/instituut voor de
Nederlandse taal/

Beleidsplan 2021

oktober 2020

Inhoudsopgave

Inleiding

Na de hervormingen van 2016 is het INT voluit aan de slag gegaan met het uitvoeren van al haar (kern)taken. In 2017 werd een meerjarenbeleidsplan voor de periode 2018-2022 geschreven en goedgekeurd. In 2018 werd meteen gestart met het uitvoeren van deze doelstellingen, en in 2019 en 2020 werden de grote lijnen van het meerjarenbeleidsplan verder uitgewerkt, met een aantal nieuwe accenten en dit in nauw overleg met de Nederlandse Taalunie. Het werkvloeroverleg (projectgericht) en het directieoverleg tussen de NTU en het INT werpen hun vruchten af en dit leidt tot duidelijkheid en prioritering van een aantal projecten en taken. Dit vertaalt zich ook in een heldere planning van projecten en in een begroting die deze taken duidelijk laat zien.

We baseren ons ook voor het beleidsplan 2021 op de hoofdlijnen van het meerjarenbeleidsplan van de NTU (2020 tot 2024). De aandachtsgebieden die de NTU benoemt, met name: 1. Standaardtaal, 2. Nederlands, taalvariëteiten en andere talen, 3. Onderwijs Nederlands binnen het taalgebied, 4. Onderwijs Nederlands buiten het taalgebied en 5. Taal en Cultuur komen terug in onze werkzaamheden. Dat vertaalt zich in diverse structurele werkzaamheden die hieronder geformuleerd staan, maar wordt ook zichtbaar in de diverse projecten met externe partijen.

Wat betreft de werking van het INT streven we ernaar om het werk in grotere gehelen op te delen, en het beleidsplan van 2021 is daar een reflectie van. Samenhang tussen projecten en synergie in de uitvoering van werkzaamheden spelen hierbij een belangrijke rol. Dit wordt nader uiteengezet in de sectie over de taalinfrastructuur van het INT.

Een belangrijke focus voor 2021 wordt het verder uitwerken van de ideeën uit de white paper The Future of Academic Lexicography -- A White Paper (Steurs et. al. (eds.), 2020) voor de lexicografische beschrijving van het hedendaags Nederlands. AI en nieuwe technologieën spelen daarbij een belangrijke rol. Maar ook het up-to-date brengen en houden van het hedendaags Nederlands corpusmateriaal is hiervoor van belang.

Het jaar 2021 zal ook in het teken staan van de visitatie. Deze is gepland voor oktober 2021 en wordt voorafgegaan door het schrijven van een zelfevaluatierapport. In dit rapport kunnen we aandacht besteden aan de hoofdactiviteiten van het INT, hoe het instituut vanaf de start in 2016 is gegroeid, wat de aandachtspunten zijn in het beleid, onderzoek en de dienstverlening.

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

Het INT streeft ernaar om het best gesorteerde en daarmee een zeer goed, toegankelijk wetenschappelijk instituut te zijn op het gebied van de Nederlandse taal en de woordenschat. Het

instituut ontwikkelt en levert data voor woordenboeken, (computationele) lexica, corpora en tools. De woordenboeken zijn online te raadplegen. Software en computerlinguïstische tools zijn open source beschikbaar.

Het instituut speelt in op de nieuwe ontwikkelingen in de Geesteswetenschappen, met name op het terrein van de Digital Humanities. Om deze rol te kunnen vervullen beheert en onderhoudt het INT een digitale infrastructuur voor het Nederlands, met aandacht voor taalvariatie (terminologie, dialecten etc.). Zowel academische als niet-academische partijen kunnen gebruikmaken van deze infrastructuur. We werken dan ook hard aan het meer zichtbaar maken van al onze taalmaterialen. In 2020 werd een totaal nieuwe website gelanceerd, met uitgebreide zoekfuncties, gericht op een divers doelpubliek en geschikt gemaakt voor mobiele telefoons.

De taalinfrastructuur van het INT

Het Instituut voor de Nederlands taal bevordert de beschrijving van de Nederlandse Taal in al haar facetten door het produceren en ter beschikking stellen van taalmaterialen als woordenboeken, corpora, lexica, grammatica's, en het aanbieden van hulpmiddelen voor de ontsluiting van deze data.

In de loop der jaren is het van steeds groter belang gebleken hierbij niet te denken in termen van afzonderlijke producten, maar vanuit een bredere visie op de samenhang van de diverse componenten van de wetenschappelijke beschrijving van de taal. Hiervoor is een infrastructuur waarbij het werk aan de diverse soorten data elkaar ondersteunt en kruisbestuift, van groot belang. Zo is het bijvoorbeeld weinig efficiënt als in meerdere projecten gewerkt wordt aan correcte spelling, paradigma-informatie en woordafbrekingen. Maar bijvoorbeeld ook het tot stand brengen van een centraal corpus waaruit de diverse projecten het nodige materiaal kunnen halen, is onderdeel hiervan.

Het instituut is inmiddels al geruime tijd bezig om daar waar mogelijk naar een zo groot mogelijke synergie toe te werken, zowel bij het plannen van activiteiten als bij de manier waarop ze uitgevoerd worden. Deze aanpak heeft een extra impuls gekregen door de in november 2019 gehouden Lorentz Workshop "The Future of Academic Lexicography" en het daaruit voortvloeiende white paper waarin plannen en mogelijkheden worden geschetst voor waar het INT naartoe kan wat betreft de academische lexicografie (Steurs et al. (eds.), 2020).

De kern van de lexicale infrastructuur is het computationeel lexicon GiGaNT, bestaande uit twee componenten, MoLex voor het hedendaags Nederlands en HiLex voor het historisch Nederlands, waaraan verschillende lexicale databases gekoppeld worden. De verbeteringen en uitbreidingen die de lexicografen uitvoeren in GiGaNT hebben zo effect in alle specifieke lexica en woordenboeken die door het INT gecreëerd worden. Door de koppelingen vloeien ook verbeteringen terug naar de centrale database. Zo wordt bijvoorbeeld zowel het ANW als het Referentiebestand Nederlands (RBN) gekoppeld aan MoLex, wat in 2021 kan resulteren in een verbeterd bereik van het ANW door middel van de definities die reeds aanwezig zijn in het RBN. De lexica en de gekoppelde datasets worden op

hun beurt weer geïntegreerd via linked open data in een internationaal meertalig netwerk (o.a. het onderdeel ELEXIS).

Hieronder zullen we de manier waarop de diverse onderdelen van de infrastructuur in elkaar grijpen aanschouwelijk maken.

GiGaNT

Het computationele lexicon GiGaNT wordt modulair opgebouwd. Enerzijds wordt er gewerkt aan de moderne lexiconcomponent, GiGaNT-Molex, de database die bijvoorbeeld de inhoud van *woordenlijst.org* levert. Anderzijds wordt er gewerkt aan de historische component, GiGaNT-Hilex, die taalmateriaal zal gaan bevatten vanaf de zesde eeuw tot ca. 1976. Beide modules gebruiken het modern Nederlandse lemma + woordsoort als ingang en zijn reeds grotendeels op lemmaniveau gekoppeld. Ze hebben een gemeenschappelijke relationele datastructuur en worden via eenzelfde platform bewerkbaar en toegankelijk gemaakt.

Aan het materiaal in de moderne lexiconcomponent wordt systematisch gewerkt via het project *spelling.* Het materiaal in GiGaNT-Molex bestaat uit gekeurmerkt materiaal. Het lexicon wordt verder aangevuld via de koppelingen met de diverse lexicale producten van het INT. De informatie die aan de lemmata in het lexicon wordt toegevoegd wordt mede bepaald door de projecten die met het lexicon gekoppeld zijn.

Het materiaal van de historische lexiconcomponent is gebaseerd op de lemmata en citaten van de historische woordenboeken, waarbij ervoor gezorgd wordt dat de koppeling met de historische woordenboeken blijft behouden.

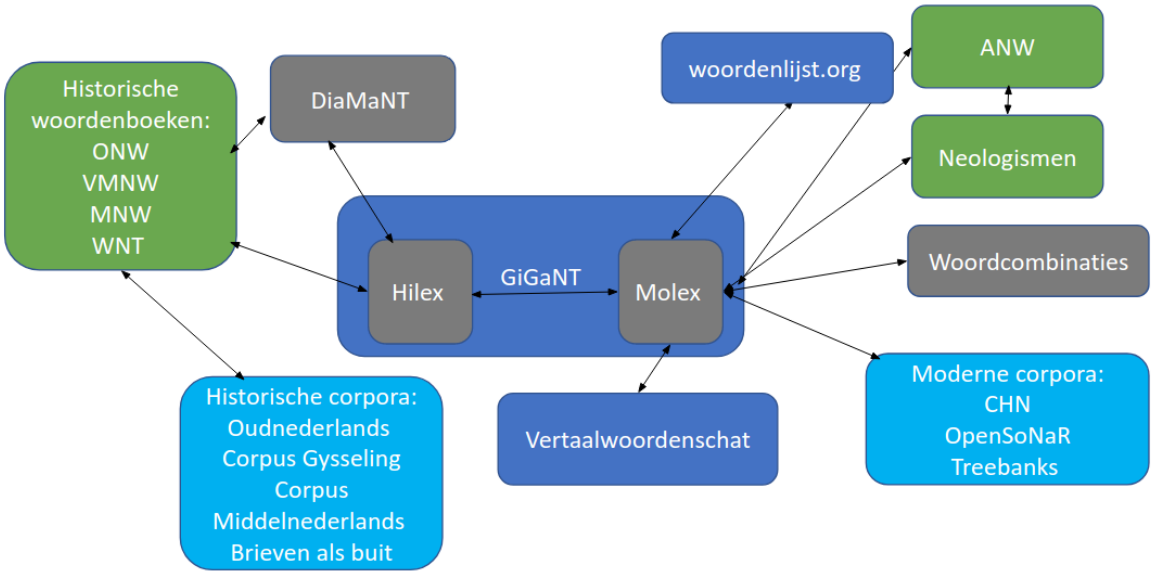
Beide computationele lexiconmodules zijn inzetbaar voor de taalkundige verrijking van corpusmateriaal. GiGaNT wordt via de lexiconservice (API) regelmatig ingezet voor query-expansie in zoeksystemen voor niet-taalkundig verrijkt materiaal, intern voor eigen corpusmateriaal maar ook extern, zoals in het zoekportaal *Delpher* van de KB, en is al gebruikt in diverse computationeel-linguïstische tools.

Lexicaal georiënteerde projecten

Hieronder vallen wat betreft het hedendaags Nederlands naast de werkzaamheden aan Molex, het werk aan het *Algemeen Nederlands Woordenboek*, neologismen, woordcombinaties, de spellingdata voor *woordenlijst.org* en de diverse terminologieprojecten.

Voor het historisch Nederlands zijn dat, afgezien van Hilex, de historische woordenboeken, het diachroon semantisch lexicon DiaMaNT en sinds 2020 de etymologiebank.

Voor het dialectmateriaal is er o.a. sinds september 2020 het DSDD-portaal en het e-WND. Ten slotte zijn er ook de diverse vertaalwoordenboeken die het INT in beheer heeft, cureert en online zet.



Corpora

De corpusinfrastructuur van het INT omvat naast data een arsenaal aan gereedschappen voor data processing en ontsluiting, zie het onderdeel software.

Hedendaags corpusmateriaal

Hier staat het Corpus Hedendaags Nederlands centraal. Het CHN is een omvangrijk monitorcorpus dat materiaal bevat uit Nederland, Vlaanderen, Suriname en de Antillen en dat iedere week aangevuld wordt met het meest recente materiaal. Het materiaal dat extern mag worden gepubliceerd wordt voor onderzoekers ontsloten door middel van de zoekapplicatie; het overige materiaal wordt intern gebruikt voor onderzoek. Het Corpus Hedendaags Nederlands wordt automatisch verrijkt met lemma en woordsoort. Daarnaast wordt in toenemende mate gebruikgemaakt van syntactische analyse, onder andere ter ondersteuning van het project Woordcombinaties. De samenstelling van het corpus wordt mede bepaald door de projecten die van het corpus gebruikmaken.

Historisch corpusmateriaal

Belangrijke historische corpora in het beheer van het INT zijn bijvoorbeeld het Corpus Oudnederlands, het Corpus Gysseling, het Corpus Middelnederlands, en het corpus Brieven als Buit. Op het corpus Middelnederlands na zijn deze corpora ook taalkundig verrijkt en geverifieerd, wat de waarde voor onderzoek vergroot. Daar worden de komende jaren meerdere corpora aan toegevoegd.

Dialectaal corpusmateriaal

Het INT heeft, afgezien van de historische corpora, geen corpus met dialectmateriaal. Sedert 2020 echter is het INT betrokken bij de totstandkoming van het gesproken corpus van Zuidelijk-Nederlandse dialecten.

Grammatica, terminologie en vertaalwoordenschat

De onderdelen grammatica en terminologie zijn relatief nieuw voor het instituut en staan nog min of meer op zichzelf; de vertaalwoordenschat neemt als enige niet-monolinguale onderdeel natuurlijk een aparte plaats in. Wel kan worden vermeld dat de terminologie-extractietool TermTreffer GiGaNT-Molex en corpusmateriaal uit het CHN gebruikt. De vertaalwoordenschat wordt momenteel via het RBN aan Molex gekoppeld.

Computationele linguïstiek en softwareontwikkeling

Voor het ontwikkelen en exploiteren van de taaldata is ondersteuning door computationeel linguïstische en andere softwaretools essentieel.

De computationele infrastructuur van het INT bestaat uit een aantal cruciale componenten:

- het corpuszoeksysteem BlackLab en het bijbehorende userinterface (corpus-frontend) voor ontsluiting van corpora;
- de pijplijn DUCT voor bestandsconversie, taalkundige verrijking en indexering van corpusmateriaal;
- het Rapid Database Application Development platform Lex'it voor bewerking van lexicale (en andere gestructureerde) data, in gebruik voor onder andere HileX, MoleX, DSDD en vele andere projecten;
- de woordenboekeditor "SwingLex" (INL-DWS), met specialisaties voor ANW, Neologismen en Woordcombinaties;
- de lexiconservice waarmee het INT GiGaNT in de vorm van een API gepubliceerd wordt;
- generieke componenten voor het publiceren van woordenboeken, onder meer toegepast bij de publicatie van woordcombinaties en het Lexicon Frisicum.

Daarnaast worden specialisaties ontwikkeld binnen specifieke projecten en voor specifieke doeleinden. Daarbij wordt er steeds naar gestreefd de ontwikkeling van de core-infrastructuur verder te zetten. Zo ontwikkelen we in CLARIAH-PLUS uitbreidingen op de workflow voor taalkundige verrijking en breiden we BlackLab uit met functionaliteit voor treebanks en parallelle corpora.

CLARIN-centrum en taalmaterialen

Het INT maakt als B-centrum deel uit van de CLARIN Europese infrastructuur en speelt een belangrijke rol bij het ter beschikking stellen van voor het Nederlands relevante taaldata. Op termijn wordt het INT ook een CLARIN Knowledge centre.

Hedendaags Nederlands

Infrastructureel

GiGaNT-Molex

De moderne lexiconcomponent is de centrale ruggengraat voor de lexicale beschrijving van het modern Nederlands, en wordt voortdurend uitgebreid met nieuwe data, niet alleen in de zin van toegevoegd vocabulaire, maar ook door uitbreiding van de features, zoals in 2020-2021 met toegevoegde werkwoordkenmerken die het woordcombinatieproject ten goede komen en het toevoegen van uitspraakinformatie (zie hiervoor het onderdeel Spelling). Om de informatie efficiënt te propageren wordt verder de koppeling met diverse datasets uitgebreid. De koppeling met het RBN zal worden afgerond en worden uitgebreid naar de vertaalwoordenschat; de synergie tussen Molex, ANW en Neologismen wordt verder geoptimaliseerd. Het lexicon is beschikbaar als dataset onder een niet-commerciële en commerciële licentie, en via de lexiconservice.

Corpus Hedendaags Nederlands

De moderne lexiconbouw en de beschrijving van het hedendaags Nederlands wordt gebaseerd op modern corpusmateriaal. De corpusdata voor het hedendaags Nederlands zitten in het *Corpus Hedendaags Nederlands*. De werkzaamheden bestaan enerzijds uit acquisitie, dataprocessing en taalkundige verrijking, en anderzijds uit het beschikbaar stellen van het materiaal in een corpusapplicatie voor intern en extern gebruik. Het materiaal is afkomstig uit Nederland, Vlaanderen, Suriname en de Antillen.

In 2020 is het corpus flink uitgebreid met nieuwe bronnen, zoals bijvoorbeeld Vlaams krantenmateriaal en forumdata uit Nederland en Vlaanderen en de Antillen. Voor 2021 staat parlementair materiaal op het programma, dat een extra impuls krijgt door deelname aan het CLARIN ParlaMint-project voor de ontwikkeling van vergelijkbare parlementaire corpora binnen Europa. Er zal ook verder gegaan worden met de acquisitie van nieuw taalmateriaal, mede gestuurd door de wensen van de diverse projecten die van het corpus gebruikmaken.

In 2021 publiceren we een nieuwe versie van het corpus in de laatste versie van de corpusapplicatie, waarbij ook het mechanisme voor regelmatige updates in productie wordt genomen, zodat gebruikers steeds in de gelegenheid zijn het meest recente Nederlands te onderzoeken.

Op iets langere termijn werken we aan de verbetering en uitbreiding van de taalkundige verrijking van het corpus, onder andere door het integreren van syntactische verrijking in de corpusworkflow en de zoekapplicatie. Het is ook de bedoeling om de huidige corpusworkflow te evalueren en te optimaliseren.

Lexicale beschrijving van het hedendaags Nederlands

Algemene woordenschat

De beschrijving van de algemene woordenschat gebeurt op drie verschillende terreinen: de spelling van woorden, de betekenis en de constructies waarin ze voorkomen. De informatie wordt beschikbaar gesteld via respectievelijk *woordenlijst.org, anw.ivdnt.org* en *woordcombinaties.ivdnt.org.* In 2021 wordt daaraan nog de neologismenapplicatie toegevoegd. Belangrijk voor 2021 wordt de uitwerking van de inzichten van de white paper *The Future of Academic Lexicography -- A White Paper* (Frieda Steurs et. al. (eds.)), en met name voor de semantische beschrijving van het Nederlands. Hieronder volgt per deelgebied een gedetailleerder beschrijving van de plannen.

Spelling

In 2021 wordt de in 2020 ingezette lijn om bij de updates voor woordenlijst.org in te zetten op kwaliteit boven kwantiteit voortgezet. Hierbij is nu naast potentiële spellingproblemen ook corpusfrequentie in recent materiaal een belangrijk criterium. Het spellingbestand zal worden uitgebreid met informatie uit de bestanden van de werkgroep BAN van de Commissie Anderstalige Namen. Daarnaast zal gestart worden met het toevoegen van uitspraakinformatie aan de online woordenlijst.

De overname van de beheer en onderhoud van de Spelling API wordt afgerond. Daarnaast wordt ingezet op het ontwikkelen van een nieuw backend op de lexicale data waarmee beter kan worden ingespeeld op de bestaande wensen voor uitbreiding van de functionaliteit. Deze ontwikkeling maakt het onder andere eenvoudiger nieuwe functionaliteit toe te voegen, zoals zoeken op meerdere kenmerken, waaronder ook woordsoort, flexibeler omgaan met de presentatie van resultaten, relevantere suggesties bij het zoeken, en het gebruik van de uitspraakinformatie.

Semantische beschrijving

Het *Algemeen Nederlands Woordenboek* (ANW) en het neologismenproject zijn verantwoordelijk voor de semantische beschrijving van het hedendaags Nederlands. De beide projecten maken gebruik van het CHN, en de beschrijving van de woorden wordt opgeslagen in dezelfde databank. Dat betekent dat bepaalde neologismen indien gewenst ook direct in het ANW gepubliceerd kunnen worden. Voor 2021 staat ook een release van de neologismenapplicatie op het programma. De semantische beschrijving van het hedendaags Nederlands wordt ook in 2021 voortgezet.

Er is al het nodige werk gedaan om de lexicografische werkomgeving voor ANW en neologismen zo in te richten dat er een koppeling is met de GiGaNT-Molex, zowel inhoudelijk als technisch. Dat betekent dat paradigmatische informatie uit de centrale database wordt gehaald zodat informatie over inflectie niet meer in de afzonderlijke projecten hoeft worden ingevoerd. Deze koppeling wordt in 2021 verder uitgebouwd.

Deze aanpak sluit aan bij de aanbevelingen uit de white paper *The Future of Academic Lexicography -- A White Paper* (Frieda Steurs et. al. (eds.)). Er is echter nog meer onderzoek nodig om te kijken welke nieuwe technologieën het lexicografisch beschrijvingsproces verder kunnen ondersteunen. Het is de bedoeling om de ideeën hierover in 2021 verder uit te werken en te toetsen.

Nederlands als vreemde taal

Hiervoor wordt gewerkt aan *Woordcombinaties*, een online taaltool die leerders van het Nederlands als vreemde taal ondersteunt bij het gebruiken van woorden in context. Met dit project wil het INT werk maken van een meer systematische inventarisatie en beschrijving van combinaties (collocaties, idiomen en patronen) in het Nederlands. In 2020 is een demo van *Woordcombinaties* verschenen. In 2021 zal de tool officieel worden gereleased.

Op dit moment ligt de focus op de bewerking van voorbeeldzinnen en combinatiemogelijkheden van werkwoorden uit de basislijst Schooltaalwoorden van het iTTA (Instituut voor Taalonderzoek en Taalonderwijs Anderstaligen) van de Universiteit van Amsterdam en het frequentiewoordenboek 'a Frequency Dictionary of Dutch'. De basislijst bevat algemene schooltaalwoorden en vakgerelateerde woorden die anderstaligen nodig hebben om de lessen in het secundair onderwijs goed te kunnen volgen. Ook voor de hogere NT2-niveaus is een goede en systematische fraseologische beschrijving van deze gangbare werkwoorden onmisbaar. In 2021 zal hier verder aan gewerkt worden. Tevens zal er tijd gereserveerd worden voor de redactionele bewerking van de patronen.

Door een systematische beschrijving van combinaties (collocaties, idiomen en patronen) werken we op termijn toe naar een ‘constructicon’ van de Nederlandse taal (een inventaris van constructies).

Terminologie

De nieuwe website van het INT heeft een performante zoekmachine om snel en efficiënt alles over het Expertisecentrum Terminologie (ENT) te kunnen vinden. In 2021 zal dit centrum verder worden uitgebouwd.

Termenlijsten

Termenlijsten documenteren de mate waarin een taal zich voorbij de gangbare dagelijkse communicatiebehoeften in meer specialistische domeinen blijft ontwikkelen. Ze vormen ook een handige vraagbaak voor de vele taalgerichte beroepsbeoefenaars zoals vertalers en technische schrijvers. We blijven daarom permanent de hedendaagse termenlijsten op de ENT-webpagina's updaten en uitbreiden. Analoog hiermee wordt in de lijn van het INT en zijn historische woordenboeken en corpora ook een luik voor lijsten met historische termen ingericht. Ook hiervoor is de internationale en grondige Library of Congress-classificatie goed bruikbaar.

Tools

Wat de terminologietools betreft staan werkzaamheden aan TermTreffer en de terminologiebeheertool Termbeheerder op het programma. Deze laatste is overeenkomstig ENT-advies aangepast en zou nu ook de import van geëxtraheerde termen uit TermTreffer toelaten. Tests hiervoor zijn mogelijk zodra de nieuwe versie van TT operationeel is.

Veldondersteuning

Klassieke veldondersteuning in de vorm van verdere updates voor de bestaande websiterubrieken, vooral de opleidingen in Nederland en Vlaanderen, blijft aan de orde. Ook de evenementenrubriek voor terminologie dient continu te worden bijgehouden en er worden jaarlijks 4 uitgebreide nieuwsbrieven Terminologie gedistribueerd.

Nu het pilotproject over hogeronderwijstermen in de HOTNeV-termbank is afgerond, wordt dit voortgezet via stages en scripties. De groei van deze databank is uiteraard afhankelijk van de stages die bij het ENT worden aangevraagd en de begeleiding daarvan. Begeleiding van studenten blijft belangrijk en ons stageaanbod voor terminologiewerk wordt blijvend gepromoot, het is ook belangrijk voor de netwerkpositie van het ENT. Afstudeerscripties van studenten kunnen eveneens bijdragen aan de verdere invulling van de HOTNeV-termbank.

Speerpunt 1: de medische vaktaal

Het streven is om in 2021 te beginnen met onderzoek voor een platform medische terminologie. Zo'n platform zou dan vooral drie taalregisters moeten koppelen: een vaktalig register, een vereenvoudigd

register voor communicatie met patiënten en een register met informele termen. De bedoeling is om daarvoor de Ontoterminology-editor in te zetten die is ontwikkeld aan de Université Savoie Mont-Blanc.

Voor het medische domein zal ook een versie van het Pinkhof geneeskundig woordenboek online worden gezet. Dit woordenboek wordt bijgewerkt met als primaire gebruiksfunctie: verklarend hedendaags medisch woordenboek en een taalbaak voor medisch Nederlands te vormen. Om dit te realiseren wordt er samengewerkt met de Stichting Beheer Pinkhof-database.

Speerpunt 2: de juridische vaktaal

Wat betreft het juridisch woordenboek van M.C. Oosterveld-Egas Reparaz en Johanna Vuyk-Bosdriesz wordt verder gewerkt aan de updating en uitbreiding van het bestand. Daarnaast zal in 2021 worden gekeken of er met Belgische rechtsfaculteiten een financiering kan worden aangevraagd om de Belgische rechtstermen toe te voegen aan deze collectie. Er wordt tevens onderzocht hoe de databank op een gebruikersvriendelijke manier online raadpleegbaar kan worden gemaakt.

Speerpunt 3: Nederlands als wetenschapstaal

Er zijn heel wat initiatieven die de aandacht vestigen op de noodzaak aan talige hulpmiddelen om voor studenten de overstap van het middelbaar naar het hoger onderwijs makkelijker te maken. We werken in dit verband samen met het Proefproject Nederlands als wetenschapstaal - van corpora naar terminologielijsten. Dit project is een samenwerking tussen Stichting Nederlands / Vlaams Platform Taalbeleid Hoger Onderwijs, KU Leuven, UGent en het INT.

In 2021 zal er in de context van het SaBeD-project (zie verderop) een corpus van gesproken academisch Nederlands worden samengesteld, dat door het INT wordt gehost. Er wordt een keuze gemaakt uit een groot aantal academische vakken op het niveau van BA1. Voor deze vakken zal het INT ook aangepaste terminologielijsten ontwikkelen.

Vertaalwoordenboeken

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

Inmiddels staan het Nederlands-Nieuwgrieks / Nieuwgrieks-Nederlands, het Nederlands-Portugees / Portugees-Nederlands, het Nederlands-Estisch en sinds eind 2020 ook het Nederlands-Fins / Fins-Nederlands online.

Aangezien deze tweetalige bestanden veelal 10 tot 20 jaar oud zijn is in 2020 gewerkt aan het bijwerken van de spelling van de Nederlandse glossen en voorbeeldzinnen. In 2021 zal verder gewerkt worden aan de koppeling van de bestanden aan GiGaNT-Molex, zodat alle informatie hieruit ook gebruikt kan worden ten behoeve van de Vertaalwoordenschat.

Om correcties en inhoudelijke updates in de toekomst op een gebruiksvriendelijke manier te kunnen realiseren zal in 2021 een redactieomgeving ontwikkeld worden. Daarnaast zal, in overleg met de verschillende uitgevers van uitgaven in boekvorm, gekeken worden naar de mogelijkheden om deze uitgeverijen online vertaalwoordenboeken betaald aan te laten bieden.

Grammatica

Vanaf 2020 valt grammatica binnen de structurele basistaken van het INT. Dit betekent dat het INT zorgt voor de ontwikkeling, het beheer en de beschikbaarstelling van de verschillende digitale grammaticaproducten. Onder deze producten vallen vooralsnog het Taalportaal en de e-ANS. Voor de langere termijn zijn er ook plannen om de taaladviezen van taaladvies.net aan deze producten te koppelen. Afgezien van de werkzaamheden aan de e-ANS en het Taalportaal (zie hieronder), zal gewerkt worden aan een functioneel en technisch ontwerp van een geïntegreerd grammaticaportaal, een webpagina die de spil moet gaan vormen van alle grammaticaonderdelen en zal fungeren als ontvangstpagina voor geïnteresseerde gebruikers, met informatie over projecten en producten, een zoekfunctie voor alle producten en een loket voor vragen.

e-ANS

In het eerste kwartaal van 2021 zal de nieuwe ANS-website met enkele herziene hoofdstukken worden gelanceerd. Daarbij wordt ervoor gezorgd dat ook de herziene klankleer (‘ANK’) eraan is toegevoegd. Er volgt verder in 2021 nog een tweede update met reeds herziene hoofdstukken van de ANS.

Daarnaast zal verder gewerkt worden aan de herziening van de rest van de ANS. Deze werkzaamheden betreffen o.a. de herschrijving van een nieuw hoofdstuk, de coördinatie van enkele andere hoofdstukken door externen, eindredactie en valorisatie.

Vanaf november 2020 tot voorjaar 2021 loopt voor de ANS ook een pilot voor de ontwikkeling van een didactische laag voor de ANS. Daarbij zal voor een van de herziene hoofdstukken didactisch materiaal worden ontwikkeld voor de doelgroep neerlandici extra muros, en een verslag worden gemaakt met aanbevelingen voor verdere ontwikkeling.

Taalportaal

Afgezien van regulier data-onderhoud is voor 2021 een vervanging gepland van de huidige Taalportaal-webapplicatie door de voor het project ‘Taalportaal Zuid-Afrika’ gebouwde

webapplicatie. Er zal verder gewerkt worden aan de conversie en publicatie van het nieuwe deel ‘Coordination’ van Hans Broekhuis.

Historisch Nederlands

Infrastructureel

Historisch woordenboekenportaal

De beschrijving van de historische woordenschat is te vinden in de historische woordenboeken van het INT. Deze woordenboeken zijn online beschikbaar in het historische woordenboekenportaal (gtb.ivdnt.org). In die applicatie zijn de belangrijkste historische woordenboeken van het Nederlands opgenomen: het *Oudnederlands Woordenboek* (ONW), het *Vroegmiddelnederlands Woordenboek* (WNT), het *Middelnederlandsch Woordenboek* (MNW) en het *Woordenboek der Nederlandsche Taal* (WNT). De data van deze woordenboeken zijn een bron voor GiGaNT en DiaMaNT (zie hierna).

In 2021 wordt het werk aan de verbetering van de woordenboekbestanden samengebracht in een nieuwe bestandsversie volgens de TEI P5-richtlijnen, waarin de structuurherkenning verbeterd is. Tevens kunnen reeds bekende errata in deze bestandsversie worden opgelost. Deze versie zal worden gepubliceerd in het portaal.

GiGaNT-Hilex

Het is de bedoeling dat GiGaNT het centrale lexicon wordt, waaraan alle lexicale producten van het INT gekoppeld zijn. Het historische onderdeel GiGaNT-Hilex is gebaseerd op de historische woordenboeken. Het huidige Hilex bevat het materiaal uit het *Woordenboek der Nederlandsche Taal* en het *Middelnederlandsch Woordenboek.* De koppeling met de online woordenboeken is behouden. Het lexicon is voor onderzoekers en applicatieontwikkelaars beschikbaar via de lexiconservice via maandelijkse updates.

Omdat binnen GiGaNT-Hilex met striktere lemmatiseringsprincipes wordt gewerkt dan de woordenboeken waarop het lexicon is gebaseerd, vinden er allerlei herschikkingen plaats van de lexiconinhoud. Deze werkzaamheden worden in 2021 voortgezet. Daarnaast zal verder gewerkt worden aan koppeling met de moderne component van het centrale lexicon.

Semantisch lexicon DiaMaNT

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

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

In 2021 zal verder gewerkt worden aan uitbreidingen die het lexicon meer geschikt zullen maken om taalkundig onderzoek naar trends in betekenisontwikkeling te ondersteunen. Het gaat met name om het expliciteren van betekenisrelaties ten behoeve van de semasiologische component en fijnmaziger koppeling van synoniemen ten behoeve van de onomasiologische component.

Historische corpora

In 2020 is verder geïnvesteerd in optimalisering van het corpus frontend. Naast taalkundig verrijkt materiaal kan er ook niet-verrijkt corpusmateriaal beter doorzoekbaar worden gemaakt door inzet van de lexiconservice. Drie corpora zijn online gezet, en de onlinegang van een aantal nieuwe corpora is reeds in voorbereiding.

Mede in het kader van het CLARIAH+-project is een nieuw voorstel voor een flexibele tagset voor diachroon Nederlands uitgewerkt (“TDN”). In 2021 zullen de beschikbare historische corpora worden omgezet overeenkomstig TDN. Verder zal nieuwe gold standaard data worden ontwikkeld voor historisch Nederlands; ook deze data zullen online worden gepubliceerd in de corpusapplicatie.

Het corpus historische kranten dat door Nicoline van der Sijs in samenwerking met het Meertens Instituut en een groep vrijwilligers gedigitaliseerd is, zal verder worden gecureerd.

Etymologie

Sinds 2020 zijn de etymologiebank (etymologiebank.nl), en de uitleenwoordenbank (uitleenwoordenbank.ivdnt.org) in het beheer van het INT. Voor 2021 zijn inhoudelijke uitbreidingen voorzien voor de etymologiebank.

Beschrijving van de Nederlandse dialecten

Sedert 2020 hebben de dialecten van het Nederlands een belangrijke plek gekregen op het INT. Een hoogtepunt was de lancering van de Database van de Zuidelijk-Nederlandse Dialecten. Het instituut heeft ook de elektronische Woordenbank van de Nederlandse dialecten onder haar hoede gekregen.

Database van de Zuidelijk-Nederlandse Dialecten (DSDD)

Op 30 september 2020 werd de DSDD gelanceerd (dsdd.ivdnt.org). DSDD staat voor de database of the Southern Dutch Dialects of de database van de Zuidelijk-Nederlandse Dialecten, een pilotproject

dat liep van 2017 tot 2020, met de bedoeling minstens 1500 concepten van de drie regionale dialectwoordenboeken (Woordenboek van de Vlaamse/Brabantse/Limburgse dialecten) aan elkaar te koppelen. Bij de lancering op 30 september bevatte de database al meer dan 10.000 concepten.

In 2021 wordt verder gewerkt aan het vullen van de database met ontbrekend materiaal. In de eerste plaats betreft het de resterende gegevens uit de drie regionale woordenboeken die ten grondslag liggen aan de database. Van de algemene woordenschat moeten nog wat ontbrekende gegevens worden aangevuld. Daarnaast zal gewerkt worden aan de toevoeging van landbouwwoordenschat en woordenschat van de vaktalen, beide beperkt vertegenwoordigd in het pilotproject.

Verder zal worden onderzocht hoe de gegevens voor het Zeeuwse taalgebied (ook een Zuidelijk-Nederlands dialect) kunnen worden toegevoegd. De uitdaging hierbij is dat het Zeeuws alleen een semasiologische beschrijving heeft. Het DSDD-portaal is onomasiologisch van opzet. Om tot een goede strategie te komen voor het toevoegen van semasiologische bronnen aan het portaal zal bij de analyse ook een aantal lokale dialectwoordenboeken uit de woordenbanken van Nederland en Vlaanderen (eWND, gehost door INT, en woordenbank.be) betrokken worden. Op lange termijn beogen we een dialectplatform voor het hele Nederlandse taalgebied te realiseren.

Atlas van het dialect in Vlaanderen

Samen met de UGent werkt het INT mee aan de Atlas van het dialect in Vlaanderen, die wordt uitgegeven bij Lannoo. Het boek (te vergelijken met de Atlas van de Nederlandse Taal) zal in het najaar van 2021 verschijnen. Het INT is verantwoordelijk voor ongeveer 20 bijdragen in het boek.

Gesproken Corpus van de Zuidelijk-Nederlandse Dialecten

Het INT is partner in het project Gesproken Corpus van de Zuidelijk-Nederlandse Dialecten, een project dat loopt van 2020 tot 2024 en gerealiseerd wordt aan de UGent. Het project beoogt de ontsluiting van een collectie van dialectopnames uit 768 plaatsen in België, Frankrijk en het zuiden van Nederland, opgenomen tussen 1963 en 1976 (te beluisteren via www.dialectloket.be en op de Nederlandse dialectenbank: https://www.meertens.knaw.nl/ndb/).

De opnames worden volgens een nieuw ontwikkeld transcriptieprotocol getranscribeerd om vervolgens met bestaande tools taalkundig verrijkt te worden. Het INT zal de audio, de transcripties en de annotaties op termijn vrij online beschikbaar en doorzoekbaar maken en duurzaam bewaren. Het INT heeft de eerste jaren van het project een adviserende rol.

Taalmaterialen

Een structurele taak van het INT is het beheren en ter beschikking stellen van taalmaterialen, te vinden op de in 2020 vernieuwde website. Onderdeel van de beheertaak is het informeren en adviseren van gebruikers. De kennis die hierbij is opgedaan zal verzameld worden en gepubliceerd ten behoeve van het geplande CLARIN Knowledge centre voor de Nederlandse taal (zie hierna).

CLARIN-ERIC; het INT als CLARIN-centrum

Het INT was betrokken bij de indiening van CLARIN-VL , als 3e partij, bij de Vlaamse FWO/EWI-call met betrekking tot International Research Infrastructures. Dit project werd ingediend met als doel steun te krijgen vanuit het Vlaamse departement Economie, Wetenschap en Innovatie (EWI) voor de oprichting van CLARIN-België. Als deze aanvraag goedgekeurd wordt staat deze oprichting op de agenda voor 2021.

Voor het INT bestaan de voorziene taken voor 2021 uit het verder organiseren van zogeheten User Involvement events, waarbij CLARIN onder de aandacht gebracht wordt van onderzoekers in de Digital Humanities; uit het opnemen van tools en datasets gemaakt door Vlaamse onderzoekers en de integratie hiervan in de Europese CLARIN-infrastructuur; en uit het delen van tools en modellen met Vlaamse (en andere) onderzoekers. Dit zijn taken die sowieso gepland staan voor 2021.

Voor 2021 wordt voorzien dat het INT een CLARIN Knowledge centre wordt met als focus het Nederlands. Dit houdt in dat het INT een expertisecentrum wordt voor CLARIN-gebruikers die informatie en raad willen omtrent het Nederlands, zijn bronnen en tools. Dit behelst het voorzien van Engelstalige webpagina's waarin internationale onderzoekers hun weg vinden, en de formele goedkeuring van het INT als CLARIN-K centre.

European Language Resources Coordination Initiative (ELRC)

Het INT is betrokken bij het ELRC-initiatief. Dit is een doorlopend initiatief dat als doel heeft tekstdata te verzamelen in alle EU-lidstaten, IJsland en Noorwegen, die gebruikt kunnen worden om CEF eTranslation verder te ontwikkelen. CEF eTranslation is een automatische vertaaldienst die door de Europese Commissie ter beschikking wordt gesteld om meertalige communicatie tussen openbare diensten, ministeries en burgers mogelijk te maken. De kwaliteit van een automatische vertaling hangt onvermijdelijk samen met de kwaliteit en kwantiteit van de taalbronnen die worden gebruikt om het systeem te 'trainen '. Grote hoeveelheden taaldata zijn dan ook nodig om de kwaliteit van de Nederlandse vertalingen te verbeteren. Dit alles is van groot belang om taalbarrières in Europa te slechten en de nationale talen, in dit geval de Nederlandse taal, te behouden in de digitale informatiemaatschappij.

In 2021 zullen er verschillende vergaderingen, conferenties en workshops georganiseerd worden om het ELRC-project verder te promoten zowel binnen Europa als binnen Nederland.

Impactcentrum en digitization.eu

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

In het najaar van 2020 is de COST action aanvraag *DIALOGUE* (DIgitisAtion knowLedge fOr diGital hUmanities and cultural hEritage) opnieuw ingediend, 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. De werkzaamheden m.b.t. digitalisering die in de context van CLARIAH plus worden uitgevoerd, worden in samenwerking met het Centre uitgevoerd.

Overige infrastructuur- en netwerkprojecten

Infrastructuurprojecten

European Lexicographic Infrastructure (ELEXIS)

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

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

Het INT leidt het werkpakket ‘Lexicographic data and workflow’. Daarnaast werkt het INT mee aan andere werkpakketten, met name de werkpakketten ‘Interoperability and Linked (Open) Data’, ‘Lexicographic data for NLP’, ‘NLP for lexicography’ en ‘Training and Education’.

CLARIAH+ Nederland

Het CLARIAH (Common Lab for Research in the Arts and Humanities) CORE-project (2015-2018) was erop gericht een gemeenschappelijke infrastructuur tot stand te brengen voor data-intensief wetenschappelijk onderzoek in de geesteswetenschappen. Vanaf begin 2019 tot en met 2023 loopt het vervolgproject CLARIAH-PLUS, waarin het accent nog meer gericht is op het concreet ondersteunen van de onderzoeker door middel van het tot stand brengen van (virtuele) onderzoeksomgevingen.

Het INT houdt zich onder andere bezig met een verbetering van de infrastructuur voor historisch Nederlands, uitbreiding op de corpuszoekmachine BlackLab naar parallelle corpora en treebanks, hulpmiddelen voor het aanbrengen van persistente gebruikersannotaties in corpuszoekresultaten, een gebruikersvriendelijker digitalisatieworkflow en curatie van dialectwoordenboekdata.

Het werk zal zich in 2021 vooral richten op het doorontwikkelen van de infrastructuur voor historisch Nederlands en begin van implementatie van het doorzoeken van treebanks in de BlackLab-corpusomgeving. Dit laatste voegt een belangrijk aspect toe aan de infrastructuur voor omvangrijke corpora, en komt ook de doorzoekbaarheid van moderne corpora en het werk aan de woordcombinaties ten goede. Daarnaast wordt gewerkt aan de verbeterde infrastructuur voor digitalisatie en conversie van tekstmateriaal.

Clariah Vlaanderen

Het INT was betrokken bij de indiening van CLARIAH-VL: Advancing the open humanities service infrastructure, als 3^{e} partij, bij de Vlaamse FWO/EWI-call met betrekking tot International Research Infrastructures. Als dit project gehonoreerd wordt dan is de hoofdtaak van het INT het voorzien van de benodigde infrastructuur voor het opzetten van het *Digital Text Analysis Dashboard & Pipeline.* Het doel van deze infrastructuur is om onderzoekers uit de Digital Humanities toe te staan teksten van automatische annotaties te voorzien, zonder van hen een technische achtergrond te verwachten, en dit d.m.v. een cloud-based systeem waarbij teksten geupload kunnen worden. Hiervoor is het noodzakelijk om, in samenwerking met de Vlaamse CLARIN/CLARIAH-partners tools zoals taggers en parsers te benchmarken, zodat de beste tools ter beschikking gesteld kunnen worden. Bij goedkeuring van het project zullen deze werkzaamheden starten in 2021.

Er wordt in het kader van CLARIAH-VL ook verder gewerkt aan een pilotproject in samenwerking met de Vlaamse Super Computer (VSC), waarbij het plan is om een BERT-taalmodel te trainen op basis van de corpora hedendaags Nederlands waarover het INT beschikt. Dit project dient als test voor

zowel de VSC als CLARIAH-VL om de gebruiksvriendelijkheid van de toegang tot de supercomputers te verbeteren, zodat deze ook makkelijker bruikbaar worden voor onderzoekers in de Digital Humanities.

SignOn-project

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

De taak van het INT bestaat hoofdzakelijk uit het opzetten van de infrastructuur voor dit onderzoek. In eerste instantie wordt een video (inclusief gebarentaal) + audio + autocue + ondertitels corpus verzameld, onder meer op basis van data van de VRT. Taak van het INT is om dit corpus beschikbaar en doorzoekbaar te maken. Een andere taak van het INT is om de infrastructuur op te zetten om Vertalen-als-een-service aan te kunnen bieden, die dan aangesproken kan worden binnen de Android- en iPhone-apps die ontwikkeld worden in de use cases, die in samenspraak met de doelgroepen ontwikkeld worden.

SABeD: Spoken Academic Belgian Dutch

Het industrieel onderzoeksfonds KU Leuven heeft het project Spoken Academic Belgian Dutch goedgekeurd, dat twee jaar zal duren en start en het voorjaar van 2021. Het project werd aangevraagd door Elke Peters van het Centrum voor Taal en Onderwijs, in samenwerking met twee onderzoeksgroepen die deel uitmaken van Leuven.AI: het Centrum voor Computerlinguïstiek en de ESAT-PSI Speech groep. Het INT is in deze aanvraag derde partij, en zal zorgen voor de opname van het corpus in de CLARIN-infrastructuur, zowel als download voor onderzoek als online doorzoekbaar, op gelijkaardige wijze als nu het geval is voor het Corpus Gesproken Nederlands in de OpenSonar-toepassing.

Hoorcolleges zijn typisch voor het hoger onderwijs. In hoorcolleges leren studenten nieuwe lesinhouden in een taalregister waarmee ze weinig vertrouwd zijn, academisch Nederlands. Het doel van dit project is (1) om een corpus academisch gesproken Nederlands te compileren en (2) hierbij de

effectiviteit van spraaktechnologie te onderzoeken voor automatische transcriptie van gesproken teksten, (3) om nadien een woordfrequentielijst academisch gesproken Nederlands en (4) een woordenschattoets academisch gesproken Nederlands te kunnen ontwikkelen. De compilatie van dit corpus laat toe leermateriaal en toetsen voor instromers te creëren. Het corpus zal een belangrijk hulpmiddel zijn voor zowel onderzoekers als beleidsmakers.

Netwerkprojecten (COST)

enetCollect

De COST-actie enetCollect (maart 2017- eind oktober 2021) gaat over de grote Europese uitdaging om de taalvaardigheden van alle burgers te bevorderen, ongeacht hun verschillende sociale, educatieve en taalkundige achtergrond. De actie is gericht op het verbeteren van de productie van educatief taalmateriaal op diverse leerniveaus door middel van crowdsourcing. Door corona zijn de fysieke netwerkbijeenkomsten on hold gezet in 2020, en we weten nog niet precies hoe dit nu verder zal gaan in 2021. Het INT heeft in samenwerking met universiteiten en instituten in Portugal, Slovenië, Servië en Israël, een aantal experimenten uitgevoerd met betrekking tot het gebruik van het opensource crowdsourcingplatform PyBossa voor het verwerven van (informatie over) taaldata. Dit onderzoek wordt in 2021 voortgezet. Ook de Taalradar, een toepassing die uit dit onderzoek kwam, wordt verder ingezet voor nieuwe projecten.

European network for Web-centered linguistic data science

Het INT neemt deel aan de NexusLinguarum COST-actie. Het thema van deze actie is ‘linguistic data science’, een deelgebied binnen de opkomende 'data science'. Taalkundige data vormen een specifiek geval en zijn tot nu toe nog grotendeels onontgonnen in een big data-context.

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

De activiteiten van de actie omvatten werkvergaderingen, conferenties en workshops, training schools, STSM’s (Short Term Scientific Missions) en andere evenementen. Vanwege de coronacrisis zullen de meeste evenementen in 2021 waarschijnlijk nog virtueel plaatsvinden.

Onderzoek, onderwijs en het algemene publiek

Het INT richt zich als toegepast wetenschappelijk instituut traditioneel op onderzoekers en taalkundigen. Bestaande contacten met onderzoekers uit binnen- en buitenland, verbonden aan wetenschappelijke instituten en universiteiten, worden al dan niet in samenwerkingsprojecten onderhouden en waar mogelijk geïntensiveerd en uitgebreid. Voor universitaire studenten verzorgt het INT twee verschillende collegereeksen over computationele lexicografie.

Daarnaast verschuift het INT zijn werkterrein, gezien de brede taakomschrijving, nadrukkelijk naar docenten en leerlingen in het voortgezet/secundair onderwijs. In dat verband is het INT aanwezig op en profileert het zich op beurzen, conferenties (HSN-conferentie), festivals (Drongo-festival) en evenementen (Neerlandistiekdagen). Het INT wil verder de banden met het Onderwijsnetwerk Zuid-Holland en Alphalab Leiden aanhalen.

De taalmaterialen van het INT zullen nog beter toegankelijk worden gemaakt voor het secundair en het tertiair onderwijs. Op de nieuwe website heeft onderwijs met een eigen menu-item een prominentere plaats gekregen. De daar te vinden beschikbare informatie en materialen zoals lesbrieven zullen bijgehouden en geregeld uitgebreid worden.

Ook het algemene publiek wordt niet uit het oog verloren. Op onze website verschijnen wekelijks populairwetenschappelijke rubrieken over woorden zoals ‘Nieuw woord van de week’ (neologismen) en ‘Terug in de taal’ (historische woorden), in 2021 uitgebreid met ‘Woordhoek’ (column Ewoud Sanders) en ‘Dialectwoord van de week’. Minimaal zes keer per jaar wordt een algemene nieuwsbrief verstuurd aan geïnteresseerden. Daarnaast is er een nieuwsbrief terminologie die vier keer per jaar verschijnt en die informatie geeft over vaktaal. Tevens worden er regelmatig publieksevenementen georganiseerd. Het INT levert jaarlijks een bijdrage aan de Week van het Nederlands in oktober, en vanaf 2021 komt daar het nieuwe festival Letterlijk Leiden bij. Medewerkers houden regelmatig voordrachten, zij zijn te horen in radioprogramma’s en schrijven boeken en artikelen voor een algemeen publiek dat belangstelling heeft voor taal in het algemeen en Nederlands in het bijzonder.

Met een nieuwe website, een podcast, webinars, livestreams van evenementen en berichten op de sociale media Facebook en Twitter brengen we ook online voortdurend (de werkzaamheden van) het instituut bij alle doelgroepen onder de aandacht.