

/instituut voor de
Nederlandse taal/

Beleidsplan 2026

Leiden, november 2025

Samenvatting beleidsplan 2026

Visitatie

In 2026 gaat het Instituut voor de Nederlandse Taal (INT) zich voorbereiden op de **visitatie**, een onafhankelijke kwaliteitscontrole door experts, die begin 2027 plaatsvindt.

Focus op large language models

AI en **large language models** (LLM's) staan centraal in het maken van plannen voor de toekomst. Niet alleen voor toepassing in het dagelijkse werk van het INT, maar ook met betrekking tot de projecten waarin het INT een partner is. Zoals het lopende **SSHOC**, het grootschalige onderzoeksinfrastructuurproject **Macroscopic (NWO)**, en ingediende projectaanvragen zoals **CLARIAH-PRIME**. Voor de aanzienlijke investeringen in opslag en rekenkracht die voor het werken met LLM's nodig zijn zal het INT het komende jaar samen optrekken met de **Taalunie**, onder andere in de werkgroep **AI voor het Nederlands**, die het recente advies van de Raad voor de Nederlandse Taal en Letteren gaat uitwerken.

Het INT maakt als **CLARIN-B-centrum** en **CLARIN K-centrum** (K staat voor Knowledge) deel uit van CLARIN. Dit Europese programma biedt een stabiele, permanente, toegankelijke en uitbreidbare onderzoeksinfrastructuur voor geesteswetenschappers. Een nieuw onderdeel is het **Knowledge Centre for Large Language Models in Social Sciences and Humanities**, waar het INT ook deel van uitmaakt.

Kerntaken

Het INT blijft in 2026 verschillende kerntaken uitvoeren die voortvloeien uit het **Taalunieverdrag**, waaronder de beschrijving van de Nederlandse woordenschat in een centrale kennisbank, verschillende historische en hedendaagse (online) woordenboeken en corpora (tekstverzamelingen aangevuld met taalkundige informatie), en de spellingwebsite **Woordenlijst.org**. Op het gebied van **grammatica** zorgt het INT voor de ontwikkeling, het beheer en de beschikbaarstelling van het **Taalportaal** en de **e-ANS** (Algemene Nederlandse Spraakkunst digitaal), het meest uitgebreide Nederlandse naslagwerk over de Nederlandse grammatica. Daarnaast maakt het INT deel uit van het samenwerkingsverband achter **Taaladvies.net**. Ook is het **Expertisecentrum voor Nederlandstalige Terminologie (ENT)**

ondergebracht bij het INT, dat terminologen, vertalers en organisaties bij hun activiteiten rond Nederlandse vaktaal ondersteunt.

Extra financiering Taalunie

Daarnaast ontvangt het INT in 2026 extra financiering van de **Taalunie**. Onder andere voor het samenstellen van de inhoud van een nieuwe uitgave van het **Groene Boekje**, voor het ontwikkelen van een digitaal platform voor **dialecten en streektalen** en voor het voortzetten van **JuriTerm NL-VL**: een databank waarin verschillen in juridische terminologie in Nederland en Vlaanderen beschreven worden. Verder worden er (vervolg)projecten ontwikkeld op het gebied van **gesproken taal**, en **duidelijke en eenvoudige taal**, met name in de medische en zorgsector. Ook wordt er een taalleerdercorpus voor het Nederlands ontwikkeld op basis van de examens van het **Certificaat Nederlands als Vreemde Taal (CNaVT)**. Dat is een Taalunieproject dat anderstaligen die het Nederlands als vreemde taal leren mogelijkheden biedt om hun taalvaardigheid in het Nederlands te laten toetsen en certificeren.

Intensivering samenwerking onderzoeksinstellingen

In december 2025 heeft NWO het INT na een toetsingsprocedure erkend als **onderzoeksinstelling**, wat betekent dat het INT voor bepaalde subsidievormen als aanvrager kan optreden. Verdere intensivering van onderzoeks- en infrastructurele **samenwerking met universiteiten en onderzoeksinstellingen** is een belangrijke doelstelling in 2026. Met LUCL en TNO is recent een netwerkorganisatie opgestart die zich zal richten op de relatie tussen LLM's, taal en taalwetenschap: het **Leiden Language Modeling Center (LLMC)**. Bestaande nationale en internationale netwerkprojecten worden voortgezet, waaronder EFNIL, European Language Data Space, IMPACT en de Nederlandse AI Coalitie.

Onderwijs en het algemene publiek

Het INT richt zich traditioneel op onderzoekers en taalkundigen, maar ook **scholieren, studenten en het algemene publiek** zijn belangrijke doelgroepen. In dat verband profileert het INT zich op (onderwijsgerichte) evenementen, maakt educatief materiaal beschikbaar op de website en in **podcasts zoals *Over taal gesproken***, en zorgt voor online zichtbaarheid op sociale media. Extra aandacht gaat in 2026 uit naar **Woordpeiler**, een tool die corpustrends visualiseert en taaldata toegankelijk maakt voor een breed publiek.

Financiën

Het meerjarenbeleidsplan 2023-2027 spreekt de ambitie uit om de taalinfrastructuur voor het Nederlands op hoog niveau te houden, maar vanwege de **financiële situatie** worden de activiteiten in het beleidsplan 2026 conservatiever gepland dan wenselijk is. Door lage of afwezige indexering van de lumpsum neemt het vaste budget jaarlijks relatief af. Dit blijft een belangrijk punt van zorg.

Inhoudsopgave

Samenvatting beleidsplan 2026.....	2
1. Het INT als kennisinstituut voor het Nederlands.....	8
2. Language Resources Repository en CLARIN-centrum	10
3. Corpusinfrastructuur.....	12
Corpora	12
Algemeen	13
Taalkundige verrijking.....	13
Infrastructuur voor taalkundige verrijking.....	14
Metadata.....	14
Informatie-extractie.....	14
4. Beschrijving van de woordenschat door de eeuwen heen.....	16
Centrale lexicale infrastructuur	16
Het centrale lexicon (GiGaNT)	16
Het Centrale Betekenisregister.....	17
Lexicografische producten.....	18
Woordenlijst.org	18
Algemeen Nederlands Woordenboek (ANW).....	18
Woordenboek van Nieuwe Woorden (WNW).....	19
Woordcombinaties	19
Historische woordenboeken.....	19
Vertaalwoordenschat	20
Lexicale data voor onderzoek, ontwikkeling en voor extern gebruik	20
5. Beschrijving van de Nederlandse dialecten.....	22
6. Expertisecentrum voor Nederlandstalige Terminologie.....	24
Veldondersteuning.....	24
Software en Tools	25
Terminologische projecten	26
7. Grammatica.....	29
e-ANS.....	29
Taalportaal	30
Grammaticaportaal.....	30
8. Nationale en internationale samenwerkingsverbanden	31

Netwerken	31
EFNIL	31
IMPACT Centre of Competence	31
European Language Data Space	31
Elexis Association	32
EURALEX.....	32
Nederlandse AI Coalitie	32
KANTL.....	32
Netwerkprojecten.....	33
European Network On Lexical Innovation (ENEOLI, 2023 – 2027).....	33
Universality, diversity and idiosyncrasy in language technology (UniDive, 2022-2026).....	33
Onderzoeks- en infrastructuurprojecten.....	33
CLARIAH-Vlaanderen+ (2025-2028).....	33
CultuR-AI: Culture Re(enrichment) & AI (2027-2028)	34
Macroscopic (2026-2030).....	34
SSHOC-NL (2024-2029)	34
Project Geparset Corpus van de gesproken Nederlandse Dialecten + (GCND+) (2024-2028)	35
Spread the News (2020-2025)	35
Namenloket (2025-2027).....	36
De Taalassistent	36
Pressmint	36
9. Projectaanvragen	37
CLARIAH-PRIME (2027-2031; ingediend september 2025)	37
ELEXAI (2026-2029; ingediend september 2025)	37
FIAT: Flemish Inclusive Infrastructure for Automatic Transcription (2026-2028; ingediend september 2025).....	37
GAITE: European Doctoral Network focused on Generative AI and Translation (herindiening november 2025).....	38
10. Overige infrastructurele dienstverlening.....	39
CNaVT.....	39
DaGeNTa	39
EenvoudigNL	39
GLAD.....	40
Infrastructuur Papiaments.....	40
Jiddisch-Nederlands woordenboek	40

Plain language en easy language	40
11. IT-infrastructuur.....	41
Hardware en systemen	41
Software-ontwikkeling.....	41
12. Disseminatie: onderzoek, onderwijs en het algemene publiek.....	43
Doelgroepen	43
Onderzoek en onderwijs.....	43
Algemeen publiek	44
Neerlandistiek, taalmaterialen en Taalunie.....	44

1. Het INT als kennisinstituut voor het Nederlands

2026 wordt voor het Instituut voor de Nederlandse Taal (INT) in verschillende opzichten een overgangsjaar. Frieda Steurs heeft na 9 jaar afscheid genomen als directeur-bestuurder van het instituut en is opgevolgd door Sjef Barbiers. Terwijl het instituut in 2026 doorgaat met de uitvoering van het meerjarenbeleidsplan 2023-2027, zal het zich ook gaan voorbereiden op de visitatie die begin 2027 plaatsvindt en die de opmaat zal zijn van het nieuwe meerjarenbeleidsplan dat in 2027 vorm moet krijgen.

Het is nu al duidelijk dat de razendsnelle ontwikkelingen op het gebied van artificiële intelligentie, en in het bijzonder large language models (LLM's), centraal zullen staan in het maken van plannen voor de toekomst. Hoe gaan we om met de nieuwe mogelijkheden en gevaren die dit met zich meebrengt, en hoe kan het INT daar het beste op inspelen? De veranderingen die LLM's teweegbrengen in taken, werkwijzen, methoden en doelstellingen zijn nu al volop zichtbaar in de activiteiten van het instituut en de plannen voor 2026. Onderzocht wordt hoe LLM's kunnen worden ingezet in de workflows van het INT, bij betekenisbeschrijving en koppeling, de detectie van neologismen, de ontwikkeling van tools, en omgekeerd, hoe onze datasets kunnen worden ingezet voor de ontwikkeling en benchmarking van LLM's.

LLM's spelen ook een grote rol in de projecten waarin het INT een partner is, zoals het lopende SSHOC, het grootschalige onderzoeksinfrastructuurproject Macroscopic (NWO) dat in 2026 van start gaat, en in de projectaanvragen die in september 2025 zijn ingediend, zoals CLARIAH PRIME, CultuR-AI, ELEXAI, FIAT en GAITE (zie hoofdstuk 9).

Voor de aanzienlijke investeringen in opslag en rekenkracht die voor het werken met LLM's nodig zijn zal het INT het komende jaar samen optrekken met de Taalunie. Twee senior-medewerkers vertegenwoordigen het INT in de recent door de Taalunie geïnitieerde werkgroep *AI voor het Nederlands*. Deze werkgroep gaat mede op verzoek van het Comité van Ministers het recente advies van de Raad voor de Nederlandse Taal en Letteren uitwerken, waarin voor het INT een sleutelrol wordt voorzien in de ontwikkeling van LLM-infrastructuur voor het Nederlands, de benchmarking van LLM's en kennisdeling over LLM's.

Speciale aandacht zal in 2026 uitgaan naar de benutting van de data en tools van het INT, door middel van disseminatie onder verschillende gebruikersgroepen. Daarnaast zal gewerkt worden aan de ontwikkeling van onderzoekspilots en projectaanvragen in samenwerking met universitaire onderzoeksgroepen, in domeinen zoals comparatieve lexicologie, prosodische analyse van dialect-opnames, grammaticale variatie en verandering. In december 2025 heeft NWO het INT na een toetsingsprocedure erkend als onderzoeksinstituut, wat betekent dat het INT voor bepaalde subsidievormen als aanvrager kan optreden.

Meer in het algemeen is verdere intensivering van onderzoeks- en infrastructurele samenwerking met universiteiten en onderzoeksinstituten een belangrijke doelstelling in 2026. Met LUCL en TNO wordt een netwerkorganisatie opgericht die zich zal richten op de relatie tussen LLM's, taal en taalwetenschap. Deze netwerkorganisatie zal onder andere nieuw onderzoek initiëren en onderzoek verrichten voor externe opdrachtgevers. De relatie met de Universiteit Leiden behoeft nieuwe aandacht nu het instituut niet meer in een gebouw van de universiteit zit.

Het INT heeft als structureel gefinancierd kennis- en onderzoeksinstituut een unieke positie en de opdracht om voor het hele Nederlandse taalgebied (Nederland en de Caribische rijkdelen, Vlaanderen en Suriname) op een wetenschappelijk verantwoorde wijze de digitale taalinfrastructuur uit te bouwen. Het INT voert daarbij een aantal taken uit het Taalunieverdrag uit (zie Taalunieverdrag, hdst. 1, artikelen 2, 3, 4 en 5). Het INT ontwikkelt enerzijds zelf corpusdata, linguïstische databanken en taalsoftware voor een aantal specifieke domeinen, of ondersteunt de ontwikkeling ervan, en anderzijds verzamelt het INT ook taalmaterialen en taalsoftware van andere kennisinstellingen. Samen met de eigen taalmaterialen worden deze duurzaam ter beschikking gesteld via repository's, websites, API's en als open source software. Het INT promoot die taalinfrastructuur bij onderzoekers, ontwikkelaars en het brede publiek om zo onderzoek, onderwijs en andere activiteiten rond de Nederlandse taal te stimuleren en te ondersteunen.

Daarnaast heeft het INT als kennis- en onderzoeksinstituut de doelstelling om de kennis en expertise over taalinfrastructuur verder uit te bouwen door eigen wetenschappelijk onderzoek. De opdracht van het INT is te vergelijken met die van andere taalinstellingen in Europa waarmee het intensief samenwerkt in Europese projecten en netwerken, al is het

INT relatief klein en blijft de financiering achter bij die van de buitenlandse tegenhangers. Door lage of afwezige indexering van de lumpsum neemt het vaste budget jaarlijks relatief af. Dit blijft een belangrijk punt van zorg. Het meerjarenbeleidsplan 2023-2027 spreekt de ambitie uit om de taalinfrastructuur voor het Nederlands op hoog niveau te houden, maar de activiteiten in het beleidsplan 2026 worden daarom conservatiever gepland dan wenselijk is.

In de volgende hoofdstukken wordt uiteengezet hoe het INT in 2026 uitvoering zal geven aan zijn vele taken, hoe daarbij wordt gestreefd naar zoveel mogelijk samenhang en synergie in de planning van die taken en de uitvoering ervan, en hoe dit alles is ingebed in de nationale en internationale taalinfrastructuren en netwerken. De kerntaken van het INT worden uitgevoerd met de structurele financiering van de Taalunie. In de hoofdstukken 2 tot en met 7 wordt toegelicht hoe het INT in 2026 met een aantal integratie-operaties de efficiëntie bij de opbouw en het beheer van de Nederlandse taalinfrastructuur wil verhogen en tegelijk met nieuwe initiatieven de blijvend hoge kwaliteit van het aanbod voor gebruikers wil waarborgen. Hoofdstuk 8 geeft een overzicht van hoe het INT als onderzoeksinstituut binnen nationale en internationale projecten zijn expertise ook in 2026 ter beschikking zal stellen en verder uitbouwen. Hoofdstuk 9 beschrijft kort de projecten die in de tweede helft van 2025 zijn aangevraagd en die, indien gehonoreerd, een belangrijke rol zullen spelen in de taken van het instituut de komende jaren. Hoofdstuk 10 geeft een overzicht van de overige infrastructurele diensten die het INT levert. Hoofdstuk 11 is een beschrijving van de situatie en ontwikkelingen met betrekking tot de IT-infrastructuur van het INT zelf, en hoofdstuk 12 rondt af met een overzicht van de activiteiten waarmee het INT in het komende jaar de taalinfrastructuur voor het Nederlands bij een breed publiek bekend wil maken.

2. Language Resources Repository en CLARIN-centrum

De integratie van de Taalmaterialen-catalogus en het CLARIN-B-centrum met bijhorend CLARIN-portal wordt in 2026 verdergezet. Na een aantal tests met Dataverse en D-Space werd er uiteindelijk voor D-Space gekozen omwille van een aantal technische vereisten die niet in Dataverse beschikbaar waren. In 2026 wordt verder gewerkt aan het testen van de D-

Space-implementatie en het overzetten van de tools en resources uit Taalmaterialen en CLARIN-portal. De huidige Taalmaterialen en het CLARIN-portal blijven functioneren en worden pas uitgefaseerd na uitgebreid testen van de D-Space-omgeving.

D-Space staat ons ook toe om toe te werken naar een gestroomlijnde depositing-procedure waarbij meer geautomatiseerd kan worden en waarbij de leveranciers van de data/tools metadata kunnen aanleveren op een manier die automatisch geïntegreerd wordt in de metadatabestanden.

Op het vlak van hardware (storage, memory en processing power) zullen de repository's voorbereid worden op het toenemend belang van multimediale content en tools die meer rekenkracht vergen. We vervangen een aantal servers en verhogen de opslag, het geheugen en de processing power t.o.v. van de huidige servers. Daarnaast worden de algemene opslagsystemen vervangen.

Op Europees vlak volgen we actief de ontwikkelingen omtrent 1) de European Language Data Space, waarvan het INT het nationale aanspreekpunt is en deel uitmaakt van de user group; 2) de Alliance For Language Technology EDIC, via het lidmaatschap van Vlaanderen en Nederland; en natuurlijk 3) CLARIN ERIC, waarin het INT als centrum voor CLARIN-België vertegenwoordigd is via het National Coordinators Forum, het Knowledge Infrastructure Committee, het Standards and Interoperability Committee en het Technical Centres Committee. Sinds 2025 levert het INT ook de CLARIN-coördinator voor Nederland, en is het INT vertegenwoordigd in de Board of Directors van CLARIN ERIC.

Op nationaal niveau is het INT met twee leden vertegenwoordigd in de Taalunie-werkgroep *AI voor het Nederlands*, die als doel heeft de samenwerking rond AI in het hele Nederlandse taalgebied te versterken. Verder blijft het INT zijn rol als expertisecentrum voor het Nederlands vervullen, zowel als CLARIN Knowledge Centre for Dutch (<https://kdutch.ivdnt.org/> K-Dutch) als als servicedesk voor mensen met vragen omtrent het Nederlands. Stroomlijnen van de servicedeskaanvragen staat hierbij op de agenda, waarbij een databank aangelegd wordt van aanvragen en bijbehorende antwoorden.

Het INT maakt hiernaast ook deel uit van drie gedistribueerde CLARIN Knowledge Centra, namelijk IMPACT-CKC (IMPACT centre of competence - [CLARIN K-centre in digitisation](#)), het

3. Corpusinfrastructuur

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

Corpora

De kern van het Corpus Hedendaags Nederlands (CHN) is het monitorcorpus van kranten. Ook in 2026 zal het corpus verder geüpdatet worden met krantenmateriaal dat we wekelijks dan wel jaarlijks ontvangen. In 2025 is gewerkt aan een fundamentele upgrade van de corpus-processinginfrastructuur, zowel van de corpusworkflowomgeving als van de conversie van corpusdata, inclusief de implementatie van het CMDI-metadatamodel. De focus daarbij was de processing van het CHN. De workflow zal in productie worden genomen in 2026 en in de nieuwe omgeving zal curatie van de CHN-metadata worden gedaan. We willen voor het historisch corpusmateriaal een vergelijkbare upgrade doen van de corpus-processinginfrastructuur, inclusief de implementatie van het CMDI-metadatamodel. Daarmee wordt gestart in 2026.

Ten slotte zal er in 2026, afgezien van het online brengen van nieuwe historische corpora, ook gestart worden met het project Macroscope (zie verder hoofdstuk 8), waarbij het INT onder andere meewerkt aan de bouw van een nationaal mediacorpus. De hierboven genoemde processing workflow zal daarbij een belangrijke rol spelen. Daarnaast zal mede in

de context van het PressMintproject (zie hoofdstuk 8), gewerkt worden aan de dataprocessing van nieuw materiaal voor het [Courantencorpus](#), dat via ATR (Automatic Text Recognition) in [Transkribus](#) wordt gedigitaliseerd en in datzelfde Transkribus door een team van vrijwilligers met redactionele ondersteuning wordt geverifieerd.

Algemeen

De INT-corpora worden samengesteld uit bestaand digitaal materiaal of, waar nodig, door digitalisering. De brondata worden geconverteerd naar eenzelfde XML-standaard (TEI) en zorgvuldig van metadata voorzien en daarna automatisch taalkundig verrijkt. Metadata en taalkundige verrijking zijn een noodzakelijke voorwaarde om zinvolle informatie uit de corpora te kunnen extraheren. De werkzaamheden voor 2026 worden hieronder gespecificeerd. Het doel is om alle corpora van het INT taalkundig te verrijken, zowel met woordsoort en lemma, als syntactisch (Universal Dependencies).

Taalkundige verrijking

In 2025 is verder gewerkt aan goudstandaardmateriaal voor het Middelnederlands en is er onderhoud gepleegd aan de bestaande datasets. Deze werkzaamheden worden afgerond in 2026, waarna de data gebruikt zal worden om de tools voor taalkundige verrijking mee te trainen. Die tools worden ook ter beschikking gesteld via het GaLaHaD-platform (zie hieronder), zodat de nog niet taalkundig verrijkte corpora verrijkt online kunnen worden gezet. In 2026 zal ook gewerkt worden aan goudstandaardmateriaal voor hedendaags Nederlands zodat ook het CHN conform de [TDN-richtlijnen](#) verrijkt kan gaan worden.

In 2025 hebben we de relatie tussen Tagset Diachroon Nederlands (TDN) en Universal Dependencies (UD) onderzocht en is met behulp van het GaLaHaD-platform een eerste evaluatie van bestaande verrijkingspipelines zoals [SpaCy](#) en [Stanza](#) uitgevoerd. Daarbij is naar voren gekomen dat de huidige UD-corpora op diverse punten zouden moeten worden aangepast om als training- en evaluatiemateriaal te kunnen dienen voor de verrijking van de corpusdata, zowel voor woordsoorttoekenning en lemmatisering als voor dependentierelaties. Deze aanpassingen voeren we uit in samenspraak met de Universiteit Groningen.

Infrastructuur voor taalkundige verrijking

In 2025 is de infrastructuur voor taalkundige verrijking gelanceerd ([GaLAHaD](#) en [LAnCeLoT](#)), ontwikkeld in het kader van CLARIAH+, voor het verrijken met woordsoort en lemma van historisch Nederlands. Het GaLAHaD-platform biedt automatische verrijking en evaluatie daarvan met behulp van diverse evaluatiecorpora, en maakt het daarmee mogelijk voor gebruikers om de verrijkingstools te kiezen die het beste resultaat voor een bepaalde dataset bieden. In het kader van SSHOC-NL is GaLAHaD in 2025 uitgebreid met dependentieparsing en named entity recognition. Er zal in 2026 worden gewerkt aan een lemmatiseerder voor hedendaags Nederlands, die gebruik maakt van het GiGaNT-lexicon en werkt conform de richtlijnen voor [lemmatisering](#) van GiGaNT en een basismatching van die lemmatisering met het GiGaNT-lexicon.

Metadata

De corpora hebben een gemeenschappelijk metadataformaat, met ruimte voor subcorpus-specifieke metadata. In 2024 is het INT tot een verdere uniformering van het metadatamodel voor historisch en modern corpusmateriaal gekomen, waardoor beide op termijn als één doorlopend diachroon corpus op metadatacategorieën doorzoekbaar kunnen worden. In 2025 hebben we gewerkt aan de implementatie van dat model voor de data van het CHN. In 2026 wordt gestart met de implementatie van het model voor de historische corpora.

Informatie-extractie

Het corpusmateriaal wordt toegankelijk gemaakt via een applicatie waarmee in de corpora gezocht kan worden. De software is opensource beschikbaar. Wanneer de IPR (Intellectual Property Rights) het toelaten, wordt het corpusmateriaal ook als dataset beschikbaar gesteld in Taalmaterialen, dan wel in de vorm van n-grammen. In 2025 zijn significante optimalisaties uitgevoerd aan zowel het [corpus-frontend](#) als het [corpus-backend](#). In 2026 wordt gestart met een verdere uitbreiding van de software, in de context van het project Macroscopic (zie hoofdstuk 8). Zo moet het bijvoorbeeld mogelijk worden gemaakt om op een gebruiksvriendelijke manier subcorpusselecties te doen, en eigen corpusmateriaal te combineren met reeds bestaand online beschikbaar corpusmateriaal. Daarnaast wordt, mede in het kader van het GCND+-project (zie hoofdstuk 5), verder gewerkt aan de implementatie van een gebruiksvriendelijke zoekfaciliteit voor syntactische patronen.

In 2025 is een applicatie voor de visualisatie van corpustrends en frequentieprofielen voor onderzoekers en een breder publiek gelanceerd ([Woordpeiler](#)) en is er gewerkt aan een integratie in de lexicale infrastructuur (zie hoofdstuk 4). De werkzaamheden aan die integratie worden verder gezet in 2026. Het is de bedoeling dat deze infrastructuurcomponent op termijn in de context van Macroscope verder wordt uitgebreid.

4. Beschrijving van de woordenschat door de eeuwen heen

De wetenschappelijke, corpusgebaseerde beschrijving van de Nederlandse woordenschat in al zijn facetten blijft een van de kerntaken van het INT (zie verdragstaak 4d van het Taalunieverdrag). Zoals uiteengezet in het Meerjarenbeleidsplan 2023-2027, zet het INT in de komende jaren versterkt in op twee belangrijke vernieuwingen voor de lexicale taalinfrastructuur:

- De integratie van alle componenten van de woordenschatbeschrijving in één centrale, modulair georganiseerde, relationele kennisbank van de Nederlandse woordenschat door de eeuwen heen, vanwaaruit bestaande en nieuwe lexicografische producten verder ontwikkeld zullen worden.
- Een versterking en explicitering van het corpusgebaseerde lexicografische proces, resulterend in een bidirectionele linking tussen primaire corpusdata en afgeleide lexicale data.

Voor de centrale kennisbank wordt bovenop het bestaande centrale lexicon GiGaNT een betekenisregister gebouwd. De eerste koppelingen tussen het *Algemeen Nederlands Woordenboek* (ANW), Woordcombinaties en het Referentiebestand Nederlands (RBN) hebben hiervoor een eerste basis gelegd. In 2026 ligt de focus op de aanpassingen die nodig zijn aan de databank-infrastructuur en de databank-editeeromgeving (Lex'it) om de centrale kennisbank te ondersteunen. Daarnaast wordt ook verder geëxperimenteerd met het wetenschappelijk verantwoord inzetten van generatieve AI om de woordenschatbeschrijving te ondersteunen. Hieronder beschrijven we de geplande werkzaamheden voor specifieke onderdelen van de kennisbank en voor de lexicografische producten die eraan gekoppeld zijn.

Centrale lexicale infrastructuur

Het centrale lexicon (GiGaNT)

In 2025 is, naast de reguliere werkzaamheden aan het centrale lexicon, gewerkt aan de implementatie van de herzieningen van het datamodel (hierboven). Deze reguliere werkzaamheden worden in 2026 voortgezet. Ze betreffen niet alleen het onderhoud van de

componenten in GiGaNT-Hilex, maar ook de uitbreiding van GiGaNT-Molex ten behoeve van de diverse lexicografische producten die een koppeling met GiGaNT-Molex hebben (zoals [ANW](#), [WNW](#), [Woordcombinaties](#), [Vertaalwoordschat](#) en [Woordenlijst.org](#)). Het doel is om in 2026 het GiGaNT-lexicon te releasen. Dit betekent niet alleen een upgrade van de koppeling van de beide lexiconmodules maar ook een upgrade van de infrastructuur, om de interne consistentie zo goed mogelijk te kunnen waarborgen tussen de lexiconmodules van GiGaNT en tussen GiGaNT en de lexicografische eindproducten die eraan gekoppeld zijn.

In 2025 zijn belangrijke stappen gezet in de ontwikkeling van de corpusgebaseerde workflow om corpusmateriaal systematisch te kunnen screenen op lacunes en op nieuwe woorden, aangevuld met een omgeving om deze data makkelijk te analyseren, te structureren en te bewerken. Focus hierbij is het hedendaags Nederlands (Molex). Frequentiegegevens worden systematisch uit het CHN gehaald, en zijn al gebruikt voor de online [Woordpeiler](#). Er is gestart met het ontwikkelen van een omgeving waarin deze frequentiegegevens worden gecombineerd met de data van GiGaNT-Molex. De frequentiegegevens worden in een bewerkingsomgeving ter beschikking gesteld zodat ze gebruikt kunnen worden om selecties te maken van nog niet in Molex opgenomen woorden, voor opname in de centrale database. Deze omgeving hopen we eind 2026 in productie te hebben. Ten slotte zal in de eerste helft van 2026 een plan ontwikkeld worden voor de morfologische component van GiGaNT.

Het Centrale Betekenisregister

Zoals uiteengezet in het Meerjarenbeleidsplan 2023-2027 is de koppeling van verschillende lexicografische databanken op betekenisniveau de belangrijkste innovatie en grootste uitdaging voor de woordenschatbeschrijving in de komende jaren. Voor een steekproef van zo'n 800 lemmata werd reeds een manuele koppeling gemaakt tussen het betekenisregister enerzijds en het ANW, Woordcombinaties en RBN anderzijds. In 2026 zal op basis van die ervaringen verder gewerkt worden aan de databank-infrastructuur en de databank-editeeromgeving die de complexiteit van deze koppelingen kan ondersteunen. Daarnaast zullen het datamodel en de koppelprincipes verder uitgewerkt worden zodat WNT-artikels in deze oefening betrokken kunnen worden, mede met het oog op de verdere integratie van de hedendaagse en historische lexiconcomponenten (zie GiGaNT hierboven).

Op basis van eerdere experimenten zal verder onderzocht worden hoe large language models (LLM's) het koppelen op betekenisniveau en het formuleren van definities en glossen kunnen ondersteunen.

Lexicografische producten

Zoals het Meerjarenbeleidsplan 2023-2027 aangeeft, worden op termijn de huidige lexicografische eindproducten, zoals Woordenlijst.org, ANW, WNW, Woordcombinaties en de historische woordenboeken, samen met eventuele nieuwe eindproducten, als afgeleide producten vanuit de centrale kennisbank verder ontwikkeld en verbeterd. Hoewel de huidige productspecifieke workflows ook in 2026 nog grotendeels aangehouden worden, zullen in de loop van het jaar verdere stappen gezet worden om de invoeging van lexicografische content vanuit de centrale kennisbank in de verschillende eindproducten mogelijk te maken.

Woordenlijst.org

Woordenlijst.org (de lijst met de officiële spelling van de Taalunie) wordt door het INT voortdurend uitgebreid en up-to-date gehouden, zoals vastgelegd in het Taalunieverdrag artikel 4b. Voor Woordenlijst.org is de publicatie vanuit de centrale kennisbank nu al gerealiseerd, met name dan vanuit GiGaNT als de module voor de woordenschatbeschrijving op lemma-niveau. Er zijn alleen een aantal technische updates van [Woordenlijst.org](https://woordenlijst.org) voorzien in 2026. De focus zal liggen op het maken van het nieuwe Groene Boekje, dat eind 2026 gepubliceerd zal worden.

Algemeen Nederlands Woordenboek (ANW)

Het [ANW](https://anw.let.rug.nl/) beschrijft de algemene woordenschat vanaf 1970. In 2026 blijft de bestaande redactionele workflow grotendeels behouden. Deze is reeds geïntegreerd met die van het *Woordenboek van Nieuwe Woorden* en met het centrale lexicon GiGaNT-Molex. Daarnaast werkt de redactie binnen het recent versterkte team voor hedendaagse woordenschatbeschrijving mee aan de verdere uitbouw van de centrale kennisbank en het betekenisregister. Voortbouwend op eerste experimenten met koppelingen tussen ANW, RBN en Woordcombinaties zal in 2026 verder prioritair gewerkt worden aan het opwaarderen van de zogenaamde kopjesartikelen (dit zijn artikelen met alleen

grammaticale gegevens en corpusvoorbeelden) met betekenisinformatie. Daarnaast wordt ook onderzocht hoe LLM's op een wetenschappelijk verantwoorde manier bij de betekenisbeschrijving kunnen ingeschakeld worden.

Woordenboek van Nieuwe Woorden (WNW)

Het [WNW](#) beschrijft innovaties in de woordenschat vanaf het jaar 2000. Het WNW behoudt in 2026 zijn bestaande workflow maar draagt binnen het recent versterkte team voor hedendaagse woordenschatbeschrijving ook bij aan de uitbouw van de centrale kennisbank en het betekenisregister. Daarnaast wordt de Woordpeiler-applicatie verder in de redactionale workflow geïntegreerd als hulpmiddel voor detectie van neologismen in het Corpus Hedendaags Nederlands. Verder bouwend op eerdere experimenten wordt bovendien onderzocht hoe LLM's bij de detectie en beschrijving van neologismen kunnen worden ingeschakeld.

Woordcombinaties

[Woordcombinaties](#) is de online taaltool die leerders van het Nederlands als vreemde taal en moedertaalgebruikers ondersteunt bij het gebruiken van woorden in context. Dit project inventariseert en beschrijft systematisch combinaties (collocaties, idiomen en patronen) in het Nederlands.

In 2026 wordt verder gewerkt aan de lexicografische beschrijving van de combinatiemogelijkheden van werkwoorden en substantieven. In de in 2025 in productie genomen aangepaste werkomgeving kunnen idiomen en conversatieformules nu efficiënter worden bewerkt. Conversatieformules worden gekoppeld aan taalhandelingen, en idiomen aan betekenis- en beeldcategorieën, waardoor omgekeerde zoekmogelijkheden (van beeld- of betekenis categorie of taalhandeling naar idioom resp. conversatieformule) in de applicatie mogelijk worden. Er wordt verder onderzocht hoe LLM's op een wetenschappelijk verantwoorde manier in de workflow geïntegreerd kunnen worden.

Historische woordenboeken

In 2023-2025 is een verregaande uniformering van de codering van de historische woordenboekdata uitgevoerd. In 2026 verbeteren we de integratie tussen de historische

woordenboekdata en het GiGaNT-lexicon, zodat in het lexicon aangebrachte verrijking ten goede kan komen aan het historische woordenboekenportaal. We bereiden de onlinegang van deze bestandsversie voor.

Vertaalwoordenschat

De [Vertaalwoordenschat](#) is een platform waarop tweetalige bestanden worden ontsloten die in de jaren 90 en begin 2000 ontwikkeld werden door de Taalunie voor commercieel niet afgedekte taalparen. Inmiddels staan het Nederlands-Nieuwgrieks / Nieuwgrieks-Nederlands, het Nederlands-Portugees / Portugees-Nederlands, het Nederlands-Estisch, het Nederlands-Fins / Fins-Nederlands, het Nederlands-Deens/Deens-Nederlands en het Nederlands-Arabisch / Arabisch-Nederlands online. Eind 2025 werd het Nederlands-Indonesisch aan het platform toegevoegd. Begin 2026 volgt een inhoudelijke update met de laatste versie van het Nederlands-Arabisch / Arabisch-Nederlands, waar in 2025 door een externe redacteur aan is gewerkt.

In 2026 zullen de taalparen die inmiddels online staan inhoudelijk verrijkt en verbeterd worden, onder andere met paradigma-informatie (m.n. voor het Nederlands) en gebruikslabellen. Daarnaast zal in 2026 verder gewerkt worden aan de release van het Kroatisch en het Servisch.

Lexicale data voor onderzoek, ontwikkeling en voor extern gebruik

De data uit de lexica en lexicografische databanken van het INT worden ook ter beschikking gesteld voor onderzoek en ontwikkeling en voor extern gebruik. Dit gebeurt zowel door middel van API's als in de vorm van uit de databanken afgeleide datasets.

De API voor de centrale kennisbank wordt in 2026 verder ontwikkeld vanuit de al bestaande lexiconservice, waarbij vooral gefocust wordt op het waarborgen van de beschikbaarheid en betrouwbaarheid van de dienst. Diezelfde informatie komt telkens ook in de nieuwe release van GiGaNT-Molex ter beschikking als dataset in de language repository.

Als hoogkwalitatieve, gecureerde data zijn de verschillende componenten uit de lexicale databanken van het INT ook bij uitstek geschikt als training- en testdata voor de ontwikkeling van large language models. In 2026 zal versterkt ingezet worden op de aanmaak van datasets op basis van de lexicale databanken in een formaat dat geschikt is

voor de ontwikkeling en de benchmarking van LLM's. Deze datasets worden zowel intern gebruikt als extern ter beschikking gesteld.

5. Beschrijving van de Nederlandse dialecten

Het INT beschrijft de Nederlandse woordenschat in al haar facetten. Sinds 2020 heeft het INT de verantwoordelijkheid gekregen voor het beheer, de ontwikkeling en de beschikbaarstelling van diverse dialectproducten zoals de Database van de Zuidelijk-Nederlandse Dialecten ([DSDD](#)) en de elektronische Woordenbank van de Nederlandse Dialecten ([eWND](#)). Het doel is om uiteindelijk een dialect- en streektaalportaal te kunnen bieden voor alle dialectvarianten van het Nederlands. Sinds eind 2024 is het INT ook verantwoordelijk voor de beschikbaarstelling van het Gesproken Corpus van de Zuidelijk-Nederlandse Dialecten ([GCND](#)). Aan dit project is een vervolg gegeven waarbij een uitbreiding naar alle dialecten van het Nederlands is voorzien (GCND+, zie hoofdstuk 8).

Elektronische Woordenbank van de Nederlandse dialecten (eWND)

In 2021 is de hosting en het onderhoud van het eWND-portaal overgenomen door het INT. Het eWND wordt zoals elk jaar ook in 2026 uitgebreid met nieuwe dialectwoordenboeken met behulp van vrijwilligers, die het materiaal voorbereiden.

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

In 2026 zullen de gegevens van het Woordenboek der Zeeuwse dialecten en het Supplement worden toegevoegd aan de Database van de Zuidelijk-Nederlandse Dialecten (DSDD). Die gegevens worden gekoppeld aan de bestaande concepten in de DSDD. Eventuele nieuwe concepten worden aan de database toegevoegd. Waar nodig worden in 2026 correcties aangebracht en hiaten aangevuld.

In 2026 zullen we voor andere onomasiologische woordenboeken, namelijk WGD (Gelderland) en WOD (Overijssel), ook de koppelingen met bestaande concepten uit de DSDD maken, als een van de stappen naar een dialectplatform voor het hele Nederlandse taalgebied.

Digitale infrastructuur voor dialecten en streektalen

Het INT is op vraag van de Taalunie een infrastructuur aan het uitbouwen waarin streektaalorganisaties hun talig cultureel erfgoed kunnen beschrijven en beschikbaar

stellen. Het ontwerp en de uitwerking van die infrastructuur gebeurt met externe financiering. In 2025 is in dit verband gewerkt aan het pilootproject voor het Bildts en het West-Overijssels, waarin West-Overijsselse (semasiologische) woordenlijsten en woordenboeken van een dorp/stad of een regio gemakkelijk bewerkt kunnen worden en online beschikbaar gesteld. Beide projecten dienen als volgende stap voor een uitgebreidere digitale dialectinfrastructuur, die vanaf 2026 beschikbaar gesteld kan worden voor nieuwe projecten, onder voorbehoud van externe financiering.

6. Expertisecentrum voor Nederlandstalige Terminologie

Het [Expertisecentrum Nederlandstalige Terminologie](#) (ENT) is sinds 1 juli 2016 ondergebracht bij het INT en voert in samenwerking met de Taalunie de terminologie-gerelateerde verdragstaken uit (zie artikel 4c en 5e van het Taalunieverdrag). Het ENT ondersteunt terminologen, vertalers en organisaties bij hun activiteiten rond terminologie en Nederlandse vaktaal. Via het ENT beschikken zij over terminologische informatie en hulpmiddelen. Enerzijds verschijnt er om de paar maanden een nieuwsbrief over ontwikkelingen en evenementen in de wereld van de terminologie. Anderzijds onderhoudt en ontwikkelt het INT binnen het kader van ENT zowel eigen terminologische software ([TermWerk](#)) als termenbanken voor een aantal focusdomeinen (terminologie van en voor het hoger onderwijs, juridische en overheidsterminologie).

In 2025 hebben de Taalunie en het INT samen de werking van het ENT geëvalueerd en een aantal mogelijke pistes voor de toekomstige uitbouw geformuleerd. Naast de voortzetting van de reguliere veldondersteuning en lopende projecten, zal in 2026 daarom vooral onderzocht worden hoe de in die evaluatie voorgestelde ontwikkeling van een volwaardig terminologieportaal kan voorbereid worden. Het *Handbook of Terminology: Volume 4* (Resi en Steurs 2025), dat handelt over terminologiebeleid in Europa en waarin het ENT de bijdrage over het Nederlandse taalgebied leverde, zal daarbij als basis dienen om te vergelijken met de terminologievoorzieningen voor andere talen.

In wat volgt worden eerst de plannen voor de reguliere werkzaamheden van het ENT beschreven (veldondersteuning, softwareondersteuning) en daarna de projecten die afhankelijk zijn van bijkomende financiering (verdere softwareontwikkeling en terminologische beschrijving)

Veldondersteuning

Updates van de bestaande websiterubrieken, vooral de opleidingen in Nederland en Vlaanderen, blijven aan de orde. In 2026 worden vier nieuwsbrieven Terminologie gedistribueerd en wordt de evenementenrubriek voor terminologie continu bijgehouden.

Ten behoeve van de Nederlandstalige terminologie op Europees niveau zal in 2026 verdere samenwerking met de Termraad worden gerealiseerd. Via dit overlegplatform voor terminologische afstemming binnen het Nederlandse taalgebied werken EU-terminologen samen met Belgische en Nederlandse partners uit overheidsdiensten, terminologieverenigingen en vertaalopleidingen. Het INT participeert in de driemaandelijke bijeenkomsten met het oog op de verzameling, beschrijving en uniformering van terminologie in specifieke vakgebieden.

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. Daarom worden zowel hedendaagse als historische termenlijsten op de ENT-webpagina's permanent geüpdatet en uitgebreid binnen een online raadpleegbare structuur waarvoor de internationale onderwerpenclassificatie van de Library of Congress wordt gebruikt. Onderzocht zal worden hoe deze lijsten door Search Engine Optimization beter online terugvindbaar kunnen worden gemaakt en hoe zij in een ruimer terminologieportaal zouden passen.

Ten slotte wordt onderzocht hoe bijkomende doelgroepen bereikt kunnen worden via organisaties zoals de beroepsverenigingen voor technische communicatie (TEKOM Belgium, TEKOM Netherlands), of vertalers en tolken (Nederlands Genootschap van Tolken en Vertalers, Beroepsvereniging Beëdigd Vertalers en Tolken).

Software en Tools

De tool [TermWerk](#) werd door het INT in opdracht van de Taalunie ontwikkeld en integreert termextractie, termenbankbeheer en termenbankpublicatie in één online applicatie die vrij toegankelijk is via de CLARIN-infrastructuur. De tool is sinds 2025 volledig operationeel en zal in 2026 verder onderhouden, ontwikkeld en gepromoot worden. Tot de geplande uitbreidingen voor een algemeen toegankelijke versie van TermWerk behoren een verbeterde taalkundige verrijking van vakteksten en meer ondersteuning voor de vergelijking van verschillende termenlijsten en termenbanken.

Daarnaast wordt in 2026 een interne versie van TermWerk verder ontwikkeld met aanpassingen voor de eigen terminologische projecten van het INT. Zo zal in de eerste helft van 2026 een aangepaste bewerkings- en publicatieomgeving voor het project JuriTerm NL-VL (cf. infra) afgewerkt worden waarbij TermWerk uitgebreid wordt met specifieke functies voor een rechtsvergelijkende terminologiebeschrijving en voor de online publicatie van juridische terminologiedatabanken. Op basis van die ervaringen zal in 2026 ook onderzocht worden of andere terminologieprojecten van het INT ook via aangepaste TermWerk-versies verder bewerkt en geüpdatet kunnen worden. Daarnaast zal ook onderzocht worden in welke mate TermWerk als startpunt voor een volwaardig termportaal gebruikt kan worden, met name door te testen hoe termenbanken binnen de TermWerk-omgeving gekoppeld kunnen worden aan de centrale kennisbank voor de Nederlandse woordenschat en het centrale lexicon GiGaNT-Molex (cf. supra).

Het INT zal in 2026 de ontwikkeling voortzetten van een geüniformiseerd online publicatieplatform voor lexicale databanken (cf. infra). Om de publicatie van termenbanken te optimaliseren, zowel uit TermWerk als van terminologiebestanden die door externen aangeleverd worden, wordt het TBX-formaat (TermBase Exchange Format) meegenomen als ondersteund inputformaat voor het publicatieplatform. Dit zal de online publicatie van zowel interne als externe termenbanken in de toekomst vergemakkelijken.

Ten slotte wordt net als bij de algemene woordenschatbeschrijving onderzocht hoe LLM's verantwoord ingezet kunnen worden bij de terminologiebeschrijving.

Terminologische projecten

Naast de reguliere werkzaamheden van het ENT ontwikkelt het INT ook terminologiebanken voor een aantal speerpunt domeinen. Hiervoor is wel telkens extra financiering nodig.

Juridische terminologie. Het project **JuriTerm NL-VL** heeft tot doel een workflow op te zetten en een eerste versie te ontwikkelen van een rechtsvergelijkende terminologiedatabank voor de Nederlandse rechtstaal. In de termenbank worden verschillen in juridische terminologie beschreven en in kaart gebracht. Daarmee draagt het project rechtstreeks bij aan een van de opdrachten van de Taalunie, met name het in kaart brengen van verschillen in overheidsterminologie tussen België en Nederland met het oog

op een mogelijke harmonisatie. Het project wordt dan ook gefinancierd door de Taalunie, samen met de Vlaamse Juristenvereniging, de Stichting Ridder René Victor en het Tijdschrift voor Privaatrecht. Het project bouwt deels voort op het *Sociaalrechtelijk Woordenboek* en het *Juridisch Woordenboek Nederlands-Spaans* van Oosterveld en Vuyk, dat ook door het INT gehost wordt, en maakt gebruik van een aangepaste versie van TermWerk (JuriTermWerk). JuriTerm NL-VL is bedoeld als hulpmiddel voor juristen, rechtslinguïsten en rechtsvertalers en deze groepen zijn via een begeleidingscommissie en disseminatieactiviteiten ook nauw betrokken bij het project. Een eerste proefversie van de termenbank, met een 100-tal termingen uit de domeinen privaatrecht en sociaal recht, zal begin 2026 gepubliceerd worden.

JuriTerm NL-VL werd in februari 2025 als eenjarig pilootproject opgestart en in het najaar van 2025 positief geëvalueerd door de externe begeleidingscommissie. Dankzij bijkomende, maar beperkte financiering van de Taalunie wordt het project in 2026 voortgezet met een focus op het afwerken van de collaboratieve bewerkingsomgeving en het publiceren van de eerste versie van de termenbank. Verder wordt ingezet op disseminatie en het versterken van de reeds opgezette samenwerkingen met het juridische werkveld om bijkomende fondsen te werven en om JuriTerm NL-VL in de toekomst als collaboratief terminologieproject verder te kunnen zetten. Ten slotte wordt in 2026 ook onderzocht of het corpus van juridische teksten, dat de empirische basis vormt voor de terminologische beschrijving in het project, ook als onderdeel van het Corpus Juridisch Nederlands voor externe onderzoekers ter beschikking gesteld kan worden.

Onderwijs-terminologie. Het project **HOTNeV**, dat hogeronderwijs termen in Nederland en Vlaanderen in de HOTNeV-termenbank beschrijft, wordt weliswaar niet meer bijkomend gefinancierd, maar toch wordt de termenbank doorlopend aangevuld via stages en scripties. De groei van deze databank is dus afhankelijk van de stages die bij het ENT worden aangevraagd. Afstudeerscripties van studenten kunnen eveneens bijdragen aan de verdere invulling van de HOTNeV-termenbank. Meer algemeen blijft de begeleiding van studenten belangrijk en het stageaanbod voor terminologiewerk wordt blijvend gepromoot. Dit draagt immers ook bij aan de netwerkpositie van het ENT. De HOTNeV-bestanden zullen in 2026 daarom ook in TermWerk geïmporteerd worden zodat de termenbank verder bewerkt en ook online gepubliceerd kan worden.

Wiskunde-, scheikunde- en natuurkundetermen. Hoewel het project **Nederlands als wetenschapstaal** geen bijkomende financiering meer krijgt, blijft het erg belangrijk om het Nederlands als taal van voortgezet en hoger onderwijs en van onderzoek te ondersteunen. De verklarende termenlijsten voor scheikunde, wiskunde en natuurkunde, die binnen het project tot stand kwamen, zijn beschikbaar als pdf bij de Taalmaterialen van de INT-website. Het scheikundebestand is bovendien online doorzoekbaar op scheikunde.ivdnt.org. In 2026 zal aan de wiskunde- en natuurkundebestanden de laatste hand gelegd worden zodat ze in de loop van 2026 ook als online-applicaties beschikbaar komen. Daarnaast zal worden onderzocht of de drie bronbestanden in een interne versie van TermWerk geïmporteerd kunnen worden zodat externe experts kunnen aangezocht worden om op vrijwillige basis de termenlijsten verder uit te breiden.

Op Europees niveau is het INT opnieuw toegetreden tot het consortium dat een nieuwe COST-actie voorbereidt: COST Action Proposal OC-2022-1-26011 **Collaborative Terminology Network**. Het voorstel is herzien en opnieuw ingediend in oktober 2025. Het ENT brengt in deze aanvraag zijn ervaring in met de boven beschreven projecten, die allemaal een sterk collaboratief aspect hebben, alsmede met de TermWerk-tool die precies opgezet is om collaboratieve terminologische projecten mogelijk te maken. Als de aanvraag gehonoreerd wordt, wil het ENT via deze COST-actie zijn Europees netwerk versterken met het oog op verdieping van de samenwerking en vervolgaanvragen bij de Europese onderzoeksfondsen.

7. Grammatica

Sinds 2020 valt grammatica binnen de structurele basistaken van het INT zoals ook vastgelegd in het Taalunieverdrag artikel 4d. Dit betekent dat het INT zorgt voor de ontwikkeling, het beheer en de beschikbaarstelling van de verschillende digitale grammaticaproducten. Hieronder vallen het [Taalportaal](#) en de [e-ANS](#). Daarnaast maakt het INT als infrastructurele partner deel uit van het samenwerkingsverband achter [Taaladvies.net](#), in lijn met artikel 5c in het Taalunieverdrag. Naast de werkzaamheden aan de e-ANS en het Taalportaal (cf. infra), wordt gewerkt aan een geïntegreerd grammaticaportaal, een webpagina die de spil moet vormen van alle grammaticaonderdelen en fungeert als ontvangstpagina voor geïnteresseerde gebruikers, met informatie over projecten en producten, een zoekfunctie voor alle producten en een loket voor vragen.

e-ANS

In 2026 gaat de herziening van de e-ANS verder. Het werk aan deze herziening bestaat uit een aantal componenten: contact onderhouden met externe auteurs, werving nieuwe auteurs, redactie en eindredactie van nieuw herziene hoofdstukken. Door het jaar heen worden één à twee publicatiemomenten georganiseerd, waarbij de herziene hoofdstukken beschikbaar gemaakt worden voor het publiek. Ook werken we op verschillende manieren verder aan de toegankelijkheid van de e-ANS, zoals met een verbeterd register, de tweewekelijkse *Grammaticasafari* en de verdere ontwikkeling van de didactische laag, die met name gericht is op het onderwijs. Ten slotte zal er ook verder worden gewerkt aan de verrijking, zoals de koppeling met [DaGeNTa](#) en de database met taalkundige termen en definities. In de eerste helft van 2026 zal een plan worden ontwikkeld voor de voltooiing in 2029 van de herziene e-ANS, inclusief de didactische laag.

Taalportaal

In 2026 staat een grote inhoudelijke update van het onderdeel Nederlandse syntaxis gepland. Voorafgaand hieraan zal er een uitgebreide, semi-automatische conversie van Word naar XML plaatsvinden. Ook zal de Nederlandse morfologie worden uitgebreid en wordt de webapplicatie verder bijgewerkt.

Grammaticaportaal

In de afgelopen jaren is er gewerkt aan een proefversie van het Grammaticaportaal, een startpagina voor de verschillende grammatica-applicaties binnen het INT: de e-ANS, het Taalportaal, Taaladvies.net, maar ook voor projecten als DaGeNTa (cf. infra) en Woordcombinaties (cf. supra). In 2026 zal deze versie verder uitgebreid worden, met een focus op het eenvoudig kunnen doorzoeken van de verschillende bronnen en het regelmatig publiceren van nieuwsberichten.

8. Nationale en internationale samenwerkingsverbanden

Netwerken

EFNIL

Het INT is lid van EFNIL, de Europese Federatie van Nationale Taalinstellingen, en participeert actief in dit netwerk. In deze federatie komen gelijkaardige taalinstellingen uit alle Europese lidstaten samen. EFNIL biedt de instellingen een forum om informatie over hun werk uit te wisselen en om informatie over taalgebruik en taalbeleid binnen de Europese Unie te verzamelen en te publiceren.

Daarnaast stimuleert de federatie het bestuderen van de officiële Europese talen en een gecoördineerde aanpak van het leren van moedertalen en vreemde talen, als middel om de taalkundige en culturele diversiteit binnen de Europese Unie te bevorderen. Ieder jaar is er een algemene vergadering en wordt er een workshop gehouden rond een actueel thema in verband met de taalinfrastructuur. EFNIL is tevens een belangrijk aanspreekpunt voor andere initiatieven op Europees niveau zoals de European Language Data Space (cf. infra).

IMPACT Centre of Competence

Het INT is voorzitter van het [IMPACT Centre of Competence](#). 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 sinds 2019 ook CLARIN Knowledge Centre.

European Language Data Space

Het INT is het nationale aanspreekpunt binnen de European Language Data Space (LDS). De LDS is door de EU officieel aangesteld om een Europees dataplatform en Europese datamarkt te ontwikkelen voor het verzamelen, creëren, uitwisselen en (her)gebruiken van meertalige en multimodale taaldata. Met het LDS-platform kunnen stakeholders hun taaldata eenvoudig delen en commercialiseren, conform de Europese waarden en EU-regelgeving. De taaldata in het LDS-platform zullen ook beschikbaar zijn voor de ALT-EDIC,

de Europese alliantie voor taaltechnologieën, waarvan Nederland en België (via Vlaanderen) lid zijn.

Elexis Association

Het INT is lid van de [ELEXIS Association](#) die uit het H2020 ELEXIS-project (2018-2022) is voortgekomen. De ELEXIS Association heeft als doel verdere onderzoeksinitiatieven en -activiteiten over lexicografie te bevorderen en te coördineren, zoals de eLex (electronic lexicography in the 21st century)-conferentie en Europese projectaanvragen. In 2025 is de vervolgaanvraag ELEXAI (cf. infra) ingediend.

EURALEX

Het INT is lid van [EURALEX](#), een toonaangevende beroepsvereniging voor mensen die werkzaam zijn in de lexicografie en aanverwante vakgebieden. In de snel veranderende wereld van taalanalyse en taalbeschrijving biedt EURALEX een forum voor de uitwisseling van ideeën. Ook organiseert de vereniging iedere twee jaar een groot congres. In 2026 zal het EURALEX-congres in Wenen gehouden worden van 29 september tot en met 3 oktober.

Nederlandse AI Coalitie

De [Nederlandse AI Coalitie](#) is een publiek-private samenwerking, waarin overheid, bedrijfsleven, onderwijs- en onderzoeksinstituten en maatschappelijke organisaties samenwerken. De coalitie heeft tot doel de Nederlandse activiteiten in AI te stimuleren, te ondersteunen en waar nodig te organiseren. Het INT is bij de use case Nederlandse AI voor het Nederlands (NAIN) betrokken en neemt deel aan het overleg hierover.

KANTL

Het INT heeft in augustus 2025 een samenwerkingsovereenkomst gesloten met de Koninklijke Academie voor Nederlandse Taal en Letterkunde (KANTL) in Gent. Daarin was een postdoc voorzien voor een longitudinaal project voor de monitoring van met name gesproken Nederlands (NELL), complementair aan de Staat van het Nederlands waarin het gaat om gerapporteerd taalgedrag. Door bezuinigingen kan NELL niet van start gaan. Er zal in 2026 gezocht worden naar alternatieve financiering.

Netwerkprojecten

European Network On Lexical Innovation (ENEOLI, 2023 – 2027)

Het INT neemt deel aan het Europese COST-onderzoeksnetwerk European Network On Lexical Innovation ([ENEOLI](#)). De belangrijkste doelstellingen van het netwerk kunnen als volgt worden samengevat: 1) een gemeenschappelijke, meertalige kernterminologie voor lexicale innovatie definiëren; 2) digitale methodes en instrumenten in kaart brengen die gebruikt worden om lexicale innovaties in Europese talen te identificeren en te verklaren; 3) vergelijkende studies uitvoeren naar lexicale innovaties in Europese talen, met speciale aandacht voor ontleningen en hun equivalenten; 4) specifieke training in neologie aanbieden aan vertalers, redacteurs, journalisten, technisch schrijvers en docenten. Het INT neemt vooral deel aan werkgroep 2 over digitale methodes en instrumenten voor de analyse van neologismen en leidt mee een taak over de toepassing van deze methodes binnen de lexicografische workflow voor neologismen.

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

Het INT neemt deel aan het Europese COST-onderzoeksnetwerk [UniDive](#) (Universality, diversity and idiosyncrasy in language technology). Het doel van deze COST-actie is om te onderzoeken hoe taaltechnologie verbeterd kan worden door meer kennis over de overeenkomsten en verschillen tussen talen. Zo beoogt deze COST-actie, theoretisch gezien, een beter begrip van taaluniversalia te krijgen en, praktisch gezien, taaltechnologie te realiseren die om kan gaan met een grotere verscheidenheid aan talen en taalverschijnselen. Het INT is nauw betrokken bij de activiteiten van de werkgroep die zich richt op de lexicon-corpus-interface. Verder werkt het INT binnen de UniDive COST-actie aan het realiseren van een PARSEME-corpus voor het Nederlands. In 2025 is hiervan een eerste versie gereleased.

Onderzoeks- en infrastructuurprojecten

CLARIAH-Vlaanderen+ (2025-2028)

Het INT is als derde partij betrokken bij het Vlaamse research-infrastructuurproject [CLARIAH-VL+](#): Advancing the open humanities service infrastructure. De hoofdtaak van het INT is het voorzien van de benodigde infrastructuur voor het opzetten van een nieuwe tool voor treebank-querying in universal dependency treebanks zodat taalkundigen kunnen

zoeken naar het voorkomen van niet-lineaire syntactische constructies, over verschillende talen heen. Daarnaast blijft het INT als coördinerend centrum voor CLARIN de link verzorgen tussen dit project en de Europese CLARIN-infrastructuur.

CultuR-AI: Culture Re(enrichment) & AI (2027-2028)

Het INT treedt op als subcontracting partner in CultuR-AI, een infrastructuurproject dat in december 2025 is toegekend door BELSPO ESFRI, waarbij INT als CLARIN-BE-centrum instaat voor het hosten van tools en data die in het project aangemaakt worden. Het gaat hierbij om verbeterde semi-automatische metadatageneratie en verbeterde optical character recognition (OCR).

Macroscope (2026-2030)

Het instituut neemt deel aan het Macroscope-project waarin het INT samen met 13 andere partijen in Nederland een consortium vormt. Door voort te bouwen op bestaande Nederlandse infrastructuren zoals ODISSEI en CLARIAH, beoogt dit project de eerste “Macroscoop” ter wereld te creëren, een krachtig hulpmiddel om complexe maatschappelijke problemen op grote schaal te onderzoeken. De Macroscoop zal geavanceerde systemen voor dataverzameling, AI-tools en een platform voor samenwerking omvatten. Het INT is betrokken bij de ontwikkeling van een Data Harvesting Ecosystem door het realiseren van het Netherlands Media Corpus (task 2.3) en de ontsluiting daarvan via de AI Collaboratory (task 3.1).

SSHOC-NL (2024-2029)

Het instituut neemt deel aan het [SSHOC-NL](#)-project (Social Science and Humanities Open Cloud for the Netherlands, 2024-2029), dat voortbouwt op CLARIAH+. Dit project vormt een consortium van Nederlandse onderzoeksinfrastructuren (ODISSEI en CLARIAH) voor de sociale en geesteswetenschappen, gericht op de ontwikkeling van toegankelijke diensten, data en tools. Binnen dit project richt het INT zich op de infrastructuur voor dataverrijking met machinelearning en AI. We implementeren automatische verrijkingsmethoden en ontwikkelen educatieve materialen die gebruikers helpen om deze tools op een methodologisch valide manier voor hun onderzoeksdoelen toe te passen.

In 2026 werken we samen met de Linked Open Data-taak binnen het project om koppelingen te realiseren tussen tools, datasets en trainingsmateriaal. Met de

projectpartners binnen de taak *Erfgoed- en culturele data* wordt een use case opgezet, gericht op het evalueren van bijvoorbeeld handschriftherkenning en de toepassing van multimodale taalmodellen voor de tekstherkenning en structureren van specifieke manuscriptcollecties. In samenspraak met de projectpartners die verantwoordelijk zijn voor training, outreach en communityvorming ontwikkelen we trainingsmateriaal voor relevante tools en methoden, en zorgen we voor duurzame beschikbaarheid op platforms als de SSH Open Marketplace en Zenodo. We verbeteren de vindbaarheid van dit materiaal door een uniform metadatamodel te ontwikkelen. We bevorderen kennisdeling op het gebied van verantwoordelijk toolgebruik door jaarlijks twee workshops te organiseren. Het GaLAHaD-platform voor automatische annotatie en evaluatie wordt uitgebreid met functies voor een documentweergave die de verschillen tussen verschillende tools inzichtelijk maakt, en met functies voor de evaluatie van annotaties op het niveau van tekstsegmenten. Daarnaast wordt binnen BlackLab een AI-module ontwikkeld die zoekvragen in natuurlijke taal omzet in complexe corpusqueries, waardoor de corpusdata laagdrempeliger toegankelijk wordt voor onderzoek.

Project Geparset Corpus van de gesproken Nederlandse Dialecten + (GCND+) (2024-2028)

Het INT is partner in het project *Geparset Corpus van de gesproken Nederlandse Dialecten*. Het project is een vervolg op het project Gesproken Corpus van de zuidelijk-Nederlandse Dialecten (GCND) en voorziet in een aanzienlijke geografische uitbreiding waardoor het corpus het hele Europees-Nederlandse gebied zal bevatten. Het project gebruikt de bestaande audio-gealigneerde transcripties en annotaties uit de eerste fase van het GCND om ASR- en NLP-tools te hertrainen en zo de transcriptie en annotatie van de nieuwe data te versnellen. Er is al een GCND-infrastructuur om de data te ontsluiten. De rol van het INT is om deze infrastructuur aan te passen en uit te breiden.

Spread the News (2020-2025)

In het onderzoeksproject [*Spread the new\(s\). Understanding standardization of Dutch through 17th-century newspapers*](#), gefinancierd door NWO Open Competition SSH en uitgevoerd aan de Radboud Universiteit en het INT, wordt onderzocht welke (socio)linguïstische factoren bepalend zijn bij de functionele implementatie van een standaardtaal. Het INT verzorgt de technische voorzieningen voor dit project, waaronder de

ontsluiting en verrijking van een corpus van 17e-eeuwse kranten dat door vrijwilligers is gedigitaliseerd.

Namenloket (2025-2027)

In dit project wordt een digitaal interactief informatieportaal ontwikkeld over het Vlaamse namenerfgoed, de eigennamen die Vlamingen vanouds zichzelf, maar ook o.a. plaatsen en dieren geven. Bezoekers vinden er een antwoord op vaak gestelde vragen over eigennamen en kunnen er nieuwe vragen kwijt. Via een dataplatform kunnen erfgoedvrijwilligers en -onderzoekers hun eigen namendataverzamelingen digitaal ontsluiten, centraliseren en doorzoekbaar maken. Het portaal bevat ook didactisch materiaal. Het INT is verantwoordelijk voor de website (ontwerp, bouw en hosting) en voor de bouw van digitale infrastructuur en de benodigde protocollen voor die integratie van namenverzamelingen. Bij wijze van aanzet wordt één dataverzameling – het monumentale *Woordenboek van Familienamen in België en Noord-Frankrijk* van Frans Debrabandere – voor het eerst volledig digitaal beschikbaar en doorzoekbaar voor het brede publiek. Namenloket is een samenwerking van Variaties vzw, de Koninklijke Commissie voor Toponymie en Dialectologie (KCDT), Histories vzw, Odisee (onderzoeksgroep onderwijs), de Universiteit Gent (vakgroep Taalkunde), Universiteit Keulen, de Antwerp School of Education (onderzoeksgroep Didactica), Meemoo, Vlaams instituut voor het archief en het INT.

De Taalassistent

Voor het pilootproject De Taalassistent ontwikkelt het INT in opdracht van Onze Taal een proefversie van een virtuele taalassistent: een AI-tool die taalvragen kan beantwoorden op basis van een dataset taaladviezen van Onze Taal. Het INT denkt mee over het ontwerp, assisteert bij het aanleveren en opschonen van de data, en ontwikkelt de tool.

Pressmint

PressMint is een CLARIN-flagship-project dat tot doel heeft een meertalige, vergelijkbare, geannoteerde, vertaalde en interoperabele verzameling corpora van Europese historische kranten uit het begin van de 20^e eeuw samen te stellen. Het INT draagt in ieder geval bij met de conversie en de taalkundige verrijking van de data van het Courantencorpus

9. Projectaanvragen

CLARIAH-PRIME (2027-2031; ingediend september 2025)

CLARIAH is de nationale digitale onderzoeksinfrastructuur voor de geesteswetenschappen en onderdeel van de Europese netwerken CLARIN en DARIAH. Het verbindt data, tools en diensten van universiteiten en erfgoedinstellingen, zodat onderzoekers teksten, beelden en audiovisuele bronnen op nieuwe manieren kunnen bestuderen. Het doel van de upgrade CLARIAH-PRIME is om de toegankelijkheid en het gebruik van deze infrastructuur te verbeteren, verantwoord gebruik van kunstmatige intelligentie in onderzoek te bevorderen en nationale en internationale samenwerking te versterken. Binnen CLARIAH-PRIME werkt het INT, samen met het KNAW Humanities Cluster en de Universiteit Utrecht, aan het werkpakket Integrating Tools for Text Analysis. Dit werkpakket moderniseert en integreert tekstanalyse-workflows — zoals corpusopbouw, annotatie en analyse — die voorheen verspreid waren over verschillende instellingen. De kerninnovatie is de ontwikkeling van herbruikbare workflowcomponenten die gedeeld kunnen worden tussen projecten en onderzoeksdomeinen. Naast een upgrade van diverse tools, o.a. met inzet van AI, zal het INT het community management binnen dit werkpakket verzorgen.

ELEXAI (2026-2029; ingediend september 2025)

Het project bouwt voort op het ELEXIS-project (2018-2022) en beoogt een belangrijke upgrade van de Europese lexicografische infrastructuur (ELEXIS) die uit dat project resulteerde. De upgrade betekent een transformatieve herstructurering van de infrastructuur, naar aanleiding van recente ontwikkelingen op het gebied van AI, met name de opkomst van grote taalmodellen (LLM's). Daarnaast streeft het project ernaar om de huidige basismodellen te verbeteren door hoogwaardige taalkundige kennis in deze modellen te integreren. Indien gehonoreerd, zal het project in 2026 van start gaan.

FIIAT: Flemish Inclusive Infrastructure for Automatic Transcription (2026-2028; ingediend september 2025)

Het INT treedt op als co-financing partner voor dit project, met als doel het kunnen hosten van een automatische transcriptieservice voor video en audio voor het Belgisch-Nederlands,

waarbij deze transcriptieservice ook getraind is op atypische spraak, zoals bijvoorbeeld. van kinderen, niet-moedertaalsprekers, ouderen, mensen met spraakstoornissen e.d.

GAITE: European Doctoral Network focused on Generative AI and Translation (herindiening november 2025)

Het INT is betrokken in GAITE, een aanvraag voor een Marie Skłodowska-Curie Doctoral Network met in totaal 15 doctoraatsstudenten, als Associated Partner, met als rol 'Data provision and hosting via CLARIN' en 'Training'.

10. Overige infrastructurele dienstverlening

CNaVT

Samen met de Taalunie wordt in 2026 een begin gemaakt met het ontwikkelen van een taalleerdercorpus voor het Nederlands op basis van de examens van het Certificaat Nederlands als Vreemde Taal (CNaVT). CNaVT is een project van de Taalunie dat anderstaligen die het Nederlands als Vreemde Taal leren mogelijkheden biedt om hun taalvaardigheid in het Nederlands te laten toetsen en certificeren. In 2026 zal de eerste data binnenkomen, die bij het INT zal worden opgeslagen.

DaGeNTa

Sinds 2023 wordt de website van DaGeNTa (Database Geschiedenis Nederlandse Taalkunde) gehost door INT. Deze database stelt zich als doel historische werken over de Nederlandse taal te ontsluiten. De gegevens kunnen worden verbonden aan de grammaticale websites van het INT. De DaGeNTa-website wordt met de hulp van stagiairs en vrijwilligers verder uitgebreid.

EenvoudigNL

In samenwerking met Stichting Expertisecentrum Oefenen.nl is in een pilootproject een infrastructuur ontwikkeld om semi-automatisch taal oefeningen te genereren uit corpusmateriaal. In 2024 is een detailplan uitgewerkt voor de verdere uitbouw van deze infrastructuur. Er wordt gezocht naar extra financiering hiervoor.

Etymologiebank

Het INT is verantwoordelijk voor het hosten van de etymologiebank van Nicoline van der Sijs. Het werk aan de etymologiebank wordt voortgezet, vooral met behulp van stagiairs en vrijwilligers van universiteiten in Nederland en België: de etymologiebank wordt met nieuwe woordenboeken en datasets verrijkt en daarnaast wordt gewerkt aan de betere en verdere ontsluiting van de bestaande gegevens. In 2026 moet gewerkt worden aan een herimplementatie van zowel de backend van de etymologiebank (inclusief voorziening voor de integratie met de centrale kennisbank) als de frontend.

GLAD

GLAD (Global Anglicism Database Network) is een internationaal project waarin de Engelse invloed op talen wereldwijd wordt geïncinventariseerd. Het INT host de database van dit project en op termijn ook de website en levert technische en inhoudelijke bijdragen aan het project.

Infrastructuur Papiaments

Het INT en Het Nationaal Taalinstituut Curaçao (NTIC) hebben een memorandum van overeenstemming (MoU) en hebben samengewerkt aan de uitbouw van de digitale lexicale infrastructuur voor het Papiaments. Daarvoor heeft het INT een eerste versie van de lexicale database en een bewerkingsomgeving ontwikkeld. Deze samenwerking wordt gecontinueerd en het project wordt versneld in 2026.

Jiddisch-Nederlands woordenboek

In 2024 en 2025 heeft het INT gewerkt aan de verduurzaming van het Jiddisch-Nederlands Woordenboek (JNW) op basis van een subsidie van onder andere Maror. Het JNW wordt nu gehost bij het INT. De Stichting Jiddische Lexicografie Amsterdam is verantwoordelijk voor dit woordenboek.

Plain language en easy language

In samenwerking met de Taalunie zal in 2026 gewerkt worden aan de ontwikkeling van (vervolg)projecten op het terrein van plain language en easy language. Met name zal gekeken worden naar de mogelijkheden in de medische sector en de zorgsector.

11. IT-infrastructuur

Hardware en systemen

De interne computationele infrastructuur van het INT dient in de nodige opslagcapaciteit en rekenkracht te voorzien voor de verschillende taalinfrastructurele opdrachten zoals de taalmaterialenrepository, de eigen productontwikkeling en de infrastructuurondersteuning die noodzakelijk is om onze functie als CLARIN-B-centrum te kunnen uitvoeren. Specifiek voor het CLARIN-B-centrum lopen er projecten om onze repositoryfunctie te verbeteren en is de servicedeskprocedure gestroomlijnd.

Om te kunnen profiteren van de ontwikkelingen op het gebied van AI en taalmodellen, is structurele beschikbaarheid van GPU's noodzakelijk. Daarom streven we naar het uitbreiden van GPU-capaciteit, onder andere door vervangingen van afgeschreven hardware zoveel mogelijk uit te rusten met GPU-capaciteit.

Daarnaast proberen we de afhankelijkheid van dure commerciële producten (vmware, netbackup) te verminderen door over te stappen op open source alternatieven. Verder vereist de positie van onze infrastructuur binnen het ISSC (ICT Shared Service Centre) van de universiteit Leiden aandacht vanwege bezuinigingen aldaar.

Software-ontwikkeling

Het INT ontwikkelt software die het maken en gebruiken van taalmaterialen ondersteunt. Doel is het ontwikkelen van generieke, herbruikbare toepassingen en componenten die het op een efficiënte manier mogelijk maken enerzijds de infrastructuur te ontsluiten voor onderzoekers en het algemene publiek, en anderzijds de opbouw van de infrastructuur te ondersteunen. We streven er zo veel mogelijk naar ook projectspecifieke functionaliteit op een generieke en herbruikbare manier te ontwikkelen.

Met toepassingen als BlackLab en diverse Solr-gebaseerde woordenboekplatforms (ANW, Historische woordenboeken) kunnen onderzoekers efficiënt zoeken in corpora en woordenboeken. Recent is een aanpasbaar Solr-gebaseerd publicatieplatform ontwikkeld waarmee woordenboektoepassingen efficiënt kunnen worden gebouwd. Tools als Woordpeiler en Woordenlijst.org maken taaldata ook toegankelijk voor een breder publiek. Met het GaLAHaD-platform kunnen onderzoekers taalkundige verrijking van

corpusmateriaal toevoegen, maar ook de kwaliteit van de tools voor taalkundige verrijking evalueren. Voor de ontwikkeling en verrijking van taalbronnen onderhoudt het INT ook systemen voor handmatige invoer en correctie. Het Rapid Application Developmentplatform Lex'it, wordt gebruikt voor gestructureerde data die in PostgreSQL opgeslagen worden, zoals bijvoorbeeld het centrale lexicon. SwingLex is een dictionary writing system voor XML-gebaseerde woordenboekdata. LAnCeLoT, ten slotte, is een omgeving voor manuele annotatie en correctie van met woordsoort en lemma verrijkt corpusmateriaal.

In 2026 wordt verder gewerkt aan BlackLab en de BlackLab-frontend, onder andere in het kader van GCND+ en Macroscopic; het GaLAHaD-platform wordt verder ontwikkeld binnen SSHOC-NL (zie betreffende onderdelen). Het publicatieplatform voor lexicale data wordt onder andere ingezet en verder ontwikkeld in het JuriTerm-project en bij de ontwikkeling van een nieuwe versie van de etymologiebank. Een andere rode draad is de inzet van AI-technologie binnen de infrastructuur (onder andere Macroscopic, SSHOC-NL, ELEXAI, lexicografische workflow).

12. Disseminatie: onderzoek, onderwijs en het algemene publiek

Aansluitend bij de algemene doelstellingen van het Taalunieverdrag, en in nauwe samenwerking met de Taalunie, levert het INT ook in 2026 zijn bijdrage aan de bevordering van de kennis over de Nederlandse taal en neemt daarbij initiatieven om de taalinfrastructuur voor het Nederlands, en het onderzoek rond de opbouw ervan, bij een ruim publiek bekend te maken. Daarbij wordt zowel aandacht geschonken aan historisch en modern Nederlands als aan standaardtaal en regionale variëteiten binnen het hele Nederlandse taalgebied (Nederland en de Caribische rijkdelen, Vlaanderen en Suriname).

Doelgroepen

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 in 2026 zowel binnen samenwerkingsprojecten als door geregelde contacten onderhouden en waar mogelijk geïntensiveerd en uitgebreid. De medewerkers van het INT maken ook in 2026 het onderzoek aan het instituut bekend via wetenschappelijke congressen en publicaties.

Onderzoek en onderwijs

Voor universitaire studenten verzorgt het INT twee verschillende collegereeksen over computationele lexicografie. Aan de Universiteit Leiden gaat het om het vak *Corpus Lexicography* binnen de researchmaster Linguistics en aan de KU Leuven om het vak *Computationele Lexicologie* binnen de master Taalkunde. Er worden ook masterproeven begeleid binnen deze masteropleidingen en binnen de Master of Artificial Intelligence aan de KU Leuven.

De communicatie naar onderzoekers en onderzoeksinstituten is in 2026 beter gefaciliteerd door de nieuwe, Engelstalige website van het INT met informatie gericht op onderzoekers uit het Nederlandse taalgebied en daarbuiten. Op deze webpagina's wordt een overzicht gegeven van het toegepaste onderzoek aan het INT en van de verschillende manieren waarop andere onderzoekers met het instituut kunnen samenwerken in het kader van

ationale of Europese projecten of via onderzoeksstages. In 2026 wordt verder gewerkt aan het synchroniseren van de informatie op de Engelstalige en Nederlandstalige website.

Voor het onderwijs blijft het INT zijn kennis en taalmaterialen nog beter toegankelijk maken, onder andere via webinars. In dat verband is het INT aanwezig op en profileert het zich op onderwijsgerichte bijeenkomsten, conferenties (HSN-conferentie) en evenementen (Neerlandistiekdagen, Taalfeest, Week van het Nederlands, Taalkundeolympiade). Op de website heeft onderwijs een vaste plaats. De daar te vinden beschikbare informatie en materialen zoals lesbrieven en posters voor in de klas worden in 2026 up-to-date gehouden en waar nodig uitgebreid.

Algemeen publiek

Ook het algemene publiek wordt niet uit het oog verloren. De website is ingericht om een zo breed mogelijk publiek te bedienen. We publiceren wekelijks populairwetenschappelijke rubrieken onder de noemer ‘Woorden van de week’: *Nieuw woord van de week* over nieuwe woorden, *Uit de streek* over dialect, *Grammaticasafari* over grammatica en de e-ANS en *Terug in de taal* over historisch Nederlands. Net zoals in de voorgaande jaren zal in 2026 minimaal zes keer per jaar een INT-nieuwsbrief verstuurd worden aan onze abonnees. Daarnaast is er een nieuwsbrief terminologie die vier keer per jaar verschijnt en specifiek informatie geeft over vaktaal. In 2026 worden er weer regelmatig publieksevenementen georganiseerd, zowel live als digitaal. Met webinars, lezingen, taalquizen en berichten op de sociale media zoals Instagram, LinkedIn en Bluesky brengt het INT in 2026 ook online (de werkzaamheden van) het instituut bij alle doelgroepen onder de aandacht. De twee podcasts in samenwerking met Onze Taal – *Waar komt pindakaas vandaan?* en *Over taal gesproken* – worden vervolgd met een nieuw seizoen. Extra aandacht gaat in 2026 uit naar publiciteit rondom Woordpeiler, de tool die corpustrends visualiseert en taaldata toegankelijk maakt voor een breed publiek.

Neerlandistiek, taalmaterialen en Taalunie

Sinds 2023 heeft het INT niet alleen de technische verzorging maar ook de hosting van Neerlandstiek.nl op zich genomen; tevens wordt het INT in de redactie van Neerlandistiek vertegenwoordigd door Nicoline van der Sijs. Medewerkers van het INT leveren elke maand een of meer bijdragen aan Neerlandistiek.

Met de Taalunie wordt samengewerkt aan het onder de aandacht brengen van taalmaterialen in het kader van het project 'benutting van digitale infrastructuur'. Er is maandelijks overleg tussen de communicatieverantwoordelijken van het INT en de Taalunie.