Il modello semantico di EuroWordNet come strumento per la strutturazione della relazione associativa nei thesauri

di Rita Marinelli, Fulvio Mazzocchi, Melissa Tiberi, Marta Motta

Introduzione

I thesauri sono strumenti che organizzano semanticamente uno specifico dominio di conoscenza per fini applicativi. La loro semantica relazionale [1] si basa su metodi attraverso i quali vengono stabiliti nessi tra termini con significati correlati. La struttura relazionale di un thesaurus è uno strumento di supporto fondamentale per il recupero dell'informazione (*information retrieval*), attraverso cui vengono aumentati il richiamo (*recall*) e la precisione (*precision*) della ricerca.

La rete delle relazioni thesaurali svolge una funzione semantica: da un lato essa fornisce una rappresentazione del significato di ciascun termine contenuto nel thesaurus; dall'altro essa fornisce una sorta di mappa della struttura concettuale del dominio di conoscenza.[1]

Il formato tradizionale di un thesaurus, così come è descritto negli standard internazionali, include tre relazioni fondamentali, due delle quali vengono definite su base concettuale (relazione gerarchica e relazione associativa), mentre la terza a livello soprattutto lessicale (relazione di equivalenza).

È opinione diffusa che, per poter meglio rispondere ai bisogni attuali in ambito di organizzazione dell'informazione, questo formato debba essere in qualche modo riconsiderato e perfezionato.

Oltre che per ottimizzare le procedure di recupero dell'informazione, questa ridefinizione appare necessaria anche per migliorare le potenzialità di utilizzo dei thesauri in ambiti quali l'intelligenza artificiale e il Web semantico.

RITA MARINELLI, CNR-Istituto di linguistica computazionale, via Moruzzi 1, 56124 Pisa, email rita.marinelli@ilc.cnr.it.

FULVIO MAZZOCCHI, CNR-Istituto dei sistemi complessi, via Salaria km 29,300, 00015 Monterotondo st. (RM), email fulvio.mazzocchi@isc.cnr.it.

MELISSA TIBERI, collaboratrice a progetto della Biblioteca nazionale centrale di Firenze (BNCF), piazza Cavalleggeri 1, 50122 Firenze, e-mail: tiberim77@yahoo.it.

MARTA MOTTA, collaboratrice a progetto della Biblioteca nazionale centrale di Firenze (BNCF), piazza Cavalleggeri 1, 50122 Firenze, e-mail: motta.marta@gmail.com.

Ultima consultazione siti web: 25 luglio 2010.

1 «a good thesaurus provides, through its hierarchy augmented by associative relationships between concepts, a semantic road map for searchers and indexers and anybody else interested in an orderly grasp of a subject field» [27].

Bollettino **AIB**, ISSN 1121-1490, vol. 50 n. 3 (settembre 2010), p. 249-263.

Il presente contributo[2] esamina la possibilità di strutturare e specificare la rappresentazione semantica di una delle relazioni thesaurali, la relazione associativa, mediante la sua differenziazione in sottotipi. A tal fine verrà utilizzato come riferimento il modello semantico di EuroWordNet (EWN), così come è stato usato in una delle sue versioni nazionali, ItalWordNet (IWN), nell'ambito di un progetto riguardante la terminologia del settore marittimo (*Mariterm*).

Vengono, inoltre, prese in considerazione alcune questioni connesse con il lavoro di articolazione della relazione associativa e, in particolare, il modo in cui essa sembra dipendere dal dominio di conoscenza in cui viene considerata.

1. La funzione della relazione associativa nei thesauri

Un thesaurus è organizzato secondo due piani distinti: una struttura verticale (classificatoria e tassonomica) e una struttura di connessioni orizzontali. Entrambe le strutture sono importanti in virtù delle funzioni complementari che (esse) svolgono.

La struttura verticale si basa soprattutto sulla relazione gerarchica *genere-specie,* che collega tra loro termini semanticamente sovraordinati (BT: *Broader Terms*) e termini semanticamente sottordinati (NT: *Narrower Terms*).

La struttura ad albero che viene così generata funziona come strumento di controllo e rappresentazione del significato dei termini: la semantica di ogni termine è infatti stabilita, almeno in parte, dalla sua posizione all'interno dell'albero gerarchico. Essa inoltre è uno strumento di supporto per la navigazione semantica: gli utenti possono selezionare, all'interno di una serie di termini con differenti livelli di specificità, quale di questi usare per esprimere un determinato concetto [2].

La struttura orizzontale si sviluppa, invece, a rete ed è composta dalle relazioni associative (RT/RT: *Related Terms* / *Related Terms*). L'importanza della relazione associativa consiste nel fatto che essa risponde al bisogno cognitivo e pragmatico di rendere esplicita la struttura delle connessioni (non gerarchiche), che caratterizzano un determinato settore del sapere e che non possono essere rappresentate mediante la struttura ad albero, che si basa sostanzialmente su criteri logici.

In tal senso, la relazione associativa contribuisce anche alla mappatura semantica dei termini fornendo connessioni concettuali potenzialmente utili a tale scopo [3]. Per esempio, mentre in una ipotetica gerarchia il termine "anidride carbonica" <*is_a*> "composto chimico inorganico", attraverso la rete degli RT è possibile evidenziare altri aspetti del suo significato, come il fatto che può essere causa (<*causes*>) di "inquinamento" o di "spegnimento del fuoco" .

La relazione associativa include diversi tipi di nessi tra termini che possono risultare utili per suggerire termini alternativi o aggiuntivi per l'indicizzazione e il recupero dell'informazione [2]. Tuttavia, essa è piuttosto difficile da specificare, anche perché le regole per la sua "implementazione" (applicazione) in un thesaurus sono piuttosto vaghe.

Gli standard internazionali per la costruzione dei thesauri, infatti, forniscono in merito soprattutto definizioni pragmatiche, stabilendo che esiste una relazione associativa quando un termine è necessariamente contenuto nella definizione dell'altro [4] o quando è possibile individuare un collegamento di natura tematica tra termini in base all'esperienza [5].

Soergel [6] sostiene che i termini sono connessi in modo associativo quando un indicizzatore (o chi compie la ricerca) nel valutare l'uso di un determinato termine è portato a richiamare alla mente l'altro, e quando quest'ultimo non è collegato gerarchicamente con il precedente.

2 Il presente articolo è stato sviluppato sulla base di [28].

Maniez [7], dal canto suo, sottolinea l'importanza dei nessi associativi tra termini che si sovrappongono semanticamente, o che hanno un alto tasso di co-occorrenza nei titoli e/o nei campi di soggetto, o che sono connessi da fattori "extrasemantici" (per esempio, da un rapporto causa-effetto).

Tuttavia, come sottolineato da Dextre-Clarke [2], non è semplice trasformare le suddette indicazioni in una regola da usare in modo coerente e costante.

L'operazione di stabilire nessi associativi fra termini è piuttosto connessa in molti casi a valutazioni soggettive e contingenti da parte di chi compila (allestisce) un thesaurus il quale, piuttosto che effettuare un'approfondita analisi semantica, stabilisce gli RT sulla base di ciò che ritiene effettivamente utile all'utente. Questo sembra il caso, per esempio, del thesaurus INSPEC in cui il termine *Ferromagnetism* ha come RT *Magnets* ma non *Magnetism*, sebbene quest'ultimo termine sia presente nel thesaurus stesso e sebbene il suo legame con il primo sia più evidente rispetto all'RT che è stato scelto.

2. Strutturazione della relazione associativa mediante il modello semantico di EuroWordNet-ItalWordNet

Una delle possibilità per definire in modo più chiaro il contenuto semantico della relazione associativa si basa sulla differenziazione della relazione standard in sottotipi. A tale argomento sono stati dedicati molti studi interessanti, considerata l'attuale tendenza verso l'aumento del grado di specificità e di formalismo, anche al fine di aumentare le possibilità di applicazione dei sistemi di organizzazione della conoscenza (*Knowledge Organization Systems - KOSs*) nell'ambito dell'intelligenza artificiale.

Lancaster [8] e Schmitz-Esser [9] sostengono, per esempio, che una articolazione in sottotipi della relazione associativa e una più chiara definizione della sua semantica possano incidere positivamente anche sul recupero dell'informazione.

Uno degli schemi proposti è basato su uno studio compiuto dall'ALA - American Library Association [10]. Scopo della ricerca era quello di analizzare le relazioni fra termini utilizzate nell'indicizzazione semantica e nella catalogazione. Sono stati identificati e ordinati gerarchicamente circa 100 sottotipi di RT. I primi due livelli sono elencati nella tab. 1.

Tab. 1. Sottotipi della relazione associativa secondo l'ALA (1997)

- Combined ideas
- Conceptually related terms
- Contiguity
  - Definitional based contiguity
  - Empirical knowledge-based contiguity
- Definitional associative relationships
- Different hierarchy associative relationships
  - Environmental relationships
  - Etymologically related RTs
  - Process issue RTs
  - Property issue RTs
- Same hierarchy associative relationships
  - Causal relationships
  - Closely related siblings
  - Considered as relationships
  - Coordinated ideas
  - Entities studied in mutual relationships
  - Partitive relationships

- Persons interacting in a social context
- Property/property pairs
- Reciprocals
- Similarity

Meaning overlap associative relationships

- Extrasemantic relationships
- Used somewhat interchangeably

Scope issues

- Generic Terms
- Polysemes
- Noun not true broader term
- Scope noted term and other possible meanings

Unspecified associative relationships

Le relazioni di tipo associativo, ovviamente, non sono di pertinenza esclusiva dell'organizzazione della conoscenza. Nell'ambito della terminologia, spesso ci si riferisce ad esse come relazioni *ontologiche*. La loro importanza nell'analisi concettuale è evidente in quanto:

> concepts have also other dimensions apart from those expressed in logical concept relations and systems. The number and the quality of their dimensions depend on the particular concept category (e.g., entity, activity, process, method, and property concepts) and the subjects fields [11, p.130].

Tali relazioni possono essere suddivise secondo due tipologie: relazioni basate sulla contiguità spaziale e temporale e relazioni che includono una componente causale (corrispondono alle relazioni di *influence* nella fig. 1).

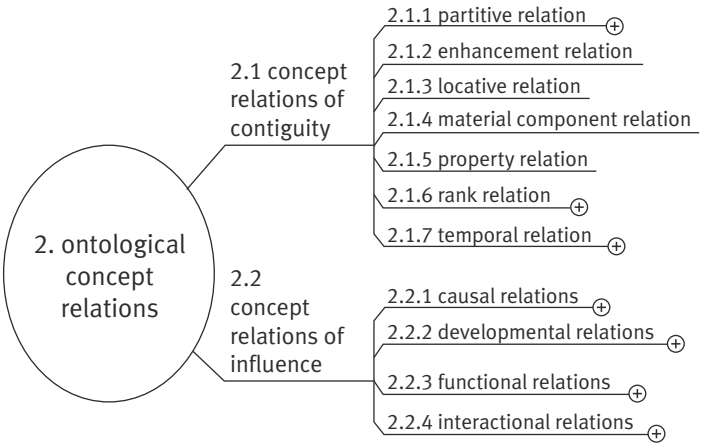

Fig. 1: Relazioni ontologiche fra concetti [11]

Un'altra possibilità di suddivisione in sottotipi della relazione associativa potrebbe derivare dalla linguistica computazionale, prendendo in considerazione, per esempio, i database semantico lessicali *EuroWordNet* (EWN) e *ItalWordNet* (IWN).

IWN è stato sviluppato nell'ambito di due progetti di ricerca distinti: *EuroWordNet* [12] e *SI-TAL* (Sistema integrato per il trattamento automatico del linguaggio). Quest'ultimo è un progetto nazionale dedicato alla creazione di risorse linguistiche su ampia scala e di strumenti software per l'elaborazione della lingua italiana scritta e parlata. Durante questo progetto, la versione del database italiano costruito per il progetto europeo EWN è stata aggiornata, aumentandone la copertura lessicale. Infatti sono stati aggiunti insiemi di nomi e verbi, che non erano presenti nella versione EWN, aggettivi, avverbi e un consistente insieme di nomi propri.

Il modello semantico di EWN-IWN viene qui analizzato in quanto è stato usato per la creazione del database terminologico *Mariterm,* che contiene termini appartenenti al dominio marittimo, in particolare al settore tecnico-nautico e a quello dei trasporti marittimi. Questo progetto ha evidenziato la possibilità di applicare il modello di EWN-IWN per strutturare la terminologia di uno specifico campo di conoscenza [13].

Mariterm fornisce informazioni codificate di tipo semantico e concettuale secondo un modello multidimensionale di significato. L'informazione lessicale è rappresentata in modo tale da poter essere utilizzata per molteplici tipi di applicazioni.

Il database terminologico è stato strutturato con gli stessi principi del wordnet italiano generico, cioè si tratta di un database di tipo relazionale, in cui le relazioni che legano i termini fra loro sono di tipo semantico lessicale. Contiene informazioni semantiche relative a 3500 lemmi, raggruppati in circa 2500 *synset*. Il concetto di *synset* è il punto focale attorno a cui EWN e IWN, e prima ancora WordNet (WN), sono costruiti: un *synset* è definito come un insieme di termini sinonimi, che appartengono alla stessa *Part-of-Speech* (PoS) e che sono intercambiabili in almeno un contesto [14-15], per esempio: *imbarcazione/natante*; *timone/governale*; *traghetto/nave traghetto*; *naufragare/ colare a picco/affondare*. Va considerato che, a differenza di quanto accade in un thesaurus, le tipologie grammaticali contenute in *Mariterm* non sono soltanto nomi o sintagmi nominali ma anche verbi, aggettivi e avverbi. È inclusa, inoltre, un'ampia rappresentanza di nomi propri, codificati come *instances* di classi di appartenenza [16].

Ogni *synset,* come in EWN-IWN, è collegato ad altri *synset* per mezzo di un ampio set di relazioni semantico-lessicali ed è collegato, in base al suo *iperonimo*, alla *Top Ontology* (TO) di EWN-IWN. Essa è costituita da un insieme di concetti (*entities*) con grande livello di astrazione: «a livelli di gerarchia più alti, più astratti» [17]; il primo livello della TO è costituito da tre fondamentali categorie (*orders*) concettuali [18, 12]:
1) entità del 1° ordine: oggetti concreti che esistono nel tempo e nello spazio;
2) entità del 2° ordine: eventi, processi, stati e situazioni che possono essere collocati in una dimensione temporale e spaziale;
3) entità del 3° ordine: idee, pensieri, teorie, ipotesi, ecc. che esistono indipendentemente dal tempo e dallo spazio e che sono "non osservabili" [17].

Queste tre categorie concettuali fondamentali sono suddivise ulteriormente in base ad una gerarchia di 63 *Top Concepts*, che vengono considerati come sostanzialmente indipendenti dal linguaggio[3]:

3 Questo schema è stato parzialmente modificato in IWN per consentire una classificazione ontologica degli aggettivi.

Tab 2. Top Ontology di EWN

Top

1st Order Entity

- Origin
  - Natural
    - Living
      - Plant
      - Human
      - Creature
      - Animal
  - Artifact
- Form
  - Substance
    - Solid
    - Liquid
    - Gas
  - Object1
- Composition
  - Part
  - Group
- Function
  - Vehicle
  - Representation
    - MoneyRepresentation
    - LanguageRepresentation
    - ImageRepresentation
  - Software
  - Place
  - Occupation
  - Instrument
  - Garment
  - Furniture
  - Covering
  - Container
  - Comestible
  - Building

2nd Order Entity

- SituationType
  - Dynamic
    - BoundedEvent
    - UnboundedEvent
  - Static
    - Property
    - Relation
- SituationComponent
  - Cause
    - Agentive
    - Phenomenal
    - Stimulating
  - Communication
  - Condition
  - Existence
  - Experience
  - Location
  - Manner
  - Mental
  - Modal
  - Physical
  - Possession
  - Purpose
  - Quantity
  - Social
  - Time
  - Usage

3rd Order Entity

In base a questa gerarchia, ciascun termine del database di dominio marittimo ha una classificazione ontologica, cioè è collegato a uno o più concetti della TO, per esempio: vento → *Dynamic, Phenomenal*; navigazione →*Agentive, Dynamic, Purpose* [19].

Mariterm è stato sviluppato sulla base del modello concettuale EWN-IWN, nella filosofia *WordNet* (WN) [14, 20], ma adotta un set di relazioni semantico-lessicali più ricco di quelle usate in WN. Tali relazioni possono essere raggruppate secondo due tipi: le *relazioni di equivalenza*, che collegano i *synset* in italiano con i concetti corrispondenti (sinonimi, o quasi-sinonimi) nel *WordNet* di Princeton, e le *relazioni interne*, in cui l'informazione è codificata nella forma di relazioni

semantico-lessicali tra coppie di *synset*. Queste relazioni sono sia di tipo gerarchico ("verticale"), espresse con <*has_hyperonym/has_hyponym*>, sia di tipo "orizzontale" (circa 40) che comprendono, tra le altre, le relazioni di causa, di ruolo, di mezzo, di luogo ecc.

Proprio queste ultime relazioni di tipo "orizzontale" contenute in *Mariterm* potrebbero essere utilizzate anche nei vocabolari controllati, come in parte già sta accadendo, per esempio per esprimere più efficacemente relazioni di causalità, di implicazione, ecc. al fine di incrementare il grado di informazione semantica contenuto in un thesaurus.

Devono essere menzionate, inoltre, le relazioni *Plug-in*. Esse vengono usate per permettere il collegamento tra un termine del dominio specialistico a un *synset* del lessico generico IWN, in termini di iper/iponimia e/o di equivalenza, per es.:

Nave has hyperonym plug-in *veicolo*
Navigazione eq-plug-in *navigazione*

Queste relazioni connettono un cluster gerarchico di termini (rappresentato dal suo *root node*) a un nodo del wordnet generico. In questo modo il termine, che è legato a un *synset* del database generico IWN come una specie di "punto di aggancio", viene visto (grazie al software di gestione del database)[4] con tutte le relazioni, sia quelle di tipo iperonimico in IWN, sia quelle di tipo gerarchico/verticale e orizzontale in Mariterm [21].

Le relazioni semantico-lessicali orizzontali di EWN-IWN che potrebbero essere usate come base per la creazione di sottotipi della relazione thesaurale associativa sono elencate nella tab. 3. Come già detto, queste relazioni sono usate per collegare non solo nomi o sintagmi nominali, come accade normalmente nei vocabolari controllati, ma anche verbi, aggettivi e avverbi.

Questa lista deve essere considerata come uno schema già sperimentato in diverse applicazioni, ma comunque parziale. Numerosi tipi di connessione classificabili come "associative" (es.: *disciplina/oggetto della disciplina)*, o relazioni attributive rilevanti da un punto di vista concettuale e potenzialmente utili anche per fini applicativi, non sono infatti inclusi in tale schema.

Le relazioni di EWN-IWN sono state preliminarmente classificate in due categorie: in relazione alla loro definizione su basi linguistiche oppure ontologiche, ed inoltre secondo il loro specifico contenuto semantico[5].

Nella tab. 3, nell'ultima colonna, vengono riportate le categorie concettuali fondamentali della TO che sono coinvolte nella formazione delle coppie di RT (1°: entità del 1° ordine; 2°: entità del 2° ordine; 3°: entità del 3° ordine).

4 I dati, le relazioni e tutte le informazioni riguardo a ogni termine sono visualizzate e aggiornate per mezzo di un software di gestione realizzato in Visual Basic, mentre i dati sono archiviati in un database SQL. Il tool di gestione permette una consultazione "integrata": cioè il termine è visualizzato con tutte le relazioni, sia quelle di tipo iperonimico in IWN, sia quelle di tipo gerarchico/verticale e orizzontale in Mariterm.

5 In alcuni casi la stessa relazione potrebbe essere considerata come definita sia su basi linguistiche sia su basi ontologiche. Per esempio, i termini *timone/timoniere* possono essere visti sia come connessi etimologicamente, sia come espressione della relazione *Co_Agent_Instrument/Co_Instrument_Agent*.

Tab. 3. Classificazione delle relazioni orizzontali di EWN-IWN in base al loro contenuto semantico

Relazioni raggruppate in base al contenuto semantico	**Esempi**	**Categorie**
LINGUISTICALLY BASED		
Antonymy	*imbarco/sbarco*	1°/1°, 2°/2°, 3°/3°
Derivation	*nave/navale*	All
ONTOLOGICALLY BASED		
<Attributive and related kinds>		
Has_Mero_Madeof/Has_Holo_Madeof (considerata come *relazione partitiva* in EWN-IWN)	*vela/canapa*	1°/1°
In_Manner/Is_Manner_For	*navigare/col vento in poppa*	2°/2°
Is_A_Value_Of/Has_Value	*tempesta (grado 10) /scala Beaufort*	2°/2°
<Related to condition>		
Be_In_State/State_Of	*mare/calmo*	1°/2°-2°/1°
<Causal and related kinds>		
Causes/Is_Caused_By	*collisione/naufragio*	2°/2°
Result_In/Is_Result_Of	*instabilità/capovolgimento*	2°/2°
For_Purpose_Of/Is_Purpose_of	*attracco/sbarco*	2°/2°
Is_Means_Of/Has_Means	*orientamento/navigazione*	2°/2°
<Involved/Role> (legano un evento con una entità coinvolta secondo il suo ruolo; la relazione sottospecificata è usata per i casi non chiari)		
Involved_Agent/Role_Agent	*stivaggio/stivatore*	2°/1°-1°/2°
Involved_Patient/Role_Patient	*naufragio/yacht*	2°/1°-1°/2°
Involved_Instrument/Role_Instrument	*posizione /sestante*	2°/1°-1°/2°
Involved_Location/Role_Location	*regata/golfo*	2°/1°-1°/2°
Involved_Direction/Role_Direction	*navigazione/porto*	2°/1°-1°/2°
Involved_Source_Direction/Role_Source_Direction	*sbarcare/stiva*	2°/1°-1°/2°
Involved_Target_Direction/Role_Target_Direction	*rotta/porto*	2°/1°-1°/2°
Involved_Result/Role_Result	*contratto/noleggio*	2°/1°-1°/2°
<Relazioni *Co_Role*> (definite in EWN e usate in IWN per codificare connessioni fra entità del 1° ordine che hanno un ruolo nella stessa situazione)		

Co_Agent_Patient/Co_Patient_Agent	*comandante/mozzo*	1°/1°
Relazioni raggruppate in base al contenuto semantico	**Esempi**	**Categorie**
Co_Agent_Instrument/Co_Instrument_Agent	*timone/timoniere*	1°/1°
Co_Agent_Result/Co_Result_Agent	*velaio/vela*	1°/1°
Co_Patient_Instrument/Co_Instrument_Patient	*carico/gru a carroponte*	1°/1°
Co_Patient_Result/Co_Result_Patient	*cima/nodo*	1°/1°
Co_Instrument_Result/Co_Result_Instrument	*barografo/diagramma della pressione*	1°/1°
<Others>		
Liable_To/Has_Liability	*navigabile/navigazione*	2°/2°
Pertains_To/Has_Pertained	*costiero/costa*	2°/2°
Fuzzynym	*ferryboat/orario*	all
Xpos_fuzzynym	*direzione/manovra*	2°/2°

3. Dipendenza dal dominio di conoscenza

Un accordo su come differenziare la relazione thesaurale associativa in sottotipi appare difficile da raggiungere. Alcuni tipi di relazione, come quella *causa/effetto*, vengono utilizzate di frequente nei thesauri (anche se non esplicitate come tali) e ricorrono spesso anche nelle differenti proposte di articolazione della relazione associativa. Tuttavia, un set di sub-relazioni che possa risultare valido in ogni contesto o circostanza sembra assai difficile da realizzare.

Questo accade per diversi motivi, incluso il fatto che la decisione su quale tipo di sub-relazione è utile includere in un sistema dovrebbe basarsi anche sulla sua capacità di contribuire ad aumentare *richiamo* e *precisione* nel recupero dell'informazione, nello specifico contesto operativo [22].

Nel presente contributo, tuttavia, analizzeremo il fatto che le relazioni di tipo associativo (inclusa la loro implementazione e l'eventuale suddivisione in sottotipi) sembrano dipendere dal dominio di conoscenza in cui sono considerate.

In effetti, la letteratura di ogni dominio è una fonte importante da cui derivare relazioni più specifiche, essendo essa una sorta di banca dati delle connessioni concettuali, comprese quelle che sono specifiche di quel dominio [1]. Tuttavia, la designazione delle relazioni di tipo associativo, che molto probabilmente costituiscono una classe intrinsecamente aperta, è un'operazione difficilmente eseguibile in modo univoco. Essa include una componente interpretativa ed è condizionata in diversi modi dalle caratteristiche dello specifico dominio in cui ci troviamo ad operare.

In primo luogo, in differenti domini possono essere richiesti differenti tipi di RT e gradi diversi di differenziazione della relazione standard [23]. In tal senso, le relazioni semantiche di EWN-IWN, elencate nella Tabella 3, devono essere considerate una base di partenza, da valutare e se è necessario da arricchire e completare per poter meglio rispondere ai bisogni informativi dei domini considerati.

In secondo luogo, il medesimo termine, se considerato a partire dalle prospettive di differenti domini di conoscenza, può essere coinvolto in differenti nessi concettuali. Il termine *cellula*, per esempio, può essere analizzato dal punto di vista della

biologia cellulare (una disciplina che riguarda soprattutto la descrizione delle componenti cellulari), dell'istologia (che pone l'attenzione a livello dei tessuti), o della biotecnologia (che è più interessata alle applicazioni industriali).

In terzo luogo, anche quando lo stesso tipo di relazione è applicabile in domini differenti, la sua implementazione può comunque portare a risultati non omogenei.

Le relazioni semantiche, incluse le relazioni associative, sono sempre sviluppate a partire dalla conoscenza del contenuto concettuale dei termini coinvolti. Tuttavia, ogni termine può essere descritto per mezzo di molteplici caratteristiche concettuali. A seconda di quali di queste caratteristiche vengono considerate rilevanti, e ciò può variare a seconda del dominio di conoscenza o del contesto, un termine può essere connesso con termini diversi, pur applicando lo stesso tipo o sottotipo di relazione [24].

Per esempio, la relazione causale, espressa in EWN-IWN come <*causes/is caused by*>, si riferisce a eventi che sono riconoscibili sul piano spazio-temporale (perché posti in successione) e sono connessi da una relazione causa-effetto. Tuttavia, applicando questa (sub-)relazione in domini diversi, può accadere che lo stesso termine, per esempio *anidride carbonica*, sia associato a differenti nessi causali, in base alle diverse prospettive suggerite dai domini considerati:

anidride carbonica <*causes*> inquinamento (dominio: Ambiente)
anidride carbonica <*causes*> effetti tossici (dominio: Salute)
anidride carbonica <*causes*> spegnimento del fuoco (dominio: Sicurezza)

4. Interoperabilità tra domini differenti

Se la situazione è quella descritta sopra, si pone il problema di come poter sviluppare un set di sub-relazioni che possa rispondere ai bisogni dei diversi domini di conoscenza, garantendo nel contempo un certo grado di interoperabilità tra differenti domini, così come la compatibilità con i thesauri esistenti, che non distinguono la relazioni associativa in base a sottotipi.

Come proposto da Tudhope *et al.* [23], una possibile soluzione potrebbe essere quella di mantenere ad un primo livello la relazione associativa standard come tale, applicare una limitata distinzione in sottotipi ad un secondo livello, e ad un terzo livello differenziare ulteriormente gli RT in base alle specifiche esigenze del dominio:

> One possible approach might be to aim for a limited extension of RTs at a second level and to expect domain specialisation at lower levels. This would permit some degree of interoperability in advanced thesaurus-based applications involving automated traversal of a richer set of relationships for term expansion. If the three standard thesaurus relationships formed the top level of a hierarchical structure then any such new applications would retain compatibility with the large number of existing thesauri (and indexed collections) where it is infeasible to augment the core relationships [23].

Tab. 4. Classificazione dei sottotipi di RT secondo Tudhope et al. [23]

RT (Plain) for undifferentiated associative relationships
Meaning connection
- Meaning overlap
- Distinguished from
- Antonyms
- conjucted terms

Causal (taken broadly)
dependency/requires
Uses
Product
patient
and possible spatial and temporal connections
Partitive (taken broadly)
constituent parts
Aggregate parts
Property/attribute

Questa soluzione potrebbe essere facilitata distinguendo il contenuto semantico dei sottotipi di RT dalla categoria concettuale a cui appartengono i termini coinvolti nella relazione.

Alcune (sub-)relazioni intercategoriali di secondo livello potrebbero essere applicabili in thesauri relativi a domini differenti. Gli stessi RT di secondo livello potrebbero a loro volta essere distinti in sottotipi più specifici, in base ai bisogni informativi specifici del settore di conoscenza considerato:

> in many systems, several subtypes of RT address various forms of relationships between categories represented by different facets, for example *Agent-Process*, *Process-Product*, *Material-Product,* etc. Rather than making an a priori distinction between intra and inter facet/hierarchy relationships as the basis for a classification of RT types, it may be useful to broaden our focus. To this end, we can distinguish the meaning of a relationship from the semantic category of the two concepts involved (...). If [...] the semantic category of a concept is taken as a dimension separate from the type of relationship then a smaller number of inter-facet RT relationships might suffice. The same *Causal*, *Uses/Requires*, *Spatial* or *Temporal* relationship might, at a high level, connect various categories of concepts. Conceivably, this might permit a restricted second level core set of RT subtypes to be applicable across some range of thesaurus domains (although this would need investigation). These relationships could themselves be refined into richer subtypes when the purposes of the thesaurus warranted [23].

Gli standard ISO fanno riferimento a categorie concettuali generali, come *Oggetti, Materiali, Eventi e Proprietà*. Queste categorie sono incluse nei thesauri, spesso in modo implicito: seppur non appaiono come tali, il contenuto semantico di un thesaurus è comunque concettualmente subordinabile ad esse. In alcuni casi, tali categorie sono invece identificate come *faccette* e utilizzate come strumenti di classificazione generale (come accade, per esempio, nell'*Art and Architecture Thesaurus* [25]).

Categorie concettuali fondamentali sono dettagliatamente rappresentate anche nella *Top Ontology* di EWN-IWN. Come si può vedere nella tab. 2, i termini connessi mediante relazioni di tipo associativo possono essere classificati secondo le categorie di 1° livello. Per incrementare il livello di specificità della classificazione è necessario, comunque, ricorrere a schemi supplementari. La gerarchia dei *Top Concepts* della *Top Ontology* di EWN-IWN appare, infatti, troppo complessa per poter essere utilizzata efficacemente a tali scopi.

Uno di questi schemi potrebbe essere il sistema di categorie previsto da Dahlberg [26], che consiste in una sorta di riproposizione delle categorie di Aristotele, ordinate secondo quattro super-categorie (o *Ur*-categories). Il secondo livello di questo schema sembra essere particolarmente adatto per la classificazione categoriale dei ter-

mini connessi mediante relazioni associative.

Tab. 5. Super-categorie e categorie secondo Dahlberg

Entities
Principles
Immaterial Objects
Material Objects
Properties
Quantities
Qualitites
Relations
Activities
Operations
Processes
States
Dimensions
Time
Place
Position

Come è anche dimostrato dall'esistenza di differenti proposte di differenziazione in sottotipi della relazione associativa, più problematico appare invece identificare un numero ristretto (stabile e applicabile in domini differenti) di classi per raggruppare i sottotipi di RT (inclusi quelli derivati dal modello EWN-IWN) di secondo livello. La classificazione riportata nella Tab. 3 potrebbe costituire, quindi, la base per un'analisi più approfondita.

Conclusioni

In questo contributo è stata analizzata la possibilità di differenziare la relazione associativa dei thesauri in un numero ristretto di sottotipi, così come le possibili implicazioni di tale studio.

Sebbene in modo preliminare, è stata valutata l'opportunità di usare un (sub-)set delle relazioni orizzontali incluse nel modello semantico di *EuroWordNet*, così come è stato applicato per la costruzione di *ItalWordNet* e per la strutturazione di *Mariterm*.

È stato sottolineato come la designazione degli RT dipenda anche dal dominio di conoscenza, nel senso che essa è condizionata in diversi modi dalle caratteristiche dello specifico dominio in cui ci si trova ad operare. Ciò pone il problema dell'interoperabilità fra domini diversi, che potrebbe essere affrontato mediante la strutturazione della relazione associativa a livelli differenti

RIFERIMENTI BIBLIOGRAFICI

[1] Elaine Svenonius. *The intellectual foundation of information organization*: Cambridge (MA): The MIT Press, 2000.

[2] Stella G. Dextre-Clarke. *Thesaural Relationships*. In: C. Bean – R. Green (Eds.). Relationships in the Organization of Knowledge. Dordrecht: Kluwer, 2001, p. 37-52.

[3] Fulvio Mazzocchi – Paolo Plini. *Thesaurus classification and relational structure: the EARTh experience*. In: *Terminology and Content Development, TKE 2005, Proceedings of the 7th International Conference on Terminology and Knowledge Engineering, 16th – 19th August 2005*, Bodil Nistrup Madsen, Hanne Erdman Thomsen (eds.). Copenhagen: Copenhagen Business School, 2005, p. 265-278.

[4] International organization for standardization. ISO 2788/1086 Documentation. *Guidelines for the establishment and development of monolingual thesauri*. Geneva: ISO, 1986, [trad. it. Ente nazionale italiano di unificazione. *ISO 2788/1993. Linee guida per la costruzione e lo sviluppo di thesauri monolingue*. Roma: UNI, 1993].

[5] International organization for standardization. *ISO 704. Terminology work-principles and methods.* Geneva: ISO, 2000.

[6] Dagobert Soergel. *Indexing languages and thesauri: construction and maintenance*. Los Angeles: Melville Publishing, 1974.

[7] Jacques Maniez. *Relationships in thesauri: some critical remarks*, «International Classification», 15 (1988), n. 3, p. 133-138.

[8] F. Wilfrid Lancaster. *Vocabulary control for information retrieval.* Arlington: Information Resources Press, 1986.

[9] Winfried Schmitz-Esser. *Thesaurus and beyond: an advanced formula for linguistic engineering and information retrieval*. «Knowledge Organization», 26 (1999), n. 1, p. 10-22.

[10] American Library Association. *Final Report to the ALCTS/CCS Subject Analysis Committee*. Subcommittee on Subject Relationships/ Reference Structures, 1997. <http://www.ala.org/ala/mgrps/divs/alcts/mgrps/ccs/cmtes/sac/inact/subjrel/index.cfm>.

[11] Anita Nuopponen, *Concept relations. an update of a concept relation classification*. In: *Terminology and Content Development, TKE 2005, Proceedings of the 7th International Conference on Terminology and Knowledge Engineering, 16th – 19th August 2005*, Bodil Nistrup Madsen, Hanne Erdman Thomsen (eds.). Copenhagen: Copenhagen Business School, 2005, p. 127-138.

[12] Piek Vossen. *EuroWordNet General Document,* 1999. <http://www.hum.uva.nl/~EWN>.

[13] Rita Marinelli – Giovanni Spadoni. *Some considerations in structuring a terminological knowledge base*. In: *Proceedings of the Third International WordNet Conference, GWC 2006, Seogwipo, Korea, January 22-26, 2006*, Petr Sojka – Key-Sun Choi – Christiane Fellbaum, Piek Vossen (eds.). Brno: Masaryk University, 2006. p. 217-224.

[14] George A. Miller – Richard Beckwith – Christiane Fellbaum [*et al.*]. *Introduction to WordNet: an on-line lexical database*. «International Journal of Lexicography», 3 (1990), n. 4, p. 235-244.

[15] George A. Miller. *WordNet: a lexical database for English*. «Communications of the ACM», 38 (1995), n. 11, p. 39-41.

[16] Rita Marinelli. *Proper names and polysemy: from a lexicographic experience*. In: *LREC 2004: Fourth International Conference on Language Resources and Evaluation, held in Memory of Antonio Zampolli. Lisbon, Portugal, 26-28 Maggio 2004*. Proceedings, Volume I. Paris: The European Language Resources Association, 2004, p. 157-160.

[17] Piek Vossen – Laura Bloksma – Horacio Rodriquez [*et al.*]. *The EuroWordNet base concepts*

and top ontology. EuroWordNet (LE-4003) Deliverable D017D034D036, University of Amsterdam, 1998.

[18] John Lyons. *Semantics*. Cambridge: Cambridge University Press, 1977.

[19] Rita Marinelli – Giovanni Spadoni. *Modeling a maritime domain ontology*. In: *ACTAS-1 of X Symposio Internacional de Comunicacion Social*, Leonel Ruiz Miyarez, Alex Munoz Alvarado and Celia Alvarez Moreno (eds.). Santiago de Cuba: Centro de Linguistica Aplicada, 2007. p. 511-515.

[20] *WordNet: an electronic lexical database*, Christiane Fellbaum ed. Cambridge, MA: MIT Press, 1998.

[21] Rita Marinelli – Adriana Roventini. *The Italian Maritime Lexicon and the ItalWordNet semantic database*. In: *Linguistics in the twenty first century*, Eloína Miyares Bermúdez and Leonel Ruiz Miyares eds., Cambridge: Scholar Press, 2006, p. 173-182.

[22] Birger Hjørland. *Semantics and knowledge organization*. «Annual Review of Information Science and Technology», 41 (2007), p. 367-405. Il riferimento è OK

[23] Douglas Tudhope – Harith Alani – Christopher Jones. *Augmenting thesaurus relationships: possibilities for retrieval*, «Journal of Digital Information» 1 (2001), n. 8, article no. 41, <http://journals.tdl.org/jodi/article/viewArticle/181/160>.

[24] Fulvio Mazzocchi – Melissa Tiberi – Barbara De Santis – Paolo Plini. *Relational semantics in thesauri: some remarks at theoretical and practical levels*. «Knowledge organization» 34 (2007), n. 4 , p.197-214.

[25] J. Paul Getty Trust. *Art and Architecture Thesaurus*. <http://www.getty.edu/research/tools/vocabulary/aat/>.

[26] Ingetraut Dahlberg. *Ontical structures and universal classification.* Bangalore: Sarada Rangathanan Endowment for Library Science, 1978.

[27] Dagobert Soergel. *The Art and Architecture Thesaurus (AAT): a critical appraisal*. «Visual Resource», 10 (1995), n. 4, p. 369-400.

[28] Fulvio Mazzocchi – Rita Marinelli – Melissa Tiberi. *Refining the thesaural associative relationship by applying the EuroWordNet semantic model*. In: *Managing ontologies and lexical resources, TKE 2008, Proceedings of the 8th International Conference on Terminology and Knowledge Engineering*, *19 – 20 August 2008,* Bodil Nistrup Madsen, Hanne Erdman Thomsen (eds.). Copenhagen: Copenhagen Business School, 2008, p. 61-77.

ABSTRACT Bollettino AIB, ISSN 1121-1490, vol. 50 n. 3 (settembre 2010), p. 249-263.

RITA MARINELLI, CNR-Istituto di linguistica computazionale, via Moruzzi 1, 56124 Pisa, e-mail rita.marinelli@ilc.cnr.it.

FULVIO MAZZOCCHI, CNR-Istituto dei sistemi complessi, via Salaria km 29,300, 00015 Monterotondo st. (RM), e-mail fulvio.mazzocchi@isc.cnr.it.

MELISSA TIBERI, collaboratrice a progetto della Biblioteca nazionale centrale di Firenze (BNCF), piazza Cavalleggeri 1, 50122 Firenze, e-mail: tiberim77@yahoo.it.

MARTA MOTTA, collaboratrice a progetto della Biblioteca nazionale centrale di Firenze (BNCF), piazza Cavalleggeri 1, 50122 Firenze, e-mail: motta.marta@gmail.com.
Ultima consultazione siti web: 25 luglio 2010.

Il modello semantico di EuroWordNet come strumento per la strutturazione della relazione associativa nei thesauri

I thesauri sono strumenti che organizzano semanticamente un dominio di conoscenza per fini applicativi. Attraverso la loro struttura relazionale vengono stabiliti nessi tra termini con significati correlati. La semantica relazionale di un thesaurus è uno strumento di supporto fondamentale per il recupero dell'informazione, attraverso cui vengono aumentati il richiamo (*recall*) e la precisione (*precision*) della ricerca. La rete delle relazioni thesaurali svolge, infatti, una funzione semantica importante, fornendo una rappresentazione del significato di ciascun termine contenuto nel thesaurus e realizzando il prototipo di mappa della struttura concettuale del dominio di conoscenza. Il formato tradizionale di un thesaurus, così come è descritto negli standard internazionali, include tre relazioni fondamentali (relazione gerarchica, relazione associativa e relazione di equivalenza).
È opinione diffusa che, per poter meglio rispondere ai bisogni attuali in ambito di organizzazione dell'informazione, questo formato debba essere in qualche modo riconsiderato e perfezionato. In questo contributo viene analizzata la possibilità di differenziare la relazione associativa in un numero ristretto di sottotipi. È stata preliminarmente valutata a tal fine la possibilità di utilizzare una serie di relazioni incluse nel modello semantico di EuroWordNet (EWN), così come è stato applicato in una delle sue versioni nazionali, ItalWordNet (IWN), nell'ambito del progetto riguardante la terminologia del settore marittimo (Mariterm). Viene, inoltre, preso in considerazione il modo in cui le operazioni di attribuzione e di articolazione della relazione associativa sembrano essere condizionate dalle caratteristiche del dominio di conoscenza in cui sono effettuate.

The semantic model of EuroWordNet as a tool for structuring the associative relation in thesaurai

Thesauri are tools which semantically organize a domain of knowledge for operational purposes. Their relational semantics is concerned with methods that connect terms with related meanings and it is important to support information retrieval, enhancing the information recall performance and contributing to improve precision. In fact, the network of relations of a thesaurus has an important semantic function, providing a representation of the meaning of each thesaurus term and a map of the conceptual structure of a subject area.
The traditional thesaurus format - as described in international standards - includes the hierarchical, associative and equivalence relationships. However, a rather widespread opinion is that this format should be refined, in order to cope with the current needs of information organization. This paper discusses the possibility of refining the associative relation into a number of sub-kinds by adopting the semantic model of EuroWordNet (EWN), as it was used, according to one of its national versions, ItalWordNet (IWN), taking into account the terminological database Mariterm, which contains terms belonging to the maritime domain. It is also stressed how RT designation and refinement appear to be domain dependent, in the sense that they are associated with the specific features of a knowledge field.