demo-webapplicatie die eind 2019 voor de auteurs toegankelijk is gemaakt. Daarnaast heeft er continue ondersteuning plaatsgevonden van de auteurs en hun werkomgeving. Een hoofdstuk dat in LaTeX was opgeleverd, is op het INT geconverteerd naar XML. De nieuwe website zal naar verwachting in de loop van 2020 worden gepubliceerd.

Naast het algemene werk aan de webapplicatie is in het najaar van 2019 met een door de Taalunie gehonoreerd subproject gewerkt aan een betere digitale ontsluiting. Met behulp van een stagiaire, Anouk Twilt, is een serie scripts geschreven waarmee de zoekmogelijkheden flink zijn verbeterd. Daarnaast zijn er koppelingen aangebracht naar externe kennisbronnen.

Zoals bij de bespreking van het Taalportaal al is vermeld, is in 2019 een visiedocument verschenen over de toekomst van grammatica op het INT. Een belangrijk deel van de werkzaamheden op het gebied van grammatica zal de verdere inhoudelijke herziening van de ANS zijn. Eind 2019 is hiervoor dr. Maaike Beliën aangenomen. Zij zal in maart 2020 met haar werk beginnen.

In november 2019 is ook het project *Algemene Nederlandse Klankleer* van start gegaan. Dit project, dat loopt tot het einde van 2020, is opgezet door prof. dr. Geert Booij en voorziet in een aanzienlijke uitbreiding van het onderdeel fonologie in de huidige ANS - een onderdeel dat geen deel uitmaakte van de lopende inhoudelijke herziening. In dit project zal de (Engelstalige) fonologie-informatie van

[www.taalportaal.org](http://www.taalportaal.org) als basis worden genomen en worden omgezet naar een Nederlandstalige versie. Daarbij zullen ook aanpassingen aan de diepgang worden gemaakt, aangezien het ANS-publiek breder is dan het Taalportaalpubliek. Het INT verzorgt binnen dit project de omzetting naar XML, databeheer, en de plaatsing op de vernieuwde ANS-applicatie. Naast prof. dr. Booij zullen in 2020 dr. Kathy Rys en Nelleke Jansen MA. aan het project verbonden zijn.

### **13. Onderwijs**

Traditioneel richtte het INT zich als toegepast wetenschappelijk instituut op onderzoekers en taalkundigen, maar door de huidige, bredere taakomschrijving heeft het zijn werkterrein het afgelopen jaar nadrukkelijk uitgebreid naar leerlingen, studenten en hun docenten. In dat verband is er contact gelegd met het Onderwijsnetwerk Zuid-Holland, dat het voortgezet onderwijs verbindt met hogescholen en universiteiten. Verder heeft het INT overleg gevoerd met de Nederlandse Taalunie over gerichte activiteiten die inspelen op de noden van docenten Nederlands in het buitenland (Nederlands als Vreemde Taal, NVT) en heeft het zijn medewerking verleend aan het testen van het online platform Mijn NVT.

Het afgelopen jaar zijn de taalmaterialen van het INT op verschillende wijzen onder de aandacht gebracht van het secundair en het tertiair onderwijs. Zo is er op de pagina Historisch Nederlands op onze website een tabblad Onderwijs gecreëerd; daar zijn onder andere lesplannen te vinden. Voor de rubriek Neerlandistiek voor de klas op *Neerlandistiek, online tijdschrift voor taal- en letterkundig onderzoek* zijn bijdragen geschreven waarin gewezen wordt op de mogelijkheden om taalmaterialen van het INT te gebruiken. Verder is er in samenwerking met de website [profielwerkstuktaalkunde.nl](http://profielwerkstuktaalkunde.nl) een toolbox over neologismen ontwikkeld. Daarnaast heeft het INT aan studenten Nederlands van de universiteit Leiden werkcolleges verzorgd over de online historische woordenboeken. Ook hebben leerlingen uit het voortgezet onderwijs en universitaire studenten informatie gekregen tijdens een bezoek aan het INT. Tot slot profileerde het INT zich in 2019 zowel met een informatiestand als met presentaties op de HSN-conferentie ('Het schoolvak Nederlands') en op het DRONGO talenfestival.

### **14. Externe communicatie & wetenschapscommunicatie**

In 2019 werd op het gebied van externe communicatie de huisstijl verder doorgevoerd, de bestaande communicatiemiddelen werden voortgezet en waar nodig verbeterd en er werden nieuwe populairwetenschappelijke activiteiten ontwikkeld. Het instituut trad in 2019 op als sponsor voor CLIN, de TABU-dag en de Digital Humanities Conferentie. Ook op het gebied van



wetenschapscommunicatie gingen we op dezelfde voet voort: in 2019 publiceerden we in populairwetenschappelijke tijdschriften en werkten we mee aan onderzoeken of artikelen van onlineplatforms of tijdschriften. Ook het begeleiden van scholieren en studenten bij het maken van profielwerkstukken, scripties en andere onderzoeken valt onder wetenschapscommunicatie, net als het verzorgen van optredens op de radio (bijvoorbeeld in De Taalstaat) en tv en het doen van interviews voor kranten, tijdschriften en blogs. Medewerkers van het INT houden regelmatig populairwetenschappelijke lezingen, vaak over neologismen, spelling, taalverandering, jongerentaal etc., bijvoorbeeld op taalcongressen, maar ook in het (middelbaar) onderwijs.

### **Huisstijl**

De huisstijl draagt de identiteit van onze organisatie uit, zorgt voor herkenning en onderscheidt ons van andere organisaties. Om ervoor te zorgen dat de huisstijl van het INT zo breed mogelijk uitgedragen wordt, is er een huisstijlhandleiding ontwikkeld, zodat alle uitingen van alle medewerkers zo veel mogelijk in lijn zijn met de huisstijl. Wegens het eerdere succes van de linnen tasjes zijn voor het DRONGO talenfestival opnieuw tasjes ontworpen in de huisstijl van het INT, dit keer in oranje. Sinds een paar jaar wordt er vanuit het INT een digitale nieuwjaarswens verstuurd naar relaties. Dit jaar is er voor het eerst een nieuwjaarskaart ontworpen die per post verstuurd werd, in combinatie met een nieuwjaarsgeschenk.

### **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. De website heeft in 2019 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 in aan de bezoekersaantallen op de website.

De populairwetenschappelijke webrubrieken zijn al jaren de grootste aandachttrekkers op de website, en dan met name *Woordbaak* (etymologie) en *Neologisme van de week* (nieuwe woorden). *Woordbaak* is geen wekelijkse rubriek meer, maar gezien de populariteit is het nuttig om de rubriek in stand te houden en af en toe een artikel te blijven plaatsen. Er is een opvallende stijging te zien in de interesse voor historisch Nederlands. Ondanks dat de rubriek *Terug in de taal* minder bekeken is dan in het voorgaande jaar, is de algemene informatie over de historische woordenboeken aanzienlijk meer bezocht. Dat kan verklaard worden doordat er in 2019 meer aandacht was voor historische projecten vanuit het INT, o.a. door het Rembrandtjaar.

Er is een verdubbeling in het aantal bezoeken van webpagina's over terminologie. Die categorie is in 2019 dan ook uitgebreid met o.a. termenlijsten en hulpmiddelen. Daarnaast is er een nieuwe populairwetenschappelijke rubriek toegevoegd: *Uitgelichte term*.

In 2019 zijn er twee nieuwe gelegenheidswoordenboekjes verschenen: *Het Groene Woordenboekje* (1.754 views) en *Woorden voor Rembrandt* (2.036 views). Het *Opvoedwoordenboekje* is in februari 2019 geüpdatet en is met 2.061 views in 2019 het vaakst bekeken van alle woordenboekjes.

Tot slot zijn er drie nieuwe themapagina's gelanceerd: *Historisch Nederlands*, *Hedendaags Nederlands* en *Terminologie*. Op die pagina's staat alle informatie die op de website over die onderwerpen te vinden is bij elkaar. Zo kan er in één oogopslag alles over het Nederlands van vroeger, over het Nederlands van nu of over vaktaal gevonden worden.

In 2020 wordt de website overgezet op een nieuw contentmanagementsysteem, van Joomla! naar Wordpress. Het ontwerp en de huisstijl blijven ongewijzigd, maar de website wordt gemoderniseerd en geoptimaliseerd. Ter voorbereiding is er in 2019 door een communicatiestagiair onderzoek naar het gebruik van de website gedaan, en op basis daarvan is een nieuwe menustructuur gemaakt die de ruggengraat vormt van de nieuwe website.

### **Nieuwsbrieven, mailings & persberichten**

De nieuwsbrieven (algemeen & terminologie) worden verstuurd om geïnteresseerden te blijven binden aan en te informeren over het instituut. In 2019 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.

Het totale abonneebestand van de nieuwsbrieven bestaat uit ongeveer 6.500 e-mailadressen, gelijk aan vorig jaar. Het abonneebestand van de algemene nieuwsbrief is in 2019 licht gedaald, maar de terminologienieuwsbrief heeft er juist iets meer abonnees bij gekregen. De cijfers voor het aantal opens en kliks zijn voor de algemene nieuwsbrief 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 iets onder het gemiddelde van die percentages.

Er zijn persberichten verstuurd over het 17<sup>e</sup>-eeuwse Nederlands van Rembrandt, het gelegenheidswoordenboekje *Woorden voor Rembrandt* en de verkiezing *Rembrandtwoord van het jaar*. Alle berichten hebben geleid tot aandacht in de media, zie voor de resultaten bijlage 3.

## **Sociale media**

Sinds 2010 heeft het instituut een Twitteraccount. Via Twitter worden links naar de website en taalweetjes/taalnieuws gedeeld, en contacten met volgers en partnerorganisaties onderhouden. Het aantal volgers groeit gestaag. In 2019 zijn er 309 nieuwe volgers bij gekomen.

Er bestond al een Facebookpagina voor de verkiezing *Weg met dat woord!* en in 2019 is er een algemene Facebookpagina van het instituut toegevoegd. Het was niet mogelijk de bestaande pagina om te vormen, dus er is een nieuwe gemaakt.

## **Populairwetenschappelijke activiteiten**

Tijdens de Week van het Nederlands in oktober organiseerden we een goedbezochte publieksmiddag (circa 80 bezoekers) rond het thema ‘groene taal’ in de Theaterzaal van het Leidse Volkshuis. Voor die gelegenheid werd er ook een gedrukte versie van *Het Groene Woordenboekje* gemaakt. Het andere gelegenheidswoordenboekje dat in 2019 werd gemaakt, *Woorden voor Rembrandt*, werd opgenomen als onderdeel van de publicatie *De Taal van Rembrandt* van het Taalmuseum.

Eind oktober stond het instituut opnieuw op het DRONGO talenfestival met een lab over medische taal in het *Algemeen Nederlands Woordenboek* en Taalradar, en dialectloket.be.

In november verscheen bij uitgeverij Scriptorum het populairwetenschappelijke boek *Klimatschieter, dwangzoen en kastelen in Spanje*. Het boek werd geschreven door Roland de Bonth en Dirk Geirnaert en is gebaseerd op de webrubriek *Terug in de taal*. Het boekje werd eind 2019 verstuurd als relatiegeschenk samen met de nieuwjaarskaart.

Eind 2019 is in het kader van het Rembrandtjaar de verkiezing *Rembrandtwoord van het jaar* georganiseerd. Op de website [rembrandtwoord.ivdnt.org](http://rembrandtwoord.ivdnt.org) konden deelnemers stemmen. Er werden 7.996 bezoekers geregistreerd en er werd 818 keer gestemd. Het winnende woord *zwadderen* werd onthuld door auteur Onno Blom in de Pieterskerk in Leiden, waar het tot eind januari 2020 in de vorm van een lichtinstallatie te zien was.

Tot slot beantwoorden we door het jaar heen veel vragen over het Nederlands aan studenten, journalisten en programmamakers. De gecombineerde communicatieactiviteiten in 2019 hebben geleid tot meer bezoekers op de website, meer volgers op de sociale media, nieuwe samenwerkingen, publicaties in vaktijdschriften en kranten, artikelen op goedbezochte nieuwswebsites en aandacht op radio en tv. Zie voor een uitgebreid overzicht de bijlage *Publicaties, lezingen, media etc.*

## 15. COST CA16105: European Network for Combining Language Learning with Crowdsourcing Techniques

De COST-actie *EnetCollect* beleefde in 2019 haar derde 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 zijn Peter Dekker (tot 1 oktober 2019) en Mathieu Fannee bij deze actie betrokken.

Van 23 tot en met 25 februari was het eerste Crowdfest van de actie in Brussel (België). Namens het INT was Peter Dekker hierbij aanwezig. Hij werkte mee aan de opzet van het project ‘Crowdsourcing corpus cleaning for language learning’. Dit project is in de loop van het jaar verder uitgewerkt en de te gebruiken crowdsourcemethode is met een aantal experimenten uitgetest. Hiervoor werd het door het INT ontwikkelde platform Taalradar ingezet.

Frieda Steurs, Tanneke Schoonheim en Peter Dekker woonden op 14 en 15 maart de bijeenkomst bij van Management Committee en werkgroepen in Lissabon (Portugal). Op 16 maart was er een extra dag gereserveerd om verder te werken aan het project ‘Crowdsourcing corpus cleaning for language learning’. Behalve Peter Dekker is inmiddels ook Tanneke Schoonheim hierbij betrokken. Dit project is gepresenteerd op EUROCALL 2019 in Louvain-la-Neuve (België) en tijdens eLex 2019 in Sintra (Portugal).

Op 5 en 6 december vond in Coimbra (Portugal) de workshop ‘Crowdsourcing examples’ plaats. Tanneke Schoonheim en Lut Colman waren hierbij aanwezig. Op 7 december was weer een extra dag gereserveerd om aan het project ‘Crowdsourcing corpus cleaning for language learning’ te werken. Tanneke Schoonheim was hierbij aanwezig. De IT-werkzaamheden van Peter Dekker met betrekking tot dit project zijn overgenomen door Mathieu Fannee.

Voor 2020 is er begin februari opnieuw een crowdfest in Coimbra (Portugal). Hierbij zijn deze keer geen deelnemers van het INT betrokken. Verder is er eind maart een bijeenkomst van het Management Committee in Napels (Italië). Frieda Steurs en Tanneke Schoonheim zullen hier namens het INT aanwezig zijn. Het project ‘Crowdsourcing corpus cleaning for language learning’ wordt gepresenteerd op EURALEX dat van 8 tot 12 september 2020 in Alexandroupolis (Griekenland) plaatsvindt.

## **16. COST CA18209: European network for Web-centered linguistic data science**

Het hoofddoel van deze nieuwe COST-actie is het bevorderen van synergieën in heel Europa tussen taalkundigen, computerwetenschappers, terminologen en andere belanghebbenden in de industrie en de maatschappij, op het gebied van ‘linguistic data science’. Taalkundige data heeft zijn eigen specifieke eigenschappen en vormt in de context van big data een nog grotendeels onontgonnen terrein. Het doel van deze nieuwe COST-actie is om hier verandering in te brengen.

De actie heeft een looptijd van 4 jaar en is eind oktober 2019 van start gegaan met een kick-offmeeting in Brussel. Carole Tiberius was namens het INT hierbij aanwezig.

## **17. European Infrastructure for Lexicography (ELEXIS)**

*ELEXIS* is een Europees Horizon 2020 project dat als doel heeft een duurzame infrastructuur voor e-lexicografie te creëren. Deze infrastructuur moet het mogelijk maken de kwalitatief hoogwaardige semantische informatie die momenteel nog veelal opgesloten zit in individuele lexicografische bronnen verspreid over Europa op grote schaal te koppelen, delen, verspreiden en op te slaan. Tevens zal deze infrastructuur bijdragen aan het verkleinen van de kloof tussen gemeenschappen met veel en weinig lexicografische expertise. *ELEXIS* komt voort uit de COST actie IS1305 European Network of eLexicography, die in oktober 2017 afliep. *ELEXIS* is een samenwerkingsverband tussen 17 partners uit Europa en Israël. Het project is in februari 2018 begonnen en heeft een looptijd van 4 jaar.

Het INT leidt het werkpakket “Lexicographic data and workflow”. Speerpunten van dit werkpakket zijn a) het maken van een inventarisatie van de behoeften van de lexicografische gemeenschap om zo het maken van woordenboeken optimaal te kunnen ondersteunen; b) het vastleggen en ondersteunen van gezamenlijke standaarden en werkwijzen voor het lexicografische proces en c) het ontwikkelen van methoden en tools voor de conversie, automatische segmentatie en identificatie van lexicografische inhoud.

In januari zijn de belangrijkste werkzaamheden met betrekking tot punt a afgerond met de publicatie van de deliverable *DL.1 Lexicographic Practices in Europe: A Survey of User Needs*.

Het zwaartepunt van de activiteiten in 2019 lag vervolgens op het maken van een inventarisatie van bestaande standaarden die gebruikt worden binnen de lexicografie en op het verder ontwikkelen van een gezamenlijk datamodel voor lexicografische materialen. Van deze werkzaamheden verschijnt

begin 2020 een tussentijdse rapportage (*Deliverable 1.2: Best practices for lexicography – intermediate report*).

Daarnaast was het INT in 2019 nauw betrokken bij de werkzaamheden in werkpakket WP6 “Integration, Evaluation and Sustainability” waar gewerkt is aan aanbevelingen voor copyright en juridische issues met betrekking tot lexicografische data. De resultaten hiervan worden begin 2020 gepubliceerd in deliverable *D6.2 Recommendations on legal and IPR issues for lexicography*. Verder heeft het INT in 2019 binnen werkpakket 4 WP4 “NLP for Lexicography” bijgedragen aan twee woordenboekeditors die binnen ELEXIS ontwikkeld worden, nl. *Lexonomy* en *Skema*. *Lexonomy* is een cloud-gebaseerd, open-source platform voor het maken en publiceren van woordenboeken. *Skema* is een specifieke editor, die speciaal bedoeld is om het bewerken van patronen en het koppelen van corpusconcordanties aan patronen te ondersteunen zodat woordbetekenis aan woordgebruik gekoppeld kan worden volgens de techniek van Corpus Pattern Analysis (ontwikkeld door Patrick Hanks). Binnen het INT wordt de *Skema*-editor gebruikt voor de bewerking van de patronen in het project *Woordcombinaties*.

In 2019 heeft het INT twee gastonderzoekers ontvangen met een ELEXIS trans-national access-beurs. Van 4-8 februari was Margarida Viegas Ramos (Centro de Linguística da Universidade NOVA de Lisboa – CLUNL, Portugal) op het INT om haar kennis van lexicografische en terminologische methoden te consolideren in het kader van haar PhD. Van 17-28 juni 2019 was Anna Tenieshvili (Batumi State Maritime Academy, Batumi, Georgia) gastonderzoeker op het INT om te werken aan een nieuw project voor het opzetten van een Engels-Georgisch maritiem woordenboek.

## **18. European Language Resources Coordination Initiative (ELRC)**

Het INT is betrokken bij het *ELRC*-initiatief. *ELRC* heeft als doel 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.

Carole Tiberius is samen met Jan Odijk (Universiteit Utrecht) Technical National Anchor Point voor Nederland voor het *ELRC*-project. In 2019 is voor het ELRC-initiatief ook een Public Sector National Anchor point benoemd, Piet van den Berg van de stichting RINIS.

Maandelijks wordt er door *ELRC* een online Q&A sessie gehouden om de voortgang van het project te bespreken. Daarnaast is er twee keer per jaar een Language Resource Board-bijeenkomst. Dit jaar vond de eerste LRB-bijeenkomst eind maart plaats in Amsterdam. Nederland stond tijdens deze bijeenkomst in de spotlights met drie presentaties, nl. *Country Profile* (door Jan Odijk en Carole Tiberius, een presentatie over *Language Services at the Ministry of Foreign Affairs* door Michel Verhagen van het ministerie van Buitenlandse Zaken en een over de stichting *RINIS* door Piet van den Berg). De tweede bijeenkomst vond eind november plaats in Helsinki. Deze bijeenkomst werd bijgewoond door Jan Odijk en Piet van den Berg.

Tijdens deze laatste bijeenkomst werd de *ELRC White Paper Sustainable Language Data Sharing to Support Language Equality in Multilingual Europe* gelanceerd. Dit rapport bevat de resultaten van een tweejarig onderzoek naar de bestaande infrastructures voor het delen van taalbronnen in Europa en geeft richtlijnen voor hoe het delen van taalbronnen kan worden verbeterd. In een annex wordt de situatie per land beschreven in een zgn country profile. Daarnaast heeft het INT in 2019 verder gewerkt aan het verzamelen van de taalbronnen voor het ELRC repository.

## **19. European Federation of National Institutes of Language (EFNIL)**

Het INT is sinds 2017 lid van de Europese Federatie van Nationale Taalinstellingen. Dit netwerk biedt een forum voor de taalinstututen en taalacademies van de verschillende Europese landen om kennis en ervaring uit te wisselen in verband met taal, taalmaterialen, taalvariatie en taalpolitiek. Frieda Steurs vertegenwoordigt het INT in dit netwerk. In 2019 werd de jaarlijkse vergadering van 9 tot 11 oktober in Tallin gehouden . Het congres werd georganiseerd door het Instituut voor de Estische Taal (Eesti Keele Instituut), de Estische Taalraad (Eesti Keelenõukogu), het Estische ministerie van Onderwijs en Onderzoek, het stadsbestuur van Tallinn, de organisatie voor de moedertaal (Emakeele Selts) en het Directoraat-generaal Vertaling (DGT) van de Europese Commissie. Het thema van het congres was **Language and Economy: Language Industries in a Multilingual Europe**. Frieda Steurs gaf er een lezing met als titel: **Language is Business! On the impact of linguistic applications in our society**.

## 20. NIAS-Lorentz-workshop “The Future of Academic Lexicography”



Van 4 tot 8 november organiseerde het INT, in samenwerking met de Universiteit Leiden ([LUCL](#)), de KU Leuven ([QLVL](#)) en [Ljubljana University/JSI](#), een vijfdaagse workshop over de toekomst van de academische lexicografie binnen het kader van het prestigieuze [NIAS-Lorentz-programma](#) dat interdisciplinaire samenwerking faciliteert en financieel ondersteunt. Een zestigtal internationale onderzoekers uit de lexicografie, linguïstiek, artificiële intelligentie, datawetenschap, softwareontwikkeling staken de koppen samen om tot een beter inzicht te komen over wat de belangrijkste uitdagingen zijn voor toekomst van de academische lexicografie en hoe die het best aangepakt kunnen worden.

Uitgangspunt daarbij was de lange traditie van academische woordenboekprojecten, zoals het WNT, ANW en hun tegenhangers in heel Europa, om de woordenschat zo nauwkeurig mogelijk te beschrijven op basis van de wetenschappelijke analyse van grote hoeveelheden geobserveerd taalgebruik. Hoewel dit type *evidentie-gebaseerde* lexicografie in de laatste decennia, net als andere wetenschappen, een digitale revolutie onderging, vertrok de workshop van de vaststelling dat de positie van de academische lexicografie in de huidige digitale kennismaatschappij en kenniseconomie op drie vlakken uitgedaagd wordt: (1) Het bedrijfs- en onderzoeksmodel van lexicografische instituten is onzeker zowel in een academisch omgeving met kortlopend projectgebaseerd onderzoek als in een online economie die gekenmerkt wordt door de concentratie van data en content bij de grote technologiebedrijven. (2) Lexicografische analysemethodes en de lexicografische inhoudsproductie hebben beide een probleem van schaalbaarheid in de context van exponentieel toenemende hoeveelheden taaldata en te beschrijven taalfenomenen. (3) Lexicografische databanken zijn onvoldoende aanpasbaar aan de noden van zowel een steeds diverser wordende groep eindgebruikers als van ontwikkelaars van taalapplicaties. De uitdagingen op elk van deze drie terreinen werden tijdens de workshop verder uitgespit aan de hand van presentaties door experts en groepsdiscussies. Interdisciplinaire werkgroepen brachten daarna mogelijke oplossingen in kaart, zoals het gebruik van technologieën uit de artificiële intelligentie, data-analyse, en crowd sourcing, Ook het effect op de lexicografische praktijk en de noodzakelijke samenwerkingsverbanden tussen lexicografen en computerwetenschappers werden onderzocht en opgelijst. Als hoofdorganisator zal het INT de inzichten en aanbevelingen uit de workshop tegen half februari 2020 samenvatten in een *white paper*. Na feedback van alle workshopdeelnemers en andere stakeholders uit de lexicografische veld zullen de workshoporganisatoren de white paper als strategisch planningsdocument gebruiken om verdere onderzoekssamenwerkingen op te zetten. Tijdens de workshop werden alvast concrete afspraken



gemaakt tussen verschillende lexicografische instituten om op Europees en internationaal vlak nog intenser samen te werken om een mooie toekomst voor de academische lexicografie te verzekeren.

## 21. CLARIAH en CLARIAH-PLUS

In 2015 ging het NWO-project CLARIAH van start: Common Lab Research Infrastructure for the Arts and Humanities. In 2018 is het INT-aandeel in dit project reeds grotendeels afgerond; in 2019 is aan een aantal onderdelen de laatste hand gelegd.

Binnen werkpakket 2 (infrastructuur) is het project in het kader van het project DiaMaNT (diachroon semantisch lexicon) een data-update gerealiseerd.

Binnen werkpakket 3 zijn de onderdelen *federated search* en *chaining search* in het eerste kwartaal van 2019 afgerond. Het eindresultaat is dat corpora, treebanks en lexica op een uniforme manier en in combinatie doorzoekbaar zijn in de CLARIAH-infrastructuur door middel van een aantal applicaties en services (zie hieronder). We noemen verbeteringen in de doorzoekbaarheid van Nederlandse corpora dankzij nieuwe versies van het op het INT ontwikkelde corpuszoekstelsel, de ontsluiting van corpora in de CLARIN federated content search, de publicatie van belangrijke lexica voor het Nederlands als linked open data, de ontwikkeling van een gefedereerde zoekinterface voor treebanks en de mogelijkheid zoekresultaten uit diverse databronnen te combineren met behulp van chaining search.

In het kader het CLARIAH-project ATHENA is gewerkt aan een extensie van de Athenaplantenthesaurus als bijdrage aan het Athena-eenplatform voor het in samenhang doorzoeken en analyseren van drie typen digitaal biodiversiteitserfgoed (tekstuele bronnen, gestructureerde data en audiovisueel materiaal). De Athenaplantenthesaurus is door koppeling met het Nederlands Soortenregister, het DiaMaNT-lexicon, de DSDD-database en de Databank Plantennamen in de Nederlandse Dialecten (PLAND) uitgebreid met o.a. dialect- en spellingsvarianten, synoniemen en dateringsinformatie. De dataset zal geïntegreerd worden in de Athenawebsite.

### CLARIAH-PLUS

In 2018 is voorts het vervolgproject CLARIAH-PLUS gehonoreerd, waarin infrastructuur wordt toegevoegd voor de disciplines die teksten niet op de gebruikte taal, maar op de inhoud bestuderen, zoals letterkunde, geschiedenis, maar ook filosofie en theologie. Het INT werkt hier aan de doorzoekbaarheid van taalmateriaal (met name treebanks en parallelle corpora, alsmede de mogelijkheid voor onderzoekers om zoekresultaten te annoteren) en op twee manieren bijdragen aan

de ontwikkeling van (verrijkt) tekstmateriaal voor onderzoek: door de ontwikkeling van een infrastructuur voor het taalkundig verrijken van historisch tekstmateriaal en het opzetten van een digitalisatieworkflow.

In 2019 is vooral aandacht besteed aan het onderdeel “verbeterde infrastructuur voor historisch Nederlands”. Er is gewerkt aan de principes van de verrijking en een tagset die enerzijds recht doet aan de eigenheid van het historisch Nederlands en anderzijds zoveel mogelijk compatibel is met reeds gebruikte tagsets op dit gebied, en aan de harmonisatie van de bij het INT onderhouden historische corpora.

#### **Online webapplicaties en services:**

- [Autosearch applicatie](#) waarmee onderzoekers hun eigen corpus doorzoekbaar kunnen maken
- [OpenSoNaR-CGN applicatie](#)
- [CLARIN Web service voor federated content search](#)
- [“Gefedereerde” instantie van Gretel](#)
- [Sparql endpoint voor lexica](#)
- [User interface voor gecombineerd zoeken in lexica](#)

#### **Open source repositories:**

- [BlackLab corpuszoekmachine en bijbehorende user interface](#)
- [Broncodevoor token-based federated content search](#)
- [Gretel](#), zoekinterface voor treebanks, bevat ook de INT-code voor het gecombineerd zoeken in Gretel en PaQu-gebaseerde treebanks
- [Extensie om PaQu toegankelijk te maken voor het gecombineerde zoeken in treebanks](#)
- [Gecombineerd zoeken in lexica, user interface](#)
- Implementatie van [chaining search](#) als Jupyter notebook

## **22. CLARIN ERIC**

Op 1 april 2019 stopte het lidmaatschap van de Nederlandse Taalunie bij CLARIN ERIC, waardoor Vlaanderen *de facto* geen lid meer is van de CLARIN ERIC en het INT niet meer als vertegenwoordiger aanwezig is in de National Coordinators Forum en andere CLARIN-organen. Niettemin werd toegang tot de CLARIN-infrastructuur voor onderzoekers uit België voorlopig gevrijwaard, omdat ervan uitgegaan wordt dat deze situatie tijdelijk is en blijven de praktische gevolgen van deze Vlexit tot een minimum beperkt.

Er werd met FWO (Fonds voor Wetenschappelijk Onderzoek Vlaanderen) en EWI (Departement Economie, Wetenschap en Innovatie van de Vlaamse regering) overlegd hoe en wat er mogelijk is om CLARIN België op te richten. Uit deze gesprekken bleek dat dit kan zodra er een Vlaams of Franstalig IRI-project (International Research Infrastructure) goedgekeurd is waardoor BELSPO (Programmatorische Federale Overheidsdienst Wetenschapsbeleid, België) CLARIN België opricht en Vlaanderen op die manier opnieuw lid kan worden van CLARIN ERIC.

Het INT zorgde er mee voor dat een aantal vergaderingen belegd werden zodat het CLARIAH-VL-consortium (samensmelting van CLARIN-Vlaanderen en DARIAH-Vlaanderen) goed voorbereid is op de nieuwe IRI-call van het FWO en EWI, om zo de kansen tot een toetreding van België tot CLARIN ERIC te verhogen.

Begin oktober vond de jaarlijkse CLARIN-conferentie plaats in Leipzig. Het INT maakte deel uit van de programmacommissie en was daar aanwezig met presentaties van Peter Dekker en Vincent Vandeghinste. Deze conferentie bood ook de gelegenheid om, nu het INT niet meer vertegenwoordigd is in de CLARIN-gremia, toch te overleggen met de mensen van CLARIN ERIC omtrent de gevolgen van de eerder vermelde vlexit.

Op 3 december 2019 werd er een seminarienamiddag georganiseerd, in samenwerking met Barbara De Cock (UCL) voor onderzoekers aan de UCL (Louvain-la-Neuve) en bij uitbreiding alle doctorandi in Franstalig België waarin CLARIN geïntroduceerd werd aan de hand van drie lezingen. In een eerste lezing introduceerde Franciska De Jongh (CLARIN ERIC-directeur) CLARIN ERIC, in een tweede lezing gaf Bob Boelhouwer toelichting bij de GDPR en hoe daarmee om te gaan bij dataverzameling en verspreiding, en in een derde lezing gaf Vincent Vandeghinste een demonstratie en hands-on sessie van een aantal verschillende CLARIN-toepassingen.

## **23. Digitization.eu**

Het INT is voorzitter van het IMPACT Centre of Competence ([www.digitisation.eu](http://www.digitisation.eu)). Dit is een non-profitorganisatie bestaande uit publieke en commerciële organisaties met als doel de digitalisering van historisch materiaal “beter, sneller, en goedkoper” te maken. Het centrum voorziet in data, tools, services en expertise op het gebied van document imaging, taaltechnologie en het verwerken van historisch tekstmateriaal.

Samen met het INT, CLARIN-VL, Dariah-BE heeft het Centre van 8-10 mei het derde Datech-congres georganiseerd in Brussel. Het IMPACT Centre of Competence is in 2019 CLARIN Knowledge centre

geworden. In september 2019 is een COST-actionaanvraag 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. Indien deze gehonoreerd wordt, zal het INT deelnemen aan het project.

## 24. Infrastructuur

Door het uitbreiden van de mankracht bij systeembeheer lukte het in 2019 om uitgesteld en achterstallig onderhoud van de infrastructuur weg te werken. Uitgestelde operationele upgrades zijn afgerond en i.s.m. ISSC is er de intentie uitgesproken om het netwerk verder in de ISSC-infra te integreren vooral om laag-niveau netwerkbeveiliging beter te kunnen overzien en beheersen.

Nu de druk van de operationele upgrades weg is zal er meer aandacht besteed worden aan planmatige verbetering van de infrastructuur. Zo is er in de organisatie een behoefte geformuleerd aan een nieuwe dataplan om de immer groeiende dataset van het INT beter beheersbaar te maken. Aspecten die hierbij aan bod komen zijn robuustheid (“data resilience”), organisatie en levenscyclus. Verder zal er gekeken worden hoe beter om te gaan met applicaties die gebruikmaken van verouderde software en een revaluatie van de functionaliteit van de aangeboden services eventuele overlapping te identificeren.

## 25. IT-afdeling: projecten en organisatie

### Projecten

Voor *CLARIAH Chaining Search* werd een API ontwikkeld die het voor wat technisch gebruikers mogelijk maakt om via de browser de resultaten van verschillende bronnen te combineren. Via de *CLARIN Federated Content Search (FCS) Lexica* werd het lexicon *DIAMANT* beschikbaar gesteld, en daarnaast werd ook een speciaal op dit project toegespitste user interface ontwikkeld. Voor het onderdeel *FCS Treebanks* werd de applicatie *Gretel* aangepast om ook zoekresultaten uit het andere treebank-zoekprogramma *PaQu* te kunnen verzamelen en gezamenlijk te presenteren.

Het overnemen van de terminologiewebsite van de Taalunie is dit jaar afgerond. *Termtreffer* en *Termbeheerder*, producten die de Taalunie aan het INT heeft overgedragen, zijn kritisch bekeken, gemoderniseerd en verbeterd. Aan *Termtreffer* bleek helaas meer werk nodig dan gedacht; hierover wordt verder overlegd met de Taalunie. *Termbeheerder* is met nog enig werk online te zetten; naar verwachting lukt dit in 2020.

Op verzoek van de Taalunie en SADiLaR is de software voor grammaticawebsite *Taalportaal* aangepast zodat deze door hen kan worden gebruikt voor een nieuwe site over Zuid-Afrikaanse talen isiZulu, isiXhosa en Setswana. Aan *Vertaalwoordenschat*, onze site voor vertaalwoordenboeken van en naar het Nederlands, is de taal Estisch toegevoegd. Zowel de website als de mobiele app zijn bijgewerkt.

Onze corpusapplicatie is sterk verbeterd, ten behoeve van de in 2020 online te brengen nieuwe release van onze historische corpora. Het product *AutoSearch*, waarin gebruikers zelf corpora kunnen uploaden, accepteert nu ook diverse documenttypes zoals Word, PDF, etc. Hiervoor werd ook een aantal verbeteringen doorgevoerd aan de zoeksoftware *BlackLab*, waarvan versie 2.0 werd uitgebracht. De bij ons gehoste applicatie voor optische tekstherkenning (OCR), *PICCL*, werd gekoppeld aan *AutoSearch*, zodat het mogelijk werd om documenten uit *PICCL* direct door te sturen naar *AutoSearch* om ze doorzoekbaar te maken.

De bewerkingsomgevingen voor de projecten *Woordcombinaties* en *Neologismen* werden verbeterd, en voor beide projecten werd ook een demonstratieversie van de zoekwebsite ontwikkeld. Zo'n website werd ook opgezet voor het *Lexicon Frisicum*, een historisch Fries-Latijn woordenboek.

Er werden diverse verbeteringen doorgevoerd aan de databases voor het historisch (*HiLex*) en modern (*MoLex*) lexicon en de bijbehorende webservices. Er is hard verder gewerkt aan het koppelen van *MoLex* aan diverse projecten. In 2020 moeten het *Algemeen Nederlands Woordenboek* en de projecten *Neologismen*, *Vertaalwoordenschat* en *Woordcombinaties* definitief aan *MoLex* gekoppeld worden. Hierdoor wordt overal eenduidige spellinginformatie gegeven en wordt dubbel werk voorkomen.

Onze veelzijdige tool voor interne databewerkingsklussen, *Lex'it*, werd verder verbeterd en voor meerdere interne bewerkingstaken ingezet. Er is verder gewerkt aan de taalkundige verrijking van historisch materiaal met behulp van machine learning. Dit werk zal in 2020 worden voortgezet.

Werk is begonnen aan een applicatie voor de dialectwoordenboeken van de Vlaamse dialecten (*DSDD*). Hierin kunnen o.a. geografische kaarten getoond worden met symbolen erop.

## **Organisatie**

Begin 2019 werd een dataconversiespecialist aangenomen, voorlopig voor 2 dagen in de week. Hierdoor maakten we sneller voortgang met diverse conversieklussen.

Tegen het eind van het jaar koos een van onze ontwikkelaars ervoor om te gaan promoveren, waardoor wij hem helaas kwijt raakten. In de laatste maanden is een nieuwe ontwikkelaar gevonden die begin 2020 aan de slag zal gaan.

In 2019 lag er meer werk dan onze ontwikkelaars uit konden voeren. Daarom is twee keer gekozen voor tijdelijke inhuur: eenmaal voor de projecten *Termtreffer* en *Termbeheerder*, en eenmaal voor het project *Taalportaal Zuid-Afrika*. Aan beide projecten is goed werk verricht, maar zulke kortdurende inhuur van ontwikkelaars heeft ook nadelen; het is relatief duur en tijdrovend. Daarom zal gezocht worden naar een extra ontwikkelaar, voorlopig voor een periode van een jaar, om te voorkomen dat er weer tijdelijk moet worden ingehuurd.

## **Technologie**

Steeds meer projecten hebben een vergelijkbaar soort onlinezoekapplicatie nodig. Er wordt gewerkt aan een generieke basis hiervoor, zodat applicaties sneller op te zetten zijn.

Traditioneel wordt er op het INT veel Java-code geschreven, maar er wordt nu ook steeds meer gewerkt met de talen Python en TypeScript. Doel is om voor elke taak de geschiktste taal te kiezen zonder dat er een wildgroei ontstaat.

Het gebruik van Vue.js waarmee in 2018 begonnen is, blijkt een productieve manier om websites en -applicaties te ontwikkelen, dus de ervaring hiermee is verder uitgebreid. Daarnaast wordt meer gebruik gemaakt van zoektechnologie Solr om diverse datasets snel en flexibel te kunnen doorzoeken.

Voor het schrijven van woordenboeken wordt nu vaak een (wat verouderende) editor gebruikt. Dit jaar werden de mogelijkheden van Lexonomy onderzocht, een web-gebaseerde, open source woordenboekeditor. Komend jaar zal in het kader van het project *ELEXIS* onderzocht worden of deze geschikter gemaakt kan worden voor sommige van onze projecten.

Er komen voortdurend nieuwe technologieën beschikbaar die de efficiëntie kunnen verhogen en de kwaliteit kunnen verbeteren, maar het komt er niet altijd van om ze te evalueren. Komend jaar zal daarom meer aandacht besteed worden aan (zelf)studie, zodat ontwikkelaars op de hoogte blijven van en kunnen experimenteren met nieuwe technologieën. Daarvoor zal o.a. een maandelijkse studiedag worden ingesteld.

## **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. Antal van den Bosch (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

Dr. Jan Theo Bakker (tot 1 december 2019) – systeemarchitect

Petra Belt (tot 1 september 2019) – administrateur

Marjolijn van Bennekom – taalkundig assistent

Dr. Bob Boelhouwer – computerlinguïst

Dr. Roland de Bonth (per 1 februari 2019) – onderzoeker/taalkundige

Lic. Lut Colman – onderzoeker/taalkundige

Iulianna Ciudin MA – onderzoeker/taalkundige

Peter Dekker MSc (tot 1 november 2019) – systeemontwikkelaar

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 (per 1 december 2019) – 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 (per 1 juni 2019) – systeemontwikkelaar

Wil de Ruyter – taalkundig assistent

Dr. Kathy Rys (per 1 november 2019) – onderzoeker/taalkundige

Dr. Tanneke Schoonheim – onderzoeker/taalkundige

Rob van Strien – programmeur/ontwikkelaar

Paulette Tacx – managementassistent

Drs. Rob Tempelaars – onderzoeker/taalkundige

Dr. Carole Tiberius – computerlinguïst

Lic. Katrien Van pellicom – projectleider Spelling

Dr. Vincent Vandeghinste – onderzoeker/taalkundige

Drs. Boukje Verheij – onderzoeker/taalkundige



Rayvano van Vliet (per 1 januari 2019) – systeembeheerder

Drs. Vivien Waszink – onderzoeker/taalkundige

Karin van Weerlee – managementassistent

Dr. Hans Westgeest (tot 1 juli 2019) – onderzoeker/taalkundige

Laetitia de Winter (per 1 oktober 2019) – administratief medewerker

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

### Publicaties

- Bonth, Roland de (2019). 'Uitslag van de Worm-en-donder-kerstquiz'. Op: [www.neerlandistiek.nl](http://www.neerlandistiek.nl)
- Bonth, Roland de (2019). 'Ik kunnen niet, meester!'. Op: [www.neerlandistiek.nl](http://www.neerlandistiek.nl)
- Bonth, Roland de (2019). "'Ionck bestaen, oud ghedaen'". Op: [www.neerlandistiek.nl](http://www.neerlandistiek.nl)
- Bonth, Roland de (2019). 'De witte wijn leppende elite'. Op: [www.neerlandistiek.nl](http://www.neerlandistiek.nl)
- Bonth, Roland de (2019). 'Het 17<sup>e</sup>-eeuwse Nederlands van Rembrandt'. Op: [www.neerlandistiek.nl](http://www.neerlandistiek.nl)
- Bonth, Roland de (2019). 'Ten enenmale te allen tijde'. Op: [www.neerlandistiek.nl](http://www.neerlandistiek.nl)
- Bonth, Roland de (2019). 'Zoekt (met reguliere expressies), en gij zult (meer) vinden'. Op: [www.neerlandistiek.nl](http://www.neerlandistiek.nl)
- Bonth, Roland de (2019). 'Max Havelaar, of de meningen van Nederlandse scholieren'. Op: [www.neerlandistiek.nl](http://www.neerlandistiek.nl)
- Bonth, Roland de (2019). 'Oefenen in stijl'. Op: [www.neerlandistiek.nl](http://www.neerlandistiek.nl)
- Bonth, Roland de (2019). 'Van test naar tekst'. Op: [www.neerlandistiek.nl](http://www.neerlandistiek.nl)
- Bonth, Roland de (2019). 'Historischewoordenboekquiz'. Op: [www.neerlandistiek.nl](http://www.neerlandistiek.nl), 27 juli 2019.
- Bonth, Roland de (2019). 'Uitslag van en uitleg bij de Historischewoordenboekenquiz'. Op: [www.neerlandistiek.nl](http://www.neerlandistiek.nl)
- Bonth, Roland de (2019). "'Dit was het (taal)nieuws'". Nieuwsbrieven voor het schoolvak Nederlands'. Op: [www.neerlandistiek.nl](http://www.neerlandistiek.nl)
- Bonth, Roland de (2019). 'Inleiding bij De Rembrandt Tutorials'. Op: [www.neerlandistiek.nl](http://www.neerlandistiek.nl)
- Bonth, Roland de & Dirk Geirnaert (2019). 'De Rembrandt Tutorials. Volledige uitgave, naar beste vermogen bezorgd'. Op: [www.neerlandistiek.nl](http://www.neerlandistiek.nl)
- Bonth, Roland de (2019). 'Wtenbogaert. Kunt u dat spellen?'. In *AdRem, Remonstrants maandblad*. Oktober 2019, p. 16-17.
- Bonth, Roland de (2019). 'Podcast, potel en prosument'. Op: [www.neerlandistiek.nl](http://www.neerlandistiek.nl)
- Bonth, Roland de (2019). Bijdragen over 'konterfeiter', 'boekenbeer', 'verschillig', 'verknecht', 'gierwolf', 'laaf', 'overvlieger', 'vrouwentuin', 'abeet', 'orewoet', 'genan', 'Sinte Lors', 'warmbier'. In: [Terug in de taal](http://www.neerlandistiek.nl)
- Bosque-Gil, Julia, Jorge Gracia, John McCrae, Philipp Cimiano, Sander Stolk, Fahad Khan, Katrien Depuydt, Jesse de Does, Francesca Frontini and Ilan Kernerman. *The OntoLex Lemon Lexicography Module*. Final Community Group Report 17 September 2019. Editors: Julia Bosque-Gil & Jorge Gracia. <https://www.w3.org/2019/09/lexicog/>
- Chambers, Sally, Wout Dillen, Tom Gheldof, Wouter Ryckbosch, Vincent Vandeghinste, Christophe Verbruggen (2019). From DARIAH to CLARIAH in Flanders. Challenges and Opportunities

- for the Development of an Open Humanities Service Infrastructure. DH Benelux 2019. Liege, Belgium.
- Coudenys, W. (Eds.), Coudenys, W. (contr.), Steurs, Frieda (contr.), Van Hecke, A. (contr.), Behiels, L. (contr.), Thomas, W. (contr.), Salaets, H. (contr.), Balogh, K. (contr.), Lanslots, I. (contr.), Peters, E. (contr.), Daniëls, H. (contr.), Vermeerbergen, M. (contr.), Brône, G. (contr.), Noreillie, A-S. (contr.), Puimège, E. (contr.), Huibrechts, M. (contr.) (2020). *Taal en de wereld*. Leuven: Leuven University Press. ISBN: 9789462702134.
- Dekker, Peter, Laura van Eerten (2019). *Ga lekker zitten*. Op: *Neerlandistiek.nl*.
- Dekker, Peter, Mathieu Fannee, Jesse de Does. (2019), *CLARIAH Chaining Search: A Platform for Combined Exploitation for Multiple Linguistic Resources*, CLARIN Annual Conference 2019, Leipzig, Germany.
- Depuydt, Katrien en Hennie Brugman, “Turning Digitised Material into a Diachronic Corpus: Metadata Challenges in the Nederlab Project.” In: *DATECH2019: Proceedings of the 3rd International Conference on Digital Access to Textual Cultural Heritage*. Association for Computing Machinery, New York 2019, p. 169-173.
- Eerten, Laura van (2019). *Het Groene Woordenboekje*. Op: [www.ivdnt.org](http://www.ivdnt.org).
- Eerten, Laura van (2019). Biogondiër, ecorexia, verticaal bos. De jongste klimaatwoorden. In: *Onze Taal* 88, 10, p. 16-17.
- Geirnaert, Dirk & Roland de Bonth (2019). ‘Woorden voor Rembrandt. Gelegenheidswoordenboekje 17<sup>e</sup>-eeuwse schilderstermen’. Op: [www.ivdnt.org](http://www.ivdnt.org) en [www.neerlandistiek.nl](http://www.neerlandistiek.nl)
- Geirnaert, Dirk (2019). Diverse bijdragen in de webrubriek *Terug in de taal* op de website van het INT.
- Geirnaert Dirk, Roland de Bonth (2019). ‘Drakenbloed, blozeren en zwaddereren. Zeventiende-eeuwse schilderstermen’. In: T. Askamp (red.), *De taal van Rembrandt. De kunst van het overtuigen*, Taalmuseum, Leiden 2019, pp. 22-41.
- Geirnaert, Dirk, Roland de Bonth, m.m.v. Katrien Depuydt (2019). *Klimaatschieter, dwangzoen en kastelen in Spanje – Woorden die we wisten*. Scriptum, Schiedam.
- Geirnaert, Dirk, Robrecht Lievens (2018). Van zondeval tot hemel: staties uit de heilsgeschiedenis. Een reeks geïllustreerde devotionele strofen van Anthonis de Roovere (Gouda, 1482). In: *Jaarboek ‘De Fonteyne’ LXIII*, pp. 7-253.
- Geirnaert, Dirk (2018). One hundred and eleven Drawings identified as the Work of Marcus Gheeraerts. Op: [www.leidenartsinsocietyblog.nl](http://www.leidenartsinsocietyblog.nl).
- Hofmeester, Karin, Ashkan Ashkpour, Katrien Depuydt en Jesse de Does, ‘Diamonds in Borneo: Commodities as Concepts in Context’ In: *DATECH2019: Proceedings of the 3rd International Conference on Digital Access to Textual Cultural Heritage*. Association for Computing Machinery, New York 2019, p. 45-50. <https://doi.org/10.1145/3322905.3322924>

- Kallas, J., Koeva, S., Langemets, M., Tiberius, C & Kosem, I. 2019. *Lexicographic Practices in Europe: Results of the ELEXIS Survey on User Needs*. In: Kosem, I., Zingano Kuhn, T., Correia, M., Ferreria, J. P., Jansen, M., Pereira, I., Kallas, J., Jakubiček, M., Krek, S. & Tiberius, C. (eds.) 2019. *Electronic lexicography in the 21st century. Proceedings of the eLex 2019 conference*. 1-3 October 2019, Sintra, Portugal. Brno: Lexical Computing CZ, s.r.o. 519-536.
- Kosem, I., Zingano Kuhn, T., Correia, M., Ferreria, J. P., Jansen, M., Pereira, I., Kallas, J., Jakubiček, M., Krek, S. & Tiberius, C. (eds.) 2019. *Electronic lexicography in the 21st century. Proceedings of the eLex 2019 conference*. 1-3 October 2019, Sintra, Portugal. Brno: Lexical Computing CZ, s.r.o.
- McCrae, J.P., Tiberius, C., Khan, A.F., Kernerman, I., Declerck, T., Krek, S., Monachini, M., Ahmadi, S. (2019). *The ELEXIS Interface for Interoperable Lexical Resources*. In: Kosem, I., Zingano Kuhn, T., Correia, M., Ferreria, J. P., Jansen, M., Pereira, I., Kallas, J., Jakubiček, M., Krek, S. & Tiberius, C. (eds.) 2019. *Electronic lexicography in the 21st century. Proceedings of the eLex 2019 conference*. 1-3 October 2019, Sintra, Portugal. Brno: Lexical Computing CZ, s.r.o. 642-659.
- Parlevliet, D. & M. Fannee (red.) (2019). *Een ambacht aan de Rijnmond. Valkenburg en Katwijk in de middeleeuwen*. Stichting Historische Publicaties Holland-Rijnland.
- Schoonheim, Tanneke (te versch.). *Words crossing Borders*. Te verschijnen in *Lexicographica*, International Annual for Lexicography.
- Schoonheim, Tanneke (2019). *Oudfries en Oudnederlands, twee verwante talen in één woordenboek*. Trefwoord 2019.
- Steurs, Frieda (2019). Technologie en integriteit. In: *Technologie, ethiek en beleid*, (81-92). Brussel: Politeia nv. ISBN: 978-2-509-03500-4. [Open Access](#)
- Steurs, Frieda (2019). De juridische toren van Babel - de (on)vertaalbaarheid van juridische teksten. In: C. Koster (Eds.), *Teksten in beweging*, (98-103). Nijmegen: Vantilt. ISBN: 9789460044441. [Open Access](#)
- Steurs, Frieda. 'Taal is Business! Over globalisering en toepassingen in de taalkunde' In: *Neerlandica Wratislaviensa XXVIII. Op reis!*. Wroclaw, Wydawnictwo. Uniwerytetu Wroclawskiego
- Tempelaars, Rob en Vivien Waszink, Rubriek 'Neologisme van de week' op de website van het INT. <http://www.ivdnt.org/>
- Tempelaars, Rob e.a. Rubriek 'Vraagbaak' op de website van het INT. <http://www.ivdnt.org/>
- Tempelaars, Rob en Vivien Waszink (2019). 'Heeft de academische lexicografie toekomst? Woordenboekmakers en andere experts bijeen in Leiden' In: *Neerlandia* 4 (2019), pp. 36-37.

- Vandeghinste, Vincent, Vanallemeersch, T., Augustinus, L., Bulte, B., Van Eynde, F., Pelemans, J., Verwimp, L., Wambacq, P., Heyman, G., Moens, M-F., Van der Lek-Ciudin, I., Steurs, F., Terryn, A.R., Lefever, E., Tezcan, A., Macken, L., Hoste, V., Daems, J., Buyschaert, J., Coppers, S., Van den Bergh, J., Luyten, K. (2019). Improving the Translation Environment for Professional Translators. *INFORMATICS-BASEL*, 6 (2), Art.No. ARTN 24. [doi: 10.3390/informatics6020024](https://doi.org/10.3390/informatics6020024) [Open Access](#)
- Vandeghinste, Vincent, Bram Bulté, Liesbeth Augustinus (2019). Wablieft: An Easy-to-Read Newspaper corpus for Dutch. CLARIN Annual Conference, Leipzig. p. 188-191.
- Vandeghinste, Vincent, Bram Bulté (2019). Linguistic Proxies of Readability: Comparing Easy-to-Read and regular newspaper Dutch. *Computational Linguistics in the Netherlands Journal*, vol. 9. p. 81-100.
- Vandeghinste, Vincent, Tom Vanallemeersch, Liesbeth Augustinus, Bram Bulté, Frank Van Eynde, Joris Pelemans, Lyan Verwimp, Patrick Wambacq, Geert Heyman, Marie-France Moens, Iulianna van der Lek-Ciudin, Frieda Steurs, Ayla Rigouts Terryn, Els Lefever, Arda Tezcan, Lieve Macken, Véronique Hoste, Joke Daems, Joost Buyschaert, Sven Coppers, Jan Van den Bergh, Kris Luyten (2019). [Improving the Translation Environment for Professional Translators.](#) *Informatics*. 6(2), 24. *Special Issue Advances in Computer-Aided Translation Technology.* MDPI.
- Waszink, Vivien (2019). Neologisms in an online portal. Op: [globalex.link](http://globalex.link)
- Waszink, Vivien (2019). Aquí met that ass, Spaanse invloeden in de nederhop. In: *Onze Taal* 6 (2019), p. 22-23
- Waszink, Vivien (2019). Smibanese: taal uit de Bijlmer. In: *Onze Taal* 11 (2019), p. 8.
- Waszink, Vivien (2019). Neologisms in an Online Portal: The Dutch Neologismenwoordenboek (NW). Geaccepteerd artikel, te verschijnen in *Dictionaries* (2020).

### **Lezingen en presentaties**

- Augustinus, Liesbeth, Vincent Vandeghinste (2019). Improving GrE TEL's English. *Computational Linguistics in the Netherlands (CLIN)* 29, Groningen
- Boelhouwer Bob (2019). GDPR -- General Data Protection Regulation. *Seminar: "Easy use of corpora and language tools. An introduction to CLARIN, a European infrastructure"* Louvain-la-Neuve.
- Bonth, Roland de. *Duurzaam Nederlands: tips om zelf een steentje bij te dragen*. Lezing tijdens de middag 'Groene Taal' in de Week van het Nederlands. Leiden, 8 oktober 2019.
- Dekker, Peter, Tanara Zingano Kuhn, Branislava Šandrih, Rina Zviel-Girshin, Špela Arhar Holdt, Tanneke Schoonheim (2019). *Corpus filtering via crowdsourcing for developing a learner's dictionary*. Presentatie op e-Lex 2019 *Smart lexicography* (1 oktober 2019).

- Depuydt, Katrien, Tanneke Schoonheim, Jesse de Does (2019). *Towards a More Efficient Workflow for the Lexical Description of the Dutch Language.* Presentatie op e-Lex 2019 *Smart lexicography* (1 oktober 2019).
- Depuydt, Katrien, *Towards a Diachronic Semantic Lexicon of Dutch.* Lezing op de *10th International Conference on Historical Lexicography and Lexicology* (13 juni 2019).
- De Tier, Veronique, Jesse de Does, Katrien Depuydt, Tanneke Schoonheim, Sally Chambers, Jacques van Keymuelen (2019). *Modeling and Analysis of Dialect Data in the Database of the Southern Dutch Dialects (DSDD).* Presentatie op e-Lex 2019 *Smart lexicography* (1 oktober 2019).
- Eerten, Laura van, Vivien Waszink. *Waarom is taal soms zo irritant?* Lezing & quiz tijdens ‘Broodje Kennis’. deBuren, Brussel, 11 oktober 2019.
- Eerten, Laura van. *Duurzaamheid & taal.* Presentatie duurzaamheidsmiddag CBRE Global Investors. Amsterdam, 31 oktober 2019.
- Geirnaert, Dirk. *Commercie en devotie. Brugs rederijkerswerk door Gheraert Leeu gedrukt (Gouda, 1482).* Lezing bij de presentatie van nummer LXIII van *Jaarboek ‘De Fonteine’*. KANTL, Gent, 27 mei 2018.
- Geirnaert, Dirk. *Aesopus in Brugge. De fabelillustraties van Marcus Gheeraerts en een verrassende vondst in Dresden.* Lezing voor de Brugse vereniging *Levend Archief*. Brugge, 9 mei 2019.
- Schoonheim, Tanneke (2019). *Words crossing Borders.* Lezing op de conferentie (*Historical Lexicography of the Landscape and the Digital Age*) (23 mei 2019).
- Schoonheim, Tanneke (2019). *Old Frisian and Old Dutch, sibling languages in one dictionary?* Lezing op de 10de International Conference of Historical Lexicography and Lexicology (14 juni 2019).
- Schuurman, Ineke, Leen Sevens, Vincent Vandeghinste, Frank Van Eynde (2019). Picto for migrants. *Computational Linguistics in the Netherlands (CLIN) 29*, Groningen
- Schuurman, Ineke, Vincent Vandeghinste, Leen Sevens (2019). Most wanted resources: imperfect written corpora. *Computational Linguistics in the Netherlands (CLIN) 29*, Groningen
- Sevens, Leen, Vincent Vandeghinste, Ineke Schuurman, Frank Van Eynde (2019). The Picto Project: Conclusions and Lessons Learned. *Computational Linguistics in the Netherlands (CLIN) 29*, Groningen
- Steurs, Frieda, Griet Depoorter, Vincent Vandeghinste (2019). Wat kan het INT voor u doen? *De Taalsector*. Gent
- Steurs, Frieda (2019). Terminology 4.0: Enabling Knowledge Exchange and Transfer. Presented at the Nordterm 2019, University of Copenhagen, 11 Jun 2019-15 Jun 2019
- Steurs, Frieda (2019). Het Expertisecentrum Nederlandstalige Terminologie. In: *Online*. Presented at the Termraad Europese Instellingen, Brussel, 25 Oct 2019-25 Oct 2019. [Open Access](#)

- Steurs, Frieda (2019). Language is Business! On the impact of linguistic applications in our society. In: *to be published by the Hungarian Academy of Sciences*. Presented at the Language and Economy: Language Industries in a Multilingual Europe Programme, Tallin, Estonia, 10 Oct 2019-11 Oct 2019. (professional oriented) [Open Access](#)
- Steurs, Frieda (2019). Taal is Business! Over de impact van taalkundige toepassingen in onze samenleving. In: *Online*. Presented at the Europese dag van de Talen, Brussel, Brussels, 26 Sep 2019-26 Sep 2019. [Open Access](#)
- Tiberius, Carole. *ELEXIS data model*. Presentatie op ELEXIS Observer event. 18 februari 2019, Wenen.
- Tiberius, Carole. *Lexicographers' needs*. Presentatie op ELEXIS Observer event. 19 februari 2019, Wenen.
- Tiberius, Carole. *How will lexicographers and new technologies interact in the lexicographic process?* Presentatie op de Lorentz-workshop *The Future of Academic Lexicography*, 6 november 2019, Leiden.
- Vandeghinste, Vincent (2019). CLARIN. A Hands-on Session. *Seminar: "Easy use of corpora and language tools. An introduction to CLARIN, a European infrastructure."* Louvain-la-Neuve
- Vandeghinste, Vincent, Bram Bulté (2019). Wablieft: A News Corpus of Easy-To-Read Text. *Computational Linguistics in the Netherlands (CLIN) 29*, Groningen
- Van pellicom, Katrien, (29 mei 2019) Algemene presentatie over spelling voor de Raad van Toezicht.
- Waszink, Vivien. Lezing *Aquí met that ass. Spaanse invloeden in de nederhop*, gehouden op de Ledenvergadering van Onze Taal, 8 juni 2019.
- Waszink, Vivien. Lezing *Dikketruierendag, ecorexia en klimaatspijbelen*, nieuwe groene woorden. Publieksmiddag Groene Taal, Leiden, 8 oktober 2019.
- Waszink, Vivien. Lezing *Irritante taal, Moeilijke Dingen Makkelijk Uitgelegd*, deBuren, Brussel, 10 oktober 2019.
- Waszink, Vivien. Lezing *Waarom is taal soms zo irritant?* Broodje Kennis, deBuren, Brussel, 11 oktober 2019.
- Waszink, Vivien. *Flitscollege hiphoptaal*, gehouden bij *KennisCafé: De Taal Verloedert!?*, georganiseerd door *de Volkskrant*, KNAW en NEMO, De Balie, Amsterdam, 21 oktober 2019.
- Waszink, Vivien. Lezing *Leuk is anders, de betekenis van het woord leuk*. Viering van het Nederlands georganiseerd door het Taalmuseum, Leiden, 13 december 2019.
- Zingano Kuhn, Tanara, Peter Dekker, Branislava Šandrih, Rina Zviel-Girshin, Špela Arhar Holdt, Tanneke Schoonheim. *Crowdsourcing corpus cleaning for language learning resource development*. Presentatie op EUROCALL 2019 'CALL and Complexity' (30 augustus 2019).

## Onderwijs

Bonth, Roland de. BA-college *De historische woordenboeken van het Nederlands in de Geïntegreerde Taalbaak (GTB)*. Leiden, 18 oktober en 19 oktober 2019.

Kinable, Dirk. MA-college *Concepts: systems, diagrams and definitions*. KULeuven – campus Antwerpen, 28 oktober 2019.

Schoonheim, Tanneke. MA-college *Woordbetekenis en definities*. KU Leuven, 25 februari.

Schoonheim, Tanneke. MA-college *Word Senses and Definitions*. Universiteit Leiden, 24 april.

## In de media

Bonth, Roland de. Interview over de Rembrandt Tutorials. NPO Radio 5, 28 februari 2019.

Bonth, Roland de. Interview over de Rembrandt Tutorials. Radio Noord-Holland, 28 februari 2019.

Bonth, Roland de. Interview over de Rembrandt Tutorials. *Jan-Willem staat op* NPO Radio 2, 28 februari 2019.

Bonth, Roland de. Interview over de Rembrandt Tutorials. *De Taalstaat* NPO Radio 1, 2 maart 2019.

Bonth, Roland de. 'Kennis over grammatica onder scholieren verslechterd'. Interview in *Metronieuws* 28 augustus 2019.

Bonth, Roland de. Interview over verkiezing Rembrandtwoord op Radio1.be, 15 november 2019.

Bonth, Roland de. Interview over verkiezing Rembrandtwoord. *De Taalstaat* NPO Radio 1, 16 november 2019.

Eerten, Laura van, Vivien Waszink. *De grote taalirritatietest, aan welke woorden erger jij je groen en geel?* Samenwerking met Quest, 27 februari 2019.

Eerten, Laura van, Vivien Waszink. Interview over taalergernissen. *Ergerpret*. In Quest Psychologie, april 2019.

Eerten, Laura van. Over de middag 'groene taal' in *de Taalstaat*, NPO Radio 1, 5 oktober 2019.

Geirnaert, Dirk & Roland de Bonth. Geïnterviewd voor Kennislink.nl over de taal van Rembrandt, 8 augustus 2019.

Geirnaert, Dirk. Interview op de persbijeenkomst georganiseerd door ING en het Rijksmuseum bij de lancering van de zes 'Rembrandt Tutorials' (schilderlessen) op Youtube. Rijksmuseum Amsterdam, 28 februari 2019.

Geirnaert, Dirk. Interview over Klimaatschieter, dwangzoen en kastelen in Spanje – Woorden die we wisten. *Lunchroom*, Radio Noord-Holland, 16 december 2019.

Instituut voor de Nederlandse Taal. Kader over 'zet u maar/ga lekker zitten', *Taalradar* in het Leidsch Dagblad, 5 maart 2019.

Instituut voor de Nederlandse Taal. Rubriek De Volkskeuken over 'Botergeile aardappelpuree' in *de Volkskrant*, 18 maart 2019.



- Instituut voor de Nederlandse Taal. Hoe groene woorden het Nederlandse taalveld binnenkomen, nieuwsbericht over *Het Groene Woordenboekje*. In Nederlands Dagblad, 28 juni 2019.
- Instituut voor de Nederlandse Taal. Nieuwsbericht over *Het Groene Woordenboekje*. In Greenpeace Magazine zomereditie, juni 2019.
- Instituut voor de Nederlandse Taal. Kadertekst over de verkiezing *Weg met dat woord!*. In De Telegraaf Leef nu!, 22 juni 2019. pp 36.
- Instituut voor de Nederlandse Taal. *We gaan allemaal dood' werkt voor sommige kiezers, maar de meesten horen liefst een positieve boodschap* – verslag van de middag 'groene taal' in *Leidsch Dagblad*, 8 oktober 2019.
- Tempelaars, Rob. Interview over het gebruik van het woord kanker in scheldwoorden en verwensingen. EditieNL (RTL-tv), 21 november 2019.
- Vandeghinste, Vincent (2019). Radio-interview op VRT Radio 1. Wat als iedereen dezelfde taal sprak? De Wereld van Sofie. Vanaf ong. 1u37.
- Waszink, Vivien (2019). Interview over onnodig Engels: *Onnodig Engels in de winkelstaat*. In Editie.nl, RTL 4 en op rtlnieuws.nl, 25 januari 2019.
- Waszink, Vivien (2019). Interview over taaltoets politie: Taaltoets politie te moeilijk. In Editie.nl, RTL 4 en op rtlnieuws.nl, 7 februari 2019.
- Waszink, Vivien (2019). Interview over verkleinwoordjes: *Boycotje na bevinkje*. In Editie.nl, RTL 4 en op rtlnieuws.nl, 28 mei 2019.
- Waszink, Vivien (2019). Interview over verengelsing: *Nederlanders zijn 'not amused' met de 'verengelsing' van onze taal*. In Editie.nl, RTL 4 en op rtlnieuws.nl, 26 juni 2019.
- Waszink, Vivien (2019). Interview over genderneutraal taalgebruik: *Eerst nieuwe mores, daarna nieuwe taal*, in: *Trouw*, 6 augustus 2019.
- Waszink, Vivien (2019). Interview over mannelijke en vrouwelijke taal in vacatureteksten. In: uitzending *Een Vandaag*, NPO Radio 1, 15 oktober 2019.
- Waszink, Vivien (2019). Interview over woorden *bemanning* en *bemensing*: *Bemanning of bemensing? 'Het geslacht maakt niet uit voor het beroep'*. In Editie.nl, RTL 4 en op rtlnieuws.nl, 28 november 2019.
- Waszink, Vivien (2019). Interview over het woord *neger*: *Mag je het woord 'neger' wel of niet gebruiken?* Op: ad.nl, 7 december 2019.
- Waszink, Vivien (2019). Interview over het woord *Wollah!*. In: uitzending *Shay!*, NPO Radio 2, 16 december 2019.
- Waszink, Vivien (2019). Interview over *boomer*, woord van het jaar bij Van Dale: *Bij het Woord van het Jaar draait het al tijdje niet meer om taal*. Op: nos.nl, 17 december 2019.
- Waszink, Vivien (2019). Interview over irritante taal: *Van welk woord krijg jij jeuk?* In: uitzending *Gaan!*, NPO Radio 1, 17 december 2019.

Waszink, Vivien (2019) Interview over de Woord van het Jaar-verkiezing, pilotuitzending *Humberto!*, NPO Radio 1, 19 december 2019.

Waszink, Vivien (2019), diverse interviews in *de Taalstaat*, NPO Radio 1.

### **Actieve deelname en organisatie congressen en workshops**

Bennekom, Marjolijn van, Peter Dekker, Laura van Eerten, Tanneke Schoonheim en Boukje Verheij.

Laborant voor Taalradar op het DRONGO talenfestival in Nijmegen, 25 en 26 oktober 2019.

Colman, Lut. Week van het Nederlands: “Groene taal”, 8 oktober 2019.

Colman, Lut, Katrien Depuydt, Jesse de Does, Kris Heylen, Dirk Kinable, Jan Niestadt, Frieda Steurs, Rob Tempelaars, Carole Tiberius, Vincent Vandeghinste, Boukje Verheij, Vivien Waszink en Ben Williams. Deelname aan Lorentz-workshop *The Future of Academic Lexicography: Linguistic Knowledge Codification in the Era of Big Data and AI*. Leiden, 4-8 november 2019.

Colman, Lut. Cost-workshop “Crowdsourcing related to corpus examples for language learning” in Coimbra, Portugal, 5-6 december 2019.

Depuydt, Katrien en Jesse de Does. Workshop ter ontwikkeling van TEI Lex-0. Berlijn. 30 januari 2019.

Depuydt, Katrien en Jesse de Does. Workshop over de *Thesaurus of Old English*. Leiden. 1 februari 2019.

Depuydt, Katrien en Jesse de Does. Workshop Ontolex-TEI lex 0 in Sintra in het kader van ELEXIS. 4 oktober 2019.

Eerten, Laura van. Organisatie middag ‘groene taal’ tijdens de Week van het Nederlands. Leiden, 8 oktober 2019.

Kinable, Dirk. Deelname aan projectbijeenkomst voor het *Platform medische terminologie* (11 juli, Université Savoie Mont-Blanc).

Kinable, Dirk. Deelname aan het congres: TiNT-dag (29 november, Leiden).

Kinable, Dirk. Deelname aan de projectbijeenkomst *Testen met patiëntvriendelijke termen* (12 november, Amsterdam).

Steurs, Frieda en Boukje Verheij. Deelname aan de bijeenkomst ‘Wat verstaan we onder een patiëntvriendelijke term’, georganiseerd door Nictiz bij Pharos in Utrecht, 17 oktober 2019.

Schoonheim, Tanneke en Boukje Verheij. Deelname aan de werkbijeenkomst ‘Betere zorg met zorg voor taal’, georganiseerd door de Taalunie, op het DRONGO talenfestival in Nijmegen, 25 oktober 2019.

Schoonheim, Tanneke. Deelname aan de enetCollectbijeenkomst in Lissabon (14-16 maart, Lissabon, Portugal).

Schoonheim, Tanneke. Deelname aan de conferentie (*Historical*) *Lexicography of the Landscape and the Digital Age* (22-25 mei, Göttingen, Duitsland).

Schoonheim, Tanneke. Deelname aan de 10de International Conference of Historical Lexicography and Lexicology (12-14 juni, Leeuwarden).

Schoonheim, Tanneke. Deelname aan e-Lex 2019 *Smart Lexicography* (1-3 oktober, Sintra, Portugal).

Schoonheim, Tanneke. Medeorganisatie van de Streektaalconferentie 2019 *Streektaal kun je leren* (11 oktober, Mechelen, België).

Schoonheim, Tanneke. Medeorganisatie van en deelname aan de Lorentz-workshop *The Future of Academic Lexicography: Linguistic Knowledge Codification in the Era of Big Data and AI* (4-8 november, Leiden).

Schoonheim, Tanneke. Deelname aan het DRONGO-festival *De impact van taal* (25 oktober, Nijmegen).

Schoonheim, Tanneke. Deelname aan de enetCollectworkshop *Crowdsourcing examples* (5-7 december, Coimbra, Portugal).

Tiberius, Carole. Deelname aan congres: *ELEXIS Observer Event*. 18 en 19 februari Wenen.

Tiberius, Carole. Deelname aan projectbijeenkomst: *ELEXIS Project Management Board*. 20 februari, Wenen.

Tiberius, Carole. Deelname aan projectbijeenkomst: *ELRC LRB meeting*. 27 en 28 maart 2019, Amsterdam.

Tiberius, Carole. Deelname aan hackaton: *Lexonomy*. 2-4 april 2019, Brno.

Tiberius, Carole. Deelname aan projectbijeenkomst *ELEXIS Technical Management Board*. 10-12 juni 2019, Leiden.

Tiberius, Carole. Deelname aan en organisatie van congres: *eLex 2019 smart lexicography*, 1-3 oktober 2019, Sintra.

Tiberius, Carole. Deelname aan en organisatie van workshop: *TEI Lex-0 & Ontolex-Lemon*, 3-4 oktober 2019, Sintra.

Tiberius, Carole. Deelname aan projectbijeenkomst: *COST CA18209 Kickoff*. 28 oktober 2019, Brussel.

Tiberius, Carole. Deelname aan projectbijeenkomst *ELEXIS Technical Management Board*. 9 en 10 december 2019, Wenen.

Waszink, Vivien. Deelname aan conferentie en lezing *Neologisms in a Dutch online portal op Globalex workshop on Lexicography and Neologisms* gehouden op de conferentie van the *Dictionary Society of North America (DSNA)*, DSNA 22, Bloomington, Indiana (US), 8-11 mei 2019.

Waszink, Vivien. Deelname aan workshop en lezing *Detecting neologisms using Neoloog* op *Neologismen: korpuslinguistische Ermittlung und lexikographische Bearbeitung* op het Leibniz-Institut für Deutsche Sprache (IDS) in Mannheim (Duitsland), 3-4 juni 2019.

Waszink, Vivien. Deelname aan conferentie en Lezing *Zieke zinnen zullen zegevieren*,  
*Hiphop in de klas* en presentatie opdrachten gehouden op HSN-conferentie op de Hogeschool  
Windesheim in Zwolle, 22-23 november 2019.

## Diversen

Bonth, Roland de:

- Columnist voor Neerlandistiek voor de klas, onderdeel van Neerlandistiek, Online tijdschrift voor taal- en letterkundig onderzoek
- Coördinator en jurylid Profielwerkstukwedstrijdprijs van Levende Talen Nederlands (LTN) in samenwerking met de Internationale Vereniging voor Neerlandistiek (IVN)
- Kolomleider van de HSN-conferentie (Het Schoolvak Nederlands)

Depuydt, Katrien

- Nam samen met Jesse de Does deel aan de verdere ontwikkeling van de modules voor lexicografie en voor frequentie, attestatie en corpusdata van OntoLex Lemon, en aan de verdere ontwikkeling van TEI lex-0, een XML-standaard voor woordenboeken.
- Samen met Vincent Vanderghinste betrokken bij de organisatie van *Datech 2019* (<http://datech.digitisation.eu/>) van 8-10 mei in Brussel.
- Verrichtte diverse activiteiten in de context van het IMPACT Centre of Competence.
- Was tot eind februari 2019 voorzitter van de Standards Committee voor ELEXIS en bezocht in die hoedanigheid de ELEXIS meetings in Wenen 18-20 februari 2019.

Schoonheim, Tanneke

- Voorzitter van de Commissie Spelling van de Nederlandse Taalunie
- Penningmeester van EURALEX, European Association for Lexicography
- Penningmeester van de Stichting Nederlandse Dialecten
- Bestuurslid van de Stichting Rechtstaal (Juridisch Woordenboek Spaans – Nederlands)
- Lid van de Raad van Advies van Hogeschooltaal (Noordhoff Uitgevers)
- Redacteur van *Trefwoord*, digitaal tijdschrift voor lexicografie

Tempelaars, Rob

- Redacteur en redactiesecretaris van het tijdschrift *Trefwoord*

Tiberius, Carole:

- Board of Trustees Adam Kilgarriff Prize

Van pellicom, Katrien

- Lid van de Commissie Anderstalige Namen)

Waszink, Vivien:

- Redactielid *Neerlandia*.
- Medewerking aan profielwerkstuktaalkunde.nl, onderdeel Woorden en Toolkit Neologismen.

## **Bijlage 4: Verslag Taalbank Nederlands**

De Taalbank Nederlands omvat corpora en computationele lexica van zowel modern als historisch Nederlands. Het lexicon is bedoeld om ingezet te worden voor onderzoek in de vorm van diverse producten die niet alleen in principes van verrijking en structuur op elkaar aansluiten, maar bovendien in toenemende mate inhoudelijk aan elkaar gekoppeld zijn. Hieronder volgt een opsomming van de werkzaamheden die in 2019 zijn verricht. De zichtbaarheid van de resultaten is in 2019 vergroot door het realiseren van de themapagina's historisch en modern Nederlands.

### **I Lexica**

#### **I.1 GiGaNT**

GiGaNT is het computationele woordvormenlexicon dat de door het INT beschreven woordenschat moet gaan bevatten van het Nederlands vanaf de 6<sup>e</sup> eeuw tot nu en de centrale database vormt van het Nederlands van het INT. In GiGaNT zit de formele beschrijving van de woorden. De semantische informatie zit in DiaMaNT, en in de diverse woordenboeken van het INT.

Werkzaamheden aan deze centrale data-infrastructuur zijn modulair opgezet. De twee grote componenten zijn GiGaNT-Molex, waarin de hedendaagse woordenschat wordt beschreven, en GiGaNT-Hilex, de historische lexiconcomponent. Voor de historische lexiconcomponent moeten de vier historische woordenboeken van het Nederlands (ONW, VMNW, MNW en WNT) de kern vormen, aangevuld met nieuw materiaal. De moderne lexiconcomponent bevat het hedendaags taalmateriaal. Lexiconinformatie wordt ter beschikking gesteld via de lexiconservice en het doel is ook om downloadbare datasets te distribueren.

Afgelopen jaar zijn belangrijke stappen gezet in het realiseren van de centrale rol van het moderne lexicon in de beschrijving van de woordenschat in de diverse producten van het INT. De structuur en de inhoud van GiGaNT-Molex is uitgebreid om de koppeling met het combinatiewoordenboek mogelijk te maken, en er is verder gewerkt aan de technische koppeling van ANW, Neoloog en Molex. De API op Molex is hiervoor aangepast, en er zijn aanpassingen gedaan in de verschillende werkomgevingen. Wat betreft de werkomgeving voor Molex, deze heeft technische aanpassingen gehad om de werkzaamheden ten behoeve van spelling efficiënter te laten verlopen en om de performance te verhogen.

##### **a. GiGaNT- HILEX: Historische lexiconcomponent**

De werkzaamheden aan de historische lexiconcomponent hebben zich in 2019 geconcentreerd op systematische verfijningen van de verrijking in MNW en WNT. Met de komst van een nieuw historisch taalkundige op 1 februari konden grote dataklussen worden aangepakt. Die hebben de voorrang gekregen boven het integreren van de ONW- en de VMNW-module. Diverse verbeteringen

zijn doorgevoerd aan de bewerkingsomgeving. De lexiconservice, waarmee de data ter beschikking wordt gesteld voor gebruik in applicaties, is geïmplementeerd in de zoekomgeving van een aantal historische corpora.

b. GiGaNT- MOLEX: Moderne lexiconcomponent

De moderne lexiconcomponent wordt continu geactualiseerd. Het grootste deel van deze werkzaamheden worden uitgevoerd in het kader van het project spelling (zie aldaar). Door de koppeling met projecten als ANW, Neoloog en het combinatiewoordenboek wordt het lexicon ook verder uitgebreid. Hetzelfde gebeurt door de koppeling van Hilex en Molex.

In 2019 is het lexicon in een eerste release verschenen als downloadbare dataset bij de Taalmaterialen van het INT.

Integratie GiGaNT-Molex - ANW

In 2019 is het resultaat van de datakoppeling van ANW en Molex toegevoegd aan het ANW. Er is gewerkt aan de implementatie van een workflow waarbij bij bewerking van nieuwe woorden in het Neologismenproject en het ANW de koppeling met Molex meteen tot stand wordt gebracht.

Paradigmainformatie en controle spelling van nieuwe woorden in deze producten wordt daardoor binnen de context van GiGaNT-Molex gerealiseerd en via de koppeling ANW/Neologismen aan GiGaNT-Molex aan deze woordenboekproducten toegevoegd. De workflowimplementatie zal in de loop van 2020 in productie worden genomen.

Integratie GiGaNT-Molex - Combinatiewoordenboek

In 2019 is ook de koppeling tussen de data van het combinatiewoordenboek en GiGaNT-Molex gerealiseerd. Het lexicon is uitgebreid met een extra tabel met werkwoordskenmerken die voor het combinatiewoordenboek relevant zijn. Door de koppeling zal de paradigmainformatie in het combinatiewoordenboek rechtstreeks uit GiGaNT-Molex gehaald worden.

c. Integratie GiGaNT-Hilex - GiGaNT-Molex

In 2019 is gestaag verdergewerkt aan de redactionele controle van de ruim 60.000 koppelingen tussen GiGaNT-Hilex (via de WNT-module) en GiGaNT-Molex. Deze koppelwerkzaamheden leiden tot substantiële verbeteringen in zowel GiGaNT-Molex als ook in GiGaNT-Hilex.

## **I.2 DiaMaNT**

In dit project wordt gewerkt aan de ontwikkeling van een Diachroon semantisch lexicon van het Nederlands (DiaMaNT). Het diachroon semantisch lexicon heeft als doel een hulpmiddel te bieden bij tekstontsluiting en bij het onderzoek naar begrippen door de eeuwen heen. Het lexicon is het derde

onderdeel van de lexicografische infrastructuur voor historisch Nederlands die in ontwikkeling is en als kern de vier historische woordenboeken van het Nederlands bevat, ONW, VMNW, MNW en WNT. Het eerste onderdeel van deze infrastructuur is het historische woordenboekenportaal, waar de gebruiker de betekenis van woorden kan opzoeken. Het tweede onderdeel is het morfosyntactische lexicon GiGaNT – ook functionerend als centrale database – dat informatie bevat over mogelijke variatie in spelling en vorm van de historische Nederlandse taal, waarmee het zoeken in historische tekst vergemakkelijkt wordt. De derde component is het DiaMaNT-lexicon. Het vormt een laag bovenop het GiGaNT-lexicon en is bedoeld om te hulp te bieden bij het omgaan met historische semantische variatie.

Het doel van het lexicon is om relaties te leggen tussen woordvormen en betekenseenheden (concepten), en deze in de tijd te plaatsen en zo diachrone onomasiologie, d.i. de veranderende uitdrukking/verbalisatie van een concept, en semasiologie, d.i. de verschuiving van betekenis(nuance) van woorden in de tijd, systematisch vast te leggen op een zodanige wijze dat de informatie voor mens en computer bruikbaar is.

Het datamodel van het uiteindelijke lexicon ligt vast. De bewerking van de data tot een gestructureerd semantisch lexicon gebeurt in stappen, waarbij geprobeerd wordt om tussentijdse resultaten ook ter beschikking te stellen. Het ontwikkelen van intelligenter strategieën voor lexiconbouw is een continue taak.

### **Data en datamodel**

Ook in 2019 is gewerkt aan een verbetering van het datamodel in RDF. Het werk daaraan is in internationaal verband uitgevoerd en resultaten van die werkzaamheden hebben hun neerslag gehad in de OntoLex Lemon module voor lexicografische producten, verschenen in het najaar van 2019 (OntoLex Lemon Lexicography Module (<https://www.w3.org/2019/09/lexicog/>). Omdat DiaMaNT attestaties met woordvormen koppelt aan lemmata en betekenissen, en attestaties geen onderdeel van de vermelde module vormden, is in 2019 het werk aan het modelleren ervan in RDF uitgevoerd binnen een nieuw te definiëren OntoLex Lemon module voor frequentie, attestaties en corpus data (FRAC; <https://github.com/acoli-repo/ontolex>).

In 2018 was na het afronden van de correctie van de synoniemdetectie in de betekenisomschrijvingen van het MNW, dat ruim 120.000 relaties tussen een hedendaags en historisch trefwoord heeft opgeleverd, het nakijken gestart van de ruim 180.000 automatisch gedetecteerde synoniemen in het WNT en was een kwart afgerond. De correctie van de resterende driekwart kon in 2019 worden afgerond. Dit betekent dat een nieuwe dataset opgeleverd kon worden aan CLARIAH (dat een gedeelte van de ontwikkeling van het lexicon had gefinancierd) en dat voor de online applicatie (zie verder) geen voorziening gemaakt moest worden voor het selecteren van al dan niet handmatig nagekeken materiaal.



## **Koppelen met externe datasets**

In de context van het CLARIAH-project ATHENA (zie aldaar) de Athenaplantenthesaurus door koppeling met het Nederlands Soortenregister, het DiaMaNT-lexicon, de DSDD-database en de Databank Plantennamen in de Nederlandse Dialecten (PLAND) uitgebreid met o.a. dialect- en spellingsvarianten, synoniemen en dateringsinformatie.

## **Zoekapplicatie**

Er is in 2019 verder gewerkt aan een gebruikapplicatie voor DiaMaNT. Eind 2019 is de DiaMaNT-applicatie versie 1 gereleased. De gebruikersapplicatie laat toe om te zoeken naar historische synoniemen van moderne concepten en kan daarvan een visualisatie zien door de tijd heen. De applicatie is te vinden op <http://diamant.ivdnt.org/diamant-ui/>.

## **II Corpora**

### **II.1 Modern corpusmateriaal: Corpus Hedendaags Nederlands**

#### **Corpusdata**

Het CHN bevat al het hedendaags Nederlands corpusmateriaal waarover het INT beschikt en wordt door diverse projecten in het INT gebruikt. Het corpus wordt continu uitgebreid. Daarbij wordt beoogd om materiaal binnen te halen van Nederland, Vlaanderen, Suriname en de Antillen. De acquisitie wordt mede bepaald door de projecten die van het corpus gebruikmaken. Een belangrijke functie van het corpus is het monitoren van de Nederlandse taal. Daarvoor proberen we data te verwerven die systematisch en continu aangeleverd worden. Kranten spelen daarbij een belangrijke rol, maar ook sociale media en parlementair materiaal leveren een continue stroom aan taaldata op. Sedert juni 2019 hebben we een dataconversiespecialist in dienst die twee dagen per week met het modern corpus materiaal aan de slag is gegaan. Mede daarom hebben we onze acquisitieactiviteiten geïntensiveerd en hebben we conversie van nieuw taalmetaal systematisch op kunnen pakken.

Voor het project spelling was er behoefte aan meer materiaal uit Suriname en de Antillen. De werving daarvan is niet evident gebleken, maar er is resultaat geboekt. Voor de Antillen hebben we *ABC Online Media* en de *Achterkant van Curaçao* aan het corpus kunnen toevoegen en voor het Surinaams het forum van *culturu.com*. Van de bestaande bronnen zijn alle materialen tot en met 2018 geconverteerd, m.u.v. *De Ware Tijd*, die pas eind vorig jaar geleverd werd En nog verwerkt moet worden.

Voor het combinatiewoordenboek is krantenmateriaal niet voldoende. Er is behoefte aan materiaal dat de gesproken taal benadert en ook behoefte aan eenvoudiger Nederlands.

Het materiaal van Open Subtitles hebben we binnengehaald en geanalyseerd en zal aan het corpus worden toegevoegd. Er zijn verder onderhandelingen gestart met de VRT. Er is ook een eerste aanzet gedaan tot het benaderen van educatieve uitgeverijen omdat we meer tekstmateriaal willen dan wat er in het Basilexcorpus zit. Vooralsnog zijn daar nog geen resultaten geboekt.

Parlementair materiaal is een welkome aanvulling op het krantenmateriaal dat we gebruiken als monitorcorpus. We krijgen het materiaal van het Tweede Kamer uit Nederland systematisch binnen via de ftp-server. Omdat de data in een ander formaat worden aangeleverd moet de data-analyse ten behoeve van de conversie herzien worden. Inmiddels is ook een eerste verkenning gedaan van de data van het Vlaams parlement en het federaal parlement van België. De toegangsmogelijkheden zijn onderzocht en een eerste analyse van de data van het Vlaams parlement is uitgevoerd. In 2020 willen we onze bevindingen toetsen met de parlementaire corpora die via CLARIN ter beschikking zijn gesteld om tot een zo groot mogelijke afstemming te komen.

Het monitorcorpus van kranten willen we graag uitbreiden met meer Vlaamse en Nederlandse kranten, en het liefst zodanig dat we het materiaal kunnen distribueren als dataset aan geïnteresseerde gebruikers. De onderhandelingen met het Vlaams Mediahuis verlopen voorspoedig. Het ziet er naar uit dat we meer Vlaamse kranten binnen zullen krijgen, zij het (vooralsnog?) onder dezelfde voorwaarden als we hebben voor de krant *De Standaard*.

Tot slot zijn de recente aan de Taalmaterialencollectie toegevoegde corpora *Morrocorp* en *Wablieft* geconverteerd en aan het CHN toegevoegd.

### **Corpusworkflow (DUCT)**

Het uitbreiden van het Corpus Hedendaags Nederlands is zoveel mogelijk geautomatiseerd. Ontvangen en opslaan van materiaal, conversies, verrijking en indexering zijn stadia in een corpusworkflow, uitgevoerd binnen een daartoe ontwikkelde tool DUCT (Data Update Creation Tool). Dit is een tool voor het converteren van bestanden in verschillende stappen. Voor 2019 was gepland om de parsing van de data met Alpino onderdeel te laten uitmaken van de automatische verwerking van corpusdata in DUCT. Deze stap is in 2019 niet gerealiseerd. Voorrang moest gegeven worden aan het migreren van de database naar een nieuwe server, hetgeen gezien de omvang van de database een complexe operatie bleek met lange doorlooptijd.

### **Corpus Front End en BlackLab/BlackLab server**

In het kader van CLARIAH is substantieel werk verricht aan het corpus-frontend, en zijn waar nodig aanpassingen gedaan aan het backend, hetgeen geleid heeft tot de release van het Open SoNaR+

corpus op 1 april 2019. Er zijn tests gedaan met het nieuwe frontend met de data van het CHN. Tot een nieuwe release van het CHN in de nieuwe applicatie is nog niet overgegaan. Voorrang is gegeven aan het inrichten van een krachtiger server en het verder uittesten van het frontend met de historische corpora (zie aldaar).

## II.2 Historisch corpusmateriaal

### Corpusdata

In 2019 zijn de data van de volgende corpora onderhanden genomen

- *Brieven als buit*  
Er is data-update gerealiseerd met correcties aangeleverd door M. van der Wal en er zijn verbeteringen gedaan aan de taalkundige verrijking.
- *Corpus Gysseling*  
Er is een data-update gerealiseerd waarin met name de taalkundige verrijking is aangepast.
- *Corpus Oudnederlands*  
Er is een verdere curatie van de data gedaan om aaneengesloten teksten te krijgen. De datacuratie wordt in 2020 afgerond
- *Eindhoven corpus en Corpus Middelnederlands*  
Bij deze corpora is nog wat verbeterwerk aan de metadata uitgevoerd
- *Brieven als buit aanvullingen*

Deze dataset is in 2019 nieuw binnengekomen. De data is geconverteerd en gekoppeld aan de afbeeldingen om in 2020 online gezet te kunnen worden.

### Taalkundige verrijking

Werkzaamheden aan een betere automatische taalkundige verrijking van historisch Nederlands en aan een tagset die door de onderzoeksgemeenschap breed gedragen wordt, zijn gestart in 2019 en worden in de context van CLARIAH plus uitgevoerd (zie aldaar). Op termijn zal de taalkundige verrijking van bovengenoemde corpora volgens de nieuwste inzichten en met de nieuwe tagset worden uitgevoerd.

### Corpusapplicatie

Na het online gaan van Open SoNaR+ in de nieuwe corpusapplicatie, is gewerkt aan een versie van de bovengenoemde corpora in het nieuwe frontend. Dat betekende dat het corpus frontend zo aangepast moest worden om het makkelijker configureerbaar te maken voor corpora met andere metadata of andere taalkundige verrijking. Ook aan het BlackLab backend zijn aanpassingen gedaan. Omdat het *corpus Middelnederlands* en de *Brieven als Buit aanvullingen* geen taalkundige verrijking hebben, is

de lexiconservice als functionaliteit aan het frontend toegevoegd. De corpora worden in 2020 aan een laatste test onderworpen en komen dan online.

## Bijlage 5: Rapport Externe Communicatie

In dit rapport worden de in 2019 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 2018 – 1 januari 2019).

#### *Cijfers*

<b>Algemeen</b>	<b>2018</b>	<b>2019</b>
Totaal aantal paginaweergaven	492.446	576.575
Bezoeken uit Nederland	184.590	203.701
Bezoeken uit België	42.269	54.045
Gemiddelde tijd op pagina	00.01.50	00.01.57

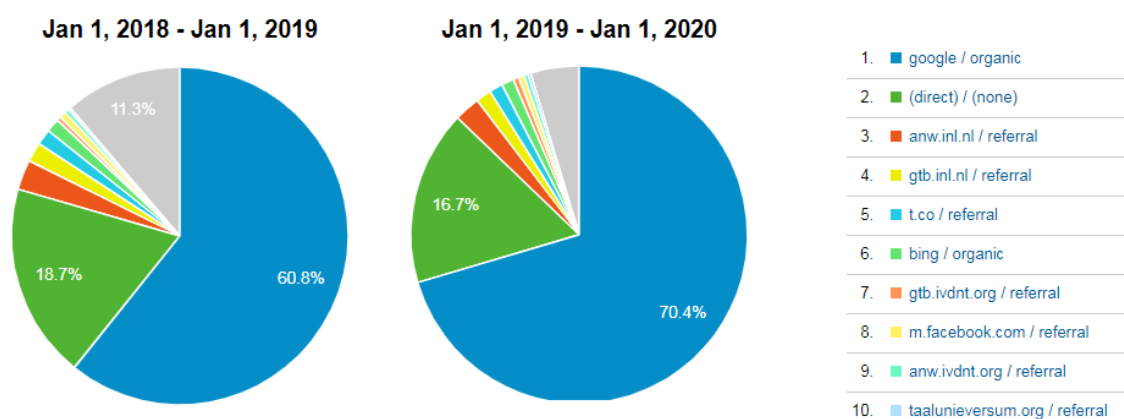
<b>Paginaweergaven rubrieken</b>	<b>2018</b>	<b>2019</b>
Woordbaak*	110.296	124.447
Neologismen (neo v/d week + neo v/d week archief + neologismen algemeen (68.223)**	77.323	77.837
Historische woordenboeken (algemeen + ONW, VMNW, MNW, WNT, WFT)	36.355	51.212
Index/home	38.997	36.542
Nieuws	13.342	13.792
Gelegenheidswoordenboekjes	9.966	11.485
Algemeen Nederlands Woordenboek	7.524	10.855
Taalmaterialen	9.243	7.143
Terug in de taal	7.909	5.886
Etymologie	3.133	5.240
Spelling	2.087	3.668
Jaaroverzicht in neologismen	3.181	3.124
Terminologie (Termenlijsten (1.226) +	1.093	2.678

evenementen (578) + ENT (496) + Hulpmiddelen (378))		
CLARIN	1.364	1.587
Contact/medewerkers	1.299	1.347
Themapagina Hedendaags Nederlands	-	1.197
Themapagina Historisch Nederlands	-	1.110
Themapagina Terminologie	-	574
Uitgelichte term		169

\* *Best bezochte artikel 'Zijn pappa en mamma wel correct?'*.

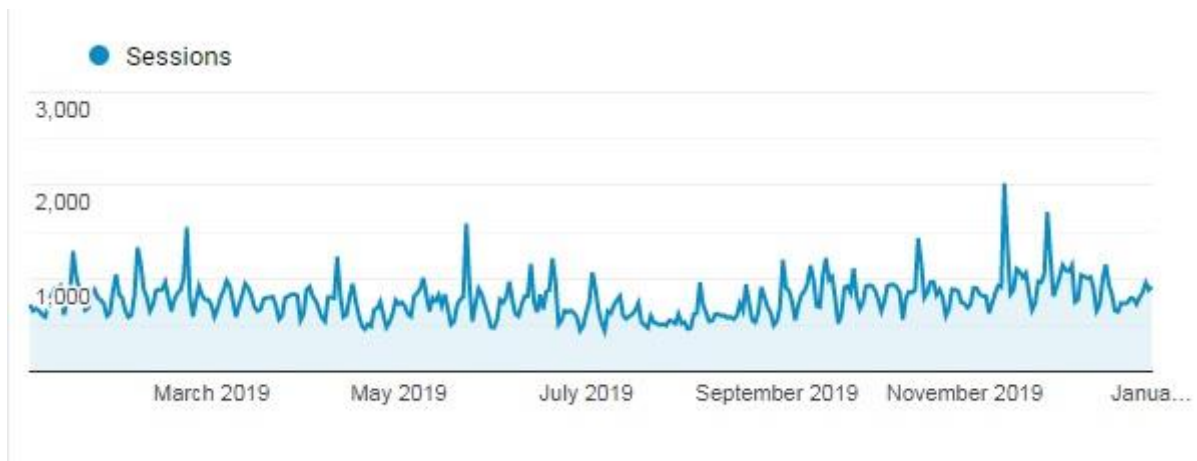
\*\**Populairste rubriek: Neologisme van de week*

### Bronnen/verwijzingen



- Via het *Algemeen Nederlands Woordenboek*, de *Historische Woordenboeken online* en de sociale media komt een relatief klein deel van de bezoekers op de website terecht. Het grootste deel vindt de website via Google.

## Verloop bezoekersaantallen



- Het verloop van de bezoekersaantallen door het jaar heen is stabiel. De opvallende pieken die er zijn hebben te maken met populaire neologismen van de week (zoals *milkshaking* (mei) *menstruatiearmoede* (november) en *gasklever* (november)).

De website heeft in 2019 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 is aan de bezoekersaantallen op de website. De verhouding bezoeken uit Nederland en België is ongeveer gelijk gebleven, en de gemiddelde tijd die een bezoeker doorbrengt op een pagina is iets gestegen.

De populairwetenschappelijke webrubrieken zijn al jaren de grootste aandachtstrekkers op de website, en dan met name *Woordbaak* (etymologie) en *Neologisme van de week* (nieuwe woorden). *Woordbaak* is geen wekelijkse rubriek meer, maar gezien de populariteit is het nuttig om de rubriek in stand te houden en af en toe een artikel te kunnen blijven plaatsen. Er is een opvallende stijging te zien in de interesse voor historisch Nederlands. Ondanks dat de rubriek *Terug in de taal* minder bekeken is dan in het voorgaande jaar, is de algemene informatie over de historische woordenboeken aanzienlijk meer bezocht. Dat kan verklaard worden doordat er in 2019 meer aandacht was voor historische projecten vanuit het INT, o.a. door het Rembrandtjaar.

Er is een verdubbeling in het aantal bezoeken van webpagina's over terminologie. Die categorie is in 2019 dan ook uitgebreid met o.a. termenlijsten en hulpmiddelen. *Uitgelichte term* is een nieuwe rubriek gebaseerd op *Term van de dag* (Week van het Nederlands 2018). De uitgelichte term verschijnt in elke Terminologienieuwsbrief (4x per jaar) en is daarin in zijn geheel te lezen (geen link om door te klikken naar het volledige artikel). Dat verklaart dat deze rubriek op de website niet vaak bekeken werd. Een beter vindbare plek op de website kan de bezoekersaantallen verbeteren, evenals een link vanuit de nieuwsbrief Terminologie.

In 2019 zijn er twee nieuwe gelegenheidswoordenboekjes verschenen: *Het Groene Woordenboekje* (1.754 views) en *Woorden voor Rembrandt* (2.036 views). *Het Groene Woordenboekje* is alleen in de nieuwsbrief aangekondigd, over *Woorden voor Rembrandt* is daarnaast ook nog een apart persbericht verstuurd. Het *Opvoedwoordenboekje* is in februari 2019 geüpdatet en is met 2.061 views in 2019 het vaakst bekeken van alle woordenboekjes.

Er zijn drie nieuwe themapagina's gelanceerd: *Historisch Nederlands*, *Hedendaags Nederlands* en *Terminologie*. Op die pagina's staat alle informatie die op de website over die onderwerpen te vinden is bij elkaar. Zo kan er in één oogopslag alles over het Nederlands van vroeger, over het Nederlands van nu of over vaktaal gevonden worden. De themapagina *Terminologie* is gelijk aan het al bestaande menu-item, en heeft ook minder bezoekers getrokken dan de andere twee.

De meeste zoektermen die op de website gebruikt worden zijn namen van taalmaterialen, zoals *sonar*, *wabliedt*, *cgn* en *gigant*.

In 2020 wordt de website overgezet op een nieuw contentmanagementsysteem, van Joomla! naar Wordpress. Het ontwerp en de huisstijl blijven ongewijzigd, maar de website wordt gemoderniseerd en geoptimaliseerd. Ter voorbereiding is er in 2019 door een communicatiestagiair onderzoek naar het gebruik van de website gedaan, en op basis daarvan is een nieuwe menustructuur gemaakt die de ruggengraat vormt van de nieuwe website.

## **2. Sociale Media**

Sinds 2010 heeft het instituut een Twitteraccount. Via Twitter worden links naar de website en taalweetjes/taalnieuws gedeeld, en contacten met volgers en partnerorganisaties onderhouden. Er bestond al een Facebookpagina voor de verkiezing *Weg met dat woord!* en in 2019 is er een algemene Facebookpagina van het instituut toegevoegd.



## Cijfers

	December 2018	December 2019
Volgers op Twitter	4.183	4.492
Volgers op Facebook (ivdnt)		255
Volgers op Facebook (wmdw)		1.221

## Conclusies & aanbevelingen

Twitter blijft gestaag groeien met dit jaar een toename van 309 volgers. De Facebookpagina voor *Weg met dat woord!* wordt af en toe nog gebruikt. Het was niet mogelijk de pagina om te vormen tot een algemene pagina voor het instituut, dus er is een nieuwe gemaakt.

## 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, waarvoor abonnees zich los van de algemene nieuwsbrief kunnen inschrijven. In 2019 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

<b>Nieuwsbrief algemeen</b>	<b>2018</b>	<b>2019</b>
Aantal inschrijvingen	3.894	3.813
Gemiddeld aantal opens	46%	48%
Gemiddeld aantal kliks (CTO)*	29%	27%
<b>Nieuwsbrief terminologie</b>		
Aantal inschrijvingen	2.660	2.676
Gemiddeld aantal opens	31%	28%
Gemiddeld aantal kliks (CTO)	16%	14%

\* 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 (52%)
- Best gelezen artikelen: *Het Groene Woordenboekje* (312 kliks), het *Opvoedwoordenboekje* (258 kliks), artikel *Ga lekker zitten* (250 kliks) en *Woordbaak* over Lesbos en *lesbisch* (220 kliks)
- Meest geopende nieuwsbrief terminologie is de maarteditie (29,7% opens)

- Links waar het vaakst op geklikt is: Workshop Terminology in Language Technology (274 kliks), termenlijsten op ivdnt.org (68 kliks) en tijdschrift edition (59 kliks)

<b>Mailings &amp; persberichten</b>	<b>opens</b>
Persbericht: Het 17 <sup>e</sup> -eeuwse Nederlands van Rembrandt	35,2%
Onderzoek website	50,1%
Persbericht: Woorden voor Rembrandt	29,1%
Uitnodiging middag ‘groene taal’	43,1%
Herinnering middag ‘groene taal’	42,1%
Herinnering TiNT-dag 2019	30,2%
Persbericht: Rembrandtwoord van het jaar	32,5%
Mailing Verkiezing Rembrandtwoord	38% %
Mailing: Zwadderen is mooiste Rembrandtwoord	42,9%
Persbericht: Zwadderen is mooiste Rembrandtwoord	30,2%
Nieuwjaarswens 2020	54,8%

*De cijfers van de mailings en persberichten zijn lastig met elkaar te vergelijken omdat ze naar verschillende doelgroepen binnen het adressenbestand worden verstuurd. De berichten worden gemiddeld iets minder vaak geopend dan de algemene nieuwsbrief.*

Het abonneebestand van de algemene nieuwsbrief is in 2019 licht gedaald, maar de terminologienieuwsbrief heeft er juist iets meer abonnees bij gekregen. De cijfers voor het aantal opens en kliks zijn voor de algemene nieuwsbrief 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 iets 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.

### **Enquête**

Begin 2019 is er een enquête gehouden onder de lezers van de algemene nieuwsbrief en de nieuwsbrief Terminologie. 136 respondenten vulden de enquête in voor de algemene nieuwsbrief. De reacties waren overwegend positief. De vaste rubrieken werden gewaardeerd en de onderwerpen die het meest aanspreken zijn etymologie, nieuwe woorden, historische woordenschat, en spelling &

grammatica. Daarnaast gaven respondenten aan dat ze graag meer zouden willen lezen over streektalen en Nederlands in het onderwijs.

Voor de nieuwsbrief Terminologie vulden 20 respondenten de enquête in. Het verschil in reacties kan verklaard worden doordat de enquête onderaan in de nieuwsbrief stond, in tegenstelling tot de algemene nieuwsbrief waar hij juist helemaal bovenaan stond. Reacties op de terminologienieuwsbrief waren overwegend positief, een enkeling vond de teksten te ingewikkeld of formeel. De selectie aan onderwerpen werd goed bevonden, alleen gaven de nieuwsbrieflezers aan minder geïnteresseerd te zijn in evenementen/congressen en verslagen van bijeenkomsten. Onderwerpen die genoemd werden waar mensen meer over zouden willen lezen zijn medische taal, weetjes & feiten en toepassing in de praktijk.

#### **4. Huisstijl, sponsoring en populairwetenschappelijke activiteiten**

Het instituut trad in 2019 op als sponsor voor CLIN, de TABU-dag en de Digital Humanities Conferentie. In februari werden in het kader van het Rembrandtjaar de Rembrandt Tutorials gelanceerd door ING als hoofdsponsor van het Rijksmuseum. Het INT werkte mee aan dit project, en maakte een hertaling van hedendaags Nederlands naar 17<sup>e</sup>-eeuws. De hertaling heeft weer voor twee afgeleide projecten gezorgd: het gelegenheidswoordenboekje *Woorden voor Rembrandt* en de verkiezing *Rembrandtwoord van het jaar*.

Tijdens de Week van het Nederlands in oktober organiseerden we een goedbezochte publieksmiddag (circa 80 bezoekers) rond het thema ‘groene taal’ in de Theaterzaal van het Leidse Volkshuis. Voor die gelegenheid werd er ook een gedrukte versie van *Het Groene Woordenboekje* gemaakt. Het andere gelegenheidswoordenboekje dat in 2019 werd gemaakt, *Woorden voor Rembrandt*, werd opgenomen als onderdeel van de publicatie *De Taal van Rembrandt* van het Taalmuseum.

Eind oktober stond het instituut opnieuw op het DRONGO talenfestival met een lab over medische taal in het *Algemeen Nederlands Woordenboek* en Taalradar, en dialectloket.be. Wegens het eerdere succes van de linnen tasjes zijn hiervoor opnieuw tasjes gemaakt in de huisstijl van het INT, in oranje.

In november verscheen bij uitgeverij Scriptum het populairwetenschappelijke boek *Klimaatschieter, dwangzoen en kastelen in Spanje*. Het boek werd geschreven door Roland de Bonth en Dirk Geirnaert en is gebaseerd op de webrubriek *Terug in de taal*. Het boekje werd eind 2019 verstuurd als relatiegeschenk en daarbij werd een speciale nieuwjaarskaart ontworpen in de huisstijl van het INT.

Eind 2019 is in het kader van het Rembrandtjaar de verkiezing *Rembrandtwoord van het jaar* georganiseerd. Op de website [rembrandtwoord.ivdnt.org](http://rembrandtwoord.ivdnt.org) konden deelnemers stemmen. Er werden 7.996 bezoekers geregistreerd en er werd 818 keer gestemd. Het winnende woord *zwadderen* werd onthuld door auteur Onno Blom in de Pieterskerk in Leiden, waar het tot eind januari 2020 in de vorm van een lichtinstallatie te zien was.

Tot slot beantwoorden we door het jaar heen veel vragen over het Nederlands aan studenten, journalisten en programmamakers. De gecombineerde communicatieactiviteiten in 2019 hebben geleid tot meer bezoekers op de website, meer volgers op de sociale media, nieuwe samenwerkingen, publicaties in vaktijdschriften en kranten, artikelen op goedbezochte nieuwswebsites en aandacht op radio en tv. Een uitgebreid overzicht van publicaties en media-optredens is te vinden in het jaarverslag 2019.