

Terminologie op het web: ‘Google unearth(ed)’

Joost Buyschaert, Bart Defrancq¹

1. De terminoloog en het web

Voor de vertaler of tekstschrijver die met vaktaal werkt is het klassieke woordenboek vaak ontoereikend. Zelfs specialistische woordenboeken, als zij al bestaan voor het betrokken vakgebied, geven niet altijd betrouwbare of actuele informatie.

Het internet biedt een schat aan primair tekstmateriaal waarin vaktermen worden gebruikt, uitgelegd en (rechtstreeks of onrechtstreeks) vertaald. Om de juiste informatie binnen een redelijk tijdsbestek op het spoor te komen moet de terminoloog echter oordeelkundig tewerk gaan. Deze bijdrage geeft een reeks principes en tips voor een terminologische zoektocht op het web.

Eerst verduidelijken we even wat is bedoeld met “terminologie”; en we zetten op een rijtje naar welke gegevens de terminoloog typisch op zoek gaat.

2. Terminologie

De terminologie is de studie van vaktaal en meer bepaald van de benamingen die binnen een vakgebied worden gegeven aan de voor het vakgebied relevante concepten. Die benamingen worden *termen* genoemd. Maar ook het geheel van termen binnen een vakgebied wordt *de terminologie* van het vakgebied genoemd. *Terminologie* is dus op zichzelf al een dubbelzinnige term. (Vgl. ook Delisle et al. 1999 : 186.)

Termen kunnen naar de vorm samenvallen met woorden of uitdrukkingen in de gewone omgangstaal maar binnen het vakgebied hebben zij een specifieke betekenis die niet altijd overeenkomt met de betekenis in de omgangstaal. Zo schrijft de krant algauw dat een gewonde “in shock(toestand)” werd afgevoerd; daarbij denkt de gemiddelde taalgebruiker aan een toestand van emotionele verwarring. Voor de medicus is *shock* echter een welomschreven syndroom met een aantal ondersorten (zie van Everdingen et al. 2006 : 1035.)

Termen kunnen bestaan uit één of meer woorden. Voor de vertaler van de Europese Unie die *Commission for Sustainable Development* moet vertalen is deze reeks van vier woorden één term. In een aantal gevallen is de grens tussen een collocatie (een veel

¹ Deze bijdrage is gebaseerd op een onderdeel van een lessenreeks ‘Principles of Terminology and Terminography’ die beide auteurs aan vertalers en terminologen van de Raad van de EU en van het Europees Parlement hebben gegeven in de periode 2004-2008. De allereerste versie van de presentatie die aan de basis van deze bijdrage ligt, is van de hand van Olaf Du Pont. Een woord van dank aan Olaf is hier dus op zijn plaats.

voorkomende woordcombinatie) en een term moeilijk te trekken. Het belangrijkste criterium is dat aan een term een specifiek, definieerbaar concept vast hangt dat binnen het vakgebied operationeel is.

Terminologen kunnen aan systematische terminologie doen – dan proberen ze de volledige woordenschat van het vakgebied in kaart te brengen en de relaties tussen de concepten duidelijk te maken. Daarnaast kunnen ze ook aan occasionele terminologie doen; dan zoeken ze naar de meest geschikte term om in een concrete tekst (bvb. een handleiding, een gebruiksaanwijzing, een webpagina) te verwerken; of naar het meest geschikte equivalent in een andere taal. Zowel de tekstschrijver als de vertaler kunnen dus aan terminologie doen.

Wie het resultaat van zijn terminologische zoektocht registreert, doet aan *terminografie* (een iets minder gebruikelijke term bedoeld als tegenhanger voor *lexicografie*.) Terminografen verzamelen hun gegevens in begrippenbomen, thesauri, glossaria, gespecialiseerde woordenboeken en digitale termenbanken. Van deze laatste maakt de tekstschrijver gebruik om de systematiek in zijn zakelijke tekst te verhogen (zodat er niet de ene keer *monitor* staat en de andere keer *beeldscherm*, wat de lezer kan verwarren). Ook de vertaler schakelt zijn termenbank aan bij het vertalen; vaak identificeert zijn software automatisch termen in de brontekst die ook in zijn termenbestand voorkomen. Termenbanken spelen eveneens een belangrijke rol bij het verbeteren van de kwaliteit van automatische vertaling.

3. Gegevens die de terminoloog wil kennen

Wie een terminologische zoektocht onderneemt, zal meestal belangstelling hebben voor een of meer van de volgende gegevens:

- het statuut van de lexemen waarmee hij werkt
- een degelijke definitie van een term
- ruimere informatie over het concept dat aan de term beantwoordt
- voorbeeldcontexten waarin de term aanwezig is
- vormvarianten, synoniemen en verwante termen
- bestaande glossaria (één of meertalig) over het vakgebied
- equivalent(en) in een andere taal (met opnieuw: vormvarianten, synoniemen, verwante termen.)

Het internet is een ongelooflijk rijke bron aan informatie die bij deze zoektocht vaak uitstekend kan helpen. Toch kampt de terminoloog vaak met een aantal problemen:

- lang niet alle webpagina's zijn betrouwbaar als bron; er moet dus worden gefilterd

- sommige thema's geven *information overload*, voor sommige andere weinig besproken terreinen en minder courante talen kan de informatie toch nog schaars zijn
- een aantal hoogwaardige bronnen is niet direct toegankelijk omdat er bijvoorbeeld een wachtwoord nodig is om toegang te krijgen tot de site. Dit soort pagina's noemt men wel eens *the invisible web*. (Cf. Sherman and Price 2001.)

In deze bijdrage proberen we alle genoemde aspecten even de revue te laten passeren.

4. Termextractie

Het identificeren van kandidaat-termen in een tekst en het beoordelen ervan is een tijdrovende bezigheid. Eerst moeten de lexemen die in aanmerking komen voor de termenbank in de tekst opgespoord worden. Vervolgens wordt in het beoordelingsproces voor elk lexeem nagegaan worden of:

- het overeenkomt met één begrip (dit is vooral belangrijk wanneer lexemen uit meer dan één woord bestaan);
- het specifiek is voor het vakgebied;
- het relevant is voor de gebruikers van de databank.

Er is software op de markt die de eerste fase van de termselectie automatisch uitvoert en terminologen flink wat tijd kan doen besparen. De verzamelterm voor dergelijke software is 'automatische termextractie'. Sommige van die systemen beperken zich niet tot het selecteren van termen in teksten uit één taal, maar zijn in staat om in parallelle teksten (teksten en hun vertaling in een andere taal) de termen op te sporen die equivalent zijn in verschillende talen. De bekendste software is ongetwijfeld Multiterm Extract van SDL, dat zowel eentalig als tweetalig werkt, maar die is nogal prijzig. Daarom kan het voor de minder bemiddelde terminoloog interessant zijn om uit te kijken naar online term extractors die gratis werk leveren. Er zijn voorlopig nog geen tweetalige extractors die tekstbestanden kunnen verwerken. Wie een vertaalgeheugen gebruikt en in de configuratie Engels-Zweeds-Fins werkt, kan dat wel tweetalig laten analyseren door TermExtractor.com (<http://www.termextractor.com/>). Goed om weten is dat het bedrijf achter TermExtractor.com het opgeloadede vertaalgeheugen bijhoudt en kan gebruiken. Eentalige software die goed werk levert, is TermoStat Web (http://olst.ling.umontreal.ca/~drouinp/termostat_web/). De webapplicatie aanvaardt documenten in vier talen (Engels, Frans, Spaans en Italiaans), maar die moeten wel eerst naar een txt-formaat geconverteerd worden. Conversie is niet nodig bij

TermExtractor² (<http://lcl2.di.uniroma1.it/termextractor>), maar deze applicatie aanvaardt dan weer alleen Engelse teksten. De resultaten zijn ook minder overtuigend dan die van TermoStat Web.

5. Een definitie zoeken

Wie een vakterm oordeelkundig wil hanteren, moet eerst zeker zijn over zijn betekenis. Over hoe de definitie er moet uitzien, lopen de meningen uiteen. In een traditionele benadering in de geest van Wüster (1968) is een definitie niet alleen een beschrijving van de betekenis van een begrip, maar tegelijk ook een middel om een begrip een unieke plaats te geven binnen een begrippensysteem. Een ideale definitie bepaalt daarom ook de verbanden die een begrip onderhoudt met andere concepten. In een streng wetenschappelijke context is dit nog steeds de gangbare benadering (zie bv. Hebenstreit 2007).

De technische definitie van de insider, hoe correct en volledig ook, is echter soms te gesofistikeerd voor de vertaler of de tekstschrijver (bij deze laatste moet men denken aan de auteur van bvb. publicitaire teksten of een licht-vulgariserende website.) Temmerman (2000) breekt daarom een lans voor de ‘understanding approach’: vertalers hebben genoeg aan definities waarmee ze de betekenis van een begrip kunnen “vatten”, aangevuld met encyclopedische informatie (bv. afbeeldingen).

Hoe vindt men definities op het net? Er zijn uiteraard ontelbare websites die online woordenboeken en glossaria aanbieden. We sommen er enkele op zonder volledigheid te willen claimen (en met de waarschuwing dat ze vaak alleen voor het Engels resultaten geven): www.onelook.com, www.websters-online-dictionary.org, www.answers.com, <http://www.babylon.com/define/>, <http://en.wikipedia.org>, www.wordreference.com, www.eurotermbank.com. Relatief nadeel van deze bronnen is dat ze definities geven vanuit online referentiewerken; die zijn geregeld net iets minder betrouwbaar dan de definities uit primaire teksten. DefinitionSearch (<http://www.definitionsearch.com/>) probeert daar iets aan te doen door gebruikers met vakkennis uit te nodigen om betrouwbare definities over wetenschappelijke termen aan te bieden. Voordeel zou moeten zijn dat de definities hier vooraf worden gescreend; nadeel is dat de zoektocht vaak zonder resultaat blijft. Ook bij Wiktionary (<http://nl.wiktionary.org/>) wordt de online gebruiker uitgenodigd om zelf zijn bijdrage te leveren, volgens het ondertussen bekende Wiki-principe. Andere glossaria komen hieronder in punt 8 nog aan de orde.

² Hoewel de naam bijna identiek is aan de eerder vermelde TermExtractor.com, is de organisatie erachter verschillend.

Google zelf biedt ook een aantal zoekfuncties waarmee definities gevonden kunnen worden. Er is vooreerst Google Glossary (<http://www.googleglossary.com>), dat zoekopdrachten beperkt tot online glossaria en daaruit relevante definities licht. Dit kan eigenlijk ook rechtstreeks in Google zelf met de *define*-functie. Open Google, typ:

define: kenniseconomie

en er worden drie definities opgegeven met hun bron. Sterker nog, typ:

define: Helicobacter pylori

(geen aanhalingstekens gebruiken) en er worden definities gevonden in acht talen, wat voor de aanmaak van een meertalige termenbank uiteraard interessant is. Google beweert zelf (<http://www.googleguide.com>) dat de *define*-functie ook kan worden opgeroepen door *what is* of *what are* (altijd in het Engels, ook al is het een zoekterm in een andere taal) te typen voor de zoekterm. De ervaring leert echter dat er dan meestal maar één resultaat is, ook voor *Helicobacter pylori*.

Wordt er op die manier geen definitie gevonden of voldoen de definities niet, dan zijn er nog andere middelen. Wie zijn verbeelding wat gebruikt, kan merkwaardig goede resultaten vinden door volgende zoekstrings in een zoekmachine in te voeren (waarbij X de gezochte term is; gebruik dubbele aanhalingstekens):

“definieert X”

“X wordt gedefinieerd”

“definitie van X”

“wat is X”,

uiteraard steeds aan te passen volgens de taal van de term. De combinatie “definieert kenniseconomie” leidt tot een tekst die begint met ‘De Wereldbank definieert kenniseconomie als ...’, gevolgd door alternatieve definities van de OESO, het Ruimtelijk Planbureau en nog enkele instanties. Zelfs

“X is”

levert verbazend vaak goede definities op. Voor Engelse definities blijkt de combinatie

“X i.e.”

overigens één van de beste detectoren.

Het spreekt voor zich dat deze zoekopdrachten niet uniek zijn voor Google. Gelijk welke zoekmachine is hiervoor geschikt.

6. De zoektocht toespitsen: Google en alternatieven

Definities kunnen op zichzelf ontoereikend zijn om het precieze gebruik en de volledige betekenis te achterhalen. De term in een originele context zien, een omstandiger uitleg kunnen nalezen met een grafische illustratie erbij: vaak leidt dit tot een duidelijker beeld over de betekenis en het gebruik van de term. Het is om die reden dat ook klassieke woordenboeken voorbeeldzinnen, citaten en soms beeldmateriaal opnemen en ook termenbanken doen dit geregeld. Bij het opzoeken van dit bijkomende materiaal kan een zoekmachine alweer heel nuttige diensten bewijzen. Toch zal de terminoloog geregeld moeten filteren.

6.1. Dubbelzinnige termen

Zo zijn er vooreerst ook bij de vaktermen homoniemen. ‘Dematerialisering’ is een concept in de ecologische economie maar geldbeleggers kennen de term ook (zij hebben het over *dematerialisering van effecten*). Bij de zoektocht naar geschikte contexten moet dus worden gedesambigueerd. Dit kan door de zoekstring uit te breiden, bijvoorbeeld met *ecologie*, of met *-effecten* (waarbij het minteken ervoor zorgt dat contexten met het woord effecten worden geweerd.)

6.2. Gericht zoeken binnen een klassieke zoekmachine

Behalve het eigenhandig aanpassen van de zoekstring, kan de gebruiker van een zoekmachine best even te rade gaan bij de geavanceerde opties van de klassieke zoekmachines. *Google advanced* biedt bijvoorbeeld een ruim aantal mogelijkheden. Zo kan er gezocht worden naar alleen recente teksten; of alleen naar powerpoints; ook de taalkeuze kan worden ingesteld (al is de werking van die optie niet helemaal waterdicht). Ook heel interessant is de mogelijkheid om alleen te zoeken binnen bepaalde sites of domeinen. Zoeken op sites met ‘.org’ als domeinnaam beperkt de zoektocht tot organisaties, wat ongewenste vindplaatsen kan uitfilteren (maar anderzijds ook andere, wel nuttige sites). Andere voorbeelden, afhankelijk van de doelstelling, zijn: zoeken binnen sites eindigend op ‘.ac.uk’ (dit zijn Britse universitaire websites); of binnen sites met ‘.europa.eu’, wat de zoektocht toespitst op de terminologie die binnen de EU toegankelijk is (in alle EU-talen nog wel.) Ook het doorzoeken van één bepaalde site kan via deze optie gebeuren; dit is handig als de webstek geen eigen zoekmachine heeft, of een minder krachtige.

6.3. Gespecialiseerde zoekmachines en meta-zoekmotoren

Wie bijvoorbeeld over het thema ‘vogelgriep’ moet schrijven of vertalen en zijn termen op Google legt, krijgt een stortvloed aan informatie uit nieuwssites, populariserende sites en blogs. Die leveren niet altijd betrouwbare informatie op. Uitwijken naar een gespecialiseerde zoekmachine is hier de beste oplossing. Om te beginnen is er Google Wetenschap (Scholar Google <http://scholar.google.be>). Deze uitstekende zoekmachine zoekt enkel in online wetenschappelijk teksten. Google Wetenschap was echter niet de eerste op dit terrein. Een eerder initiatief was Scirus, nog steeds een voortreffelijke wetenschappelijk zoekmachine (www.scirus.com); bij een zelfde zoektocht geeft zij weliswaar meestal minder hits dan Google Wetenschap maar er zijn bijkomende filtermogelijkheden. Ten eerste kan er geselecteerd worden op vindplaatsen uit ‘Journal sources’, vindplaatsen uit ‘Preferred web sources’ maar ook ‘Other web sources’. Als de wetenschappelijke contexten niets opleveren, blijft het ruime web dus ook nog altijd bevroegbaar (bij Google Wetenschap moet men in een dergelijk geval overschakelen naar de gewone Google). Ten tweede geeft Scirus stevast een kadertje met ‘Refine your search’. Daarin staan trefwoorden die uit de gevonden teksten zijn gedistilleerd. Door een passend trefwoord aan te klikken, worden alleen de hits opgeroepen die ook dit trefwoord bevatten, waardoor de zoektocht nog nader wordt toegespitst.

Deze verfijningsmethode is ontleend aan weer een andere categorie zoekmachines, namelijk de meta-zoekmotoren. Vivísimo is daarin een voortrekker geweest. Hun online meta-zoekmotor heet nu Clusty (<http://clusty.com/>). Als we in Clusty naar het letterwoord TeDIF zoeken, dan komt er aan de linkerzijde een tabelletje met “clusters”; dat meldt ondermeer dat 18 hits te maken hebben met “Terminology, Format”, en 11 hits met “PAP” – dat laatste blijkt een onderwerp uit de ruimtevaart te zijn. Als het ons alleen om het concept binnen de terminologie te doen is, dan klikken we op ‘Terminology, Format’ en alleen de echt relevante hits komen op het scherm.

Ook de meeste andere meta-zoekmotoren (*turbo10*, *metacrawler*, *dogpile* ...) hanteren een vergelijkbare verfijningsmethode maar zij hebben alle het nadeel dat ze heel breed blijven zoeken: zij leggen de zoekstring op een aantal andere klassieke zoekmachines (om die reden heten ze ook ‘meta’). Voor de terminoloog zijn Google Wetenschap en Scirus daarom meestal nuttiger.

Wie lange tijd rond een bepaald thema werkt, vindt soms ook een specialistische zoekmotor voor zijn vakgebied; SearchMedica, bijvoorbeeld, voor Engelstalige (Britse) medische terminologie (<http://www.searchmedica.co.uk/>). Het onderdeel *Moteurs de recherche spécialisés* van <http://urfist.univ-lyon1.fr/risi/outils.htm> geeft een aantal van die gespecialiseerde zoekmotoren.

Voor het Nederlandse taalgebied valt er vaak minder te rapen dan voor Engels en Frans. Toch zoeken bijvoorbeeld Google Wetenschap en Scirus ook in Nederlandstalige bronnen. Wel kan er in hun geavanceerde zoekopties niet op taal worden ingesteld. Als dit dreigt tot dubbelzinnigheden te leiden, volstaat het echter om aan de zoekstring enkele Nederlandse woorden toe te voegen (*het een met*).

Voor het Nederlands kan het nuttig zijn om een gezaghebbend tijdschrift op het spoor te komen binnen het vakgebied. Het *Nederlands Tijdschrift voor Geneeskunde* (www.ntvg.nl), bijvoorbeeld, is erg nauwgezet in taalzaken; de site heeft een zoekfunctie die op zijn minst de mogelijkheid geeft om na te gaan of een term ooit binnen het tijdschrift is gebruikt en hoe vaak. De volledige tekst is meestal alleen voor abonnees toegankelijk.

Nog een andere bron voor wetenschappelijke contexten, in vele talen, is Google Books (<http://books.google.com/>). Er kan gezocht worden binnen het boek en dat levert vaak relevante voorbeeldcontexten op uit belangrijke (en actuele) primaire werken.

Het trio Google Wetenschap - Scirus - Google Books zal zeker voor de belangrijke talen vaak volstaan en opent een snelle weg naar relevante bronnen voor de terminoloog. Wie meer geduld heeft, kan ook nog de 'subject directories' consulteren. Dat zijn thematisch geordende verzamelingen van betrouwbare websites. Een gezaghebbend voorbeeld hier is de Librarians' Internet Index (<http://lii.org>).

6.4. Het onzichtbare web

Een beperking van de zoekmachines is dat ze zelden doordringen tot het zogenaamde 'invisible web'. Termen zitten bijvoorbeeld geregeld ook verborgen in achterliggende databases bij websites. De webpagina's zelf worden door de zoekmachines geïndexeerd, de achterliggende databank niet. De inhoud komt bijvoorbeeld pas naar boven als de gebruiker een zoekopdracht invoert op de webpagina.

Er zijn sites die beweren dat zij toch tot het onzichtbare web doordringen (zoals <http://www.freepint.com/gary/direct.htm>) maar hun claim lijkt bij een concrete zoektocht vaak wat twijfelachtig. De site <http://www.inciwincy.com> is wellicht nog de meest succesrijke. Wij zochten op deze site naar TeDIF en vonden een aantal referenties (met TeDIF in de ruimtevaartbetekenis) waar de zoekterm in een achterliggende database voorkwam; via Google bleken deze referenties onvindbaar.

7. Synoniemen en verwante termen

Eugen Wüster, grondlegger van de wetenschappelijke terminologie (Wüster 1968), ging ervan uit dat vaktalen geen synoniemen dulden. Aan één concept hoort één technische term te beantwoorden. In de vakgebieden die worden genormaliseerd is dit inderdaad de bedoeling maar in de praktijk bestaan er binnen de vaktalen wel degelijk synoniemen. Een voorbeeld is alweer de geneeskunde, die ondanks zijn imago van 'exacte' wetenschap heel vaak tal van synoniemen hanteert voor een zelfde begrip.

Google beweert synoniemen te zoeken van een woord als het woord als zoekterm wordt ingevoerd met een tilde ervoor (~). Een test met '~flu' spoort inderdaad ook hits van 'influenza' op (bovenop de hits van 'flu'), maar ook een aantal hits van 'stomach'. Helemaal accuraat is de zoekfunctie dus niet. Bovendien blijkt ze alleen te werken voor het Engels: het Nederlandse '~griep' leidt niet naar hits van 'influenza', die er nochtans zijn. Voor het Nederlands kan de combinatie van de zoekterm en 'of' uitkomst brengen. Voor de volgende combinatie (met aanhalingstekens):

“griep of”

levert Google bij de tweede hit al het synoniem 'influenza'.

Van vermoedelijke synoniemen moet de terminoloog nagaan of ze wel volwaardig synoniem zijn en een zelfde gebruikswaarde hebben. Dit kan door definities en contexten met elkaar te vergelijken. Soms vermoedt de terminoloog dat er toch nog een klein betekenisverschil is tussen ogenschijnlijke synoniemen. Er bestaat een kans dat dit ook een ander is opgevallen maar ook dat valt soms via het web te traceren. Wie op Google Wetenschap invoert “difference between knowledge economy and knowledge society”, krijgt twee teksten waarin op het onderscheid wordt ingegaan. Op Google zelf zijn er ook twee hits voor het omgekeerde (“difference between knowledge society and knowledge economy”).

Blijken de synoniemen 'echt' te zijn, dan blijft nog de vraag aan welk ervan de voorkeur moet worden gegeven. Eén argument hierbij is de relatieve frequentie van de concurrerende termen en ook hier bieden de zoekmachines weer interessante mogelijkheden.

Zo is er Google Fight (<http://www.googlefight.com/>), waarop de frequentie van twee zoekstrings tegenover elkaar kan worden afgewogen. Maar Google Fight gebruikt de ruimere Google, niet Google Wetenschap en heeft geen geavanceerde opties. Voor de terminoloog is het daarom veel nuttiger om rechtstreeks op Google Wetenschap of Scirus de frequentie te vergelijken en daarbij een oordeelkundig gebruik te maken van de geavanceerde zoekopties en andere middelen om de zoektocht voldoende toe te spitsen. De ruwe cijfers die de zoekmachines geven, zijn niet in alle opzichten betrouwbaar. Er gebeuren bijvoorbeeld vertekeningen als een zelfde bron meervoudige hits geeft. Soms ook levert de zoektocht een

hele reeks hits op waarin de zoekterm alleen als trefwoord voor indexering voorkomt maar waarin hij niet echt in de tekst wordt gebruikt en dat kan het beeld aardig vertekenen. Anderzijds is het feit dat de term als trefwoord werd gehanteerd ook een indicatie van zijn status. In elk geval kunnen er pas duidelijke conclusies worden getrokken als het frequentieverschil vrij aanzienlijk is. Als blijkt dat Google Wetenschap ongeveer 2000 hits geeft voor ‘pulpectomy’ en slechts ongeveer 500 voor concurrent ‘pulp extirpation’, bijvoorbeeld, kan er gesproken worden van een significant verschil.

Bij het maken van vergelijkingen moeten echter alle waarschuwingen in acht worden genomen waarop eerder al is gealludeerd, zoals mogelijke homonymie binnen de taal of over de talen heen. Voor de Nederlandse equivalenten van de zopas genoemde termen kwam een studente tot de bevinding dat Google Wetenschap 121 hits gaf voor ‘pulpectomie’ en slechts één voor ‘pulpa-extirpatie’. Zij verloor uit het oog dat de ‘pulpectomie’ ook het Franse equivalent is. Wie ‘pulpectomie het een van met’ op Google Wetenschap invoert, krijgt ook hier slechts één hit.

8. Bestaande glossaria opzoeken

Primaire teksten – geschreven door iemand uit het vakgebied en bedoeld voor vakgenoten (of aspirant-vakgenoten, denk aan leerboeken) – zijn de meest betrouwbare bronnen voor vaktermen, met dit verschil dat de vakliteratuur soms een loopje neemt met de taalregels en met name met de spellingsregels. Voor het al of niet aaneenschrijven van samenstellingen heeft onze Nederlandse taal bijvoorbeeld ingewikkelde regels die niet altijd nauwgezet worden opgevolgd (‘account executive’ komt massaal vaker voor in originele bedrijfsteksten dan ‘accountexecutive’ maar alleen de tweede optie volgt de huidige spellingregels.) In een dergelijk geval moet de terminoloog soms afwijken van wat hij of zij in de primaire bronnen vindt.

Goed bijgehouden secundaire bronnen – naslagwerken met name – kunnen wat dat betreft een streepje voor hebben op de primaire bronnen, omdat zij vaak door een taalkundige zijn samengesteld en die hoort meer oog te hebben voor de spelregels van de taal. Het nadeel van de referentiewerken is dat hun samenstellers niet altijd specialisten uit het betrokken vakgebied zijn, wat wel eens leidt tot inhoudelijke fouten.

Bij naslagwerken denken we niet langer uitsluitend aan de papieren woordenlijsten laat staan steekkaartencollecties maar ook aan online bronnen, al of niet gratis te raadplegen. We noemen er hieronder enkele op.

8.1. Bronnen met enig gezag

Grote instellingen zoals de Europese Unie, de Verenigde Naties, de Werelgezondheidsorganisatie, de Canadese overheid ... zijn erom bekend dat zij teksten in verschillende talen moeten aanleveren en zij hebben dan ook al diverse terminografische initiatieven genomen. Sommige van die bronnen vallen vlot online te consulteren.

IATE (<http://iate.europa.eu/>) is de waardige opvolger van Eurodicautom en bevat naast de termenbank van de Europese Commissie (waartoe Eurodicautom was beperkt) ook die van de Raad van de EU en van het Europese Parlement, naast termenlijsten van nog een reeks andere instanties, alle ondergebracht binnen één systeem. Aangezien de Europese instellingen over een breed aantal onderwerpen teksten en vertalingen aanmaken, levert een zoektocht op IATE vaak nuttige informatie op. Toch moet de gebruiker kritisch blijven. Voor een EU-vertaler is een tekst over een relatief nieuw onderwerp als de klimaatsverandering ook een harde dobber. Ook hij/zij zal als niet-specialist een zoektocht hebben moeten ondernemen naar correcte equivalenten en definities. De gevonden suggestie nog eens extra nagaan in de primaire bronnen is vaak geen overbodige luxe.

IATE biedt 24 talen maar uiteraard zijn ze nog niet alle even sterk vertegenwoordigd. Met name voor de talen van de landen die maar recent zijn toegetreden, is de informatie in verhouding nog geringer. Een initiatief dat in de aanloop tot de toetreding is genomen door een consortium dat die lacune op haar eigen wijze wilde opvullen, heeft geresulteerd in de Eurotermbank (www.eurotermbank.com). Dit is, voor alle duidelijkheid, niet een initiatief van de EU maar het is wel een rijke bron van informatie die 27 talen bestrijkt en voor een aantal Oost-Europese en Baltische talen voorlopig nog meer biedt dan IATE zelf.

Wel weer van de Europese Unie zelf is Eurovoc (<http://europa.eu/eurovoc/>). Wie op zoek gaat naar een beknopte omschrijving en een vertaling van (vooral) rechtskundige termen binnen de EU-sfeer (zoals 'proportionality'), zal in deze thesaurus zijn gading vinden. De term moet worden opgezocht in de brontaal, waarna men moet overschakelen naar de corresponderende pagina in de doeltaal.

Een ander belangrijk instrument maar dan voor wie alleen met Engels en Frans werkt, is de Canadese *Grand Dictionnaire Terminologique* (<http://www.granddictionnaire.com/>).

8.2. Verzamelbronnen

Behalve de gezaghebbende bronnen vermeld in 8.1. zijn er ook tal van andere glossaria, woordenlijsten, lexica en gespecialiseerde woordenboeken online te vinden, aangemaakt door de meest uiteenlopende instanties of individuen (zie ook 8.3.). Ze allemaal

apart op het spoor komen en dezelfde terminologische vraag voorleggen, zou erg tijdrovend worden. Gelukkig kan men via een portaal als OneLook (www.onelook.com) in één zoekbeweging een hele reeks bronnen, gezaghebbende en minder gezaghebbende, tegelijk bevragen. Zowel definities als vertalingen worden op die manier toegeleverd. Voor de vertalingen van meer technische termen maakt OneLook geregeld gebruik van de lijsten van Babylon (<http://dictionary.babylon.com/>); ook wordt de meertaligheid van de Wikipedia uitgebuit: wie een term zoekt in deze (soms omstreden) encyclopedie, ziet meestal in de linkerkolom de mogelijkheid om naar een andere taal over te schakelen; op die manier wordt de Wikipedia ook een vertaalbron en daar maakt OneLook gebruik van.

OneLook heeft de loffelijke gewoonte zijn bronnen te citeren. Dat is niet het geval voor concurrent Webster's Online Dictionary (<http://www.websters-online-dictionary.com/>). Ook dit portaal doorzoekt tal van online bronnen tegelijk maar zegt slechts sporadisch waar het de mosterd haalt. Vaak wordt 'various references' opgegeven als bron; wie die melding aanklikt, wordt doorverbonden met de site van Amazon. Bovendien creëert de benaming Websters verwarring met het bekende Amerikaanse woordenboek Merriam-Webster, waarmee de webstek geen uitstaans heeft. Toch moet men het Webster's Online nageven dat zij vaak een zeer informatieve insteek geven aan een terminologische zoektocht. Naast definities en vertalingen bezorgen zij de gebruiker ook etymologieën, afbeeldingen, synoniemen enzovoort.

Ook bij OneLook en Webster's Online (en vele andere gelijkaardige portaal sites) geldt dat het veilig is om de gevonden oplossingen nog ter bevestiging na te trekken in primaire bronnen.

8.3. Zelf een glossarium opzoeken

Wie veel terminologie uit een specifiek vakgebied nodig heeft, kan terecht bij de duizenden glossaria die organisaties, bedrijven en individuen allerhande online aanbieden. De kwaliteit ervan is, hoeft het gezegd, niet altijd gegarandeerd. Er zijn websites die een ruwe selectie maken met het betere materiaal en er een thematische indeling in aanbrengen. Glossarist (<http://www.glossarist.com>) en Lexicool (<http://www.lexicool.com>) zijn er twee van. Glossarist werkt met een standaardrubricering van sites die kan aangeklikt worden en met een invulvak. Het invulvak is bedoeld om een trefwoord op te geven waarmee een glossarium gevonden kan worden, maar blijkt vooral handig om de zoekopdracht te beperken tot bepaalde talencombinaties. Glossarist biedt namelijk geen andere mogelijkheid om de

taalkeuze te beperken. Lexicool biedt die wel en levert zelfs voor exotische combinaties resultaten (al gaat het niet altijd om glossaria).

Als deze sites geen uitkomst brengen, kan men nog altijd zijn geluk beproeven met een combinatie van zoektermen en ‘glossary’, ‘dictionary’ of ‘lexicon’ in Google. De eerste hits voor ‘welding glossary’, bijvoorbeeld, leiden naar verscheidene Engelse websites met glossaria over lastechniek. Ook hier is het Engels uiteraard bevoordeeld. Zoeken naar ‘lastechniek glossarium’ of ‘lastechniek woordenboek’ levert niets op.

8.4. Gebruik maken van een vertaalprogramma

Goede meertalige glossaria zijn vrij zeldzaam, ook op het web, in het bijzonder wanneer een van de talen een zogenaamd ‘kleine’ taal is of wanneer termen nog vrij recent zijn. Zo is er bijvoorbeeld geen Nederlandse vertaling in IATE voor ‘ocean acidification’, een begrip uit het domein van de klimaatverandering. Google translate (<http://translate.google.com>) is in dat geval een alternatief. De vertaalmachine van Google mag dan nog heel wat problemen hebben met het vertalen van teksten en niet altijd beter scoren dan andere online vertaalmachines (zie o.a. Kit & Wong 2008), met termen zijn de resultaten betrekkelijk goed. In het geval van ‘ocean acidification’ stelt het ‘verzuring van de oceaan’ voor. Helemaal accuraat is dat niet, omdat in het Nederlands vaker het meervoud ‘oceanen’ gebruikt wordt, maar het komt aardig dicht in de buurt. Voor dit soort gevallen biedt de zoekmachine overigens de mogelijkheid om een bijdrage te leveren. Rechts op het scherm met de resultaten staat een link waarmee ontevreden een betere vertaling kunnen voorstellen.

9. Een vertaling opzoeken in primaire bronnen

Een beproefde methode voor het opzoeken van een vertaling is de ‘gismethode’. Van een Engelse medische term als ‘enterocyst’ kan men vermoeden dat zijn Nederlandse equivalent ‘enterocyste’ zal zijn. Bevestiging voor dat vermoeden kan men vervolgens zoeken binnen primaire Nederlandstalige medische teksten, gebruik makend van de zoekmachines vermeld in punt 6 (met name Google Wetenschap, Google Books, Scirus en in dit specifieke geval ook het Nederlands Tijdschrift voor Geneeskunde.) Toch zou het nog altijd kunnen dat enterocyste slechts een minder gebruikelijk synoniem is. Daarom zoekt men de term ook beter nog eens op via een verklarende Nederlandse bron, die mogelijk doorverwijst naar een couranter synoniem.

Interessant zijn ook de twee- of meertalige websites. Door <http://europa.eu/> te openen in de brontaal, via de zoekfunctie de bronterm in te vullen en dan over te schakelen naar de doeltaal, vindt men wel eens wat, al kan men aanvoeren dat het hier ten minste in één van de talen om een vertaalde tekst gaat en niet om een originele tekst. Hetzelfde kan met <http://eur-lex.europa.eu> (vooraf registreren). De Zwitserse site <http://www.multilingual.ch> biedt een mogelijkheid om meertalige sites over een onderwerp op het spoor te komen, al is hun zoekmethode niet helemaal waterdicht. Wie verstand heeft van vertaalgeheugensoftware, kan het uitgebreide vertaalgeheugen van het Directoraat-Generaal Vertaling van de EU gratis downloaden (<http://langtech.jrc.it/DGT-TM.html>); binnen deze meertalige versie van het *Acquis Communautaire* kan men naar hartenlust terminologische equivalenten opsporen.

Maar ook voor wie dit technisch geweld niet aankan, bestaan er nog handige middeltjes. Vertalingen waarvoor de gismethode niet lijkt te werken, kunnen soms met een gewone zoekopdracht in Google achterhaald worden. Met de Google Advanced Search (zie eerder) kan namelijk de taal van de op te zoeken documenten gedefinieerd worden. Auteurs die de realiteit van een taalgebied in de taal van een ander taalgebied moeten beschrijven, zijn soms geneigd om termen uit de brontaal te gebruiken in de doeltaal en die te voorzien van een vertaling (meestal tussen haakjes). Met behulp van de Google Advanced Search kunnen deze vreemde eenden in de bijt gemakkelijk opgespoord worden. Het volstaat het woord waarvan de vertaling gezocht wordt in Google in te voeren en de resultaten via de Advanced Search te beperken tot documenten in de doeltaal. Een zoekopdracht naar het notoir moeilijk te vertalen 'bedrijfsrevisor' beperkt tot pagina's in het Engels levert zo al gauw een aantal waardevolle voorstellen: 'statutory auditor', 'auditor accountant', 'registered auditor'.

10. Besluit.

Een volledig overzicht geven van de methodes die het web biedt om terminologie op te sporen, is ondoenlijk en riskeert meer verwarring dan verheldering te brengen. Het is net zaak om uit de veelheid van mogelijke invalshoeken er enkele uit te kiezen die in vele gevallen relatief snel tot een betrekkelijk betrouwbaar resultaat leiden. Wie rond een specifiek onderwerp werkt, zal meestal na enige tijd de meest relevante sites en methodes voor dit onderwerp weten te vinden. In veel gevallen is intelligent giswerk, te bevestigen via oordeelkundig gebruik van zoekmachines, de meest doeltreffende methode.

Bronnen

- Delisle, Jean, Lee-Jahnke, Hannelore, Cormier, Monique (red.), *Terminologie de la traduction. Translation terminology. Terminología de la traducción. Terminologie der übersetzung*. John Benjamins, Amsterdam/Philadelphia 1999.
- Hebenstreit, Gernot, 'Defining Patterns in Translation Studies: Revisiting two classics of German *Translationswissenschaft*', in: *Target*, 19 (2007), nr. 2, pp. 197-215.
- Kit, Chunyu & Wong, Tak Ming, 'Comparative Evaluation of Online Machine Translation Systems with Legal Texts', in: *Law Library Journal*, 100 (2008), nr. 2, pp. 299-321. [http://www.aallnet.org/products/pub_llj_v100n02/2008-16.pdf]
- Sherman, Chris, Price, Gary, *The invisible web: uncovering information sources search engines can't see*. CyberAge Books 2001.
- Temmerman, Rita, *Towards new ways of terminology description: The sociocognitive approach*. Benjamins, Amsterdam 2000.
- van Everdingen, J.J.E., van den Eerenbeemt, A.M.M., Klazinga, S., Pols, J. (red.) Pinkhof geneeskundig woordenboek, elfde, herziene en uitgebreide druk. Houten, Bohn Stafleu van Loghum 2006.
- Wüster, Eugen, *Dictionnaire multilingue de la machine-outil. Notions fondamentales, définies et illustrées, présentées dans l'ordre systématique et l'ordre alphabétique. Volume de base anglais-français = The Machine Tool. An Interlingual Dictionary of Basic Concepts comprising an Alphabetical Dictionary and a Classified Vocabulary with Definitions and Illustrations. English-French Master Volume*, London, Technical Press 1968.

Selectie van vermelde online bronnen

- Answers.com www.answers.com/
- Babylon <http://dictionary.babylon.com/> , <http://www.babylon.com/define/>
- Clusty <http://clusty.com/>
- DefinitionSearch <http://www.definitionsearch.com/>
- Eurotermbank www.eurotermbank.com/
- Eurovoc <http://europa.eu/eurovoc/>
- Glossarist <http://www.glossarist.com>
- Google Books <http://books.google.com/>
- Google Fight <http://www.googlefight.com/>
- Google translate [http://translate.google.com\)](http://translate.google.com)
- Scholar Wetenschap <http://scholar.google.be>

Grand Dictionnaire Terminologique <http://www.granddictionnaire.com/>
IATE <http://iate.europa.eu/>
Lexicool <http://www.lexicool.com>
Librarians' Internet Index <http://lii.org>
OneLook www.onelook.com
Scirus www.scirus.com
TermExtractor <http://lcl2.di.uniroma1.it/termextractor>
TermExtractor.com <http://www.termextractor.com/>
TermoStat Web http://olst.ling.umontreal.ca/~drouinp/termostat_web/
Webster's Online Dictionary <http://www.websters-online-dictionary.com/>
Wikipedia <http://en.wikipedia.org/>
Wiktionary <http://nl.wiktionary.org/>
WordReference www.wordreference.com/