

Joost Buysschaert

Projecten medische terminologie aan de Hogeschool Gent

1 De Gentse vertaalopleiding

In deze bijdrage gebruik ik gemakshalve de omschrijving 'de Gentse vertaalopleiding'. Hiermee bedoel ik de vertaalopleiding die in 1968 in Gent is opgestart, maar die naar aanleiding van fusies of andere wijzigingen diverse namen heeft gekend: Hoger Instituut voor Vertalers en Tolken (HIVET), Provinciale Hogeschool voor Vertalers en Tolken (PHVT), Departement Vertaalkunde van de Mercator Hogeschool, Departement Vertaalkunde (later Faculteit Toegepaste Taalkunde) van de Hogeschool Gent. Omdat dit artikel een historisch overzicht wil geven, is het van belang te weten dat het hierbij in wezen altijd om dezelfde instelling en opleiding gaat.

De naam 'vertaalopleiding' is overigens te eng. Binnen de huidige Faculteit Toegepaste Taalkunde bestaat er immers niet alleen een Master of Arts in het vertalen, maar ook een Bachelor in de toegepaste taalkunde, een Master of Arts in het tolken, een in de meertalige communicatie, en een drietal postgraduate opleidingen. In al die opleidingen speelt terminologie een belangrijke rol en wat hierna beschreven wordt, heeft dus niet alleen betrekking op de opleiding van vertalers in strikte zin. Onderzoek en onderwijs met betrekking tot terminologie zijn momenteel vooral geconcentreerd in het Centrum voor Terminologie (CvT) enerzijds – dit is een samenwerkingsverband tussen de taalvakgroepen – en de Vakgroep Taaltechnologie (LT3) anderzijds.

Tot slot moet ik ook nog meegeven dat de genoemde opleidingen in het najaar van 2013 zullen worden overgeheveld naar de Faculteit Letteren en Wijsbegeerte van de Universiteit Gent en er zullen worden ondergebracht binnen een 'Vakgroep vertaalkunde en meertalige communicatie'.

2 Het drietalig vocabularium

De Gentse vertaalopleiding heeft een lange traditie van medisch terminologisch onderzoek. Het startschot werd gegeven met het nu nog moeilijk te verkrijgen rode boekje met de lange titel *Drietalig vocabularium van wetenschappen*.

lijke en populaire medische termen Nederlands/Frans/Duits uit 1986.¹ De uitgever was het Belgische ministerie van Volksgezondheid en de redacteur was Paul van Hauwermeiren, neerlandicus van het toenmalige HIVET (en later nog Algemeen Directeur van de Mercator Hogeschool). Het werk bevatte ongeveer 1400 medische begrippen, geselecteerd uit bestaande geneesmiddelenbijsluiters, en gaf voor elk begrip zowel de technische term als een meer populaire variant voor het Nederlands, Frans en Duits (de drie landstalen van België). De bedoeling van het vocabularium was bij te dragen tot patiëntvriendelijke bijsluiters. In de jaren '80 kregen geneesmiddelen in België al bijsluiters mee, maar het ging om 'wetenschappelijke' teksten in technische taal. In 1984 nam minister Wivina Demeester het initiatief om de bijsluiters leesbaar te maken voor wie tot zijn zestiende naar school was geweest. Het drietalige vocabularium moest dit mogelijk helpen maken. Tussen 1988 en 1991 werden 6000 wetenschappelijke bijsluiters vervangen door patiëntenbijsluiters. Het initiatief trok de aandacht van de Europese overheden en de kiem was gelegd voor een veel breder project.

3 Multilingual Glossary

In de jaren tachtig van de vorige eeuw was patiënteninformatie bij geneesmiddelen niet overal in Europa vanzelfsprekend. In 1992 vaardigde de Europese Economische Gemeenschap een Richtlijn uit (92/27/EEG) die voorschreef dat vanaf januari 1994 alle geneesmiddelen in de lidstaten bijsluiters moesten bevatten in een taal die voor de patiënt "duidelijk en begrijpelijk" is (artikel 8 van de Richtlijn).

Medicus Robert Vander Stichelen, die al een belangrijke rol had gespeeld bij het tot stand komen van het Drietalig Vocabularium, nam naar aanleiding van de nieuwe Europese regelgeving weer contact op met de Gentse vertaalopleiding, nu met een groter doel: een uitbreiding van het vocabularium naar de 9 talen die de EEG toen rijk was. Het project kreeg subsidie van de Europese Commissie (DG III) en kreeg de naam "Multilingual Glossary of Technical and Popular Medical Terms". Er kwam ook steun vanuit het Heymans Instituut voor Farmacologie van de Gentse Universiteit.

Vermeldenswaard hierbij is dat het project in oorsprong uitging van Nederlandse termen, hoewel het Nederlands niet de belangrijkste taal van het toenmalige Europa was. Een nieuwigheid in die tijd was echter dat er Nederlandstalige bijsluiters beschikbaar waren gekomen op cd-rom. Met de toen al bestaande software werden frequentielijsten gegenereerd op basis van dit Geneesmiddelencompendium. Uit de frequentste medische vaktermen

werden er 1830 geselecteerd die relevant werden geacht voor het project. Het ging om technische termen die mogelijk de leesbaarheid in de weg stonden, en waarvoor dus een tegenhanger in leken taal wenselijk was.

De 1830 termen werden eerst vertaald naar het Engels, omdat die taal de nieuwe brugtaal zou worden voor het project. Al gauw bleken er problemen te ontstaan met polysemie en daarom werd er voor elk begrip een Engelse betekenisomschrijving opgezocht en genoteerd. De volgende stap was de vertaling van de Engelse termen naar het Frans, Duits, Spaans, Italiaans, Portugees, Nieuwgrieks en het Deens. De laatste stap bestond uit het voorstellen van equivalente lekentermen in alle negen talen.

Het ging al met al om een omvangrijk project en aangezien de Gentse opleiding niet alle genoemde talen in eigen huis had, moesten er ook medewerkers van buitenaf worden ingezet. Er stond grote druk op alle medewerkers om de beloofde deadlines te halen. Alle werkzaamheden werden in niet veel meer dan twee jaar tijd (van 1993 tot 1995) voltooid.

Wat de publicatie van het multilinguaal glossarium betreft, werd voor een toen innovatieve vorm geopteerd, namelijk een publieke website. De site is ondertussen nog steeds beschikbaar op <http://users.ugent.be/~rvdstich/eugloss/welcome.html> en wordt nog druk geconsulteerd. De webstek wordt ook op tal van andere sites vernoemd, en is overigens al een aantal keren schaamteloos gekopieerd.

De site mag dan wegens haar rijkdom populair zijn, toch loopt ze op een aantal punten mank. In de vroege dagen van het internet slaagden de webmasters er bijvoorbeeld niet in om het Nieuwgriekse materiaal te incorporeren. Wie het materiaal van naderbij onderzoekt, kan ook niet naast een aantal fouten kijken die haastwerk verraden. De Engelse definities die voor eenduidigheid moesten zorgen, blijven in een aantal gevallen toch polyseem. Niet alle medewerkers hebben overigens de definities secuur geraadpleegd toen ze hun vertaling voorstelden, zodat de equivalentie in een aantal gevallen te wensen overlaat. Soms kan men ook kanttekeningen maken bij de voorgestelde populaire varianten. Een beperking die te maken heeft met de gebruikte selectiemethode, is dat er uitsluitend is uitgegaan van eenwoordtermen; moderne termextractie had ook meerwoordtermen opgeleverd. Verder is de site ook niet meer bijgewerkt sinds het jaar 2000 en is de handige snelzoekfunctie ondertussen inactief. Ten slotte zou men nog kunnen aanvoeren dat er ondertussen 23 EU-talen zijn, zodat het project ook nog sterk in de breedte kan worden uitgebreid.

In de loop der jaren werd wel vaker het plan opgevat om de bestaande collectie bij te stellen. Het onderzoek naar leesbaarheid van bijsluiters, dat onder tussen op diverse fronten is gevoerd en ook aanleiding heeft gegeven tot een gereviseerde *Readability Guideline* van de EU (ENTR/F/2/SF/jr (2009)D/869), was daarbij zeker een stimulans. Alleen is het nu aanzienlijk moeilijker om een Europese (of andere) subsidie te verkrijgen. In 2007 kwam er toch een aanzet: met de hulp van een tijdelijke kracht, gefinancierd door de faculteit, werd een nieuwe lijst met eenduidige Engelse definities opgesteld – een belangrijke voorwaarde voor een revisie van de bestaande collectie. De definities werden – mét bronverwijzing – uit publiek toegankelijke standaardwerken gehaald en geselecteerd op basis van duidelijkheid en eenduidigheid. Een aantal termen werd geschrapt, andere werden toegevoegd. Enkele opvallende fouten werden alvast weggewerkt. Het Nederlandse vertaalbedrijf Medilingua, gespecialiseerd in medische vertaling, werd bereid gevonden om *pro bono* de lijsten van enkele talen te reviseren aan de hand van de nieuwe definities. In ruil zou het bedrijf een grote zichtbaarheid krijgen op de nieuwe site. Hier ontstond echter al snel een probleem: de partner die – eveneens gratis – de nieuwe site zou verzorgen, liet verstek gaan. Er kwam verder uitstel en de bezigheden vielen andermaal stil. Er wordt nu gehoopt dat het werk in een ander kader kan worden voortgezet (zie paragraaf 6).

4 MeSH Termbase Project

Een project van een andere orde, dat lang heeft kunnen overleven zonder noemenswaardige subsidiëring, is het MeSH Termbase Project. Het acroniem MeSH staat voor “Medical Subject Headings”, een zeer uitgebreide systematische trefwoordenlijst van de National Library of Medicine in de Verenigde Staten (<http://www.nlm.nih.gov/mesh/MBrowser.html>). Oorspronkelijk werden de MeSH-trefwoorden al gebruikt voor het indexeren van de jaarlijkse medische bibliografie Index Medicus, maar ondertussen kennen ze een veel ruimere toepassing: de MeSH-codes worden bijvoorbeeld ook geregeld gebruikt om medische bibliotheken te ordenen.

Robert Vander Stichele, bezieler van de andere medische projecten, kwam in 1987 met het voorstel om de MeSH-termenlijst ten minste gedeeltelijk te voorzien van Nederlandse equivalenten. Nederlandstalige onderzoekers uit de medische sector zouden dan ook via Nederlandse trefwoorden Engelstalige publicaties kunnen opzoeken.

Twee medewerkers van de Gentse vertaalopleiding, Paul Robberecht en Joost Buysschaert, voelden zich door het voorstel aangetrokken. Er werd beslo-

ten om de uitgebreide lijst op te delen in onderwerpen voor afstudeerscripties (eerst heetten die licentiaatsverhandeling, nu heten ze masterproef). De terminologie bleek een ideaal oefenterrein voor aankomende vertalers, maar ook voor tolken en studenten in de meertalige communicatie. Door mee te doen aan het project, zouden zij zich niet alleen bekwamen in de praktijk van het terminologisch onderzoek, ze zouden ook leren bij te dragen tot een grote, door vele terminologen samen gemaakte termenbank.

Vanaf 1992 werd voor het project gewerkt met het GenTerm invulfiche, in huis ontwikkeld door neerlandicus Paul Callewaert. Het fiche respecteert in grote mate de principes van de zogenaamde Vienna School (die voortbouwt op de ideeën van de grondlegger van de terminologieleer, de Oostenrijker Eugen Wüster) en van ISO 1087. Het terminologiewerk is conceptgestuurd, m.a.w. termen die verwijzen naar eenzelfde concept, worden binnen één fiche behandeld. Toch heeft GenTerm de nodige aandacht voor synoniemen (die in de Wüsteriaanse traditie ongewenst waren) en voor het expliciet aangeven van de subtiele nuances tussen synoniemen en equivalenten. Voor medische terminologie is dit erg belangrijk: niet-ingewijden staan vaak versteld over het grote aantal termen dat bestaat voor eenzelfde medisch concept²: ziektenamen bijvoorbeeld hebben vaak een beschrijvende naam naast een eponiem, of een Latijnse naam naast een echt Engelse, respectievelijk Nederlandse benaming; bovendien zijn er zeker in het Nederlands veel medische concepten die naast een technische naam ook nog een lekenterm hebben.

Het GenTerm-fiche is bijzonder gedetailleerd en dwingt de student dan ook om aandacht te hebben voor zowel betekenisnuances als spellingvarianten, grammaticale kenmerken, uitspraak, collocaties, elkaar tegensprekende definities en veel meer. Een ingevuld fiche kan worden bekeken op <http://cvt.hogent.be/mesh.htm>. Het fiche is zo gemaakt dat het kan worden omgezet naar Multiterm en van daaruit ook naar andere formaten, zoals TBX. De volledige termenbank kan enkel door personeel en studenten worden geraadpleegd, maar een (heel) beperkte smaakmaker (beperkt zowel in het aantal concepten als in het aantal gegevens per concept) is voor een breder publiek raadpleegbaar op <http://gtt.hogent.be/mesh>.

Scriptiestudenten verwachten geen subsidie. Het MeSH Termbase Project heeft hierdoor kunnen genieten van een bijzonder lange adem en heeft ondertussen al meer dan 150 scripties opgeleverd. Het scriptieonderwerp wordt nog elk academiejaar aangeboden, want de MeSH-thesaurus is erg omvangrijk: de versie 2013 omvat 26.853 'descriptor' (voorkeurtermen) en bijna 200.000 'entry terms' (te vergelijken met synoniemen). Dezelfde descriptor komt wel op meer dan één plaats in de thesaurus voor en het aantal unieke descriptor'en ligt dus een stuk lager. Eind 2010 waren 4.819 van de

descriptoren aan bod gekomen in het MeSH Termbase Project, goed voor 6.198 Engelse termen, 4.827 Nederlandse equivalenten en 1004 Latijnse equivalenten. Minder voorkomende synoniemen die in een veld "ExtraSyn" worden opgenomen, zijn hier niet meegerekend.

Eén bedrijf heeft een betalende licentie op het materiaal genomen, een tweede toont belangstelling. Beide bedrijven willen de termenbank gebruiken binnen hun systemen van *federated search*.

Naar aanleiding van een opmerking van een medisch vertaalster dat de MeSH-terminologie weliswaar medisch is, maar dat er toch heel wat farmaceutische termen in ontbreken die voor medische vertalers belangrijk zijn, is ook een aanvullend project 'Farma' opgestart. Dit deelproject gebruikt hetzelfde stramien als het MeSH Termbase Project, maar gaat uit van bijkomende termen, geselecteerd uit referentiewerken over farmacie en uit documenten die worden gebruikt voor registratie van geneesmiddelen.

Het MeSH Termbase Project heeft ook aanleiding gegeven tot een aantal onderzoekspublicaties en een nog lopend doctoraatsproject (Buysschaert & Robberecht 2001, Buysschaert 2006a, Buysschaert 2006b, Vanopstal et al. 2011, Vanopstal et al. 2012).

5 ABOP en OptiPIL

Voor descriptief onderzoek – zoals de aanmaak van termenbanken – is het (ten onrechte) heel moeilijk geworden om subsidies te verkrijgen. Voor de ontwikkeling van softwarematige toepassingen echter kan er met enig geluk wel nog financiële steun worden verkregen. Binnen de faculteit is sinds een aantal jaren een aparte Vakgroep Taaltechnologie actief die zich naar buiten profileert als 'LT3' (Language and Translation Technology Team) en die op dit terrein werkzaam is. Ook het LT3-team heeft zich in zijn projecten al een aantal keren toegelegd op medisch taalgebruik en medische terminologie, al gaat het in hun geval niet om terminografisch werk maar om ontwikkeling van software.

Met hun project Automatic Leaflet Optimizer (ABOP), gesubsidieerd door het IWT-TETRA-Fonds, deden LT3-medewerkers van 2007 tot 2010 onderzoek naar de haalbaarheid van een softwarematige redactieomgeving waarmee de leesbaarheid, begrijpelijkheid en gebruiksvriendelijkheid van patiëntenbijsluiters kan worden gecontroleerd en waarbij het systeem correcties kan suggereren. Naast het suggereren van alternatieven voor moeilijke medische termen (helemaal volgens het stramien van de *Multilingual Glossary*, vgl. punt 3) moest het systeem ook kunnen wijzen op onnodige herhaling of

andere redundantie, en op onduidelijkheden over risico's bij het gebruik van geneesmiddelen. Waar ABOP de grondslagen legde, moest het vervolgproject OptiPIL (Optimizer for Patient Information Leaflets) in 2010 de concrete redigeeromgeving uitwerken. Meer details over beide projecten vallen na te lezen op <http://lt3.hogent.be>.

6 Verdere plannen

Zoals eerder al is aangeklaagd, is het momenteel moeilijk om voor descriptief onderzoek fondsen te werven. Toch zijn er bij het CvT en LT3 plannen om in samenwerking met externe partners nieuwe gesubsidieerde initiatieven op te starten waarin medische terminologie een cruciale rol zal spelen. Het zal hierbij de bedoeling zijn om gegevens uit verschillende projecten te bundelen – onder meer gegevens uit het MeSH Termbase Project en uit de verbeterde versie van het Multilingual Glossary – met TMF (Terminology Markup Framework) als gemeenschappelijke onderliggende structuur. In dit synthese-project zou er ook telkens een verband kunnen worden gelegd met andere gezaghebbende medische classificatiesystemen dan MeSH (zoals Snomed-CT, ICD en ICPC). Het project is extern opgestart onder de naam MERITERM en er zijn ook plannen voor een afgeleide, dat voor de Belgische markt zou zijn bestemd. Dit “Belgisch terminologisch referentiesysteem” zou, net als het historische *Drietalig Vocabularium*, de talen Frans, Nederlands en Duits omvatten. In hetzelfde kader komt er hopelijk ook ruimte om eindelijk een nieuwe versie van de *Multilingual Glossary* online te zetten.

De verdere evolutie kan worden gevolgd op de sites <http://cvt.hogent.be> en <http://lt3.hogent.be>. Na de integratie in de Faculteit Letteren en Wijsbegeerte van de Universiteit Gent (najaar 2013) zullen deze sites migreren naar een ander adres. Vermoedelijk worden de URLs dan www.cvt.ugent.be respectievelijk www.lt3.ugent.be.

Literatuur

- Bowker, Lynne & Shane Hawkins (2006). “Variation in the organization of medical terms: exploring some motivations for term choice”. In: *Terminology* 12: 79-110.
- Buysschaert, Joost & Paul Robberecht (2001). “Enkele informatiseringsaspecten van het MeSH-Vertaalproject”. In: Willy Vandeweghe et al. (red.), *Polyfonie. Opstellen voor Paul van Hauwermeiren*, 55-64. Gent: Mercator Hogeschool.

- Buysschaert, Joost (2006a). "The development of a MeSH-based biomedical termbase at Hogeschool Gent". In: P. Zweigenbaum et al., eds. *LREC 2006 Satellite Workshop W08. Acquiring and representing multilingual, specialized lexicons: the case of biomedicine*. 39-43. Genoa.
- Buysschaert, Joost (2006b). "Exploiting an English-and-Dutch biomedical termbase: the search for an ideal format". In: *Equivalences* 33: 33-42.
- Guideline on the readability of the labelling and package leaflet of medicinal products for human use. Revision 1, 12 January 2009. ENTR/F/2/SF/jr (2009)D/869. http://ec.europa.eu/health/files/eudralex/vol-2/c/2009_01_12_readability_guideline_final_en.pdf
- Hauwermeiren, P. van (red.) (1986). *Drietalig vocabularium van wetenschappelijke en populaire medische termen Nederlands/Frans/Duits*. Brussel: Ministerie van Volksgezondheid en het gezin.
- ISO 1087 (1990). *Terminology – Vocabulary = Terminologie – Vocabulaire*. Genève: Organisation internationale de normalisation (ISO/TC 37).
- Richtlijn 92/27/EEG van de Raad van 31 maart 1992 betreffende de etikettering en de bijsluiter van geneesmiddelen voor menselijk gebruik. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31992L0027:nl:NOT>
- Vanopstal, Klaar, Robert Vander Stichele, Godelieve Laureys & Joost Buysschaert (2011). "Vocabularies and Retrieval Tools in Biomedicine: Disentangling the Terminological Knot". In: *Journal of Medical Systems* Volume 35, Number 4 (2011), 527-543. (<http://www.springerlink.com/openurl.asp?genre=article&id=doi:10.1007/s10916-009-9389-z>)
- Vanopstal, Klaar, Robert Vander Stichele, Godelieve Laureys & Joost Buysschaert (2012). "PubMed searches by Dutch-speaking nursing students: the impact of language and system experience". In: *Journal of the American Society for Information Science and Technology* 63 (2012): 1538-1552. (<http://onlinelibrary.wiley.com/doi/10.1002/asi.22694/pd>)

Websites

<http://cvt.hogent.be/mesh.htm> *

<http://gtt.hogent.be/mesh> *

<http://lt3.hogent.be> *

<http://users.ugent.be/~rvdstich/eugloss/welcome.html>

<http://www.nlm.nih.gov/mesh/MBrowser.html>

* De hogent-sites zullen in het najaar van 2013 migreren naar equivalenten op de servers van de Universiteit Gent.

Noten

- ¹ Er bestond ook een Franstalige tegenhanger van het werk onder de titel *Vocabulaire trilingue de termes médicaux scientifiques et populaires. Français-Néerlandais-Allemand*.
- ² Over variatie binnen medische terminologie: zie ook Bowker & Hawkins (2006).