

La struttura del Thesaurus: monogerarchie e poligerarchie

Marta Motta

Collaboratrice del progetto

“Il thesaurus è uno strumento basilare per trasformare le idee in parole”.

Roget's International Thesaurus, 1977 (1. ed. 1852)

La struttura generale del *Nuovo soggettario* è prevalentemente monogerarchica, ma con una apertura verso soluzioni poligerarchiche, ossia ammette che alcuni termini possano avere più di un termine sovraordinato.

La relazione di base che distingue tipicamente un thesaurus da una semplice lista di termini non strutturati, e ne rivela le fondamenta sistematiche, è la relazione classificatoria per eccellenza, ossia la *relazione gerarchica*, come previsto dagli standard sui thesauri. Essa esprime il concetto ed il grado di subordinazione o sovraordinazione fra termini appartenenti allo stesso albero; in questo albero, il termine sovraordinato rappresenta una classe o un tutto, e il termine subordinato rappresenta un suo elemento o parte.

È importante ricordare che la relazione gerarchica si può averare solo tra termini che fanno riferimento a concetti appartenenti alla stessa categoria. Per esempio, un termine che rappresenta un concetto di attività non può avere come BT un termine che rappresenta un concetto di materia. Come è noto, possiamo avere tre tipi di relazione gerarchica:

1. la *relazione generica*, che stabilisce una relazione tra un genere e le sue specie, per es. Pompelmi BT Agrumi;
2. la *relazione parte-tutto*, per es. Stomaco BT Apparato digerente;
3. la *relazione esemplificativa*, che rappresenta il legame esistente tra una classe, espressa da un nome comune, e un esempio individuale di questa classe, espresso da un nome proprio: per es. Natale BT Feste cristiane.

La struttura di un thesaurus può essere monogerarchica o poligerarchica. Mentre nei thesauri poligerarchici un termine può appartenere a più di una gerarchia, quindi avere più di un termine sovraordinato, nei thesauri monogerarchici ogni termine può avere uno ed un solo BT, ossia può essere inserito in una sola catena gerarchica. La differenza principale, quindi, è che in un thesaurus poligerarchico siamo in presenza di un maggior numero di alberi e di una maggiore complessità strutturale rispetto ad un thesaurus monogerarchico.

Un breve esempio:

THESAURUS POLIGERARCHICO

Pianoforti

BT1 Strumenti a corda
BT1 Strumenti a tastiera
BT2 Strumenti musicali

THESAURUS MONOGERARCHICO

Pianoforti

BT1 Strumenti a corda
BT2 Strumenti musicali
RT Strumenti a tastiera

Per quanto riguarda il *Nuovo soggettario*, sostanzialmente un thesaurus monogerarchico, si è deciso di non escludere la creazione di poligerarchie per alcuni termini. Si dovevano stabilire dei criteri per assumere la poligerarchia con moderazione, per “temperarla” come era già stato ipotizzato dallo studio di fattibilità.

I principi cardine che si devono seguire nella eventuale creazione di una gerarchia sono:

1. l'*appartenenza dei termini alla stessa categoria*, ossia non si crea una poligerarchia fra **Astronomia** e **Stelle**, in quanto il primo termine appartiene alla categoria Discipline e il secondo alla categoria Spazio (fra i due termini, ovviamente, ci sarà una relazione associativa);
2. il criterio della *relazione quasi generica*, ossia non ci può essere una relazione di BT quando il legame tra il termine più specifico e ciascun termine più generale non risulta valido al test “*tutti e alcuni vs. alcuni e alcuni*”, per es. non si può creare un legame di BT fra **Cani** e **Animali domestici**.

Cani

BT Canidi
RT Animali domestici

Nonostante questi criteri generali, all'inizio del lavoro si è verificata una sorta di “esplosione” dell'uso della poligerarchia, che in quel momento sembrava un arricchimento della struttura del Thesaurus. Le poligerarchie creavano problemi di gestione del Thesaurus: si costruivano delle gerarchie complicate e intrecciate, piene di ramificazioni che risultavano poco leggibili. Inoltre si verificavano problemi di economia del lavoro in prospettiva dell'implementazione futura del Thesaurus.

Ci si è resi conto di queste difficoltà nella seconda fase del lavoro, quando è stata abbandonata l'ottica della costruzione dal basso verso l'alto e si è iniziato a controllare la struttura del Thesaurus a partire dai suoi livelli apicali.

Nel controllo finale si è deciso di smontare molte poligerarchie presenti, trasformando la relazione gerarchica “soccumbente” in relazione associativa. Si è scelto, quindi, di assumere una *poligerarchia ulteriormente temperata*, rispetto a quanto era stato ipotizzato durante la stesura dello studio di fattibilità.

Nella strutturazione dei termini del Thesaurus, e in particolare nella scelta del BT, in presenza di una possibile poligerarchia abbiamo elaborato alcune linee guida da seguire nell'ordine:

1. accertare l'ammissibilità di ciascuna gerarchia, rifiutando le relazioni che non rispondono ai criteri indicati precedentemente;
2. tenuto conto che ad ogni termine è assegnata una categoria/faccetta, verificare l'eventuale “prevalenza” di

una delle gerarchie ammissibili rispetto alle altre, sulla base di un ordine di precedenza stabilito tra le faccette cui appartengono i termini.

Un'altra delle attività avviate al momento del controllo dall'alto del Thesaurus è stata l'elaborazione delle etichette di nodo all'interno di ogni faccetta, in modo da rendere più chiara e leggibile la struttura dei vari termini. La creazione delle etichette di nodo rendeva necessario stabilire *ordini di precedenza* fra le sottofaccette o classi.

Per esempio nel caso degli **Oggetti**: la struttura gerarchica presenta le etichette di nodo [Oggetti secondo l'uso e lo scopo] e [Oggetti secondo il materiale]. Molti termini potrebbero avere come BT entrambe queste etichette, pertanto si è deciso di dare la precedenza all'etichetta [Oggetti secondo l'uso e lo scopo], mentre, se ritenuto opportuno, di costruire una relazione associativa con il termine più ampio che indica il materiale.

Esempio:

Decorazioni in bronzo

BT [Oggetti secondo l'uso e lo scopo]

RT Bronzi

A sua volta, se si presenta la possibilità di una poligerarchia all'interno della stessa etichetta di nodo [Oggetti secondo l'uso e lo scopo], in questo caso si privilegia il termine che identifica l'ambito d'uso più specifico, stabilendo una relazione associativa con il termine che esprime l'uso più generico.

Esempio:

Arredi sacri

BT Oggetti sacri

RT Arredi

Tuttavia, l'analisi a faccette, pur essendo sempre presente come riferimento metodologico generale, non è applicata nel *Nuovo soggettario* in modo rigido. Le etichette di nodo non sono state inserite sempre, ma il più delle volte quando era necessario dare un ordine più chiaro e leggibile alla struttura gerarchica. Di conseguenza anche quanto stabilito in merito all'ordine di precedenza fra le varie sottofaccette potrà essere ripensato in corso d'opera con l'implementazione del Thesaurus.

Passiamo ora ad analizzare in quali casi abbiamo invece usato la poligerarchia. Ad esempio quando era impossibile stabilire la prevalenza di un concetto rispetto ad un altro.

Esempio:

Lavoratori anziani

Nel termine sono rappresentati due concetti: il primo la condizione lavorativa, il secondo l'età, riconducibili a due sottofaccette [Persone secondo la condizione lavorativa] e [Persone secondo l'età]; entrambe hanno come BT l'etichetta di nodo [Persone secondo la condizione].

Ci troviamo di fronte ad un caso in cui non c'è un concetto che prevale sull'altro, ecco perché è ammessa la poligerarchia.

Lavoratori anziani

BT Lavoratori

BT Anziani

Un caso in cui è stato sperimentato l'uso della costruzione poligerarchica è in un particolare gruppo di termini, che costituiscono un'eccezione rispetto ai criteri della *poligerarchia ulteriormente temperata*, ossia i termini che fanno riferimento ai generi letterari:

Commedia latina avrà come BT sia Letteratura drammatica latina sia Commedie

Tragedia greca avrà come BT sia Letteratura drammatica greca sia Tragedie

Romanzi italiani avrà come BT sia Narrativa italiana sia Romanzi

Per limitare ulteriormente la poligerarchia sono stati adottati alcuni criteri particolari, applicabili a particolari gruppi di termini:

1. per i termini con più differenze, l'ordine di precedenza è basato sull'ordine in cui sono enunciate le differenze nel sintagma: **Animali marini velenosi** ha come BT Animali marini, ma non Animali velenosi, che però è legato al termine con una relazione di tipo associativo;
2. quando si ha una cumulazione di differenze geografiche e differenze di genere, non si costruisce la relazione gerarchica neppure con i termini sovraordinati che esprimono lo stesso genere: **Atlete italiane** non ha come BT né Atleti italiani né Donne italiane, ma solo Atlete; con i termini Atleti italiani e Donne italiane si crea una relazione di tipo associativo;
3. per i termini con differenze cosiddette "ricorrenti", come quelle geografiche, etniche e nazionali, non si costruisce la relazione gerarchica con la classe sovraordinata che contiene una di queste differenze: **Coppe fenicie** ha come BT Coppe, ma non Oggetti di scavo fenici; **Xilografia italiana** ha come BT Xilografia, ma non Incisione italiana.

Come già detto, un'importante eccezione a questo criterio è costituita dai generi letterari, per i quali è stato scelto di sperimentare la poligerarchia. Concludo ricordando che, oltre ai criteri di precedenza all'interno delle sottofaccette, abbiamo stabilito delle precedenze, una sorta di *meta-ordine*, anche ad un livello superiore, tra le varie faccette. Durante questa operazione sono state rilevate delle sovrapposizioni tra faccette come è esemplificato nella fig. 1. Rimandiamo ad altre sedi l'illustrazione di questi numerosi e interessanti casi. ■

Fig. 1 – Rappresentazione grafica di sovrapposizione di faccette

