

Programma « AGORA »

Eindverslag – deel « Synthese van het onderzoek »¹

ONDERZOEKSCONTRACT: AG/LL/169

ACRONIEM VAN HET PROJECT: MULTITA

TITEL: **MULTILINGUAL TERMINOLOGICAL RESEARCH (FRENCH, DUTCH AND ENGLISH) FOR THE DEVELOPMENT AND INTEGRATION OF SEMANTICALLY ENRICHED SCIENTIFIC THESAURI (MULTITA)**

ONDERZOEKSPLOEG:

COÖRDINATOR: **EVA COUDYZER (KMKG)**

PROMOTOR(EN): **ERIK BUELINCKX (KIK)**

ONDERZOEKER(S): **VERONIQUE LHEUREUX, EVA COUDYZER, ERIK BUELINCKX**

DATUM: 30/01/2015

¹ Zie art. 5.5.2 van het basiscontract.

Links naar de websites waar de werken van de onderzoeksploeg in het kader van het project worden vermeld:

- Website KMG: [KMG-MULTITA](#)
- Website KIK: [KIK-MULTITA](#)
- Linked Heritage, Seminar Multilingual Terminologies, April 13 2013 (Paris, France), presentation " Recommendations and Guidelines for Terminologies": [MULTITA-Linked Heritage](#)
- AthenaPlus, Access to Cultural Heritage Networks for Europeana (Rome, Italy), Signed Memorandum of Understanding between MULTITA and AthenaPlus: [MULTITA-AthenaPlus](#)
- FRANTIQ, Fédération et Ressources sur l'Antiquité (Nanterre, France), PACTOLS-thesaurus:
 - [FRANTIQ_01](#)
 - [FRANTIQ_02](#)
 - [FRANTIQ_03](#)
 - [FRANTIQ_04](#)
- Mémoires des archéologues et des sites archéologiques (MASA-consortium), Réunion de travail « Ontologies et redocumentarisation » , October 13 2014 (Nanterre, France) : [MULTITA-MASA](#)
- RKD, Rijksbureau voor Kunsthistorische Documentatie (Den Haag, Nederland), Art and Architecture Thesaurus Nederlandstalig: [MULTITA-AAT](#)
- Getty Institute,(Los Angeles, USA), Art and Architecture Thesaurus. Mention of MULTITA-project in presentation RICH: [MULTITA-Getty Institute](#)
- l'UMR8546 CNRS/ENS Archéologie et Philologie d'Orient et d'Occident – AOROC, Presentation project MULTITA (meeting November 13 2013): [MULTITA-AOROC](#)
- Fédération Wallonie-Bruxelles, AICIM-thesaurus: [MULTITA-AICIM](#)
- University of Savoie, TOTH-workshop, December 5 2014 (Brussels, Belgium), presentation and paper project: [MULTITA-TOTH](#)
- VIAA, Vlaams Instituut voor Archivering, Unified Thesaurus Haalbaarheidsstudie uitgevoerd door iMinds (Media Innovation Center), Eindrapport, augustus 2013-januari 2014: [MULTITA-VIAA](#)

MULTITA

Multilingual Terminological Research for the Development of Semantically Enriched Thesauri
AG/LL/169

Véronique l'Heureux, Eva Coudyzer

Koninklijke Musea voor Kunst en Geschiedenis
Brussel, januari 2015

Samenvatting

MULTITA is een project over thesauri en meertaligheid dat ontstond vanuit de noodzaak om gestandaardiseerde, verrijkte en meertalige thesauri te gebruiken binnen de federale wetenschappelijke instellingen. Veel culturele instellingen in België gebruiken een thesaurus of meerdere thesauri die intern worden ontwikkeld. Het kan gaan om eenvoudige, alfabetisch geordende woordenlijsten of volwaardige, semantische thesauri. De thesauri zijn vaak specifiek ontwikkeld voor de collecties van de instelling en verschillen onderling naargelang de instelling en collectie – wat betekent dat de collecties (online) moeilijk met elkaar in verband kunnen worden gebracht.

De belangrijkste activiteiten van het MULTITA project zijn de afstemming en verrijking van lokale thesauri en het ontwikkelen van een aantal wetenschappelijke, meertalige micro-thesauri over een bepaald domein, ontwikkeld volgens de internationaal heersende thesaurusstandaarden. In het project konden ook een aantal moeilijkheden en uitdagingen bestudeerd worden die ontstaan als men een thesaurus ontwikkelt of vertaalt.

Trefwoorden: thesauri, meertaligheid, thesaurus management tool, SKOS

1. Inleiding

MULTITA is een onderzoeksproject gesubsidieerd door het Federaal Wetenschapsbeleid, een departement voor wetenschappelijk onderzoek binnen de Belgische Federale Overheid. Het project ging van start in juni 2012 en eindigde in december 2014. MULTITA werd gecoördineerd en gepromoot door twee federale wetenschappelijke instellingen: de Koninklijke Musea voor Kunst en Geschiedenis (KMKG) en het Koninklijk Instituut voor Kunstpatrimonium (KIK) in Brussel.

Versillende Belgische en internationale organisaties werkten aan het project mee: het Koninklijk Museum voor Schone Kunsten (Brussel), de terminologische dienst en de Raad voor Cultureel Erfgoed van de Fédération Wallonie-Bruxelles (Brussel), het Museum voor het Kostuum en de Kant (Brussel), het Rijksbureau voor Kunsthistorische Documentatie (Den Haag) en de Fédération et Ressources sur l'Antiquité (FRANTIQ, Parijs). Het project werd begeleid door een opvolgingscomité met leden Jef Malliet (Provincie Limburg, Erfgoed Plus), Els Angenon (Digital Collections, KMKG) en Bert Lemmens (PACKED).

De idee achter MULTITA ontstond nadat de Koninklijke Musea voor Kunst en Geschiedenis een aantal jaren geleden begonnen met het digitaliseren van de collecties. Toen het museum een collectie management systeem (CMS) aankocht en begon met het invoeren van de collecties, werd duidelijk dat er nood was aan degelijke thesauri om hun collectie doorzoekbaar te maken. Tijdens de zoektocht naar goede kunsthistorische of archeologische thesauri die ze in hun CMS konden gebruiken werd duidelijk dat de beschikbare, meertalige, thesauri erg varieerden.

2. De data

De partners in het MULTITA-project gebruiken gecontroleerde vocabulaires voornamelijk om hun collecties te beheren. De thesauri variëren echter in structuur, domein, aantal concepten en niveau van meertaligheid. Sommige instellingen hebben in het verleden inspanningen gedaan

om een gestructureerde, op zijn minst hiërarchische, thesaurus te ontwikkelen volgens internationale standaarden zoals ISO (International Standard Organization). Andere gebruiken een eenvoudige, louter alfabetisch geordende woordenlijst zonder conceptuele relaties. Vaak zijn deze woordenlijsten ook eentalig.

De vocabulaires waarmee gewerkt werd in MULTITA zijn:

- De thesauri van de Koninklijke Musea voor Kunst en Geschiedenis (behalve de thesaurus muziekinstrumenten): materialen, technieken, objectnamen en geografische referenties.
- De thesauri van het Koninklijk Instituut voor Kunstpatrimonium (KIK): materialen, technieken en objectnamen.
- PACTOLS-thesauri van FRANTIQ: persoonsnamen, chronologie, plaatsen, kunstwerken, volkeren en onderwerpen.
- AICIM-thesauri van de Fédération Wallonie-Bruxelles: materialen, technieken en objectnamen.
- De thesaurus van het Koninklijke Museum voor Schone Kunsten: objectnamen, materialen en technieken.
- Nederlandstalige Art and Architecture Thesaurus (AAT): abstracte begrippen, activiteiten, fysieke kenmerken, materialen, objecten, actoren en organismen, stijlen en perioden.

De Art and Architecture thesaurus werd beschouwd als referentiethesaurus, terwijl de andere thesauri geïmporteerd en bewerkt konden worden in het project.

3. Projectresultaten

3.1 Afstemmen van concepten uit verschillende thesauri

We stemden zoveel mogelijk concepten tussen de verschillende thesauri af. We kozen de Art and Architecture Thesaurus, ontwikkeld door het Getty Research Institute, als een referentiethesaurus voor dit werk omdat de thesaurus beschouwd wordt als een standaard in kunsthistorische thesauri en bovendien zijn de concepten vertaald in meerdere talen, waaronder het Nederlands. Deze Nederlandstalige thesaurus wordt beheerd door RKD, de Nederlandse partner in het MULTITA-project.

Voorbeeld van afstemming: de objectnaamthesaurus van het KMKG werd zoveel mogelijk afgestemd op de AAT. Indien ze in beide bronnen aanwezig waren, werd de structuur, de scope note en de vertaling van de individuele concepten vergeleken met de concepten uit de AAT. De URI's van de AAT – de “unique identifiers” die worden toegekend aan elk online concept – werden overgenomen in een Excel-file om verder hergebruik (linken van collecties) te garanderen. De scope notes van de AAT werden – waar toepasbaar – overgenomen in het CMS van de KMKG.

3.2 Verrijking van lokale thesauri

Vaak ontbrak er in de thesauri van de partners bepaalde belangrijke informatie, zoals scope notes, vertalingen, synoniemen en relaties (hiërarchisch, associatief of equivalent). Een deel van het werk bestond erin om deze ontbrekende informatie aan de thesauri toe te voegen.

Voorbeeld van verrijking: de AICIM-thesauri en de PACTOLS-thesauri werden vertaald naar het Nederlands tijdens de afstemmingsfase. De vertalingen voor de PACTOLS-thesauri kon rechtstreeks gedaan worden in de CMS van FRANTIQ. Dit betekent dat de vertalingen rechtstreeks online zichtbaar werden. De thesauri van de KMKG werden verrijkt met scope notes van de AAT. Wanneer concepten uit de Belgische thesauri niet aanwezig waren in de AAT, wat bijvoorbeeld het geval was voor veel archeologische concepten, werden deze rechtstreeks toegevoegd als kandidaattermen via de webservice van het RKD in de

3.3 Uitwisselen van expertise

In MULTITA zochten we ook naar manieren om ons netwerk uit te breiden, zodat we meer knowhow en inzichten konden krijgen in het ontwikkelen of beheren van een semantische en structureel correcte thesaurus.

Voorbeeld van uitwisseling : het project sloot een samenwerkingscontract met AthenaPlus, een Europees project dat digitale culturele content wil creëren en ontsluiten, waardoor het mogelijk werd om vergaderingen bij te wonen en geïnformeerd te blijven door mailings. Door dit partnerschap werd het ook mogelijk om binnen MULTITA het thesaurus management systeem xTree te gebruiken, een onmisbare tool voor het ontwikkelen van semantische thesauri. xTree is een webservice waar thesauri kunnen ontwikkeld worden, gelinked, geredigeerd en geëxporteerd in SKOS, een formaat dat toelaat om thesauri online te publiceren en te linken met andere thesauri (collecties) op het web.

3.4 Hergebruik van URI's

We konden URI's (*Uniform Resource Identifier*) van de AAT hergebruiken en implementeren in onze lokale thesauri. Op die manier kunnen we een link garanderen die bestaat tussen gelijke of eerder gelijke concepten in verschillende thesauri.

Voorbeeld van hergebruik van URI's: in de thesauri van het KIK en deze van het KMKG (objectnaam) werden de URI's van de AAT overgenomen voor elk thesaurusconcept. Dit werd rechtstreeks gedaan in het CMS van het KIK en in een Excel-file voor het KMKG. De URI's van de AAT werden ook toegevoegd aan het veld "URI" in xTree. Dit betekent dat de concepten van onze micro-thesauri automatisch de URI van de AAT zullen verkrijgen.

3.5 Ontwikkelen van micro-thesauri

Tenslotte hebben we in MULTITA ook geprobeerd een eigen gestructureerde, gestandaardiseerde en meertalige micro-thesaurus over een aantal specifieke thema's te ontwikkelen, zodat elke partner deze zou kunnen gebruiken.

Het ontwikkelen van gestandaardiseerde, kunsthistorische en archeologische thesauri in het Engels, Nederlands en Frans was één van de opdrachten in dit project. Vanuit de partners was er een specifieke vraag naar thesauri in bepaalde domeinen (micro-thesaurus). Door het ontwikkelen van deze micro-thesauri wilden we hen tevens een model aanbieden van een wetenschappelijke, meertalige thesauri, rekening houdend met de geldende standaarden. Het spreekt voor zich dat het niet mogelijk was om een volledig inclusieve thesaurus te ontwikkelen, zoals de AAT. Daarom kozen we ervoor om ons te richten tot twee specifieke interessedomeinen die nuttig kunnen zijn voor alle partners. We ontwikkelden twee micro-thesauri:

- a. *Textieltechnieken, materialen, actoren en objecten*: textiel is een vrij ingewikkeld, technisch onderwerp. De textielthesauri van de partners waren allemaal zeer gevarieerd, bijvoorbeeld in de manier waarop textielconcepten werd gestructureerd of gecategoriseerd. Er werden ook veel fouten en inconsistenties gevonden.
- b. *Stijlen en periodes*: geen enkele partner beschikte over een uitgebreide kunsthistorische thesaurus van stijlen en periodes. Bovendien vonden we ook buiten het project geen goede, meertalige stijlen- en periodethesaurus.

De relevante termen/concepten, relaties en scope notes uit de partner-terminologieën werden zoveel mogelijk in de thesauri opgenomen zodat we een verbeterde, rijkere en "holistische" versie verkregen. In de micro-thesaurus over textiel, werden ook de textieltermen uit de thesaurus van het Textielmuseum in Tilburg (Nederland) opgenomen. De AAT werd beschouwd

als een referentie, alhoewel we geregeld fouten of inconsistenties opmerkten. Ook deze werden verbeterd in onze versie en doorgegeven aan de AAT.

Het Duitse thesaurus management systeem xTree was het beste systeem voorhanden om de thesauri te ontwikkelen. Het systeem biedt functionaliteiten aan zoals:

- Ontwikkelen van nieuwe thesauri of importeren van bestaande thesauri
- *Preferred* en *non-preferred terms* voor elk concept, in alle ISO-talen beschikbaar
- *Labelled concepts*: keuze tussen concepten, gidstermen (zgn. “node label”), hiërarchieën en facetten
- Unieke ID's voor elk concept en de mogelijkheid om andere externe URI's toe te voegen
- Hiërarchische structuur en relaties: generische (*kind of*), partitieve (*part of*) or instantiële (*instance of*)
- Linken van concepten uit verschillende thesauri
- *Notes*: redactionele notities (*editorial*), historische notities (*historical*), scope notes, definities (*definitions*)

Het ontwikkelen van deze twee micro-thesauri werd afgerond. De thesauri worden aangeboden als wetenschappelijk referentiewerk of om te importeren in lokale collectiebeheersystemen. De thesauri zijn beschikbaar in het Nederlands, Frans en Engels zijn, alsook in SKOS-formaat.

4. Over schema's en standaarden

In het MULTITA-project werden verschillende standaarden bestudeerd, waarvan de belangrijkste **SKOS** en **ISO 25964-2**.

Simple Knowledge Organization System of SKOS is een uitwisselingsstandaard ontwikkeld door het World Wide Web Consortium (W3C) om thesauri – en vooral de semantische informatie inherent aan thesauri – online te representeren. SKOS maakt gebruik van het *Research Description Framework* (RDF), waardoor informatie kan uitgewisseld worden door computerapplicaties. Grondbeginsel van SKOS/RDF is het gebruik van URI's in plaats van hypertext (HTML) om informatie online te plaatsen.

SKOS bestaat uit *classes* en *properties* die gedefinieerd worden in een SKOS handleiding. *SKOS:concept* bijvoorbeeld werd ontwikkeld om thesaurustermen voor te stellen.

Property:prefLabel werd ontwikkeld om *preferred terms* te representeren. De MULTITA microthesauri werden in xTree ontwikkeld in SKOS.

Naast SKOS werd in MULTITA ook de ISO 25964-2 standaard geconsulteerd. Deze ISO-uitgave biedt een leidraad voor het ontwikkelen en beheren van thesauri en ter bevordering van de interoperabiliteit tussen verschillende thesauri.

Enkele bevindingen bij het bestuderen van deze standaarden:

- De ISO-norm 25964-2 is niet altijd duidelijk en laat ruimte voor interpretatie. Een voorbeeld is het verschil tussen een *microthesaurus* (p. 8) en een *facet* (p. 5). Een *microthesaurus* wordt beschreven als een “designated subset of a thesaurus that is capable of functioning as a complete thesaurus”. Een *facet* is een “grouping of concepts of the same inherent category”. Volgens beide definities zou “architectuur” zowel een *facet* als een *microthesaurus* kunnen zijn.
- Ondanks het feit dat er in ISO een onderscheid wordt gemaakt tussen een *facet* en een *microthesaurus*, is er in SKOS niets voorzien om deze informatiestructuren afzonderlijk te presenteren.

- Gidstermen of *node labels* kunnen volgens de SKOS-handleiding worden voorgesteld als *skos:Collection*. Toch wordt in de realiteit *skos:Collection* vaak gebruikt om *facetten* of *microthesauri* te representeren.
- De Art and Architecture Thesaurus structureert de thesaurus door middel van *facetten*, *hiërarchieën* en *gidstermen*. De *hiërarchieën* kunnen niet naar SKOS worden vertaald omdat deze structuur niet voorzien is in SKOS.
- In xTree bestaat de mogelijkheid om het concept of structuur-type te selecteren: *category OF concept OF structureel element*. Binnen de structurele elementen kan men vervolgens kiezen voor *hierarchy labels*, *nonindexing concepts*, *node label* en *facet label*. Het is niet altijd duidelijk waar deze labels precies voor staan. De terminologie is verwarrend.
- De ISO-norm voorziet drie types van hiërarchische relaties (p. 25): *generic* (boek is “een soort van” informatiedrager), *instance* (het Louvre is “een voorbeeld van” een museum) en *whole-part* (Brussel is “een deel van” België). In SKOS kunnen deze relaties echter niet gerepresenteerd worden: alle relaties worden voorgesteld als generische *broader of narrower*. Een deel van de informatie aanwezig in een relationele thesaurus kan in SKOS dus verloren gaan, bijvoorbeeld geografische thesauri.
- In SKOS kunnen geen bronvermeldingen worden opgenomen, wat belangrijk kan zijn voor de scope notes.

Het is dus niet altijd eenvoudig om standaarden te volgen of te implementeren. Vaak worden er nog bepaalde subjectieve keuzes gemaakt. Het kan problematisch zijn wanneer SKOS-elementen op verschillende manieren worden geïnterpreteerd, bijvoorbeeld in applicaties waar SKOS-bestanden automatisch worden opgeladen.

5. Problemen en uitdagingen van meertaligheid

Tijdens het ontwikkelen van de meertalige thesauri en het vertaalwerk in onze eigen thesauri werden bepaalde problemen en uitdagingen vastgesteld die specifiek te maken hebben met de problematiek van de meertaligheid. Het is belangrijk om zich bij het ontwikkelen van nieuwe, meertalige thesauri, of bij het mappen van concepten tussen thesauri van deze moeilijkheden bewust te zijn.

Enkele voorbeelden illustreren dit:

Ex. *Entablature* kan vertaald worden als *hoofdgestel* (KIK-IRPA) of als *entablement* (AAT) in het Nederlands.

- Een term kan verschillende, potentieel correcte, equivalente vertalingen hebben in één taal. Het is belangrijk om de meest toepasbare synoniemen op te nemen als *alternative term*, zodat de eindgebruiker een aantal mogelijkheden heeft om tot het juist zoekresultaat te komen.

Ex. *Pichet* in het Frans kan niet vertaald worden naar het Nederlands of naar het Engels.

- Er is een gebrek aan equivalentie tussen deze talen: de Franse term heeft geen exact synoniem in het Engels of in het Nederlands. Men kan vertalen door te parafaseren: *small jug*, *kleine kruik*.
- *Factoring* of het combineren van meerdere concepten is eveneens een mogelijkheid, bijv. *gros bétail* (FR) kan in het Engels vertaald worden als een combinatie met *cattle* (EN) + *horses* + *goats* etc.

Ex. *Cup*, *trophy*, *bowl*, *champagne glass* kunnen vertaald worden als *coupe* in het Frans

- Deze situatie wordt afgeraden omdat het een ambivalentie creëert in de thesaurus. In een thesaurus is het belangrijk om te vermijden dat één term wordt gebruikt om verschillende concepten te representeren. In dergelijk geval kan men creatief zijn: *trophy* kan vertaald worden als *trophée*, *champagne glass* wordt *coupe à champagne* etc.

Deze problemen en discrepanties tussen talen hebben te maken met de intrinsieke eigenschappen van talen, maar kunnen ook met de natuur of de cultuur in het algemeen in verband gebracht worden. Een taal reflecteert altijd de kijk van een bepaalde cultuur op de wereld. Een object kan gebruikt worden in één cultuur en er op die manier een specifiek concept representeren, maar kan compleet afwezig zijn in een andere cultuur, zodat er bijgevolg geen woord voor bestaat.

Ex. *Daubière* in het Frans: dit is een type kookpot gebruikt om *daube* te maken, een klassiek Provençaals stoofpotgerecht. Het object en het typisch aardewerk werd enkel in Frankrijk gebruikt en is dus niet bekend in andere culturen, vandaar een non-equivalentie voor het concept in het Engels en in het Nederlands. In dit geval is het raadzaam om de Franse term in de thesaurus gewoon over te nemen.

Hieraan gerelateerd is het opvallend dat culturele concepten vaak de precisie en de duidelijkheid missen van concepten uit de (exacte) wetenschappen. Een mathematisch concept, bijvoorbeeld, heeft een duidelijke, begrensde betekenis. De definitie is logisch en kan slechts op één en dezelfde manier geïnterpreteerd worden. Eén concept vertegenwoordigd één term. Culturele concepten daarentegen, zijn moeilijk te definiëren en de grenzen zijn vaak niet duidelijk. Dit is al vaak het geval voor mensen die zich in dezelfde culturele setting bevinden, en vaak nog meer voor mensen van verschillende nationaliteiten, culturen en talen. Bijv. *Trapezium*. De term 'trapezium' refereert aan een geometrische figuur die aan volgende criteria voldoet:

- heeft twee aanliggende hoeken die aanvullend zijn, dit betekent dat ze samen 180 graden vormen.
- de hoek tussen een zijde en een diagonaal is gelijk aan de hoek tussen de tegenovergestelde zijde en dezelfde diagonaal.
- de diagonalen snijden elkaar in wederzijdse verhouding (deze verhouding is dezelfde als deze tussen de lengten van de parallelle zijden).
- etc.

Een figuur kan niet "min of meer trapeziumachtig" zijn. Ofwel voldoet de figuur aan alle criteria en is het een trapezium, ofwel voldoet het niet aan deze criteria en is het geen trapezium. Er is geen grijze zone, het concept heeft duidelijke betekenisgrenzen.

Bijv. *Stoel*. De term "stoel" refereert aan een bepaald meubeltype dat ontworpen werd om op te zitten. Een stoel kan verschillende eigenschappen bezitten, maar het voldoet meestal aan volgende criteria:

- het heeft een zitfunctie.
- het heeft meestal vier poten.
- het heeft meestal een rugvlak.
- het kan armleuningen hebben.

Een stoel moet echter niet aan al deze criteria voldoen om een stoel te zijn. Sommige stoelen zijn meer "stoelachtig" dan anderen. Er zijn geen specifieke grenzen aan dit concept. Bovendien kunnen ze ook cultureel verschillend zijn. Wat als een stoel beschouwd wordt in één cultuur, kan in een andere beschouwd worden als een sofa.

Veel van voorgaande voorbeelden met betrekking tot meertaligheid en vertalingen werden reeds geïdentificeerd en benoemd door Aitchison en Gilchrist (London, 1987):

- Exacte equivalentie: *mosaïque* (FR), *mosaic* (EN), *mozaïek* (NL) ; *amulette* (FR), *amulet* (EN), *amulet* (NL)
- Inexacte equivalentie: *chapeau* (FR), ***top hat*** (EN), *hoed* (NL) ; *crucifix* (FR), ***cross*** (EN), *kruisbeeld* (NL)
- *Single to multiple* equivalentie: *bol* (FR), ***small bowl*** (EN), ***kommetje/kleine kom*** (NL) ; *javeline* (FR), *javelin* (EN), ***kleine werpspeer*** (NL)
- Gedeeltelijke equivalentie: *carriage* (EN), ***voiture*** (FR), *rijtuig* (NL) ; ***model*** (EN), *mannequin* (FR), *paspop* (NL)
- Non-equivalentie : *statuette* (FR), *statuette* (EN), ***beeldje (van bepaalde hoogte)*** ; *figurine* (FR), *figurine* (EN), ***(menselijk) beeldje***.

6. Case study: *Kandelaar* (NL), *Chandelier* (FR) and *Candlestick*, *Candelabrum* (EN)

De vier termen uit de titel refereren aan een bepaald object dat kaarsen kan houden. Er zijn echter twee termen in het Engels en slechts één term in het Frans en in het Nederlands. *Kandelaar* en *chandelier* kunnen verwijzen naar een standaard die één kaars houdt of naar een standaard die twee of meer kaarsen houdt. Het woordenboek Van Dale definieert het Nederlandse kandelaar duidelijk als: "Standaard waarop één OF meer kaarsen geplaatst kunnen worden". In het Engels bestaat er niet één term die naar een kaarshouder verwijst, ongeacht het aantal kaarsen deze kan houden. Het Engels beschikt over twee concepten: een *candlestick* die één kaars houdt en een *candelabrum* die twee of meer kaarsen houdt.

Het probleem is conceptueel van aard. Het feit dat het Nederlands of Frans geen aparte term voorziet voor een object dat één kaars houdt en een object dat meerdere kaarsen houdt, toont aan dat de sprekers van het Nederlands en Frans geen verschil zien: ze identificeren een houder voor kaarsen als één object, één concept, ongeacht het aantal kaarsen het object kan houden. Aangezien er geen verschil is, is er ook geen terminologisch onderscheid. Eén woord/term volstaat om het concept te beschrijven. In het Engels daarentegen, worden de twee types als verschillende concepten beschouwd en bijgevolg zijn afzonderlijke termen nodig.

Wanneer een meertalige thesaurus wordt ontwikkeld, of wanneer vertalingen aan een bestaande thesaurus worden toegevoegd, zal dit problemen opleveren. Laten we naar de AAT kijken, een thesaurus ontwikkeld door Amerikanen en bijgevolg sterk geankerd in de Angelsaksische cultuur. De AAT voorziet twee concepten voor twee termen: *candlestick* en *candelabrum*. De scope notes beschrijven duidelijk dat *candlestick* één kaars houdt en *candelabrum* twee of meer. Als men deze termen naar het Nederlands vertaald, is het niet mogelijk om *kandelaar* te gebruiken voor beide concepten (zoals in het Nederlands gangbaar is). De Nederlandstalige AAT lost het probleem als volgt op:

- *Candlestick* = *Kandelaar*
- *Candelabrum* = *Kandelaber*

Deze oplossing is echter niet volledig correct. Eerst en vooral kiest de AAT ervoor om een deel van de betekenis van *kandelaar* te negeren: in de AAT verwijst het enkel naar het object dat één kaars houdt. Het is echter mogelijk dat eindgebruikers de beperkte betekenis van het woord in de AAT niet kennen en met deze term op zoek gaan naar alle types kaarshouders, d.w.z. deze die één en deze die meerdere kaarsen houden. Het Nederlandse kandelaar wordt in de Van Dale beschouwd als een hyperoniem van kandelaber. Dit betekent in thesaurustermen: kandelaar NT kandelaber of kandelaber BT kandelaar. Een kandelaar is immers een "standaard

voor één of meer kaarsen". In the AAT worden beide termen echter als narrower terms beschouwd van het concept candleholder.

De term is ook vrij formeel en kan een verouderde betekenis hebben. Het is weinig waarschijnlijk dat eindgebruikers deze term als zoekterm zullen gebruiken.

Het is dus belangrijk om bijzondere aandacht te vestigen op twee zaken wanneer een thesaurus wordt vertaald:

- De discrepantie tussen concept vs. term en de verschillen tussen talen: sommige talen hebben geen term om bepaalde concepten uit te beelden (er is in het Nederlands geen term die een *candlestick* uitdrukt, i.e. een kaarshouder met maar één arm). Of, andersom, wat beschouwd wordt als één concept in één taal of cultuur, wordt in een andere taal/cultuur gezien als twee verschillende concepten (de Engelse taal ziet bijv. een verschil in *candlestick* en *candelabrum*).
- Het is van belang dat er rekening wordt gehouden met de eindgebruiker wanneer men een thesaurus vertaalt of ontwikkelt. Het is niet voldoende om taalproblemen op te lossen met kunstmatige ingrepen: de gebruiker zal de taal gebruiken zoals in het dagelijkse leven.

7. Besluit

Het MULTITA-project heeft ons toegelaten om de lokale thesauri van de partners te verbeteren en te verrijken, alsook om twee nieuwe micro-thesauri te creëren over textiel en stijlen/periodes. Het project omvatte een vooronderzoek, waarbij onder meer de thesauri van de partners werden geanalyseerd en er gezocht werd naar een thesaurus management tool; een aanleverfase, waarbij de thesauri in een CSV-file of in SKOS-formaat aan de coördinator werden geleverd en vervolgens geïmporteerd (automatisch of manueel) in een Excel of in het thesaurus management systeem xTree. Daarna volgde het terminologisch onderzoek, dat grotendeels samenviel met de implementatie-fase: hierin werden de concepten uit de verschillende thesauri vergeleken, vertaald en verrijkt met semantische relaties. Nieuwe scope notes werden gecreëerd en uitgewisseld met de AAT. In de *content aggregation*-fase werden twee nieuwe micro-thesauri ontwikkeld in xTree op basis van content (concepten) uit de partnerthesauri. In het MULTITA-project werden ook de problemen en uitdagingen van meertaligheid onderzocht.

Er werd een *memorandum of understanding* ondertekend met het Europese best practice-project AthenaPlus en nauwe samenwerkingsverbanden met FRANTIQ-CNRS (Parijs), RKD (Den Haag) en de Belgische partners. Er namen in totaal achtentwintig vergaderingen plaats, waarvan er vijf met alle nationale en internationale partners samen. Daarnaast werden er drie vergaderingen georganiseerd met het opvolgingscomité. Op 15 december 2012 werd een MULTITA-symposium gehouden voor de partners en op 5 december 2014 werden de resultaten van het MULTITA-project voorgesteld op de wetenschappelijke TOTh-workshop over het onderwerp *Terminology and Multilingualism*.

De praktische resultaten van het project zijn:

- Nederlandstalige vertaling van de PACTOLS-thesaurus (> 5000 concepten)
- Nederlandstalige vertaling van de AICIM-thesaurus (> 5000 concepten)
- Aanlevering van 200+ kandidaattermen in de AAT, in Frans, Nederlands en Engels
- Verrijking van de KIK-thesauri: toevoegen van (AAT- of eigen) scope notes, ontbrekende vertalingen in het Nederlands en Frans, structurele verbeteringen, toevoegen in CMS van AAT-URI's.

- Verrijking van de KMKG-thesauri: toevoegen van (AAT- of eigen) scope notes, ontbrekende vertalingen in het Nederlands en Frans, structurele verbeteringen, bijhouden van AAT-URI's.
- Ontwikkeling van een textielthesaurus (objecten, materialen, technieken) in Engels, Nederlands en Frans, beschikbaar in SKOS
- Ontwikkeling van een stijlen en perioden-thesaurus in Engels, Nederlands en Frans, beschikbaar in SKOS

Het was niet mogelijk om alle problemen van de partnerinstellingen met betrekking tot hun thesauri op te lossen in het MULTITA-project, maar het bracht zeker nieuwe inzichten over thesaurus-ontwikkeling in de culturele sector. Er werden grote delen van de thesauri verbeterd, verrijkt en voorbereid om in de toekomst te kunnen uitbreiden met nieuwe concepten of om de thesauri te hergebruiken in andere applicaties.

8. Bronnen

Boeken en artikels:

AITCHISON et al. 1987

Jean Aitchison, Alan Gilchrist, *Thesaurus Construction*, London 1987².

AITCHISON et al. 2005

Jean Aitchison, Alan Gilchrist, David Bawden, *Thesaurus construction and use : a practical manual*, London 2005.

CAGNOT et al. 2011

Stéphane Cagnot, Johann Holland, Marie-Véronique Leroi, Michael Culture Aisbl, *Your Terminology as Part of the Semantic Web. Recommendations and Guidelines for Design and Management*, Linked Heritage 2011.

DE BOER et al. s.d.

Victor de Boer, Jan Wielemaker, Judith van Gent, Marijke Oosterbroeck, Michiel Hildebrand, Antoine Isaac, Jacco van Ossenbruggen, Guus Schreiber, *Amsterdam Museum Linked Open Data*, EuropeanaConnect research project, s.d.

HUNTER 2001

Jane Hunter, *MetaNet: A metadata term thesaurus to enable semantic interoperability between metadata domains*, *Journal of Digital Information*, 1(8), 2001.

MAY et al. 2011

Keith May, Ceri Binding, Doug Tudhope, Stuart Jeffrey, *Semantic Technologies Enhancing Links and Linked Data for Archaeological Resources*, United Kingdom 2011.

MILLER 2003

Miller U., *Thesaurus and new information environment*. In: M. Drake and M. N. Maack (Eds.); *Encyclopaedia of library and information science*, 2nd ed. Boca Raon: Taylor & Francis Group.

MORSHED et al. 2011

Ahsan Morshed, Caterina Caracciolo, Gudrun Johanssen, Johannes Keizer (2011), *Thesaurus Alignment for Linked Data Publishing*, Proc. International Conference on Dublin Core and Metadata Applications 2011.

SHIRI 2012

Ali Shiri, *Powering Search. The Role of Thesauri in New Information Environments*, New Jersey

2012.

SOERGEL 2003

Dagobert Soergel, *Functions of a Thesaurus/Classification/Ontological Knowledge Base*, University of Maryland 2003.

VAN DALE, Elektronisch Woordenboek, versie 6.10/14.10, Van Dale Uitgevers 2014.

Thesaurusstandaarden:

ANSI/NISO Z39.19-2005, *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*, 2005.

ISO 25964-1, *Information and documentation — Thesauri and interoperability with other vocabularies — Part 1: Thesauri for information retrieval*, Geneva 2011.

ISO 25964-2, *Information and documentation — Thesauri and interoperability with other vocabularies — Part 2: Interoperability with other vocabularies*, Geneva 2013.

9. Relevante websites:

AICIM-thesauri: [AICIM thesauri](#)

Art and Architecture Thesaurus: [AAT Getty](#)

Art and Architecture Thesaurus Nederlandstalig: [AAT Ned](#)

AthenaPlus: www.athenaplus.eu

British Museum Object Names Thesaurus: [British Museum](#)

Carmentis: www.carmentis.be

Erfgoedplus: [Erfgoedplus](#)

Inventarisatie Terminologiebronnen, DEN (Digitaal Erfgoed Nederland) 2010, [Inventarisatie terminology sources DEN](#)

Glossarium CEST: www.projectcest.be

Mondeca – Linked Open Vocabularies: [MONDECA](#)

Multilingual websites and multilingual thesauri, Minerva survey 2004-2005, [Minerva survey](#)

PACKED: [PACKED](#)

PACTOLS: [PACTOLS](#)

TextielMuseum-TextielLab Tilburg: [TextielMuseum](#)

TOTh-workshop: [TOTh](#)

10. Contactpersonen:

e.coudyzer@kmg-mrah.be

erik.buelinckx@kikirpa.be