

INL (Nieuwsbrief)

maart 2007



Vroegmiddelnederlands Woordenboek (VMNW) online!

Wil je een lijst van plantennamen uit de dertiende eeuw, wil je weten of je familienaam reeds in de dertiende eeuw voorkwam, ben je een bijbelse naam tegengekomen en je weet niet wie er bedoeld wordt, wil je precies weten wat een bepaalde 13^e-eeuwse rechtsterm inhoudt, dan kan je vanaf april 2007 terecht in het VMNW online!

Op 26 januari jl. kwam het WNT online, als eerste deel van de woordenboekencomponent van de Geïntegreerde Taalbank, waarin de grote woordenboeken van de Nederlandse taal geïntegreerd opvraagbaar zullen worden gemaakt. In april 2007 wordt het VMNW eraan toegevoegd en is integratie van meerdere woordenboeken een feit.

Lezingen op het INL

In maart start de lezingencyclus *Capita Selecta Lexicologie*. Het thema is 'Collocaties in lexicologie en lexicografie: theorie en praktijk'. De sprekers en de titels van hun lezingen kunt u vinden op onze website: www.inl.nl. Op vrijdag 13 april wordt de Kiliaan-lezing gehouden. Deze vindt plaats in het Verbarium van het INL. Spreker is dr. Wil Pijnenburg. De titel van zijn lezing is 'Herkomst bekend: Marten Toonder als woordjesschepper'. Alle belangstellenden zijn welkom!

Het VMNW beschrijft het Nederlands van de dertiende eeuw en is gebaseerd op het zogenaamde *Corpus Gysseling*, dat naast literaire teksten en fragmenten ruim 2200 ambtelijke documenten bevat. Het is een bron van informatie voor taalkundigen, filologen en letterkundigen. Door de uitgebreide beschrijving van de woordenschat van ambtelijke documenten is het ook een handig instrument voor historici, zeker als zij werken op meer specifieke terreinen als de rechtsgeschiedenis, de nederzettingsgeschiedenis, de ontwikkeling van het sluis- en dijkwezen etc. Interessant voor naamkundigen is de opname van antroponiemen en toponiemen. Ook literaire namen, inclusief Bijbelse, zijn van betekenis en commentaar voorzien.

Het VMNW biedt nog meer zoekmogelijkheden dan het WNT, en kan zowel afzonderlijk als samen met het WNT doorzocht worden. Ook bestaat de mogelijkheid om bij elk lemma dialectvariatie zelf in kaart te brengen. Het VMNW is straks toegankelijk via vmnw.inl.nl en via wnt.inl.nl.



Descriptieve Historische Projecten (DHP)

***iWNT*.** Het kan bijna niemand ontgaan zijn: het Woordenboek der Nederlandsche Taal is op internet raadpleegbaar. Na de officiële presentatie op 26 januari jl., barstte het mediageweld los. Het NOS-achtuurjournaal meldde een overbelasting van de servers door het grote aantal belangstellenden. Binnen vier dagen konden we 10.000 gebruikers tellen; op dit moment zijn dat er al ruim 30.000!

Na deze eerste succesvolle module, wordt er de komende tijd onder andere gewerkt aan het dateren van citaten en het aanvullen van de bronnenlijst. De vrijwilligers die zich hebben ingezet voor het koppelen van externe informatie, gaan naarstig verder. Na een groot aantal afbeeldingen van planten volgen nu de afbeeldingen van dieren. Ook zullen het *Etymologisch Woordenboek van de Nederlandse Dialecten* van prof. Weijnen, het supplement op het *Etymologiewoordeboek van Afrikaans* en het in november 2007 te verschijnen derde deel van het *Etymologisch Woordenboek van het Nederlands* aan het WNT worden gekoppeld.

Oudnederlands Woordenboek (ONW)

Een aantal woorden uit het ONW kennen we alleen dankzij plaatsnamen uit het gebied waar ooit Oudnederlands werd gesproken. Dat geldt bijvoorbeeld voor het woord *alvit*, dat 'zwaan' betekent en voorkomt in de Noord-Hollandse plaatsnaam Elft (Onl. *alvitlo*) en voor *fagar* 'mooi' dat voorkomt in de inmiddels verdwenen plaatsnaam Uagara felda, die eveneens in Noord-Holland is aangetroffen. Alle plaatsnamen waarbij een Germaanse herkomst vermoed wordt, worden verzameld. De Oudnederlandse woorden waarmee zij zijn gevormd, worden als afzonderlijke trefwoorden opgenomen en beschreven. Dit levert aardig wat trefwoorden op die anders niet of in elk geval niet zo vroeg al in het Nederlands zouden zijn gevonden.

Etymologisch Woordenboek van het Nederlands (EWN)

De letters K en L zijn inmiddels zo goed als klaar; aan de M, de N en de O wordt gewerkt, en de lemma-schrijvers hebben inmiddels al de eerste porties van de P geschreven. Na verwerking van eventueel commentaar, afkomstig van een team van vakspecialisten, zorgt de hoofdredactie vervolgens voor de afronding.

Automatisering

De werkzaamheden van de afdeling Automatisering stonden de afgelopen tijd grotendeels in het teken van het online brengen van het Woordenboek der Nederlandsche Taal. De data, programmatuur en ondersteunende systemen werden volgens planning opgeleverd. Uitgebreide tests van de juiste werking van de gebruikersinterface met diverse browsers werden uitgevoerd. De performance werd bestudeerd, waarbij verschillende aantallen gelijktijdige gebruikers werden gesimuleerd. De resultaten waren bevredigend en de eerste release was stabiel. Enkele bugfixes werden doorgevoerd op 1 februari 2007.

TST-centrale

Vanaf het einde van 2006 zijn er verschillende nieuwe taalmaterialen beschikbaar via de TST-centrale. Naast een lijst met de nieuwste Nederlandse woorden en uitdrukkingen, een referentiebestand met typisch Belgisch-Nederlandse woorden en een taalkundige bibliografie, zijn er ook een tweetal historische taalbronnen online toegankelijk gemaakt; het *Corpus Oudnederlands* en het *Woordenboek der Nederlandsche Taal*.

Neologismen online versie 2

In *Neologismen online* zijn ongeveer 18.000 nieuwe Nederlandse woorden en uitdrukkingen opgenomen, afkomstig uit de periode juni 2000 tot en met december 2005. De online woordenlijst is alfabetisch gerangschikt en ieder neologisme is voorzien van een bronvermelding. Bij een beperkt aantal neologismen is gedetailleerde informatie beschikbaar zoals de definitie, woordsoort, woordtype, en etymologie.

Referentiebestand Belgisch-Nederlands (RBBN)

Het RBBN is een lexicale databank met ongeveer 4.000 woorden en uitdrukkingen die typerend zijn voor het Nederlands in België. De trefwoorden zijn voorzien van een beperkte syntactische, semantische en pragmatische beschrijving. De beschrijving van het Belgisch-Nederlands wordt gerelateerd aan de beschrijving van Noord-Nederlandse varianten. Het bestand geeft daarnaast ook informatie over de vindplaats van de woorden en het gebruiksoordeel van informanten. Het RBBN bevat alleen die woorden die in het Belgisch-Nederlands een eigen betekenis hebben en vormt daarmee een aanvulling op het *Referentiebestand Nederlands* (RBN).

Corpus Oudnederlands

De oudste woorden van het Nederlands zijn vanaf nu online raadpleegbaar in het *Corpus Oudnederlands*. In het corpus is al het overgeleverde Nederlandstalige woordmateriaal uit de periode vanaf circa 475 tot en met het jaar 1200 opgenomen. Er kan gezocht worden naar de Oudnederlandse woordvormen die in de tekstedities voorkomen, maar ook naar gestandaardiseerde Oudnederlandse en moderne Nederlandse vormen. Aan elke woordvorm is informatie gekoppeld over de woordsoort, de verbuiging en de bron met de bijbehorende lokalisering en datering. Ook is er voor elk woord een stukje context beschikbaar. De in het corpus opgenomen bronnenlijst is eveneens raadpleegbaar.

Alle materialen zijn online toegankelijk via de productenlijst van de TST-centrale: www.tst.inl.nl/producten

Bibliographie Linguistique online

Sinds 1949 verschijnt onder auspiciën van het Comité International Permanent des Linguistes (CIPL) de internationale taalkundige bibliografie *Bibliographie Linguistique* (BL). De bibliografie werd vanaf 1983 geproduceerd door de Koninklijke Bibliotheek in Den Haag en per 1 januari 2007 is deze taak overgedragen aan het INL. De TST-centrale heeft daarmee ook het beheer over de publieksdatabank BOnline overgenomen. De databank behoudt dezelfde naam als voorheen en is voor iedereen vrij toegankelijk.



Algemeen Nederlands Woordenboek (ANW)

De aanleg van een eigen corpus, het ‘ANW-corpus’, en de conceptvorming voor het ANW, een onlinewoordenboek van het eigentijdse Nederlands, zijn afgerond. Er zijn proefartikelen gemaakt en de redactionele handleiding is voltooid. De redactie is begonnen met de feitelijke bewerking. In de eerste fase richt zij zich daarbij met name op de betekenisbeschrijving van ongelede substantieven uit de klasse van de stofnamen en voorwerpsnamen.

De nomenclatuurlijst van woorden (de lijst van in een woordenboek op te nemen woorden) die met een frequentie van 15 of meer in het ANW-corpus voorkomen, heeft zijn beslag gekregen. Die lijst bevat ruim 70.000 trefwoorden. Uit de lijsten van woorden met een lagere frequentie zullen naar schatting nog ongeveer 10.000 hoofdtrefwoorden en ca. 250.000 subtreffwoorden voor uitvoerige behandeling resp. vermelding in het ANW in aanmerking komen.

In voorbereiding is de automatische invoer van data uit het Groene Boekje voor de categorieën spelling en woordsoort. In het eerste kwartaal van 2007 zullen de gegevens voor ruim 50.000 woorden ingevoerd kunnen worden. Verder staat de automatische invoer van de data voor ongeveer 1000 afkortingen op stapel.

Het verzamelen van neologismen en de bouw van het neologismencorpus gaat onverminderd voort. Via de TST-centrale zijn de neologismen online raadpleegbaar.

Het ANW-corpus, een corpus dat voornamelijk op het geschreven Nederlands is gebaseerd en dat ca.104 miljoen ‘tokens’ (afzonderlijke woordvormen) bevat, hopen we in de eerste helft van 2007 op internet beschikbaar te stellen. Voorts streven we ernaar de eerste bewerkingsresultaten van het ANW rond de jaarwisseling 2007/2008 op internet te presenteren.



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