

Costruzione del *Thesaurus*, analisi a faccette, categorie¹

Manuela Grillo

Collaboratrice del progetto

Termini contenuti nel Prototipo

I termini presenti nel Prototipo si possono distinguere, sulla base delle modalità di raccolta, in due gruppi:

1. termini della lista di partenza, o Elenco di base, in cui sono stati cumulati:
 - termini degli *Aggiornamenti BNI 1956-1998* (pubblicati);
 - termini introdotti dalla BNI dal 1999 al 2005 (mai pubblicati);
 - termini di biblioteconomia, bibliografia, ecc., individuati nel *Soggettario* e *Aggiornamenti BNI*;
 - termini provenienti dal *Soggettario* e dagli *Aggiornamenti BNI* ritenuti particolarmente obsoleti e bisognosi di revisione o, comunque, interessanti nel quadro dell'elaborazione del Prototipo;
2. termini aggiunti in corso di elaborazione del Prototipo:
 - termini introdotti per la strutturazione dei termini dell'Elenco di base (ossia tutti i termini sovraordinati o associati a ciascun termine dell'Elenco di base);
 - termini derivanti dalla scomposizione di termini composti;
 - termini impiegati nella sperimentazione della sintassi.

Metodo di costruzione della gerarchia

Nella costruzione del vocabolario del *Nuovo soggettario* sono stati usati in combinazione due diversi metodi di costruzione delle gerarchie: il metodo induttivo (dal basso verso l'alto) e il metodo deduttivo (dall'alto verso il basso). In una prima fase le gerarchie sono state costruite a partire da ciascun termine dell'Elenco di base, individuando i suoi termini sovraordinati fino al Top term, con metodo induttivo, escludendo quindi la costruzione delle relazioni di un termine con termini più specifici, limitando l'espansione della rete verso il basso. Tuttavia a questo processo si è affiancata la preparazione di una macrostruttura classificatoria, costituita da categorie e faccette, a loro volta strutturate in sottofaccette, individuate e modificate via via che nuovi termini e le relative gerarchie venivano accolti all'interno di una classe. Dunque il metodo induttivo e deduttivo sono due modalità distinte e complementari della costruzione del vocabolario.

Le relazioni che si stabiliscono nella strutturazione di un termine costituiscono la sua rete semantica, laddove i termini rappresentano i nodi della rete e le relazioni semantiche i fili che li collegano. Il modello di rete (fig. 1 e 2) è costituito da:

1. il termine di partenza;
 2. i suoi termini sovraordinati, fino al Top term (relazioni gerarchiche BT/NT);
 3. i suoi termini associati (relazioni associative RT);
 4. per ciascun termine associato, i suoi termini sovraordinati, fino al Top term (relazioni gerarchiche).
- Si avrà quindi la strutturazione completa del termine di partenza del quale si stabiliscono tutte le relazioni semantiche, tranne quelle con i termini subordinati.

Fig. 1– Modello di rete semantica

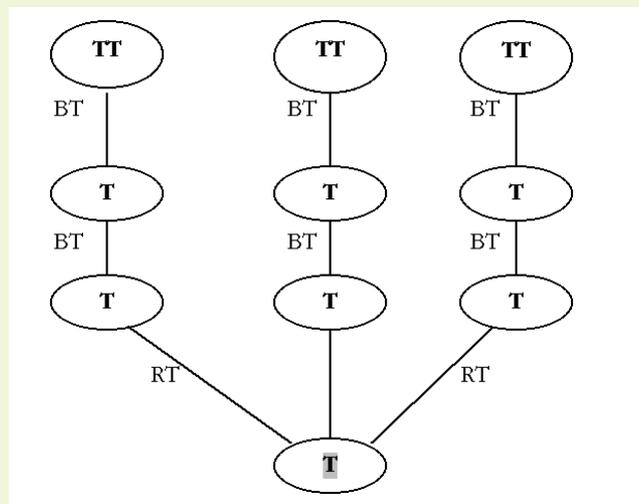
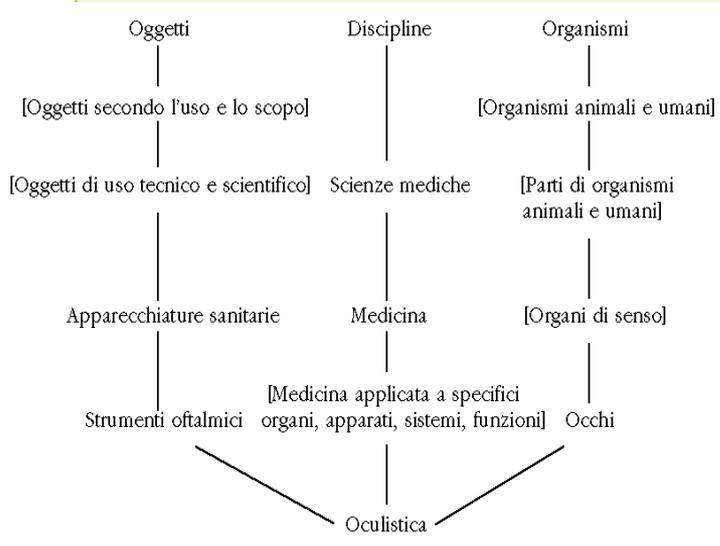


Fig. 2– Esempio di rete semantica di un termine



Il modello di rete non prevede orfani, ossia termini non legati al proprio termine sovraordinato; tuttavia nella fase iniziale di costruzione del Prototipo è stato adottato un modello semplificato, più economico, denominato modello ad àncora (fig. 3 e 4), per la somiglianza con la forma di un'àncora.

Fig. 3 – Modello di rete “ad àncora”

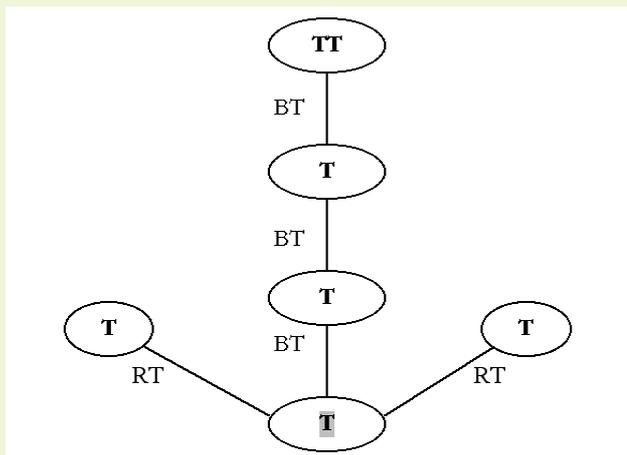
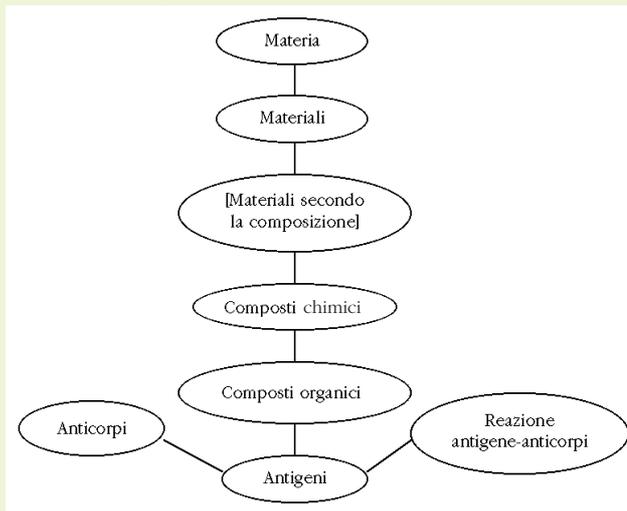


Fig. 4 – Esempio di rete semantica “ad àncora”



I criteri che hanno la funzione di controllare il numero delle relazioni stabilite nella strutturazione di un termine sono:

- *criterio di modulazione*: nella costruzione di una gerarchia non si dovrebbe omettere nessuna classe intermedia tra la più specifica e la più generale;
- *criterio definitorio*: la relazione associativa, pur essendo reciproca, viene attivata a partire dal termine la cui definizione ha come componente l'altro termine (esempio: a partire da Veterinaria è attivato RT ad Animali e non viceversa).

Operazione fondamentale e preliminare per la strutturazione del vocabolario è l'*analisi categoriale* o categorizzazione dei termini. Le categorie e le faccette infatti assolvono a tre funzioni essenziali:

- a) definire le strutture classificatorie e le loro articolazioni;
- b) permettere la suddivisione del vocabolario in classi più piccole, quindi più maneggevoli, facilitando la costruzione della struttura semantica;
- c) guidare l'analisi in alcune procedure di controllo terminologico, come la scelta tra forma singolare o plurale di un

termine (ad esempio, il termine Gestione verrà preferito nella forma singolare perché appartenente alla categoria Azioni, faccetta Attività, tendenzialmente singolare).

Le categorie e le faccette sono le classi generali in cui si suddivide il vocabolario: ogni termine appartiene a una – e a una sola – categoria all'interno della quale si stabiliscono le sue relazioni gerarchiche. Vi sono casi di sovrapposizione che si risolvono con gli strumenti della classificazione: uso della poligerarchia o della relazione associativa, criteri di precedenza tra le faccette, ecc. Ad esempio, il termine Eroi potrebbe essere attribuito alla faccetta Persone (Agenti) o, più correttamente, a quella di Forme (Cose), in quanto esseri semidivini ai quali si attribuiscono gesta prodigiose, quindi Esseri immaginari.

Profilo delle categorie e delle faccette

La macrostruttura del vocabolario è definita:

– dalle *categorie*: costituiscono i quattro maggiori raggruppamenti in cui sono organizzate le classi: Agenti, Azioni, Cose, Tempo; non coincidono con i Top term, ad eccezione della categoria Tempo; definiscono il contesto delle possibili gerarchie;

– dalle *faccette*: sono le classi generali che individuano caratteristiche condivise da un gruppo di termini; corrispondono ai Top term;

– dalle *sottofaccette*: sono le specifiche sottoclassi; possono essere introdotte da termini preferiti o da etichette di nodo; possono prevedere ulteriori articolazioni, costituite anch'esse da termini preferiti o da etichette di nodo.

Le connessioni e le eventuali sovrapposizioni tra le varie categorie e faccette sono gestite attraverso relazioni semantiche, sia gerarchiche che associative.

I nomi delle categorie hanno un carattere intuitivo, ossia definiscono in generale il contenuto delle categorie, indicando in modo sintetico ed intuitivo la caratteristica definitoria condivisa da ogni termine che appartiene a ciascuna categoria.

Vediamo nel dettaglio le categorie e le loro faccette (fig. 5).

Fig. 5 – Schema delle categorie e delle loro faccette

Categorie	Faccette
Agenti	Organismi Organizzazioni Persone e gruppi
Azioni	Attività Discipline Processi
Cose	Forme Materia Oggetti Spazio Strumenti Strutture
Tempo	Tempo

Agenti: contiene tutti i concetti di entità che presentano la caratteristica di poter essere agenti di azioni, singolarmente o collettivamente. Comprende gli organismi (contiene tre tipi di concetti: gli organismi viventi, sia animali che vegetali; le loro parti, vale a dire l'anatomia degli organismi viventi; i raggruppamenti di esseri viventi, come Greggi, Mandrie, ecc.); le organizzazioni (raccolge i concetti di organizzazioni o parti di organizzazione, cioè qualunque complesso organizzato, in maniera più o meno formale, di risorse e persone, dotato di regole per il conseguimento di uno scopo, sia esso economico, politico, sociale, ecc.); le persone e i gruppi (contiene tutti quei concetti che rappresentano insiemi di persone che non danno luogo ad alcuna organizzazione, né permanente né temporanea).

Azioni: contiene tutti i concetti che rappresentano tipologie dell'agire. Comprende le attività (contiene tutti quei termini che rappresentano azioni dotate di un agente diretto chiaramente individuabile); i processi (contiene azioni che non sono provocate da un singolo agente – persona o organizzazione – ma da una causa, naturale o sociale, oppure da azioni o movimenti collettivi); le discipline (azioni che presentano come caratteristiche principali una forte componente di ricerca nella loro attività, una finalità di accumulo e diffusione di conoscenze, una struttura organizzativa più o meno formalizzata che fa da supporto alle attività).

Cose: contiene i concetti che si riferiscono sia a entità concrete, tangibili (materia, oggetti, strutture) sia a entità astratte (strumenti e forme). Comprende quindi la materia (intesa nel senso più astratto di struttura, proprietà, stato fisico e i materiali, cioè gli elementi, le sostanze considerate in base alla composizione chimica o fisica e tutti quei materiali che hanno subito una qualche trasformazione ma che non sono ancora oggetti); gli oggetti (comprende ciò che ha una forma definita, tangibile, finita, anche di dimensioni molto grandi, purché almeno potenzialmente mobili); lo spazio (contiene i termini che individuano concetti di estensione, ma non le entità che occupano tale estensione, vale a dire che lo spazio deve essere l'elemento fondamentale e non deve essere sovrapposto ad altri aspetti, che diventano, eventualmente, quelli predominanti nell'analisi categoriale); gli strumenti (comprende cose, elementi astratti che costituiscono il mezzo per ottenere qualcosa, non comprende tutti i concetti che hanno funzione strumentale come gli oggetti, ma solo gli strumenti astratti o quelli la cui componente oggettiva è secondaria); le forme (comprende l'aspetto esteriore, la rappresentazione, la manifestazione di entità oggettuali, astratte o frutto della creatività e della fantasia, comprese le forme della comunicazione); le strutture (comprende tutti quei manufatti che occupano uno spazio in maniera stabile, che non sono cioè mobili, anche se l'occupazione può essere temporanea come nel caso dei Cantieri).

Tempo: la categoria è individuata da uno svolgimento cronologico come elemento principale o esclusivo dei termini che vi appartengono.

L'analisi a faccette

L'*analisi a faccette*² è l'operazione di analizzare e classificare i termini per faccette. La *faccetta* è un *principio di divisione* di una categoria o di una classe, ossia una *caratteristica* che guida l'attività di raggruppare e separare (classificare) i concetti. Una classe si dice organizzata per faccette quando i termini che ne fanno parte sono raggruppati secondo caratteristiche di divisione proprie della classe.

L'analisi a faccette applicata ad un vocabolario controllato è lo strumento attraverso il quale tutti i termini subordinati ad una determinata classe sono organizzati secondo le diverse caratteristiche di divisione della classe, garantendo così lo sviluppo ordinato delle gerarchie semantiche e la loro coerenza con i principi della gerarchia. L'analisi a faccette, che opera in modo deduttivo, dal generale al particolare, è lo strumento attraverso cui si completa, si razionalizza e si rende coerente la costruzione dal basso. Inoltre essa permette di stabilire un ordine di precedenza tra le faccette o sottofaccette ai fini della scelta della gerarchia "prevalente" di un termine potenzialmente poligerarchico, evitando così la proliferazione delle poligerarchie.

Tre regole fondamentali guidano l'analisi a faccette: la prima regola prescrive che la suddivisione di una classe debba avvenire applicando *una sola caratteristica alla volta*. Ad esempio, se si sta procedendo alla divisione delle Organizzazioni in base al settore di attività, tutte le altre caratteristiche devono passare in secondo piano; solo quando l'operazione di divisione sarà stata completata, si potrà applicare un altro principio di divisione.

La seconda regola prescrive di stabilire l'*ordine nel quale si applicano le caratteristiche di divisione*, ovvero, in un thesaurus come il *Nuovo soggettario*, stabilire quale sia l'ordine di precedenza che guida nella scelta della gerarchia prevalente, nel caso di termini potenzialmente poligerarchici.

La terza regola definisce i termini che appartengono alla stessa sottofaccetta come *mutualmente esclusivi*, ossia i termini che appartengono ad una stessa sottofaccetta non possono dar luogo a sovrapposizione o intersezione di classe (p. e., Coniugi e Divorziati all'interno dell'etichetta di nodo [Persone secondo lo stato civile e la parentela]).

Le etichette di nodo e la loro elaborazione

Le etichette di nodo sono termini o locuzioni la cui funzione è quella di indicare una caratteristica di divisione. Sono a tutti gli effetti termini di gerarchia che hanno sia BT che NT, ma non possono essere usati come termini di indicizzazione.

L'elaborazione delle etichette di nodo è stata operata a partire dai livelli più alti della gerarchia, concentrando l'attenzione sul primo livello e via via scendendo di grado, a seconda della necessità.

Esempio: nel primo livello gerarchico della categoria Or-

ganizzazioni si hanno i seguenti termini di gerarchia ed etichette di nodo (queste ultime indicate tra parentesi quadre):

- *[Organizzazioni secondo la forma]
- *[Organizzazioni di particolari categorie di persone]
- *[Parti di organizzazioni]
- *[Organizzazioni illegali]
- *Organizzazioni temporanee
- *Organizzazioni istituzionali
- *[Organizzazioni in base al settore di attività]

Solamente [Organizzazioni in base al settore di attività] è ulteriormente sfaccettata con l'ausilio di termini di gerarchia ed etichette di nodo, precisamente:

- **Organizzazioni religiose
- **Organizzazioni culturali
- **[Organizzazioni a carattere sociale]
- **[Organizzazioni ricreative e sportive]
- **[Organizzazioni delle arti e dello spettacolo]
- **Organizzazioni educative
- **Organizzazioni economiche
- **Organizzazioni militari
- **[Organizzazioni a carattere sanitario]
- **[Organizzazioni scientifiche]

Ciò perché l'analisi a faccette nel vocabolario del *Nuovo soggettario*, pur essendo sempre presente come riferimento metodologico generale, non è stata applicata sempre in modo sistematico: le etichette di nodo vengono introdotte il più delle volte empiricamente, quando è necessario cioè dare un ordine più chiaro e leggibile alla struttura gerarchica.

Le diverse categorie e faccette non presentano le stesse dimensioni per numero di termini né la stessa profondità per grado di specificità dei termini; inoltre la modalità di costruzione dal basso, interagendo con la modalità dall'alto, determina un processo continuo di revisione delle gerarchie, in cui l'applicazione dell'analisi a faccette avviene per successive fasi di sviluppo del vocabolario. Quindi l'analisi a faccette potrà essere utilmente applicata anche per gradi, quando l'eterogeneità semantica di un gruppo significativo di termini di una certa classe evidenzierà l'opportunità di una migliore organizzazione della classe stessa. Ad esempio, mentre per le Persone è stata applicata un'analisi a faccette assai approfondita, per i Processi si è messo in atto un procedimento molto meno analitico, in ragione della minore necessità di articolare nel dettaglio, allo stato attuale, il vocabolario relativo a questo ambito.

Quando si suddivide una faccetta in sottofaccette, può essere opportuno creare un profilo anche per le sottofaccette, introdotte talvolta da etichette di nodo, soprattutto quando, in presenza di caratteristiche sovrapponibili, si vuole indicare l'ordine di precedenza stabilito. Ad esempio, per i termini che appartengono sia alla classe identificata dall'etichetta [Persone secondo le attività] sia alla classe [Persone secondo il comportamento], si costruisce la gerarchia all'interno della prima e si esprime, se op-

portuno, il residuo gerarchico con la relazione associativa (esempio: Sacerdoti cattolici avrà come BT solamente Sacerdoti e non anche Cattolici).

L'ordine di preferenza stabilito all'interno della categoria Organizzazioni è il seguente: i termini sono normalmente ordinati secondo gerarchie che esprimono il settore o l'ambito di attività e non secondo gerarchie riferibili alla forma dell'organizzazione. Ciò ha portato a intervenire su alcune gerarchie, inizialmente elaborate sulla base di criteri non omogenei.

Esempio: Associazioni + aggettivo che esprime ambito di attività. La gerarchia di questi termini era stata costruita sulla base della loro forma, cioè Associazioni pacifiste, ecologiste, militari, politiche, religiose, culturali aveva come BT il termine strutturato Associazioni. Quando si è preferito l'altro criterio di costruzione delle gerarchie, i termini Associazioni + aggettivo sono stati spostati sotto etichette di nodo esprimenti l'ambito di attività.

Conseguenze della scelta della poligerarchia temperata sulle gerarchie preesistenti

Inizialmente la costruzione delle gerarchie è stata condotta in modo molto dettagliato, inserendo il termine controllato sui repertori, indipendentemente dalla garanzia letteraria, ed attribuendo ad un termine anche più BT (poligerarchia). La conseguenza di questo *modus operandi* è stata la proliferazione di gerarchie, particolarmente intricate e ridondanti, problema divenuto evidente con il processo deduttivo di revisione delle gerarchie dall'alto. In linea con il procedimento sperimentale che ha animato tutte le fasi della redazione del Prototipo, una volta esplosa la questione delle poligerarchie pressoché ingestibili, si è cercato di individuare una soluzione imperniata su due punti cardine: ricorso alla *poligerarchia temperata* (ossia applicata con moderazione, entro determinati limiti e con particolari cautele. Questo argomento verrà illustrato più approfonditamente nell'intervento successivo) e maggiore *attenzione alla garanzia letteraria*.

La scelta della poligerarchia temperata ha comportato la necessità di smontare numerose poligerarchie, con la conseguente necessità di trasformazione di uno dei due termini che fungevano da BT (in particolare il termine che esprime il residuo poligerarchico). Ciò non è avvenuto ovviamente nel caso in cui il termine in questione fosse un termine strutturato, pertanto la modifica ha riguardato solamente la relazione, trasformata da BT ad RT. Nel caso di un termine introdotto per la gerarchia ma non completamente strutturato Status di lavorazione MIN, i criteri – opportunamente specificati nel file “Decisioni generali cumulate” – sono stati quelli indicati nello schema A.

Generalmente i termini hanno subito una trasformazione da Termine di gerarchia (S), status di lavorazione MIN, a Termine strutturato, status di lavorazione MED, quando la loro fonte erano in primo luogo gli strumenti.

Esaminando il termine Architettura gotica nelle due fasi, precedente (fig. 6) e successiva (fig. 7) alla decisione sulla poligerarchia temperata, vediamo che la gerarchia è sta-

Schema A

[Poligerarchie: eliminazione in fase di revisione]	<p>BT che diventano RT Quando si smontano poligerarchie i termini che erano BT, ora non più tali, vengono trasformati in RT e le possibilità sono due:</p> <p>il termine diventa: Tipo di termine: <i>Termine senza BT</i>; Status lavorazione: BASE. Vengono sganciati eventuali termini di gerarchia precedenti,</p> <p>oppure</p> <p>il termine diventa: Tipo di termine: <i>Termine con relazioni</i>; Status lavorazione: MED. Vengono conservate tutte le sue relazioni gerarchiche e vengono attivate le eventuali relazioni associative.</p> <p>Valutare di volta in volta la soluzione migliore.</p>
--	--

ta semplificata e sono state introdotte le etichette di nodo ritenute opportune. Architettura gotica ha un solo BT, BT [Architettura secondo lo stile, il movimento artistico]), e il residuo poligerarchico (RT Arte gotica) è stato espresso attraverso la relazione di RT. Il termine di gerarchia (S) Arte gotica è stato strutturato, sono state quindi conservate le sue relazioni gerarchiche ed attivate le relazioni associative pertinenti.

In alcuni casi, si è avuta la necessità di eliminare dei termini di gerarchia la cui relazione con il termine di par-

Fig. 6 – Termine Architettura gotica nella prima fase

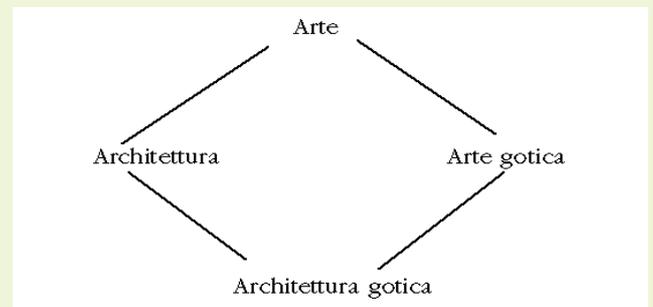
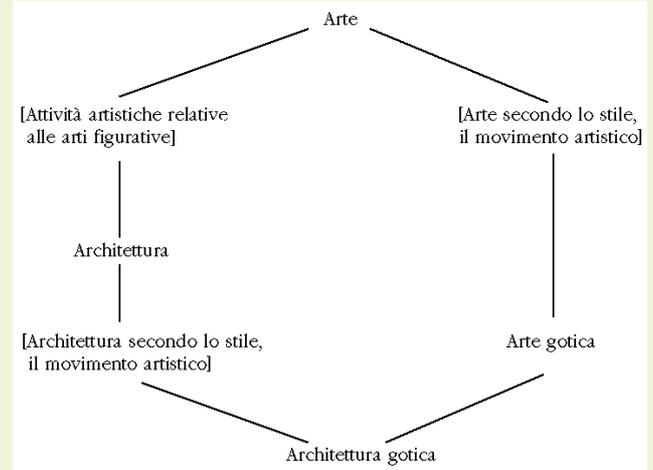


Fig. 7 – Termine Architettura gotica nella seconda fase



tenza non poteva essere trasformata in RT. Ciò è avvenuto quando sono stati inseriti più termini di gerarchia, uno sopra all'altro: in questo caso, per il rispetto del modello ad àncora, solo il primo partendo dal basso può dare luogo al legame di RT.

Un valido esempio è la strutturazione del termine Miniatura ottoniana (fig. 8 e 9): la gerarchia è stata in seguito significativamente snellita e sono state introdotte le etichette di nodo ritenute opportune.

Fig. 8 – Termine Miniatura ottoniana nella prima fase

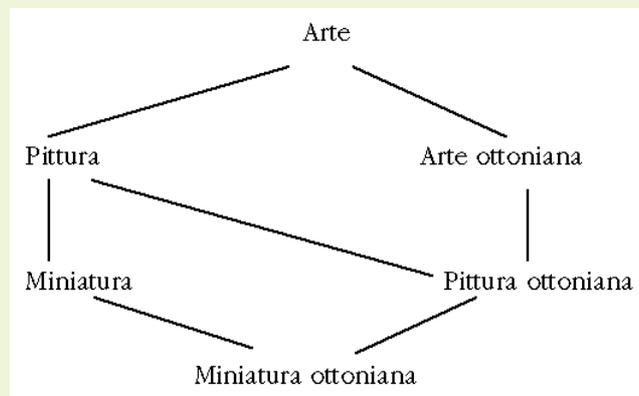
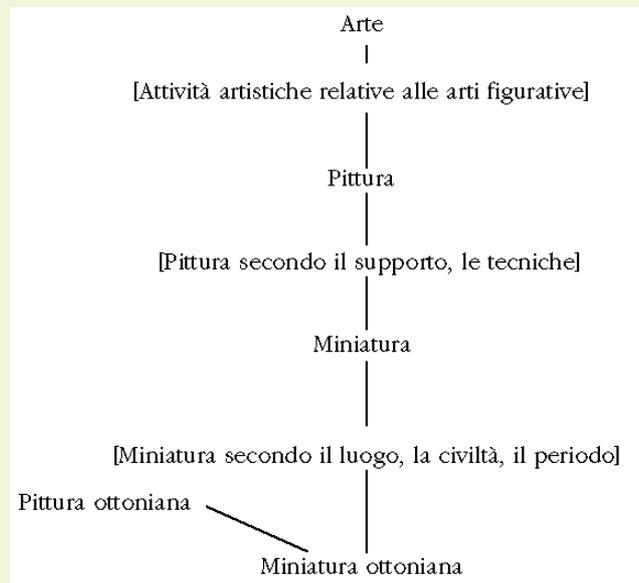


Fig. 9 – Termine Miniatura ottoniana nella seconda fase



Miniatura ottoniana ha una sola BT, [Miniatura secondo il luogo, la civiltà, il periodo], e il residuo poligerarchico (pittura ottoniana) viene espresso attraverso una relazione di RT. Il termine di gerarchia Pittura ottoniana era a sua volta poligerarchico, ma le due relazioni di BT sono state eliminate; nel caso del BT Pittura ciò è avvenuto senza conseguenze, poiché il termine era già strutturato, mentre l'altro BT Arte ottoniana è stato di necessità elimi-

nato, poiché il modello ad àncora prevede relazioni di RT solo per i termini completamente strutturati.

Questi ed altri cambiamenti sono stati operati durante tutta la realizzazione di questo nuovo strumento, in linea con il procedimento sperimentale che ha animato tutte le fasi della redazione del Prototipo. ■

Note

¹ Il testo, che riproduce l'intervento letto in occasione della presentazione del *Nuovo soggettario*, è stato in alcune parti lievemente modificato e ampliato per esigenze di chiarezza e approfondimento. Si è tuttavia cercato di mantenere il taglio discorsivo e non eccessivamente analitico richiesto dall'occasione. La fonte principale dell'intervento è BIBLIOTECA NAZIONALE CENTRALE DI FIRENZE, *Nuovo soggettario. Guida al sistema italiano di indicizzazione per soggetto. Prototipo del Thesaurus*, Milano, Editrice Bibliografica, 2006. Un sentito ringraziamento va ad Alberto Cheti per il suo aiuto e per le sue sempre preziose indicazioni.

² Il termine "faccetta" costituisce la trasposizione italiana dell'inglese *facet*, il quale contiene in sé i concetti di "aspetto, lato, fase"; nello standard BS 8723 l'analisi a faccette è definita come: "analysis of complex subjects or *concepts* into constituent *concepts* grouped into *facets*"; mentre la faccetta è definita come: "high-level grouping of *concepts* of the same inherent category [...] Examples of categories that may be used for grouping *concepts* into *facets* are: activities, disciplines, people, materials, objects, places and times". L'idea di esprimere la classe di un documento per combinazione di *faccette*, vale a dire di diversi aspetti del suo contenuto, è dovuta all'indiano Shiyali Ramamrita Ranganathan, che elaborò uno schema di classificazione, noto con il nome *Colon classification*, basato sul principio delle faccette. Ranganathan, nella sua opera *Prolegomena to library classification*, edita nel 1967, definisce *facet* come un termine generico utilizzato per denotare qualsiasi componente di un soggetto composto. L'analisi a faccette consiste nella scomposizione di un soggetto in diverse coordinate, ciascuna delle quali esprime un aspetto; essa consiste essenzialmente nell'analisi di concetti composti nelle loro parti semplici, dette isolati. Ogni isolato può essere espresso da una notazione derivante da un'apposita tavola: in questo modo i diversi frammenti di notazione vengono quindi ricomposti, sintetizzati, secondo un ordine di citazione stabilito. Pertanto il metodo dell'analisi a faccette è detto anche analitico-sintetico. Sull'argomento si veda CLAUDIO GNOLI, *Classificazione a faccette*, Roma, AIB, 2004. La teoria dell'analisi a faccette, nata nell'ambito della classificazione bibliografica, è stata applicata da tempo con successo anche a strumenti di indicizzazione semantica quali thesauri e soggettari. Attualmente, analizzando gli standard e i principi esistenti, la norma ISO 2788 si limita ad elencare le possibili e principali "categorie" di riferimento nella strutturazione di un vocabolario controllato, senza dare però indicazioni circa l'effettiva operazione di analisi, mentre lo standard BS 8723, che dedica uno spazio più ampio alla definizione delle varie categorie e faccette applicabili ad un vocabolario controllato, fornisce, in un breve paragrafo, esempi pratici di applicazione di analisi a faccette al fine di illustrare come le varie classi e sottoclassi individuabili all'interno delle faccette scelte, permettano di pervenire a procedure uniformi di gerarchizzazione, assicurando correttezza formale e strutturale alle relazioni in tal modo stabilite.