

## **Intelligenza artificiale e semantica: il problema degli atteggiamenti proposizionali\***

MARCELLO FRIXIONE

*Dipartimento di Filosofia, Università di Genova  
e Istituto Internazionale per gli Alti Studi Scientifici - Vietri sul Mare (Salerno)*

### **Sommario**

I cosiddetti paradossi epistemici hanno posto in evidenza conseguenze difficilmente prevedibili per le teorie semantiche di tipo cognitivo. Alcune assunzioni intuitivamente non problematiche circa la formalizzazione degli atteggiamenti proposizionali non sono conciliabili con il requisito delle rappresentabilità finita dei significati linguistici, requisito irrinunciabile per ogni teoria semantica di impostazione cognitiva e computazionale. In particolare, l'assunzione di onniscienza logica, se associata con una nozione di competenza semantica idealizzata, risulta inconsistente. Questo risultato è stato utilizzato come argomento contro la plausibilità stessa del progetto di una semantica cognitiva. Thomason ha sostenuto che, sebbene i soggetti epistemici reali non siano logicamente onniscienti, ciò è dovuto a cause fattuali e contingenti, e non pare plausibile che possa dipendere da fattori di tipo esclusivamente logico. D'altro canto, Fagin e Halpern hanno messo in luce come la mancanza di onniscienza logica nei soggetti cognitivi reali sembri essere un fenomeno complesso, dovuto al concorso di più cause differenti. In questa prospettiva, si può ipotizzare una distinzione tra fattori "qualitativamente" differenti che determinano la non onniscienza logica: accanto a fattori di tipo fattuale, possono esistere anche elementi "a priori" che escludono che un soggetto cognitivo finito possa essere logicamente onnisciente rispetto alla logica classica.

---

\* Questo lavoro è stato parzialmente finanziato dal Contratto quinquennale C.N.R. - I.I.A.S.S e dal progetto Esprit BRA SECOND. Ringrazio Dario Palladino per la sua accurata lettura del manoscritto e per le sue osservazioni.

## 1. Atteggiamenti proposizionali, semantica formale e onniscienza logica

A partire da Russell (1921; 1940), l'espressione *atteggiamenti proposizionali* (*propositional attitudes*) viene usata in filosofia per indicare quegli atteggiamenti di tipo mentale che vertono su una proposizione, o che comunque hanno un contenuto che può essere descritto mediante un enunciato linguistico. Sono esempi di atteggiamenti proposizionali credere, pensare, temere, sperare, e così via. In tutti questi casi si crede (o si pensa, si teme, si spera) che ###, dove ### è un certo enunciato. Individuare un trattamento adeguato degli enunciati di atteggiamento proposizionale costituisce una delle principali fonti di difficoltà per le semantiche filosofiche del linguaggio naturale di tradizione fregeana, e in particolare per quelle teorie del significato che vanno sotto il nome di *semantica modellistica* o *model teoretica* (*model theoretic semantics*).

Dal punto di vista formale, la semantica modellistica è caratterizzata dall'impiego di strumenti logico-insiemistici. Nata con i lavori di Tarski sulla semantica estensionale per il calcolo dei predicati del primo ordine, essa ha raggiunto il suo punto di massimo sviluppo con la teoria di Montague (1974), che è basata su una generalizzazione della semantica intensionale dei mondi possibili sviluppata da Saul Kripke per le logiche modali. La tradizione modellistica si fonda su alcuni assunti filosofici di ascendenza fregeana. In particolare, viene accettato l'assunto antipsicologista di Frege, in base al quale i significati linguistici non vanno intesi come entità di natura psicologica o soggettiva. In questo senso, la semantica modellistica si configura come una teoria del significato intesa come teoria delle relazioni fra espressioni linguistiche e stati del mondo (reali o possibili), e non come una teoria della rappresentazione del significato da parte dei parlanti.

Il problema principale che gli atteggiamenti proposizionali pongono alla semantica modellistica può essere ricondotto alla cosiddetta assunzione di *onniscienza logica*. Consideriamo gli atteggiamenti proposizionali di tipo epistemico, come "credere che" e "sapere che". Il loro trattamento modellistico basato sulla semantica dei mondi possibili comporta che ogni soggetto di credenza o di conoscenza sia logicamente onnisciente, cioè che creda (o che conosca) tutte le verità logiche e tutte le conseguenze logiche delle proprie credenze. L'idea intuitiva alla base della semantica dei mondi possibili per gli atteggiamenti epistemici (come è sviluppata ad esempio in Hintikka 1969) è la seguente. A un soggetto epistemico è associato un insieme di mondi, che rappresentano i possibili stati di cose compatibili con ciò che è creduto dal soggetto (qui mi concentrerò sulla credenza; problemi del tutto analoghi si pongono tuttavia per le logiche del conoscere). Così, un enunciato  $\alpha$  è creduto da un soggetto epistemico se e soltanto se  $\alpha$  è vero in tutti i mondi compatibili con le credenze del soggetto stesso. Poiché si assume che i mondi possibili siano strutture semantiche complete e consistenti (in base alle quali cioè a ogni enunciato del linguaggio è assegnato uno ed un solo valore di verità) questo tipo di soluzione comporta che ogni soggetto epistemico sia logicamente onnisciente. Si ha cioè che ogni soggetto crede tutte le formule logicamente valide e tutte le conseguenze logiche delle sue credenze. Consideriamo una logica proposizionale della credenza. Il linguaggio sia un linguaggio proposizionale classico esteso con un operatore modale di credenza  $B$  che assume formule come argomenti: se  $\alpha$  è una formula, allora  $B\alpha$  è a sua volta una formula. Intuitivamente,  $B\alpha$  può essere letta come "è creduto che  $\alpha$ ". In ogni logica della credenza basata sulla semantica a mondi possibili di Kripke vale che:

$$AD. B\alpha \wedge B(\alpha \rightarrow \beta) \rightarrow B\beta$$

per ogni formula  $\alpha$  e  $\beta$  del linguaggio. Risulta inoltre corretta la seguente regola:

$$Nec. da \alpha segue B\alpha,$$

dove  $\alpha$  è un teorema della teoria logica. AD è generalmente assunto come schema di assiomi per la credenza, e viene detto *assioma distributivo*. La regola Nec è detta *regola di necessitazione* (per analogia con la regola corrispondente dei sistemi di logica modale aleatica). Assioma distributivo e regola di necessitazione sono i responsabili, a livello di teoria della dimostrazione, dell'onniscienza logica (la regola di necessitazione fa sì che siano credute tutte le formule valide, e l'assioma distributivo fa sì che le credenze siano chiuse rispetto alla conseguenza logica). In termini semantici, le difficoltà poste dagli atteggiamenti epistemici alla teoria dei mondi possibili possono essere formulate come segue. La teoria dei mondi possibili comporta che gli oggetti della credenza siano *proposizioni*, dove (nella terminologia della logica intensionale) la proposizione associata a un enunciato  $\alpha$  è un insieme di mondi possibili: l'insieme dei mondi in cui  $\alpha$  è vero. Questa caratterizzazione degli oggetti di credenza risulta di grana troppo "grossolana" per un trattamento adeguato degli atteggiamenti proposizionali. Ad esempio, da essa consegue che a due enunciati logicamente equivalenti sia associata sempre la stessa proposizione.

L'onniscienza logica costituisce certamente una idealizzazione difficilmente accettabile ai fini di una teoria del significato. Se ad esempio l'enunciato  $\beta$  segue logicamente dall'enunciato  $\alpha$ , non per questo si vuole accettare che "Giovanni crede che  $\beta$ " segua logicamente da "Giovanni crede che  $\alpha$ ". Si noti che queste conseguenze della semantica a mondi possibili per gli atteggiamenti proposizionali sono indesiderabili anche qualora si assuma che la semantica sia la teoria delle relazioni astratte che intercorrono fra il linguaggio e il mondo, e non una spiegazione della competenza semantica dei parlanti. Infatti, accettare le forti idealizzazioni imposte dalla semantica dei mondi possibili per gli atteggiamenti proposizionali implica non solo che la semantica sia la teoria del significato del linguaggio "come se fosse parlato da Dio", ma anche che tutti i soggetti di credenza siano dei soggetti dotati di capacità logiche "soprannaturali" (Partee 1982).

Il trattamento degli atteggiamenti proposizionali (assieme ad altri aspetti del significato problematici per la semantica modellistica, come ad esempio il significato lessicale) ha indotto alcuni filosofi del linguaggio ad abbandonare almeno in parte le posizioni antipsicologiche di Frege, per avvicinarsi a concezioni del significato orientate in maniera più psicologica. Particolare attenzione è stata dedicata alla cosiddetta *semantica cognitiva*, e, in generale, alle teorie e agli strumenti sviluppati nell'ambito dell'intelligenza artificiale (IA) e delle scienze cognitive (per i rapporti fra teorie filosofiche del significato e semantica cognitiva si veda ad esempio Marconi 1992; per i rapporti fra logica, semantica e IA, si veda Frixione 1994). Si noti che non tutti gli sviluppi dell'IA sono caratterizzati da una forte connotazione psicologica in senso stretto. In ogni caso, anche nelle versioni meno orientate psicologicamente dell'IA, ci si pone comunque il problema di individuare una rappresentazione finita dei significati linguistici, esigenza legata appunto ai fini computazionali della disciplina. Lo scopo è quello di elaborare modelli delle rappresentazioni e delle procedure impiegate da un agente finito (naturale o artificiale), al fine di spiegare o di progettare il suo comportamento (in particolare, per quanto qui ci concerne, il suo comportamento linguistico). Utilizzando una terminologia di origine chomskiana, si può dire che lo scopo della semantica cognitiva è elaborare modelli della *competenza semantica*. Per Chomsky, la *competenza linguistica* consiste in ciò che un parlante conosce della propria lingua. Viene ipotizzato un parlante *idealizzato*, che non è sottoposto a vincoli di natura extra-linguistica quali le limitazioni di memoria, le distrazioni, i cambiamenti di attenzione e di interesse, e così via (in questo senso, la competenza si contrappone all'*esecuzione*, che concerne invece la pratica linguistica effettiva dei parlanti reali). Per quanto idealizzata, tuttavia, la competenza è sempre competenza *di un parlante*. In questo, la nozione di competenza semantica differisce dalla nozione model teoretica di significato, che prescinde completamente dal ruolo dei parlanti.

In IA l'assunzione di onniscienza logica è altrettanto inaccettabile che in semantica. Il ragionamento di tipo epistemico è centrale in molti aspetti della disciplina. D'altro canto, poiché lo scopo dell'IA è quello di modellare computazionalmente le prestazioni inferenziali di soggetti finiti e con risorse limitate, l'assunzione di onniscienza logica risulta del tutto irrealistica.

Nell'ambito dell'IA di impostazione logica il problema è stato quindi variamente affrontato (per una rassegna si vedano, ad esempio, Frixione 1994 e McArthur 1988). Queste proposte presentano varie affinità con le soluzioni prospettate in filosofia (per una rassegna di queste ultime si veda Bäuerle e Cresswell 1988). In generale, esse possono essere viste come tentativi per individuare candidati più adeguati rispetto alle proposizioni modellistiche classiche come oggetti degli atteggiamenti proposizionali.

Una prima classe di proposte prevede l'inserimento di elementi sintattici nella semantica, assumendo ad esempio che gli oggetti della credenza siano formule di un linguaggio di rappresentazione, come in (Konolige 1986). Questi modelli presentano analogie con le proposte di tipo *citazionale* sviluppate in filosofia. La differenza principale consiste nel fatto che le teorie citazionali filosofiche assumono come oggetti della credenza enunciati del linguaggio pubblico, mentre le proposte sviluppate in IA assumono come oggetti di credenza espressioni di un sistema di rappresentazione interno. Questo tipo di soluzioni va incontro a ben note difficoltà di tipo filosofico (si veda ad esempio Cresswell 1980). Una soluzione di tipo sintattico meno radicale e più soddisfacente consiste nell'introdurre una sorta di "filtraggio sintattico" sugli oggetti semantici, come avviene nella *logica della consapevolezza generalizzata* di Fagin e Halpern (1988). Un approccio totalmente diverso rispetto alle soluzioni sintattiche consiste nel sostituire ai mondi possibili classici "mondi" non classici, che si assume possano essere incoerenti (nel senso che in essi una formula può risultare sia vera che falsa) oppure incompleti (nel senso che una formula può risultare né vera né falsa). Questa è la linea adottata da Levesque (1984) nella sua logica della credenza implicita ed esplicita. In generale, questo tipo di soluzioni equivalgono ad assumere che, all'interno dei contesti di credenza, valga qualche tipo di logica non classica (nel caso di Levesque, una logica della rilevanza). Una diversa alternativa consiste nell'adottare versioni "indebolite" della semantica a mondi possibili, come la semantica delle logiche modali minimali (si veda Chellas 1980, cap. 7). Questa via è adottata ad esempio nella *logica del ragionamento locale* di Fagin e Halpern (1988), o, in ambito filosofico, in una logica elaborata da Rescher e Brandom (1980).

Nessuna di tali proposte riesce a fornire una soluzione pienamente soddisfacente e definitiva al problema. Particolarmente interessanti a questo proposito risultano le osservazioni di Fagin e Halpern (1988), secondo i quali le varie proposte avanzate non devono essere viste necessariamente come alternative incompatibili per affrontare un unico problema, ma come tentativi di risolvere aspetti diversi di un problema complesso. Per Fagin e Halpern infatti la mancanza di onniscienza logica nei soggetti epistemici reali è un fenomeno complesso, che dipende da una molteplicità di fattori diversi. Essi individuano i quattro aspetti seguenti.

- *Mancanza di consapevolezza.* Può accadere che un soggetto non abbia opinioni circa la verità o la falsità di un enunciato che segue logicamente dall'insieme delle sue credenze per la semplice ragione che nella formulazione di quell'enunciato compaiono termini che non conosce, di modo che egli non è consapevole della verità o falsità dell'enunciato stesso.
- *Risorse limitate.* Un soggetto razionale può ignorare certe verità logiche, o non conoscere alcune delle conseguenze logiche delle sue credenze perché non dispone del tempo o delle risorse di memoria per dedurle. Oppure perché sono formulate in maniera troppo complessa perché possa comprenderle.
- *Ignoranza di regole di derivazione.* Spesso i ragionatori reali non conoscono o non sanno applicare alcune regole di ragionamento. Ad esempio, alcune ricerche in psicologia cognitiva hanno messo in luce le difficoltà di molti soggetti nell'utilizzare la regola di contrapposizione.
- *Molteplicità dei contesti mentali.* Nel ragionamento i soggetti non utilizzano contemporaneamente tutte le informazioni di cui dispongono. Sembra che gli esseri umani abbiano difficoltà nell'utilizzare contemporaneamente informazioni che provengono da ambiti diversi. Sembra ragionevole pensare la memoria umana come strutturata in contesti diversi, in diversi "quadri mentali", che difficilmente comunicano fra loro.

Ciascuno di questi aspetti può richiedere soluzioni specifiche. In questa prospettiva, Fagin e Halpern presentano tre logiche, nelle quali aspetti diversi del problema vengono affrontati con tecniche diverse: la mancanza di cosapevolezza viene modellata mediante mondi non classici incompleti; per la limitatezza delle risorse vengono introdotti elementi sintattici nella semantica; per la molteplicità dei contesti mentali viene utilizzata una semantica dei modelli minimali.

## 2. L'insorgere dei paradossi epistemici

Esiste tuttavia un secondo ordine di problemi connesso al trattamento dei contesti di atteggiamento proposizionale, che affligge in maniera specifica le teorie del significato di impostazione cognitiva e computazionale. Esso è legato all'insorgere dei cosiddetti paradossi epistemici, e, come vedremo, è, in un certo senso, "di segno opposto" rispetto al problema dell'onniscienza logica. Infatti, l'emergere dei paradossi epistemici sembra comportare che, in una prospettiva cognitiva, un soggetto epistemico, per quanto idealizzato, non possa essere logicamente onnisciente.

I paradossi epistemici sono connessi alla scelta di utilizzare un linguaggio predicativo del primo ordine anziché un linguaggio modale per formulare una logica degli atteggiamenti proposizionali. Ad esempio, anziché introdurre nel linguaggio un operatore modale di credenza  $B$  che assuma come argomenti formule del linguaggio (per cui "è creduto che  $\alpha$ " si esprime mediante la formula  $B\alpha$ ), la credenza viene espressa mediante un predicato ad un posto (ad esempio  $bel(x)$ ), che assume come argomenti *nomi* degli oggetti di credenza. Così, a prescindere dai dettagli tecnici, se " $\alpha$ " è il nome dell'oggetto di credenza espresso dalla formula  $\alpha$ , "è creduto che  $\alpha$ " si esprime mediante la formula  $bel(" \alpha ")$ . Analogamente, la conoscenza viene formalizzata mediante un predicato ad un posto  $know(x)$ . In IA questo modo di procedere è stato sostenuto ad esempio da McCarthy (1979). Le formalizzazioni predicative al primo ordine degli atteggiamenti proposizionali sono, di norma, più espressive dei linguaggi proposizionali o del primo ordine modali. Infatti, poiché gli oggetti degli atteggiamenti proposizionali sono rappresentati mediante termini del linguaggio, diventa possibile quantificare su di essi. Così, ad esempio, è possibile scrivere formule del tipo:

$$\exists x(\alpha(x) \wedge bel(x))$$

oppure:

$$\forall x(\alpha(x) \rightarrow bel(x))$$

(dove  $\alpha(x)$  è una formula qualunque del linguaggio, in cui compare libera la variabile  $x$ ), le quali non hanno corrispettivo in un linguaggio modale del primo ordine.

Tale potenza espressiva comporta però l'insorgere di inconsistenze. Nel caso si utilizzi la tecnica della gödelizzazione come strumento per far riferimento alle formule di una teoria dall'interno della teoria stessa, Montague (1963) ha dimostrato che teorie del primo ordine che includano assiomi corrispondenti a quelli degli usuali sistemi modali per la conoscenza assieme agli assiomi per l'aritmetica elementare sono inconsistenti (Montague faceva riferimento al trattamento delle modalità aleliche, ma il suo risultato è automaticamente estendibile ai contesti epistemici). Consideriamo una teoria del primo ordine  $P_T$  il cui linguaggio comprenda i simboli per l'aritmetica elementare e un predicato a un posto  $know$  (il nome  $P_T$  sta a indicare che questa logica è il "corrispettivo" predicativo della ben nota teoria modale  $T$ ). Assumiamo che, per ogni formula  $\alpha$  del linguaggio di  $P_T$ , " $\alpha$ " sia il numerale che corrisponde al numero di Gödel di  $\alpha$ .  $P_T$  comprenda inoltre gli assiomi per l'aritmetica di Robinson, e in essa, per qualsiasi formula  $\alpha$  e  $\beta$ , siano dimostrabili:

- (i)  $know("α") \rightarrow α$ ;  
(ii)  $know("α \rightarrow β") \rightarrow (know("α") \rightarrow know("β"))$ .

Valga infine la regola:

PNec: da  $α$  segue  $know("α")$ .

Montague dimostra che, a queste condizioni,  $P_T$  è inconsistente. Si noti che (i) è ciò che caratterizza  $P_T$  come una teoria della conoscenza anziché della semplice credenza: se qualcosa è conosciuto, allora deve essere vero. Inoltre, (ii) corrisponde all'assioma distributivo, e PNec è l'equivalente della regola di necessitazione. La fonte dell'inconsistenza sta nel fatto che la rappresentabilità dell'aritmetica in  $P_T$  fa sì che il linguaggio sia autoreferenziale, per cui è possibile derivare in  $P_T$  una formula "paradossale" analoga a quella utilizzata da Tarski per dimostrare l'impossibilità di introdurre in maniera consistente un predicato di verità in un linguaggio dotato di autoriferimento. La dimostrazione del teorema di Tarski si basa sul fatto che in un sistema formale del primo ordine che comprenda il proprio predicato di verità e che sia in grado di far riferimento alle sue stesse espressioni, può essere dimostrata una formula equivalente al paradosso del mentitore, ossia un enunciato che asserisce la propria falsità (vale a dire, qualcosa del tipo: "questo enunciato è falso"). Analogamente, in una formalizzazione al primo ordine di una logica della conoscenza che comprenda l'aritmetica di Robinson, esiste una formula  $α$  per la quale è possibile dimostrare:

$$(iii) \alpha \leftrightarrow know("¬\alpha").$$

Intuitivamente,  $α$  afferma che è conosciuta la sua stessa negazione. Si tratta di quello che è usualmente noto come *the knower paradox*<sup>1</sup>.

La formula (iii), assieme a (i), (ii) e PNec, consente di derivare una contraddizione. In particolare, per derivare l'inconsistenza da (iii), è sufficiente disporre di (i) e di PNec. Infatti, mediante la logica proposizionale, da (i) e da (iii) si ottiene  $α \rightarrow ¬α$ , e quindi:

$$(*) ¬α,$$

da cui, applicando PNec, si ottiene:

$$know("¬\alpha").$$

$know("¬\alpha")$ , in base a (iii), è equivalente ad  $α$ , ed è quindi in contraddizione con (\*). Alternativamente, Montague dimostra che si ottiene l'inconsistenza qualora in una teoria del primo ordine si disponga, oltre che dell'aritmetica di Robinson, di (i), (ii) e di:

- (iv)  $know("α")$ , dove  $α$  è un assioma della logica del primo ordine;  
(v)  $know("know("α") \rightarrow α")$

(dove (iv) e (v), assieme, equivalgono a un indebolimento di PNec).

---

<sup>1</sup>La possibilità di dimostrare (iii) in  $P_T$  è una conseguenza del cosiddetto *lemma diagonale* (*diagonal lemma*) (si veda ad esempio il cap. 15 di Boolos e Jeffrey 1974). Secondo il lemma diagonale, in una teoria del primo ordine che comprenda l'aritmetica di Robinson (e nella quale quindi sono rappresentabili tutte le funzioni ricorsive) vale quanto segue: per ogni formula  $β(x)$  della teoria, con  $x$  come unica variabile libera, esiste una formula chiusa  $α$  tale che è dimostrabile  $α \leftrightarrow β("α")$ .

Il risultato di Montague non dipende dal fatto di utilizzare la gödelizzazione. E' possibile derivare il paradosso ogni qual volta si disponga in una teoria del primo ordine di termini che denotano gli oggetti della conoscenza, e si disponga inoltre di un apparato di operatori sintattici che garantiscano la possibilità di effettuare le necessarie manipolazioni formali sulla struttura di tali termini, come ad esempio un operatore di sostituzione che operi sui nomi delle formule. (Un tale apparato è necessario ad esempio per mettere in relazione un termine come " $\alpha$ " - ossia, il nome della formula  $\alpha$  - con il termine " $\alpha \rightarrow \beta$ " - il nome della formula  $\alpha \rightarrow \beta$ . Nel caso che si adotti una tecnica di gödelizzazione, tali strumenti sono garantiti dalla presenza degli assiomi dell'aritmetica, e dalla conseguente possibilità di rappresentare le funzioni ricorsive nella teoria).

Thomason (1980) ha mostrato che un risultato analogo a quello di Montague può essere esteso a taluni casi di teorie per la credenza espresse mediante un linguaggio del primo ordine. In particolare, sia  $P_{WS5}$  una teoria del primo ordine, il cui linguaggio comprenda i simboli per l'aritmetica e un predicato a un posto *bel*. (Anche in questo caso, il nome  $P_{WS5}$  suggerisce che questa teoria è analoga alla nota logica modale per la credenza detta **S5** debole, o **WS5**). Per ogni formula  $\alpha$  e  $\beta$  del linguaggio, in  $P_{WS5}$  sia dimostrabile quanto segue:

- (i)  $bel("Q")$ , dove  $Q$  è la congiunzione degli assiomi per l'aritmetica di Robinson;
- (ii)  $bel("\alpha \rightarrow \beta") \rightarrow (bel("\alpha") \rightarrow bel("\beta"))$ ;
- (iii)  $bel("\alpha") \rightarrow bel("bel("\alpha)")$ ;
- (iv)  $\neg bel("\alpha") \rightarrow bel("\neg bel("\alpha)")$ ;

e inoltre valga la regola:

PNec': da  $\alpha$  segue  $bel("\alpha")$ .

$P_{WS5}$ , a differenza di  $P_T$ , non è inconsistente, ma, per ogni formula  $\beta$ , si può derivare  $bel("\beta")$ : essa è, per così dire, inconsistente "dentro" i contesti di credenza. E, ovviamente, diventa inconsistente non appena si aggiunga, ad esempio, un assioma come  $bel("\alpha") \rightarrow \neg bel("\neg \alpha")$ .

### 3. E' possibile una competenza semantica?

Prima di procedere ad analizzare la portata dei risultati sopra enunciati per il progetto di una semantica cognitiva è opportuna una precisazione terminologica. I trattamenti logici degli atteggiamenti proposizionali basati sull'uso di predicati anziché di operatori modali vengono talvolta detti trattamenti *di tipo sintattico* (si vedano ad esempio i titoli dei lavori di Montague del 1963 e di Thomason del 1980 riportati in bibliografia). Questo crea qualche ambiguità con quei modelli cui abbiamo fatto cenno nel par. 1, in cui si introducono oggetti di tipo sintattico (ad esempio enunciati) nella semantica di una teoria per gli atteggiamenti proposizionali (come nel caso di Konolige 1986). Anche in questo caso si parla talvolta di *modelli sintattici* degli atteggiamenti proposizionali. Tuttavia, le proposte come quella di Konolige hanno a che fare con il problema di *come interpretare* una teoria degli atteggiamenti proposizionali, e non del tipo di linguaggio adeguato per esprimerla, che è invece il problema posto dalla contrapposizione fra linguaggi modali e linguaggi del primo ordine non modali. Le due istanze, quella che verte sul tipo di linguaggio da adottare e quella che verte su come interpretare gli enunciati di atteggiamento proposizionale, sono fra loro ampiamente indipendenti (Konolige 1986, McArthur 1988). Ad esempio, la logica di Konolige (1986), pur interpretando gli oggetti di credenza come entità di tipo sintattico, utilizza un linguaggio modale (e non incorre quindi nel problema del paradosso).

D'altro canto, utilizzando un linguaggio del primo ordine con predicati per gli atteggiamenti epistemici non è necessario assumere che gli oggetti degli atteggiamenti proposizionali siano oggetti sintattici. Non è necessario assumere cioè che i termini cui si applicano ad esempio predicati come *know* o *bel* denotino enunciati o formule. In ogni caso, il paradosso insorgerebbe anche qualora si assumesse che tali termini denotano oggetti di tipo intensionale, ad esempio proposizioni (Thomason 1980). E' sufficiente che tali oggetti abbiano una struttura compositiva analoga a quella delle formule, o che sia possibile esprimere nel linguaggio a quali formule essi corrispondono. Quindi, l'insorgere del paradosso non è legato direttamente agli approcci di tipo sintattico o citazionale, nel senso dei modelli come quello proposto da Konolige.

I linguaggi modali, proposizionali e del primo ordine, non presentano i problemi legati all'insorgere del paradosso, in quanto il loro potere espressivo è più limitato. E' possibile dimostrare che neppure le formalizzazioni predicative al primo ordine degli atteggiamenti proposizionali danno luogo a inconsistenze qualora il loro potere espressivo venga ridotto a quello dei linguaggi modali (des Rivieres e Levesque 1986). Come abbiamo visto, nelle assiomatizzazioni al primo ordine di tipo modale non è possibile quantificare sugli oggetti degli atteggiamenti proposizionali. Il punto è che per poter quantificare al primo ordine sugli oggetti della credenza (e, in generale, di un atteggiamento proposizionale) si deve poter disporre di termini che denotano gli oggetti della credenza stessa, di "nomi" degli oggetti dell'atteggiamento proposizionale (qualunque cosa tali nomi denotino). Inoltre, perché sia garantita la possibilità di effettuare le necessarie manipolazioni formali sulla struttura di tali termini, deve essere disponibile nel linguaggio un apparato di operatori sintattici opportuni. Ciò dà origine alla possibilità di costruire formule autoreferenziali, e quindi dà luogo al paradosso.

Il problema è che, per fornire una semantica delle espressioni del linguaggio naturale (e per le esigenze della formalizzazione del ragionamento di senso comune), è necessario disporre di un potere espressivo tale da consentire la quantificazione sugli oggetti degli atteggiamenti proposizionali. Ad esempio, si deve poter rappresentare il significato di enunciati del tipo "Mario crede tutto quello che crede Giovanna", o "C'è qualcosa di vero in quello che crede Giorgio" (in quest'ultimo caso si pone anche l'esigenza di rappresentare il predicato di verità all'interno del linguaggio, con tutti i problemi legati al paradosso del mentitore). Una maniera per ottenere il potere espressivo necessario senza incorrere nel paradosso consiste nell'utilizzare una logica modale (che esprima quindi i costrutti epistemici non come predicati, ma come operatori proposizionali), il cui linguaggio comprenda un insieme di variabili (di un tipo sintattico distinto da quello delle variabili individuali), i cui valori varino sull'insieme degli oggetti degli atteggiamenti proposizionali (ad esempio sull'insieme delle proposizioni). In questo modo è possibile quantificare direttamente su proposizioni, e non è necessario introdurre nel linguaggio termini che denotino gli oggetti degli atteggiamenti proposizionali, e meccanismi che consentano l'autoriferimento. Questa è la soluzione adottata nelle logiche intensionali sviluppate nell'ambito della semantica modellistica, come ad esempio nella teoria di Montague. Tuttavia, così facendo si esce dalla logica del primo ordine, e si passa ad una logica di ordine superiore, con le note difficoltà che ne derivano. In particolare, si esclude in questo modo la possibilità di un'assiomatizzazione completa finita della logica.

Ciò non è stato percepito come problematico nell'ambito delle teorie filosofico-formali del significato di tradizione modellistica, dove la semantica viene intesa come la teoria delle relazioni astratte che intercorrono fra linguaggio e mondo, a prescindere da come i significati possano essere accessibili e rappresentabili da parte di un parlante finito. Questa soluzione, viceversa, pone problemi cruciali se ci si pone dal punto di vista di una semantica, in senso lato, cognitiva, per cui diventa parte centrale della teoria rendere conto della *competenza semantica*, intesa come il modo in cui il significato linguistico possa essere rappresentato nella mente di un parlante che, seppure idealizzato, sia comunque finito. Abbandonare la logica del primo ordine comporta di fatto essere costretti a rinunciare alla possibilità di associare una teoria della dimostrazione al trattamento

modellistico dei significati. Comporta, in altri termini, rinunciare al fatto che l'insieme delle conseguenze logiche di una teoria sia generabile in maniera ricorsiva.

Dunque l'esigenza di una rappresentabilità finita dei significati linguistici incontra difficoltà che, dal punto di vista intuitivo, non erano facilmente prevedibili. Thomason (1979, 1986) interpreta i problemi connessi all'insorgere dei paradossi epistemici come un sintomo della difficoltà di conciliare gli obiettivi della semantica filosofica e del programma modellistico da un lato, con gli scopi della semantica cognitiva e dell'IA dall'altro. In particolare, secondo Thomason, è problematico conciliare la tesi cognitivista dell'afferrabilità (*graspability*) dei significati da parte dei parlanti, e la concezione di significato di tipo modellistico. Sembra impossibile interpretare le teorie semantiche alla Montague come modelli della competenza di un parlante idealizzato (ma, ovviamente, finito). Non solo, ma il concetto stesso di competenza idealizzata sembra problematico, in quanto sembra condurre a contraddizione.

In questo senso, Thomason (1979) sembra considerare i paradossi epistemici come una sorta di *reductio ad absurdum* delle ipotesi alla base del programma della semantica cognitiva (e delle scienze cognitive in generale). I risultati di Montague e di Thomason sui paradossi epistemici si basano sull'assunto che valga l'onniscienza logica: essi sono ottenuti a partire dall'ipotesi di disporre degli equivalenti, formulati al primo ordine, dell'assioma distributivo (lo schema (ii) del paragrafo 2) e della regola di necessitazione (PNec), che (come abbiamo visto nel primo paragrafo) sono i responsabili dell'onniscienza logica. Quindi, abbandonando l'onniscienza logica, verrebbe meno anche la possibilità di derivare il paradosso. Ora, è certamente vero che i soggetti epistemici reali non sono logicamente onniscienti. Tuttavia, sostiene Thomason, questo è un fatto puramente contingente. Sembra difficile poter accettare che l'impossibilità dell'onniscienza logica derivi dalla teoria semantica stessa. Sembra difficile cioè accettare che sia una questione di logica, e non una semplice questione fattuale, che non si possa assumere che esistano soggetti epistemici logicamente onniscienti. Lo stesso dicasi per le capacità aritmetiche che consentono di ottenere l'autoriferimento. Non sembra ammissibile che dalla teoria semantica segua che un parlante ideale deve sapere meno aritmetica di quella necessaria per derivare il paradosso. Sembrerebbe dunque messa in discussione la plausibilità di una posizione che tenti di conciliare gli obiettivi della semantica modellistica e della semantica cognitiva, e il progetto stesso di una semantica di tipo cognitivo.

Tuttavia, in anni recenti, sono stati sviluppati diversi tentativi per superare i problemi posti dall'assiomatizzazione delle modalità e delle logiche epistemiche. Lo stesso Thomason, in un lavoro più recente (Thomason 1986), presenta una sintetica rassegna delle diverse strategie possibili: indebolire la logica utilizzata adottando qualche forma di logica non classica, limitare le forme di autoriferimento consentite, e così via. Nello stesso lavoro, Thomason sembra mostrarsi meno drastico rispetto a quanto scriveva nel 1979. Egli afferma ad esempio:

Penso che la reazione corretta sia che la metafora del parlante ideale debba essere riconsiderata quando si assume che la semantica sia una parte della competenza del parlante ideale. Mi sembra che le relazioni fra teorie del linguaggio e parlanti siano molto più problematiche di quanto la metafora suggerisca. In ogni caso, credo che se le conseguenze di aderire all'obbiettivo dell'afferrabilità [del significato] verranno elaborate con attenzione, condurranno a forme forti di costruttivismo, o ad approcci paraconsistenti, o ad altri approcci che comportano una revisione dei fondamenti logici. Ogni valutazione delle forme di cognitivismo forte dovrebbe tenere conto di queste conseguenze. [...] Possiamo aspettarci [...] che verranno sviluppate varie soluzioni tecniche, e che queste ci aiuteranno a chiarire le migliori alternative. (Thomason 1986, pp. 234-5).

Particolare interesse hanno suscitato le soluzioni sviluppate a partire dalle teorie della verità di Kripke (1975) e di Gupta (1982) e Herzberger (1982). Abbiamo visto che i paradossi epistemici presentano forti analogie con i paradossi di tipo semantico, connessi all'introduzione in una teoria di un predicato verità e alla conseguente possibilità di formulare l'antinomia del mentitore. E'

quindi comprensibile che si siano affrontati i paradossi epistemici adottando tecniche analoghe a quelle utilizzate per i paradossi semantici. Dalla teoria di Kripke (nella versione sviluppata da Feferman 1984) prende ad esempio le mosse la proposta di Perlis (1985, 1988). Ai lavori di Herzberger e Gupta si sono ispirati Asher e Kamp (1986). In generale, sui rapporti fra questo tipo di teorie della verità e i paradossi epistemici si veda (Turner 1990).

Tutte queste proposte si traducono di fatto in qualche indebolimento dell'assunzione di onniscienza logica (ad esempio adottando forme indebolite della regola di necessitazione). In altre parole, in queste logiche i soggetti epistemici (per quanto idealizzati) non sono logicamente onniscienti rispetto alla logica classica, anche se risultano onniscienti rispetto a qualche logica non classica più debole (ad esempio, utilizzando le teorie ispirate alla teoria della verità di Kripke, rispetto alla logica di Kleene forte). Abbiamo visto che, secondo Fagin e Halpern, la mancanza di onniscienza logica nei soggetti epistemici reali è un fenomeno complesso, dovuto al concorso di cause differenti. Sembra che, alla luce delle conseguenze dei paradossi epistemici, questa constatazione risulti confermata, e, in un certo senso, rafforzata. Tutti i fattori individuati da Fagin e Halpern (mancanza di consapevolezza, molteplicità dei contesti mentali, e così via) sono fattori di tipo contingente e fattuale, che non riguardano soggetti epistemici idealizzati. D'altro canto, se un soggetto idealizzato non può essere logicamente onnisciente, questo non è dovuto alle stesse ragioni per cui non lo è un soggetto reale. Il fatto stesso che i soggetti epistemici modellati dalle teorie che si ispirano a Kripke e a Gupta ed Herzberger siano logicamente onniscienti rispetto a qualche logica non classica indica che sono comunque soggetti idealizzati (i soggetti reali non sono onniscienti rispetto a una logica non classica più di quanto non lo siano rispetto alla logica classica). Il modo corretto di interpretare le conseguenze dei paradossi epistemici sembra essere quello di ipotizzare che esistano aspetti qualitativamente differenti della mancanza di onniscienza logica, alcuni dei quali sembrano valere "a priori". In base a questi ultimi, un soggetto razionale finito, per quanto idealizzato, non sembra possa essere logicamente onnisciente, almeno rispetto alla logica classica.

## Bibliografia

- Asher, N.M. e Kamp, J.A. (1986). The knower paradox and representation theories of attitudes. *Proc. TARK-86*, 131-147.
- Bäuerle, M.J. e Cresswell, M.J. (1988). Propositional attitudes. In D. Gabbay e F. Guenther (a cura di), *Handbook of Philosophical Logic*, vol. IV, Reidel, Dordrecht e Boston.
- Boolos, G.S. e Jeffrey, R.C. (1974). *Computability and Logic*. Cambridge University Press, Cambridge, UK.
- Chellas, B.F. (1980). *Modal Logic. An Introduction*. Cambridge University Press, Cambridge.
- Cresswell, M.J. (1980). Quotational theories of propositional attitudes. *Journal of Philosophical Logic*, 9, 17-40.
- des Rivieres, J. e Levesque, H. (1986). The consistency of syntactical treatments of knowledge. *Proc. TARK-86*, Monterey, CA, 115-30.
- Fagin, R. e Halpern, J.Y. (1988). Belief, awareness and limited reasoning. *Artificial Intelligence*, 34, 39-76.
- Feferman, S. (1984). Towards useful type-free theories. *Journal of Symbolic Logic*, 49, 75-111.
- Frixione, M. (1994). *Logica, significato e intelligenza artificiale*. Franco Angeli, Milano.
- Gupta, A. (1982). Truth and paradox. *Journal of Philosophical Logic*, 11, 1-60.
- Herzberger, H. (1982). Naive semantics and the liar paradox. *The Journal of Philosophy*, 79, 479-497.
- Hintikka, J. (1969). Semantics for propositional attitudes. In J.W. Davis *et al.* (a cura di), *Philosophical Logic*, Reidel, Dordrecht.
- Konolige, K. (1986). *A Deduction Model of Belief*. Morgan Kaufmann, Los Altos, CA.

- Kripke, S. (1975). Outline of a new theory of truth. *Journal of Philosophy*, 72, 690-715.
- Levesque, H.J. (1984). A logic of implicit and explicit belief. *Atti del Convegno AAAI-84*, Austin, TX, 198-202.
- Marconi, D. (1992). Semantica cognitiva. In M. Santambrogio (a cura di), *Introduzione alla filosofia analitica del linguaggio*. Laterza, Bari.
- McArthur, G.L. (1988). Reasoning about knowledge and belief: a survey. *Computational Intelligence*, 4, 223-243.
- McCarthy, J. (1979). First-order theories of individual concepts and propositions. In J.E. Hayes, D. Michie e L.I. Mikulich (a cura di), *Machine Intelligence*, 9, 120-47, Halsted Press, New York, NY.
- Montague, R. (1963). Syntactical treatments of modality, with corollaries on reflexion principles and finite axiomatizability. In Montague (1974), 286-302.
- Montague, R. (1974). *Formal Philosophy*. Yale University Press, New Haven - Londra.
- Partee, B. (1982). Belief sentences and the limits of semantics. In S. Peters e E. Saarinen (a cura di), *Processes, Beliefs and Questions*. Reidel, Dordrecht e Boston.
- Perlis, D. (1985). Languages with self reference I: Foundations. *Artificial Intelligence*, 25, 301-322.
- Perlis, D. (1988). Languages with self reference II: Knowledge, belief and modality. *Artificial Intelligence*, 34, 179-212.
- Rescher, N. e Brandom, R. (1980). *The Logic of Inconsistency*. Basil Blackwell, London.
- Russell, B. (1921). *The Analysis of Mind*. Oxford University Press. London.
- Russell, B. (1940). *An Enquiry into Meaning and Truth*. George Allen and Unwin, London.
- Thomason, R. (1977). Indirect discourse is not quotational. *The Monist*, 60, 340-354.
- Thomason, R. (1979). Some limitations to the psychological orientation in semantic theory. Manoscritto inedito.
- Thomason, R. (1980). A note on syntactical treatments of modality. *Synthese*, 44, 391-5.
- Thomason, R. (1986). Paradoxes and semantic representation. *Proc. TARK-86*, 225-239.
- Turner, R. (1990). *Truth and Modality for Knowledge Representation*. Pitman, London.