

La gestione della polisemia nei thesauri: il caso dei termini filosofici

di Melissa Tiberi e Fulvio Mazzocchi

*Le parole [...] sono fuochi di memoria, segnali di trasmissione,
transiti tra passato e presente, ancoraggi per evitare derive, non certo
approdi definitivi, ma porti sicuri nel mare aperto delle verità.*

Salvatore Natoli

La polisemia è il fenomeno per il quale è possibile associare sensi molteplici alla stessa entrata lessicale. Essa pervade ogni lingua conosciuta di cui contribuisce a determinare plasticità e polivalenza, caratteristiche che in quanto tali garantiscono alla lingua, come sistema semiotico, quelle capacità adattative che sottendono alla sua stessa vitalità.

In determinate esigenze pratiche, tuttavia, la polisemia diviene un fattore che può ostacolare la comunicazione. Nasce, quindi, l'esigenza di *controllarla*, senza però disconoscerne la funzione, al fine di facilitare una gestione ottimale dell'informazione.

È noto che nei vocabolari controllati di indicizzazione, il controllo semantico abbia come obiettivo la trasformazione di ciascuna entrata lessicale in monosemica. Come vedremo, l'applicazione di tale principio presenta qualche elemento di complicazione se applicato alla terminologia del dominio filosofico.

I termini della filosofia, infatti, presentano una notevole stratificazione di significati, dovuta al fatto che essi mantengono vivo il rapporto con il loro passato. Essi sono, cioè, dei veri e propri depositi di tradizione e conoscenza filosofica. La loro frequente polisemia è traccia di quei percorsi di indagine che si sono sedimentati nella nostra cultura e che hanno anche contribuito a determinarne lo sviluppo storico.

La realizzazione di un thesaurus di filosofia, cosa che sarebbe tanto auspicabile quanto irta di difficoltà, dovrebbe quindi confrontarsi con le problematiche che saranno qui introdotte e che speriamo di poter approfondire in future pubblicazioni.

1. La polisemia

1.1 Polisemia come caratteristica generale del linguaggio

La polisemia indica la possibilità di associare allo stesso segno più significati che mantengono una relazione tra loro [6]. Tale fenomeno deve essere distinto dall'altra possibile fonte di ambiguità lessicale, l'omonimia, dove i differenti significati riferibili alla stessa parola non sono invece in relazione ¹.

MELISSA TIBERI – FULVIO MAZZOCCHI, CNR-Istituto sull'inquinamento atmosferico - EKOLab, via Salaria km 29,300, 00016, Monterotondo stazione; e-mail tiberim77@yahoo.it, mazzocchi@iia.cnr.it.

Ultima consultazione siti web: dicembre 2006.

¹ C'è da dire che alcuni studi recenti hanno teso a ridimensionare una dicotomia in termini così netti tra omonimia e polisemia, collocandole ai due estremi opposti di un medesimo gradiente [11].

Storicamente, la polisemia era molto diffusa nel linguaggio mitico. Tale linguaggio era, infatti, caratterizzato da una forte valenza simbolica in base alla quale ogni cosa oltre ad essere se stessa poteva anche essere altro. Il *principio di non-contraddizione* e le *categorie* di Aristotele² contribuiranno ad introdurre una nuova fase storica del pensiero occidentale, in cui verrà a ridursi fortemente la vocazione verso l'oscillazione semantica e l'uso polivalente dei segni propri invece del linguaggio simbolico [10, p. 242-245]. Non per questo tali caratteristiche verranno eliminate ma permarranno, anche se non necessariamente con la stessa intensità, sia nel linguaggio ordinario e, con diverse gradazioni, anche nei linguaggi speciali.

In effetti, la polisemia è riscontrabile apparentemente in ogni lingua conosciuta ed investe indistintamente sostantivi, aggettivi e verbi. Ci sono buone ragioni per credere che la polisemia sia endemica nei linguaggi naturali; dovrebbe, in altri termini, essere considerata come un fenomeno "normale" dalla semantica lessicale [23].

Secondo alcuni autori, inoltre, la polisemia rende un sistema di comunicazione "ecologicamente più vitale" rispetto ad un ipotetico sistema "rigidamente isomorfo", dove ogni significante è associato ad un'unica rappresentazione semantica [23]. Essa, infatti, garantisce flessibilità semantica, polivalenza nell'uso e in un certo senso ricchezza al sistema di segni.

Come sottolinea Paul Ricoeur, la polisemia rappresenta una «caratteristica capace di dar conto della buona salute delle nostre lingue». Una lingua senza polisemia, infatti, «contravverrebbe anche al principio di economia ed estenderebbe il vocabolario all'infinito». Per comunicare ed esprimere la varietà dell'esperienza umana ciò di cui abbiamo bisogno è proprio di «un sistema lessicale economico, flessibile e sensibile al contesto» [19, p. 153-155].

D'altro canto, soprattutto in determinati contesti operativi, la polisemia viene considerata come fattore di ambiguità lessicale, che può generare problemi ai fini della comprensione e della comunicazione dell'informazione. Come vedremo, è importante capire come *gestire* la polisemia, da un lato riconoscendo la sua funzione all'interno di un sistema di segni, dall'altro cercando di rispondere alla necessità di trovare soluzioni in quei *giochi linguistici* che richiedono per finalità operative un elevato livello di univocità semantica.

1.2 La polisemia nei linguaggi speciali

Abbiamo detto che ci occuperemo qui del lessico filosofico, ovvero di un linguaggio speciale. In che modo si configura, dunque, il fenomeno polisemico nei linguaggi speciali?

Chiariamo, innanzi tutto, che cosa si intende per linguaggio speciale. In base alla definizione di Cortelazzo, esso è: «una varietà funzionale di una lingua naturale, dipendente da un settore di conoscenze o da una sfera di attività specialistiche, utilizzata, nella sua interezza, da un gruppo di parlanti più ristretto della totalità dei parlanti la lingua di cui quella speciale è una varietà, per soddisfare i bisogni comunicativi (in primo luogo quelli referenziali) di quel settore specialistico; la lingua speciale è costituita a livello lessicale da una serie di corrispondenze aggiuntive rispetto a quel-

2 «Non avere un determinato significato equivale a non avere alcun significato, e se le parole non hanno alcun significato, allora non hanno luogo neppure la possibilità di discorso e di comunicazione reciproca, e, in verità, non ha luogo neppure la possibilità di un discorso con se stessi. Infatti non si può pensare nulla se non si pensa una determinata cosa; ma se si può pensare, allora si può anche dare un preciso nome a questo determinato oggetto che è pensato. Rimanga, dunque, stabilito, come si è detto all'inizio, che il nome esprime un determinato significato e uno solo», Aristotele, *Metafisica*, 1006 b.

le generali e comuni della lingua e a quello morfosintattico da un insieme di selezioni, ricorrenti con regolarità, all'interno di forme disponibili nella lingua» [8, p. 8].

Una lingua speciale, quindi, è un sottosistema linguistico che viene utilizzato dai membri della comunità linguistica del settore di conoscenza. È presumibile che, in corrispondenza del restringimento del campo in cui opera il sistema di segni e dello stabilirsi di vincoli comunicativi più rigidi, all'interno di un linguaggio specialistico il fenomeno polisemico venga a ridursi. Ciò accade in modo più evidente nei linguaggi tecnico-scientifici.

Se ciò è vero anche per il lessico filosofico, è altrettanto vero che quest'ultimo sembra presentare caratteristiche del tutto peculiari rispetto al fenomeno polisemico. Ci sembra che tali caratteristiche meritino uno sforzo di approfondimento, i cui esiti potranno risultare utili anche nella prospettiva in cui ci siamo posti, relativa alla progettazione di linguaggi controllati e strutturati di documentazione per il recupero dell'informazione.

2. La gestione dell'ambiguità semantica nei thesauri

2.1 Linguaggio controllato e univocità di significato

Un thesaurus è un vocabolario controllato di indicizzazione. Il controllo semantico all'interno del thesaurus viene attuato soprattutto mediante l'inserimento di ciascun termine nella struttura delle relazioni semantiche. In genere, tale struttura si snoda secondo un percorso ad albero gerarchico, cui si affianca il reticolo di nessi associativi. A ciò si aggiunge il controllo della sinonimia, dell'omonimia e della polisemia, così come l'uso di note che chiariscono il significato dei termini all'interno del thesaurus ove ciò risulti necessario. L'obiettivo di tale articolazione è, quindi, l'univocizzazione del significato dei termini per finalità operative (indicizzazione e ricerca dell'informazione). I thesauri, quindi, propongono delle descrizioni di significati, sviluppate per fini pratici.

In tal senso, la polisemia, così come l'omonimia, costituisce un elemento di ambiguità da risolvere, o meglio da *gestire*, ai fini del controllo del significato dei termini. La capacità di risolvere tale ambiguità definisce il livello di *precisione* del linguaggio controllato.

Va anche detto, però, che il controllo del linguaggio, come suggerito da alcuni autori, non dovrebbe essere portato ad un "eccesso", ad un livello di artificializzazione cioè del linguaggio stesso, tale da risultare poco utile per gli usi pratici del sistema [22].

2.2 Metodi di disambiguazione utilizzati nei thesauri

2.2.1 Metodi di disambiguazione semantica

I metodi di disambiguazione possono essere distinti in semantici e sintattici. Essi vengono usati per il trattamento dei casi di polisemia e di omonimia.

Tra i primi, il metodo classico e ampiamente più usato è la qualificazione mediante elementi parentetici [1,3,4,7,9,12,22,24]:

- Sistema (Matematica)
- Sistema (Astronomia)

il qualificatore è una parola o un sintagma che viene aggiunta al termine, da cui è distinto mediante l'inserimento tra parentesi o l'uso di diversi caratteri grafici, e assume funzione di disambiguamento, specificando il dominio di significato a cui si riferisce il termine. Esso diviene parte integrante del termine. Come risultato della

applicazione dei qualificatori parentetici si ottengono termini distinti e univoci, ciascuno con il suo corredo sintattico.

Lo Standard ANSI/NISO Z.39.19.2005 illustra una serie di raccomandazioni relative alla qualificazione parentetica, compresa quella basilare di evitarne l'uso, ove possibile, per facilitare l'archiviazione e il recupero dell'informazione. Il problema principale, soprattutto in ambiente digitale è, infatti, la creazione di forme in qualche modo artificiali o comunque non corrispondenti a quelle di uso comune [22]. Per ovviare a tale inconveniente, nelle linee-guida editoriali dell'*Art & Architecture Thesaurus* [1] viene consigliato di usare in fase di ricerca il termine senza l'aggiunta del qualificatore (che comunque già assolve la sua funzione di disambiguazione nella struttura del thesaurus).

Di seguito, le regole dello standard in estrema sintesi:

- a) al fine di risolvere l'ambiguità semantica, all'uso della qualificazione **dovrebbe** essere preferito quello di termini composti, se ciò è possibile;
- b) il qualificatore **dovrebbe** essere aggiunto ad ogni termine da disambiguare, anche in quei casi in cui uno dei due termini viene usato nell'accezione prevalente del dominio ed il secondo in modo diverso;
- c) qualora il termine venga usato all'interno di un determinato dominiosoltanto in una delle accezioni possibili e quest'ultima risulti evidente alla comunità di utenti di quel dominio, il qualificatore **potrebbe** essere omissso. Il qualificatore **dovrebbe** essere inserito, però, nel caso in cui il thesaurus venga usato in più domini, per esempio in ambito multidisciplinare. Ciò faciliterebbe la ricerca integrata dell'informazione su diverse base di dati (*cross-database searching*) e la mappatura di termini provenienti da diversi domini;
- d) la qualificazione **non dovrebbe** essere usata per rappresentare concetti composti, per esempio: tavolo (legno): va usato tavolo di legno;
- e) il qualificatore può essere un termine; spesso è un termine sovraordinato a quello da qualificare. **Dovrebbe** essere il più breve possibile (nella situazione ideale consiste di una sola parola) ma **non dovrebbe** essere esso stesso un termine ambiguo;
- f) i qualificatori **dovrebbero** essere standardizzati all'interno del vocabolario controllato;
- g) i qualificatori e note d'uso sono cose diverse. Un termine qualificato **potrebbe** avere, però, una nota d'uso ad esso associato;
- h) i qualificatori **dovrebbero** essere aggiunti anche ai non-descrittori qualora siano presenti situazioni di ambiguità semantica.

Un secondo metodo di disambiguazione consiste nella differenziazione dei termini ambigui, rappresentando uno o entrambi i significati mediante termini composti [3, 9]. Esso è applicabile per disambiguare termini polisemici:

– thesauri linguistici

– thesauri documentari

utilizzando il primo metodo, avremmo ottenuto thesauri (linguistica) e thesauri (documentazione).

Ovviamente, il metodo aumenta il numero di termini composti, ma ha il pregio di esprimere con chiarezza e mediante espressioni del linguaggio naturale i significati intesi. Non sempre, però, è applicabile e soprattutto non andrebbe introdotto facendo ricorso a forzature a livello terminologico (si veda avanti l'esempio del termine categoria).

Un terzo metodo di disambiguazione semantica è il ricorso allapoligerarchizzazione dei termini ambigui [22]. In questo caso, è il posizionamento del termine nella struttura gerarchica a risolvere l'ambiguità semantica (anche le relazioni associative possono dare un contributo in tal senso).

In altre parole, è la collocazione in un contesto, così come accade nella lingua naturale, a fornire la corretta chiave di lettura della semantica del termine;

in questo caso, abbiamo un solo termine posizionato in diverse gerarchie. Proprio per tale motivo, non ci pare il metodo più opportuno per i casi di omonimia, dove occorre una distinzione netta dei termini. Inoltre, la disambiguazione rischia di essere, soprattutto nel listato alfabetico, meno evidente per chi userà il thesaurus.

Altri metodi di disambiguazione semantica citati in letteratura sono il ricorso alla sostituzione del termine ambiguo con un suo sinonimo [24] e l'utilizzo di note d'uso [22].

2.2.2 Metodi di disambiguazione sintattica

I metodi di disambiguazione sintattica vengono poco usati nei thesauri (non vengono menzionati fra l'altro dagli Standard in materia). Sono utilizzati soprattutto nei sistemi pre-coordinati. Le *Library of Congress Subject Headings*, ad esempio, sono provviste di una sintassi molto articolata e fanno largo uso delle *subdivisions*, termini posti di seguito all'*heading* che forniscono informazioni sullo stesso (*topic, place, time e form*). Uno degli usi delle *subdivisions* a livello di singoli termini è quello di indicare i diversi aspetti in base ai quale il soggetto in questione va considerato [5, 22]:

- Reading - Ability testing
- Reading - Case studies
- Reading - Computer programs

analogamente, alcuni autori ipotizzano l'utilizzo della disambiguazione sintattica o per *contestualizzazione* [9] anche nei thesauri. Tornando all'esempio iniziale, avremmo:

- Sistema - Matematica
- Sistema - Astronomia

in questo caso il termine rimane uno solo e generico, ma viene considerato però in due modi diversi in virtù dell'azione esercitata su di esso dalle discipline.

Come abbiamo già detto, la disambiguazione sintattica è stata finora poco utilizzata nei thesauri; non ci sono quindi dati empirici comparabili con quelli disponibili per i metodi semantici. A livello teorico, possiamo però dire che la contestualizzazione riproduce parzialmente il metodo semantico della poligerarchia facendo uso di una struttura sintattica [22]. Alcune considerazioni fatte per la poligerarchizzazione valgono, quindi, anche per questo metodo.

Naturalmente, in tal caso l'elemento contestuale non è rappresentato dal termine immediatamente sovraordinato (che connette poi il termine da disambiguare all'intera gerarchia fino al livello più alto), ma da un termine, spesso generale, che indica il punto di vista con cui va considerato in termine ambiguo, circoscrivendone il significato inteso. Inoltre, a differenza della poligerarchizzazione, l'aggiunta dell'elemento di contestualizzazione di seguito al termine rende immediatamente comprensibile (anche nell'ordinamento alfabetico e senza dover fare riferimento alle relazioni semantiche del termine) l'aspetto del termine polisemico a cui ci si sta riferendo. Proprio per questo motivo, tale metodo può essere applicato anche ai casi di omonimia.

2.2.3 Altri metodi di disambiguazione

Svenonius [22] indica come altro possibile metodo la disambiguazione a livello di post-coordinamento tramite operatori come *AND*. Per esempio, l'omonimia che si riscontra nel linguaggio naturale per il termine *mercurio* verrebbe così risolta:

- Mercurio *AND* Pianeta
- Mercurio *AND* Elemento chimico.

Sono stati messi a punto anche metodi di disambiguazione automatici o semiautomatici. Gli algoritmi di disambiguazione si basano soprattutto sull'analisi delle

parole nell'ambito testuale in cui esse compaiono. Questo è il caso, per esempio, del metodo utilizzato nel Biological Sciences Information Services (BIOSIS) per disambiguare gli omonimi [26].

2.2.4 Considerazioni conclusive

Uno degli aspetti da sottolineare è che, attualmente, gli stessi metodi di disambiguazione vengono applicati indistintamente ai casi di polisemia e a quelli di omonimia.

Alcuni studi a carattere sperimentale hanno però mostrato che la possibilità di distinguere tra polisemia e omonimia ha delle implicazioni positive per l'*information retrieval*, aumentando il tasso di recupero [14,21]. È ipotizzabile, quindi, che in futuro metodi capaci di effettuare tale distinzione siano utilizzati anche nei thesauri.

Potremmo dire che, da un certo punto di vista (naturalmente, vanno considerati anche altri aspetti), il metodo della qualificazione sembra la migliore soluzione per il trattamento dei casi di omonimia, in quanto vengono creati termini distinti e ciò ben rappresenta la situazione degli omonimi.

Gli altri metodi (uso di termini composti, ricorso alla poligerarchia e disambiguazione sintattica) potrebbero essere più indicati per i casi di polisemia, poiché conservano una traccia dei sensi correlati ma distinti (perché i termini distinti ottenuti condividono in genere lo stesso termine sovraordinato o perché si considerano i diversi aspetti di uno stesso termine).

Come vedremo, nel contesto di questa situazione generale riguardante il trattamento dell'ambiguità semantica nei thesauri, più problematico appare il caso di cui ci occuperemo nel presente contributo: i termini della disciplina filosofica. Tali termini, che come intuitivamente possiamo immaginare sono caratterizzati da una elevata densità concettuale, sembrano essere contraddistinti da una più estesa presenza del fenomeno polisemico, la cui natura presenta, come abbiamo già detto, delle caratteristiche del tutto specifiche.

3. Il lessico filosofico

I termini che compongono il lessico filosofico esprimono concetti, in quasi la totalità dei casi, non soltanto complessi e astratti ma anche fortemente stratificati.

Naturalmente, il processo di stratificazione in quanto tale non è caratteristica esclusiva del lessico filosofico. Ma ciò che qui lo contraddistingue è la sua natura. Da cosa deriva, infatti, tale stratificazione? Ci sembra di poter dire che il processo sia riconducibile al continuo "riassetto di senso" che tali termini subiscono nel corso del tempo. Ciò può implicare sia la sovrapposizione del nuovo significato con quello precedente, sia la sua rideterminazione parziale o totale. Quando i filosofi successivi utilizzano un certo termine dandogli una valenza diversa da chi li ha preceduti, lo fanno tenendo conto e confrontandosi con la sedimentazione della vicenda concettuale anteriore.

Salvatore Natoli coglie sensibilmente quelle che sono le peculiarità del lessico della filosofia quando dice che le parole della filosofia: «non si sono affatto sgravate del passato, ma sono più che mai feconde in forza di quel passato: eccedono se stesse per un sovraccarico di storia che mettono a disposizione senza ipotetiche per la più ampia e libera interpretazione [...], variamente definite, hanno acquisito spessore e pur rimanendo le stesse nel corso del tempo sono divenute polisemiche, in taluni casi anche equivoche. Una stratificazione di significati tutta da indagare» [17, p. 6].

Un termine filosofico, quindi, continua a portare con sé la sua storia. I significati non si estinguono diventando obsoleti (come accade ad esempio per i termini della scienza la cui validità deve essere sempre comprovata all'interno del paradigma vigen-

te)³ma si accumulano stratificandosi. Paradigmi nati in contesti storico-culturali ben determinati, benché cronologicamente superati, si ri-generano nel presente. E sono le parole a darne testimonianza.

Nel momento della comprensione di un termine, infatti, non possiamo riferirci ad un unico paradigma interpretativo o solamente a quello attuale. Nonostante si ridisegni, ogni volta, un nuovo contesto semantico e speculativo, l'orizzonte di senso rimane allacciato al passato.

Va anche considerato il fatto che l'origine di gran parte dei termini filosofici è greca. Non è molto frequente, infatti, il ricorso a neologismi nell'ambito della disciplina filosofica.

L'origine dell'elevato numero di termini polisemici nel dominio filosofico sembra, pertanto, derivare dal processo di continua *rideterminazione semantica* che essi subiscono⁴.

La polisemia dei termini filosofici è, quindi, una polisemia forte, propriamente semantica poiché contraddistinta dalla presenza non univocizzabile di una pluralità di sensi lessicali (si veda avanti l'esempio del termine categoria).

Il tipo di polisemia riferibile ai termini filosofici sembra essere quella descritta da Paul Ricoeur come polisemia *di tipo cumulativo*: non solo il termine possiede, entro un dato sistema, più accezioni, ossia delle varianti che appartengono a diverse classi concettuali, ma risulta essere in grado di acquisire un nuovo senso (significato) senza perdere quello anteriore [19].

Possiamo forse meglio comprenderne la portata, raffrontandola con la *polisemia analitica* di termini in realtà monosemici, ovvero non connessi a diversi percorsi di senso ma riconducibili ad unicità intuitiva [25]. È questo il caso di termini come *benzene* o *uccello*, caratterizzati da una molteplicità di proprietà (in base alla quale essi sono analizzabili e classificabili secondo diverse prospettive) confinata però all'interno d'un unico percorso di significato [16].

Tab. 1 - Tratto semantico del termine *benzene*. Il *benzene* è una sostanza organica aromatica che può essere anche considerata sotto il profilo della sua tossicità o pericolosità, o delle proprietà fisico-chimiche (infiammabilità)

Benzene
 sostanza
 organica
 aromatica
 tossica
 infiammabile
 pericolosa

³ Per Thomas S. Kuhn uno degli indicatori del passaggio da un impianto teorico all'altro nella storia della scienza è proprio il cambiamento di significato dei termini, inteso dapprima come parte integrante del cambiamento di paradigma, cfr. *La struttura delle rivoluzioni scientifiche*, Torino: Einaudi, 1999, e successivamente come esito del mutamento di tassonomia lessicale, cfr. *Dogma contro critica*, Milano: Raffaele Cortina, 2000. Si veda in proposito, per esempio, la discussione in *Che cosa sono le rivoluzioni scientifiche?* in: T.S. Kuhn, *La tensione essenziale e altri saggi*, Torino: Einaudi, 2006, p. 159-184.

⁴ La nozione di *rideterminazione semantica* viene qui intesa in senso molto generale. Esistono, tuttavia, interpretazioni della stessa in ambito linguistico, che ne circoscrivono il significato in termini di impiego di unità appartenenti alla lingua comune (o a lingue speciali), inserite in un altro lessico con un diverso significato [8].

4. Rappresentare la polisemia: la *somiglianza di famiglia*

Al fine di comprendere in modo più approfondito i vari aspetti della questione, accenneremo qui brevemente ad una delle modalità di trattamento e rappresentazione della polisemia sviluppate all'interno della tradizione semantica e filosofica del XX secolo.

In generale il metodo più adeguato sembra essere quello della *somiglianza di famiglia*. La nozione di *somiglianza di famiglia* è stata introdotta dal filosofo austriaco Ludwig Wittgenstein per spiegare la natura del significato. Per Wittgenstein, l'idea d'una essenza comune che riunisca i vari "pezzi di significato" di un termine è frutto della proiezione sulla realtà del nostro desiderio che effettivamente sia così. Ma ciò non si verifica (se non in pochi casi). I diversi elementi del significato di un termine, invece, si ricongiungono attraverso una rete aperta di analogie. Non è più l'identificazione di un nucleo centrale stabile di significato, quindi, che garantisce la semantica di un termine ma una connessione che avviene a livello periferico.

Per Wittgenstein, forse più di ogni altra, la parola «gioco» si presta bene a spiegare questa situazione. Se guardiamo ai giochi, infatti, non è certo facile cogliere qualcosa che sia realmente comune a tutti e che corrisponda all'essenza del gioco stesso. Ogni tentativo di individuare ciò che è comune a tutti i giochi verrebbe contrastato attraverso contrapposizioni significative. Se, per esempio, diciamo che una delle caratteristiche essenziali del gioco è il divertimento, potremmo tuttavia trovare casi in cui parliamo di giochi senza caratterizzarli come divertenti (per esempio la roulette russa). Se, invece, diciamo che è determinante per il gioco la competizione, è comunque possibile individuare giochi in cui tale caratteristica non compare (per esempio i giochi d'azzardo). E così via. Questo è ciò che afferma Wittgenstein, il quale dice anche: «Ciò che invece si può ammettere è che tra un gioco e l'altro vi siano somiglianze, e che ciò che ci induce a chiamare cose tanto disparate "giochi" sia una rete complicata di somiglianze che si sovrappongono ed incrociano a vicenda» [27, § 66].

Wittgenstein, inoltre, ricorre ad un'analogia per renderci ancora più chiara intuitivamente l'espressione somiglianze di famiglia. Quest'ultima, infatti, viene messa in relazione alle varie somiglianze che sussistono tra i membri di una famiglia, che si sovrappongono e si incrociano nello stesso modo: corporatura, tratti del volto, colore degli occhi, modo di camminare, temperatura ecc. «I giochi formano una famiglia - essi sono processi variamente imparentati tra loro» [27, § 67]. È questo, quindi, un altro modo, secondo Wittgenstein, per ribadire che i termini non hanno un significato definibile in astratto, che rimanda a cose o a strutture ontologiche. Il significato è nell'uso, all'interno di un linguaggio mobile e mutevole che rimanda, in ultima analisi, alle forme del vivere sociale⁵.

Se Wittgenstein *sfonda* la possibilità di stabilire confini naturali al significato (è come se tutti i termini divenissero in potenza intrinsecamente polisemici), è pur vero che per alcuni termini, per esempio quelli tradizionalmente considerati polisemici, tale caratteristica si manifesta in modo più marcato. L'ambiguità lessicale che deve essere risolta per finalità operative è riconducibile soprattutto alla presenza di questi termini.

⁵ «Il linguaggio non può essere descritto se non in riferimento all'uso, o meglio agli usi, considerando che esiste una pluralità di usi alternativi e complementari che si modificano e continuamente aumentano di numero, perché è il mutare delle esigenze espressive a determinare l'insorgere di sempre nuovi giochi linguistici. È qui che trova spazio la vaghezza, proprio in relazione alla varietà dei giochi linguistici e al loro fornire un repertorio di possibilità alternative potenzialmente illimitate nell'allargare i confini dei significati a nuovi sensi» [15, p. 27].

Ecco allora che, al di là della sua potenzialità esplicativa della natura del significato, il modello della *somiglianza di famiglia* diventa utile, mediante un'applicazione più ristretta, anche per le possibilità che offre di "modellizzare" e raffigurare il fenomeno polisemico inteso in tali termini.

La polisemia non prevede un caso centrale. Ciò è ben illustrato dallo schema suggerito da Givon (fig. 1), che ci mostra appunto l'assenza di proprietà comuni, ovvero condivise da tutti i membri di una stessa famiglia. Il nesso di somiglianza è sempre periferico, locale, parziale (anche se può esistere un *sensu di base*, che possiede una maggiore rilevanza rispetto agli altri sensi).

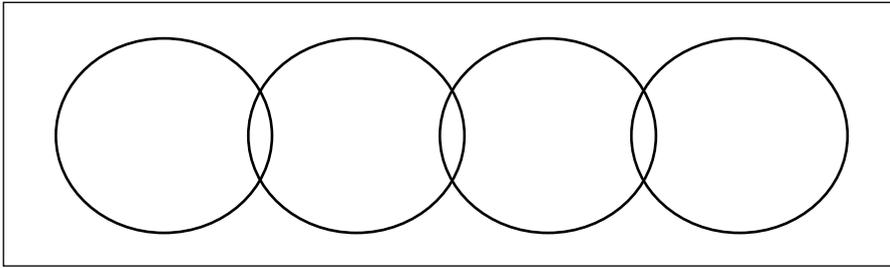


Fig. 1 - Schema di Givon per la rappresentazione della *somiglianza di famiglia* [25]

C'è da dire, comunque, che nella polisemia, almeno per alcune delle modalità attraverso cui essa si manifesta, è rintracciabile un principio di motivazione che, sulla base di schemi cognitivi profondi, regola il passaggio da un senso all'altro. Seppur tale principio non consenta di prevedere a priori quale potrà essere il percorso di sviluppo del termine, esso ne orizzonta il processo, che quindi non sarà un accumularsi casuale di accezioni diverse [25]⁶.

Per quanto riguarda la semantica dei termini filosofici, sembra avere un ruolo di notevole importanza la prima formulazione storica del termine, che ne condizionerà, poi, in moltissimi casi, anche gli sviluppi futuri.

Nel prossimo paragrafo ci occuperemo delle problematiche finora discusse a partire dall'analisi di un esempio concreto che ci renderà, crediamo, consapevoli anche delle difficoltà dell'organizzazione concettuale del dominio filosofico.

5. Verso un thesaurus di filosofia: riconoscere e gestire la polisemia dei termini filosofici⁷

5.1 Analisi di un caso studio: il termine *categoria*

Ci occuperemo qui di un caso tanto rappresentativo quanto complesso, la cui presenza in termini di concetto è rintracciabile praticamente in tutte le fasi della storia

⁶ Va comunque ricordato, come ci dice anche Violi che «il concetto di somiglianza famiglia non impone che vi siano esclusivamente somiglianze locali, ma lo permette solamente, nulla vieta di considerare una categoria classica come una forma particolare di somiglianza in cui tutti i membri hanno in comune le stesse proprietà» [25, p. 198].

⁷ Le riflessioni che seguono hanno il limite di riferirsi ad un solo caso studio e di non essere state sviluppate all'interno della prassi di sviluppo di un thesaurus per il dominio specifico. Ci proponiamo, comunque, in futuri contributi, di porre rimedio, almeno parzialmente, a tale limite.

della filosofia: il termine *categoria*. Una definizione parziale del termine potrebbe includere i seguenti passaggi:

- Categoria (Aristotele): determinazioni ultime della realtà sensibile. Sono le dieci fondamentali e più generali strutture della realtà sensibile (sostanza, quantità, qualità, relazione, luogo, tempo, azione, passione, posizione, condizione), che trovano poi i loro corrispondenti sul piano logico e linguistico.
- Categoria (Kant): ciascuna delle forme a priori dell'intelletto umano (sono pertanto funzioni del pensiero).
- Categoria (Hegel): determinazioni dell'Idea o realtà assoluta nel suo progressivo svolgimento dialettico (si perde, quindi, la connotazione puramente logica).
- Categoria (Russell): il tipo logico impiegato nella logica matematica.

In che modo potrebbe o dovrebbe essere rappresentata in un thesaurus tale polivalenza semantica?

Uno dei pochi thesauri disponibili per il dominio filosofico è il thesaurus *Library of Congress Subject Headings in philosophy*. In effetti, si tratta di un thesaurus "derivato", ovvero sia non concepito come tale dal momento della progettazione a quello della realizzazione, ma prodotto della sovrapposizione di un formato thesaurale su un vocabolario derivato dalle *LCSH*, in base ad una selezione di *headings* in materia filosofica (ciò ha causato non pochi problemi di trasformazione da un formato all'altro). La copertura concettuale e terminologica del dominio filosofico è limitata a quella presente nelle *LCSH*. Non è stata aggiunta, infatti, terminologia supplementare. Il thesaurus include un listato alfabetico, in cui vengono però elencati anche i *nodes*, divisioni standard del dominio filosofico su base sostanzialmente disciplinare, all'interno dei quali è possibile rintracciare una organizzazione sistematica dei termini [5, p. 11-16].

Vediamo ora come viene presentato il termine *categoria* nell'elenco alfabetico del thesaurus:

Tab. 2 - Relazioni semantiche del termine *categories* nel thesaurus delle *LCSH in philosophy*

Categories (Philosophy)	
Nodes:	LOGIC
	METAPHYSICS
Broader terms:	Logic, Ancient
	Predicate (Logic)
Narrower terms:	Dialectical Materialism, Categories of Modality (Logic)
	Place (Philosophy)
	Quality (Philosophy)
	Quantity (Philosophy)
	Relation (Philosophy)
	Situation (Philosophy)
	Substance (Philosophy)
	Time
Related term:	Tattvas (Sankhya)

La semantica del termine *categoria* viene, quindi, solo parzialmente, ma legittimamente, rappresentata rispetto allo schema definitorio da noi sopra proposto. Ad essere raffigurata è, infatti, solo l'accezione aristotelica, soprattutto in chiave logica. I termini subordinati (NT: *narrower terms*) corrispondono a parte delle categorie aristoteliche (ma il qualificatore annesso sembra rinviare al solo aspetto logico). I ter-

mini sovraordinati (BT: *broader terms*) rimandano ad una delle discipline di appartenenza, la logica, ma ne viene ristretta la portata temporale (logica antica). Il rimando ai *nodes*, che come abbiamo detto corrispondono ai settori disciplinari principali attraverso cui viene strutturata la filosofia, include oltre alla logica anche la metafisica.

Qualche incongruenza o possibile ambiguità sembra in ogni caso emergere. Soprattutto in considerazione del fatto che, come indicano gli NT, vengono ricondotte ad un unico termine, *categories* (philosophy), sia l'accezione logico-ontologica di origine aristotelica sia le categorie del materialismo dialettico di Marx. In tal modo, infatti, viene espressa la valenza polisemica di *categories*, che come tale andrebbe adeguatamente trattata. La polisemia dei termini filosofici, quindi, va innanzitutto riconosciuta.

5.2 Disambiguazione dei termini filosofici polisemici

Riassumiamo brevemente le problematiche connesse al caso studio. Siamo di fronte ad un termine polisemico, dove la polisemia si manifesta secondo le specifiche modalità che sono state illustrate nel paragrafo 3.

Il riconoscimento della polisemia, insieme a quello delle sue caratteristiche specifiche, è il primo passo da fare. Nel processo di sviluppo di un linguaggio controllato, tale ambiguità viene risolta mediante l'utilizzo di tecniche di disambiguazione semantica.

Naturalmente, il posizionamento del termine o dei termini nella struttura semantica del thesaurus, avverrà coerentemente con ciascuno dei sensi implicati (in alcuni casi, come abbiamo visto, è il posizionamento stesso a garantire la disambiguazione).

Nella situazione ideale di un thesaurus dedicato fin dalla sua progettazione al dominio filosofico, si dovrebbe cercare di includere tutte le accezioni possibili utili di *categoria*, facendo in modo di distinguere i diversi percorsi di senso del termine, attraverso i metodi di disambiguazione.

Per esempio, facendo ricorso alla qualificazione parentetica e tentando di rappresentare tutti i percorsi di significato contenuti nelle definizioni iniziali, otterremmo i seguenti termini:

- Categoria (Metafisica o Ontologia)
- Categoria (Logica)
- Categoria (Linguistica)

Dal punto di vista semantico, è una soluzione dotata di chiarezza, che crea le condizioni per una più completa rappresentazione della semantica del termine.

In termini operativi, il problema di questo metodo lo abbiamo già menzionato precedentemente: la creazione di forme artificiali [22]. Considerata l'elevata estensione del fenomeno polisemico nel lessico filosofico, il problema potrebbe non essere di poco conto.

Soluzioni alternative nell'ambito dei metodi di disambiguazione semantici appaiono, comunque, meno praticabili. Per esempio, il ricorrere alla formulazione di termini composti in cui attribuzioni aggiuntive (che funzionerebbero come *modificatori*) andrebbero a specificare il significato del termine. Nel caso di "categoria", avremmo, per esempio: *categoria ontologica*, *categoria logica*, *categoria linguistica*. Ma tale soluzione rischia di essere, da un certo punto di vista, innaturale ed artificiosa. ella maggior parte di casi, infatti, non sono questi i termini filosofici. L'esempio più emblematico è forse proprio in Aristotele, il quale utilizzava un unico termine, *categoria*, nei tre sensi suddetti. Questi erano, però, interconnessi in base ad un principio di corrispondenza: le categorie sono articolazioni dell'essere, del pensiero e del linguaggio.

Dei metodi che utilizzano contesti ai fini della disambiguazione, poligerarchizzazione e contestualizzazione sintattica, il secondo appare più appropriato per il trattamento della polisemia filosofica.

Entrambi i metodi prevedono la presenza di un solo termine all'interno del thesaurus. Nel primo caso *categoria* sarebbe collocato nel thesaurus sotto diversi percorsi gerarchici che farebbero capo ai tre snodi disciplinari di riferimento per il termine: metafisica/ontologia; logica; linguistica. Nel secondo, potrebbero essere gli stessi snodi disciplinari a identificare l'aspetto inteso del termine.

È la stessa presenza di un solo termine a preservare indirettamente il nesso che collega i diversi sensi del termine polisemico. Ma il metodo sintattico, mediante l'aggiunta dell'elemento di contestualizzazione di seguito al termine, esplicita in modo più diretto il senso accettato del termine polisemico, oltre al fatto di renderlo subito comprensibile anche nell'ordinamento alfabetico e senza dover fare riferimento alle relazioni semantiche del termine. Considerata la stratificazione e la complessità semantica che contraddistingue i termini filosofici, dovrebbero essere comunque garantite una certa chiarezza e facilità d'uso. Resta comunque il fatto che i metodi di disambiguazione sintattica non sono molto applicati ai thesauri.

Come abbiamo già detto, tutti questi metodi hanno però il limite di non consentire distinzioni tra polisemia e omonimia. In tal senso, si potrebbe forse sperimentare la plausibilità d'una distinzione tra metodi utilizzati per la disambiguazione dell'omonimia e metodi usati per il trattamento dei casi di polisemia (si veda il paragrafo 2).

Inoltre, per recuperare e valorizzare la connessione tra le diverse entrate lessicali (oltre alla mera vicinanza alfabetica) potrebbero essere stabilite delle relazioni associative (RT), in cui venga esplicitato il senso polisemico della connessione (ammesso che ciò sia permesso dalla struttura relazionale del sistema).

Va infine anche considerato un problema di carattere estetico-funzionale: considerata l'elevata polisemizzazione del lessico filosofico, quanti elementi di disambiguazione dovranno essere introdotti in un thesaurus per questo dominio?

Conclusioni

Un thesaurus nasce per finalità pratiche, allo stesso tempo esso propone e realizza una strutturazione del dominio di conoscenza. Può, quindi, essere considerato come una sorta di prototipo di mappa concettuale, semantica e terminologica del dominio [13,16].

Proprio per questo motivo, esso deve poter incorporare, oltre a quelle più immediate dell'ambito operativo, le caratteristiche specifiche ed essenziali del dominio stesso.

In questo articolo abbiamo analizzato una di queste caratteristiche per il dominio della filosofia, il ricorrere del fenomeno polisemico nella sua terminologia, tentando di abbozzare anche alcune soluzioni operative.

Riguardo il trattamento dell'ambiguità semantica e quindi della polisemia nei thesauri, uno dei nodi principali da sciogliere è come conciliare il raggiungimento della comprensibilità dei significati, con la praticabilità del metodo di disambiguazione in ambito operativo (si veda, per esempio, il caso dei qualificatori parentetici).

Un altro aspetto che dovrà essere approfondito, sia in ambito teorico che applicativo, è la distinzione tra omonimia e polisemia, oggi non consentita dai metodi disponibili, ma che sembra avere rilevanza nel recupero dell'informazione [14, 21]. Anche in tal caso, si dovrà tener conto sia di problematiche di carattere semantico che metodologico-applicativo.

Va detto, comunque, che la “richiesta” di disambiguazione è in forte crescita, in corrispondenza dell’ampliamento (sia in senso “spaziale” che verso una maggiore interdisciplinarietà) del contesto di riferimento in cui si sviluppano i linguaggi controllati e di una ricerca dell’informazione spesso orientata verso l’interrogazione simultanea di più base di dati [22]. Tale modalità di ricerca richiede, infatti, una maggiore compatibilità tra sistemi e si fonda sull’interoperabilità a livello semantico e tecnologico. Ciò che poteva considerarsi non ambiguo all’interno di un determinato dominio, rischia di non esserlo più dopo un ampliamento di confini.

In questo contesto, l’aumento dell’universalità di un vocabolario diviene pertanto una caratteristica auspicabile, attualizzando un criterio già espresso precedentemente da alcuni autori in base al quale il livello di utilizzabilità di un vocabolario è proporzionale alla sua universalità [20].

Da questo punto di vista, potrebbe anche essere utile un confronto con i metodi di disambiguazione che vengono utilizzati all’interno delle ontologie [18].

Ci auguriamo, infine, che la scarsissima disponibilità di strumenti per l’organizzazione della conoscenza per il dominio filosofico possa essere presto colmata.

RIFERIMENTI BIBLIOGRAFICI

- [1] AAT. *AAT editorial guidelines: editorial rules: terms*. <http://www.getty.edu/research/conducting_research/vocabularies/guidelines/aat/_3_3_terms_names.pdf>.
- [2] *Dizionario di filosofia*. Milano: Rizzoli, 1976.
- [3] Jean Aitchison – Alan Gilchrist – David Bawden. *Thesaurus construction and use: a practical manual*. London: ASLIB, 1997.
- [4] ANSI/NISO Z.39.19.2005. *Guidelines for the construction, format and management of monolingual controlled vocabularies*. Bethesda, MD: NISO Press, 2005.
- [5] Barbara L. Berman. *Library of Congress subject headings in philosophy: a thesaurus*. Charlottesville, VA: Philosophical Documentation Center, 2001.
- [6] Giorgio Raimondo Cardona. *Dizionario di linguistica*. Roma: Armando, 1988.
- [7] Maria Pia Carosella – Maria Valenti. *Documentazione e biblioteconomia: manuale per i servizi di informazione e le biblioteche speciali italiane*. Milano: Franco Angeli, 1984.
- [8] Michele Cortellazzo. *Lingue speciali: la dimensione verticale*. Padova: Unipress, 1994.
- [9] Daniele Danesi. *Le variabili del thesaurus: gestione e struttura*. Firenze: Ilnia, 1990.
- [10] Umberto Galimberti. *Segno*. In: *Idee: il catalogo è questo*. Milano: Feltrinelli, 2001.
- [11] Dirk Geeraerts. *Vagueness’s puzzles, polysemy’s vagaries*. «Cognitive linguistics», 4 (1993), n. 3, p. 223-272.
- [12] International Organization for Standardization, *ISO 2788-1986. Documentation. Guidelines for the establishment and development of monolingual thesaurus*. Ginevra: ISO, 1986.
- [13] Branka Kosovac. *Internet/Intranet and thesauri*. Project report - BELCAM roofing thesaurus. National research council Canada, Institute for research in construction, 1998, <http://www.nrc.ca/irc/thesaurus/roofing/report_b.html>.
- [14] Robert Krovetz. *Homonymy and polysemy in information retrieval*. In: *Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics, Madrid, July 1997*. [Ed. by] Philip R. Cohen, Wolfgang Wahlster. Somerset, MD: Association for Computational Linguistics, 1997, p. 72-79.

- [15] Sabrina Machetti. *Uscire dal vago: analisi linguistica della vaghezza nel linguaggio*. Roma-Bari: Laterza, 2006.
- [16] Fulvio Mazzocchi – Paolo Plini, *Thesaurus classification and relational structure: the EARTH experience*. In: *Terminology and content development - Proceedings of the 7th International conference on terminology and knowledge engineering, Copenhagen, 17-18 August 2005*. [Ed. by] Bodil Nistrup, Hanne Erdman Thomsen. Copenhagen: Litera, 2005, p. 265-278.
- [17] Salvatore Natoli. *Parole della filosofia o dell' arte del meditare*. Milano: Feltrinelli, 2004.
- [18] Domenico M. Pisanelli – Aldo Gangemi – Massimo Battaglia – Carola Catenacci. *Coping with medical polysemy in the semantic Web: the role of ontologies*. In: *MEDINFO 2004. Proceedings of the 11th World congress on medical informatics*. [Ed. by] Marius Fieschi, Enrico Coiera and Yu-Chuan Li. Amsterdam: IOS press, 2004, p. 416-419.
- [19] Paul Ricoeur. *La metafora viva*. Milano: Jaca Book, 2001.
- [20] Dagobert Soergel. *Indexing languages and thesauri: construction and maintenance*. Los Angeles, CA: Melville, 1974.
- [21] Christopher Stokoe. *Differentiating homonymy and polysemy in information retrieval*. In: *Proceedings of human language technology conference and Conference on empirical methods in natural language processing (HLT/EMNLP), Vancouver, October 2005*. Somerset, MD: Association for Computational Linguistics, 2005, p. 403-410.
- [22] Elaine Svenonius. *The intellectual foundation of information organization*. Cambridge, MA: MIT Press (Digital libraries and electronic publishing), 2000.
- [23] John R. Taylor. *Polysemy's paradoxes*. «Language sciences», 25 (2003), n. 6, p. 637-655.
- [24] Marisa Trigari. *Come costruire un thesaurus*. Modena: Franco Cosimo Panini, 1992.
- [25] Patrizia Violi. *Significato ed esperienza*. Milano: Bompiani, 1996.
- [26] Natasha Vleduts-Stokolov. *Concept recognition in an automatic text –processing system for the life-sciences*. «Journal of the American Society for Information Science», 38 (1987), n. 4, p. 269-287.
- [27] Ludwig Wittgenstein. *Ricerche filosofiche*, Torino: Einaudi, 1999.

Management of polysemy in thesauruses: the case of philosophical terms

by Melissa Tiberi and Fulvio Mazzocchi

Polysemy is the phenomenon whereby it is possible to associate a number of meanings to the same lexical entry. It exists in every known language and it contributes to determining their plasticity and versatility, characteristics which guarantee a language, as a semiotic system, those capacities to adapt that give it its vitality.

In some practical requirements, however, polysemy becomes a factor that can hinder communication. Therefore a need to *control it* arises, while not however disclaiming its function, in order to facilitate optimal information management.

It is well known that in controlled indexing vocabularies, semantic control has as its objective the transformation of each lexical entry into a single meaning entry. As we will see, the application of this principal has some complications if applied to the terminology of the area of philosophy.

In fact, philosophical terms have considerable layers of meanings, due to the fact that they maintain a certain strong relationship with the past.

The creation of a philosophy thesaurus, something that would be very desirable even if full of difficulties, should therefore be considered in the light of the problems that will be introduced here.

This article analyzed one of the characteristics for the area of philosophy, the recurrence of the phenomenon of polysemy in its terminology, while trying to at least draft some operational solutions.

With regard to the treatment of semantic ambiguity and therefore of polysemy in thesauruses, one of the main problems to be solved is how to reconcile the achievement of understanding the meanings with the practicability of the method of eliminating ambiguity when operational.

Another aspect that has to be investigated, both at theoretic and application level, is the distinction between homonym and polysemy, not allowed today by available methods, but which seems to be of importance in information recovery.

In this context, the increase of the universality of a vocabulary thus becomes a desirable characteristic, implementing a criteria that was already previously expressed by some authors on the basis of which the level of usability of a vocabulary is proportional to its universality.

MELISSA TIBERI – FULVIO MAZZOCCHI, CNR-Istituto sull'inquinamento atmosferico - EKOLab, via Salaria km.29,300, 00016 Monterotondo stazione; e-mail tiberim77@yahoo.it, mazzocchi@iia.cnr.it.

Bollettino **AIB**, ISSN 1121-1490, vol. 47 n. 1/2 (March-June 2007), p. 93-107.