

INL (Nieuwsbrief)

september 2006



TST-centrale

Het groeiende aandeel van taal- en spraaktechnologie in de maatschappij heeft ervoor gezorgd dat de beschikbaarheid van Nederlandse en Vlaamse digitale taalkundige bronnen steeds belangrijker wordt. De TST-centrale geeft gehoor aan deze ontwikkeling door het beheren, onderhouden en distribueren van door de overheid gefinancierde digitale taalmaterialen. De centrale fungeert zodoende als een kennis- en distributiecentrum waar digitale taalkundige bronnen opgevraagd en geraadpleegd kunnen worden.

Het Corpus Gesproken Nederlands (CGN) is één van de producten die de TST-centrale onder haar hoede heeft. Na grondig onderhoud is er begin dit jaar een nieuwe versie uitgebracht met onder andere toevoeging van ontbrekende geluidsbestanden en verbeteringen in een aanzienlijk aantal tekstbestanden. Met de lancering van de nieuwe versie wordt door middel van het aanbieden van workshops het gebruik van het CGN verder gestimuleerd. Naast het CGN worden ook andere corpora en lexica beheerd, zoals het Eindhoven Corpus, het Parole Corpus en het Parole Lexicon. Nieuwe taalmaterialen die in ontwikkeling zijn zullen in de toekomst toegevoegd worden aan de huidige productenlijst.

In mei dit jaar is op het internationale congres *Language Resources and Evaluation* (LREC) in Genua onder andere de productiestraat van de TST-centrale gepresenteerd. Daarnaast trad de centrale op als bronzen sponsor van het congres. Het streven is het hergebruik van de beschikbare taalbronnen maximaal te stimuleren. Overigens: de TST-centrale brengt ook een eigen nieuwsbrief uit.

Geïntegreerde Taalbank (GTB)

Alle betekenissen van ‘natuur’ in de Nederlandse taal op een rij, alle varianten van het woord ‘pelgrim’ – mét dialectkaarten!– die tot nu toe in onze taal zijn voorgekomen, citaten van Vondel waar ‘god’ in voorkomt, een overzicht van woorden met ‘walvis’ in de betekenisomschrijving, het kan straks allemaal worden opgevraagd in de internetapplicatie van de grote historische woordenboeken van de Nederlandse taal.

Het systeem vormt een onderdeel van de GTB, waarin naast de woordenboekencomponent (met ONW, VMNW, MNW en WNT) ook een gekoppeld diachroon corpus van het Nederlands opgenomen wordt. In Leiden (International Conference on Historical Lexicography and Lexicology) en Leeds (International Medieval Congress) zijn voor een internationaal gezelschap van vooraanstaande historische lexicografen en mediëvisten, succesvolle demonstraties gegeven. Hoewel aan data en programmatuur nog veel te doen is, kon reeds een systeem gepresenteerd worden dat de belangrijkste functies van de woordenboekapplicatie laat zien. Bovendien kon met behulp van het Voegmiddelnederlands materiaal ingegaan worden op de mogelijkheden om woordenboeken en tekstmateriaal aan elkaar te koppelen. Zo zal een gebruiker de complete tekst van de bron van een citaat kunnen inzien, of van een corpustekst direct bij relevante betekenisomschrijvingen in de woordenboeken uit kunnen komen. “Een adembenemend vooruitzicht”, zo concludeerde Ewoud Sanders in zijn Woordhoek in NRC Handelsblad.

INL Nieuwsbrief

Het Woordenboek der Nederlandsche Taal en het Groene Boekje zijn de bekendste schatten van het Instituut voor Nederlandse Lexicologie. Maar er gebeurt veel méér. Veel van de ontwikkelingen binnen het INL blijven voor de buitenwereld verborgen. Met een nieuwsbrief willen we daar verandering in brengen. Naast de informatie op de website brengen we u voortaan halfjaarlijks op de hoogte van de gebeurtenissen op het INL.

In deze eerste uitgave krijgen alle afdelingen de gelegenheid om zich voor te stellen en de meest recente ontwikkelingen binnen de afdeling met u te delen. Reacties naar aanleiding van de nieuwsbrief kunt u sturen naar secretariaat@inl.nl.

Colofon

Deze nieuwsbrief is een uitgave van de Stichting Instituut voor Nederlandse Lexicologie en verschijnt tweemaal per jaar.
Postbus 9515, 2300 RA Leiden
t 071-5141648
www.inl.nl
Redactie: secretariaat@inl.nl
Ontwerp: Swantje Haage Ontwerp, Amsterdam

INL
INSTITUUT VOOR
NEDERLANDSE
LEXICOLOGIE

INL
INSTITUUT VOOR
NEDERLANDSE
LEXICOLOGIE



Descriptieve Historische Projecten (DHP)

Met de projecten Etymologisch Woordenboek van het Nederlands (EWN), Oudnederlands Woordenboek (ONW) en Woordenboek der Nederlandsche Taal op internet (iWNT) beschrijft en onderzoekt de afdeling DHP taalfasen van de Nederlandse Taal.

EWN

In november 2005 is deel 2 van het EWN verschenen. Deel 3 verschijnt eind 2007 en zal lemma's tot en met de letter R gaan bevatten. De trefwoordenlijst voor de resterende twee delen is bijgewerkt en ingedeeld in hoofd- en sublemma's. Een regelmatige stroom van nieuwe kopij vindt inmiddels zijn weg weer langs de hoofdredactie en de vakspecialisten.

ONW

Het werk aan het ONW vordert eveneens gestaag. In totaal zijn er inmiddels zo'n 2000 lemma's in enig stadium van bewerking. De tot nu toe min of meer definitief geredigeerde lemma's voor het ONW staan op een afgeschermd gedeelte van de INL-website en worden op verzoek voor onderzoek aan derden ter beschikking gesteld. Belangstellenden kunnen hiervoor contact opnemen met Dick Wortel (wortel@inl.nl).

Sinds kort maakt de redactie gebruik van een lexicografisch werkstation. Dit werkstation, dat is gebaseerd op het reeds eerder ontworpen werkstation van het ANW, bevat de beide corpora met Oudnederlands materiaal, een database voor de in bewerking zijnde en bewerkte ONW-lemma's, de digitaal beschikbare naslagwerken en een administratief systeem waarin de voortgang van de bewerking van het materiaal nauwkeurig wordt bijgehouden.

iWNT

Een nieuwe digitale standaardversie van het WNT is vervaardigd met een verbeterde structuurcodering. Dankzij deze codering wordt het onder andere mogelijk om een overzicht van betekenisomschrijvingen op scherm te presenteren, waardoor de gebruiker van de online-applicatie van het WNT straks de structuur van een lemma kan bekijken zonder gehinderd te worden door het citatenmateriaal. Uiteraard kunnen de citaten vervolgens wel naar believen zichtbaar worden gemaakt.

Nieuw in deze versie is ook de toevoeging van een volgens de laatste spelling omgespeld equivalent aan de tot nog toe alleen in de spelling De Vries-Te Winkel beschikbare trefwoorden. Hiermee kan de gebruiker straks ook op basis van de hedendaagse spelling zoeken in het WNT. Verder is de woordsoortinformatie eenduidig opvraagbaar gemaakt.

Algemeen Nederlands Woordenboek (ANW)

Het ANW is een elektronisch woordenboek van de moderne standaardtaal in Nederland en België (periode 1970–2018), dat via internet ter beschikking zal worden gesteld. Het ANW is zodanig vormgegeven dat zowel de leek als de vakman snel over de gevraagde informatie kan beschikken.

De conceptfase van het woordenboek is nagenoeg afgerond, waarbij van meet af aan rekening is gehouden met de mogelijkheden, eisen en problemen die aan het maken van een nieuw elektronisch woordenboek verbonden zijn. Speciaal voor het ANW werd een elektronisch bewerkingsformulier ontwikkeld. De daarin vastgelegde microstructuur van het ANW omvat ongeveer 208 informatiecategorieën of elementen die verdeeld zijn over de rubrieken lemma, spelling, uitspraak, woordsoort, woordvorming, betekenis en gebruik.

Nieuwkomer bij 'betekenis en gebruik' is het semagram. Een semagram is een soort frame, een structuur met 'slots' en 'fillers' waarin kennis van een bepaald concept wordt gepresenteerd. Bij een drankbenaming als wijn fungeren kenmerken als 'samenstelling', 'bereiding', 'grondstof', 'kleur', 'smaak' en 'gebruik' als slots. In feite zijn het de abstracte elementen die de eigenschappen en relaties aangeven die relevant zijn voor woorden uit de semantische klasse waartoe het trefwoord behoort. De 'fillers' zijn de bij die labels ingevulde concrete gegevens die voor het trefwoord in kwestie gelden. Die structuren zijn niet uit de lucht komen vallen. Op basis van analyses van definities voor een selectie van ongelede substantieven, werkwoorden en adjectieven onderscheidt het ANW verschillende semantische klassen. Voor die klassen zijn 'typesjablonen' ontworpen, die vervolgens bij de bewerking worden ingevuld. In die technische zin kan men het semagram ook beschouwen als een 'ingevuld typesjabloon'. Het structuurmodel bij wijn is dus een structuurmodel dat bij alle drankbenamingen terugkeert.

Het semagram zal in het ANW naast de definitie deel uitmaken van de betekenisbeschrijving. Om een aantal redenen is het semagram een vernieuwing die niet alleen voor professionele, maar ook voor 'gewone' taalgebruikers interessant is: het expliciteert kennisgegevens die in de definitie alleen impliciet besloten liggen; het kan op een coherente, gestructureerde wijze veel meer gegevens bevatten dan de definitie; het zorgt voor veel meer en genuanceerdere zoek- en opvraagmogelijkheden van inhoud naar vorm; het vormt met andere semagrammen een grote verscheidenheid aan lexicale en conceptuele relaties; elementen ervan spelen onder meer ook een rol bij collocaties, polysemie en betrekkingen tussen de leden van samenstellingen van het trefwoord. Mede op basis van de semagrammen kunnen systematisch bepaalde, correcte en consistente definities worden gegeven, zoals vereist voor een nieuw te vervaardigen elektronisch woordenboek.

Op de website van het INL (onder Projecten, ANW) zijn de interne nota's en publicaties voor intern gebruik ook voor externe geïnteresseerden beschikbaar. Deze documenten zijn thematisch geordend en gaan over: het corpus, definitieanalyse, semantische klassen, sjablonen, semagrammen, lexicaal-semantische en syntactische combinaties, labels, betrekkingen tussen leden van samenstellingen, de microstructuur, polysemie, neologismen, morfologie, woordselectie en de planning.

Een voorbeeld van de resultaten die het onderzoek voor het EWN oplevert is het lemma *lijp* 'dwaas, dom'. Terwijl er tot nog toe meestal verband werd gezocht met mnl. *leep*, vmnl. *lijp* 'scheel, schuin', blijkt uit de aanhalingen in het EWN duidelijk dat het om een leenwoord uit het Jiddisch moet gaan. *lijp* bn. 'dwaas, dom'

Eerst als zn. 'scheldwoord voor een Jood', in *wat 'n leelijke lijp is dat* [1917; van Looy], als algemener scheldwoord *lijp* 'lelijk, slapjanus' [voor 1941; van Dale 1970], pas later ook als bn. *lijp* 'onnozel, niet goed wijs' [1974; Koenen]. Als zn. ontleend aan de Jiddische eigennaam *Leib*, *Leip*, die teruggaat op Hoogduits *Löwe*, als substituut voor de Joodse eigennaam *Levi*. Later werd het woord ook als bn. gebruikt.

Lit.: J. van Looy (1917), *Jaapje*, Amsterdam, 245.