

2. Dalla Semantica Modellistica alla Semantica Cognitiva

2.1 I Limiti del Paradigma Model Teoretico

Esamineremo in questo paragrafo alcune delle difficoltà e dei limiti della semantica di tipo modellistico. Da un lato, esistono infatti problemi semantici rilevanti che sono al di fuori degli scopi della semantica dei mondi possibili (o, più in generale, della semantica modellistica). Essi cadono al di fuori dell'ambito dei fenomeni che il paradigma può spiegare, e la semantica modellistica non può, neppure in linea di principio, fornire una soluzione. Vi sono poi altri problemi che, in linea di principio, rientrerebbero nell'ambito dei fenomeni che la teoria dei modelli dovrebbe spiegare, ma che risultano tuttavia estremamente difficili da affrontare con gli strumenti della teoria dei mondi possibili. Un esempio del primo tipo è costituito dal problema della *semantica lessicale*, ossia il problema di rappresentare il significato dei simboli primitivi extra-logici di un linguaggio. L'esempio più rilevante del secondo tipo è costituito dal ben noto problema degli *atteggiamenti proposizionali*: si consideri, ad esempio, il problema di un trattamento logico adeguato dei contesti epistemic (credere e sapere).

Prendiamo in considerazione per primo il problema del *significato lessicale*. Per semplicità, partiamo dagli aspetti estensionali del significato. Abbiamo visto che la teoria di Tarski consente di calcolare il valore della funzione interpretazione (ossia il riferimento) di espressioni complesse a partire dal riferimento dei simboli primitivi extra-logici. Ad esempio, una volta che la funzione interpretazione φ associa come riferimento alla costante individuale *Fido* un elemento del dominio, e una volta che la funzione interpretazione associa come riferimento al predicato *cane* un sottoinsieme del dominio, allora la semantica tarskiana ci consente di stabilire il riferimento di una formula come *cane(Fido)*. In particolare, si avrà che: $\varphi[\textit{cane}(\textit{Fido})] = v$ sse $\varphi[\textit{Fido}] \in \varphi[\textit{cane}]$. La teoria provvede poi regole per stabilire il riferimento di formule complesse a partire dal riferimento delle formule costituenti. Così, una formula del tipo $\alpha \wedge \beta$ è vera in una interpretazione (φ, D) se e solo se α e β sono entrambe vere in (φ, D) . Ne risulta ad esempio che $(\textit{cane}(\textit{Pluto}) \wedge \textit{nero}(\textit{Pluto}))$ è vera se e solo se $\varphi[\textit{Pluto}] \in \varphi[\textit{cane}] \cap \varphi[\textit{nero}]$. Le regole che concernono le formule quantificate sono più complesse, e le abbiamo omesse per semplicità. Intuitivamente, tuttavia, una formula come $\forall x (\textit{cane}(x) \rightarrow \textit{nero}(x))$ ("tutti i cani sono neri") è vera in (φ, D) se e solo se $\varphi[\textit{cane}] \subseteq \varphi[\textit{nero}]$.

Si noti che una regola semantica come quella per la congiunzione stabilisce quale sia il significato del simbolo " \wedge ". Quindi, le regole della teoria semantica definiscono il significato dei simboli primitivi di tipo logico, come ad esempio i connettivi. Tuttavia, per i simboli primitivi extra logici la funzione interpretazione è assunta come data, e, in ogni caso, caratterizzabile solo in maniera puramente "estensionale", come una tabella che, ad esempio, associ un sottoinsieme del dominio a ogni simbolo predicativo a un argomento. Gli strumenti della semantica modellistica non dicono nulla, ad esempio, su quale sottoinsieme di D φ dovrebbe associare a *cane*, né su quale sia la differenza fra *cane* e un altro predicato a un posto come, ad esempio, *gatto* o *aeroplano*. Tutto quello che si può fare è assegnare esplicitamente a φ i valori per i simboli primitivi extra logici. Se *Fido* e *Marco* sono costanti individuali del linguaggio L , e *cane*, *umano* e *maschio* sono simboli predicatori di L a un argomento, possiamo stabilire ad esempio:

$$\begin{aligned}\varphi[\textit{Fido}] &= \textit{Fido} \\ \varphi[\textit{Marco}] &= \textit{Marco} \\ \varphi[\textit{cane}] &= \{\textit{Fido}, \textit{Pluto}, \textit{Ugo}\} \\ \varphi[\textit{umano}] &= \{\textit{Marco}, \textit{Gianni}, \textit{Maria}\} \\ \varphi[\textit{maschio}] &= \{\textit{Fido}, \textit{Pluto}, \textit{Marco}, \textit{Gianni}, \textit{Ugo}\}\end{aligned}$$

dove "Ugo", "Pluto", "Marco", eccetera (in tondo) non sono simboli del linguaggio L , ma sono simboli che usiamo nella metateoria semantica per indicare oggetti del dominio D . Così, la teoria dei modelli stabilisce il concetto di verità "dato un modello", ma non dice nulla su quali modelli bisogna scegliere per interpretare i simboli primitivi del linguaggio, ossia, quale interpretazione deve essere associata alle costanti predicative e individuali.

Questo tipo di limite della teoria tarskiana è stato messo chiaramente in luce da Field (1972). Field sottolinea come la teoria di Tarski non fornisce una riduzione dei concetti semantici (come quello di verità) ad altri concetti di tipo non semantico. Tarski di fatto riduce la nozione semantica di verità ad altre nozioni semantiche, quali ad esempio quella di denotazione di un termine singolare o di un predicato. La semantica tarskiana spiega le proprietà semantiche di espressioni complesse nei termini delle proprietà semantiche dei loro componenti primitivi. Ma su queste ultime non ha nulla da dire¹. Partee (1979a) distingue fra *semantica lessicale* e *semantica strutturale*. Quest'ultima assolve appunto il compito di ricondurre il significato di espressioni complesse a partire dal significato dei loro componenti primitivi. La teoria tarskiana, e, in generale, vedremo, la semantica modellistica, affrontano il problema della semantica strutturale ma non di quella lessicale.

Da questo punto di vista infatti, la semantica modellistica intensionale (ossia la teoria dei mondi possibili) non offre assolutamente nulla di più rispetto alla semantica modellistica estensionale di Tarski. Infatti la semantica dei mondi possibili non dice alcunché circa l'intensione dei simboli primitivi extra logici. Nel caso di un simbolo primitivo s , la

¹Su questi problemi si veda anche (Bonomi 1983).

funzione Ψ_S che ne costituisce l'intensione è semplicemente assunta come data. Di solito si assume che sia data estensionalmente, in forma di tabella. Cioè, se P è una lettera predicativa a un argomento, allora l'intensione di P è qualcosa del genere:

$$\begin{aligned}\Psi_P \equiv & w_1 \rightarrow \{a, b, c, \dots\} \\ & w_2 \rightarrow \{b, c, d, \dots\} \\ & w_3 \rightarrow \{a, c, d, \dots\} \\ & \dots\end{aligned}$$

dove "a", "b", "c", "d", sono elementi del dominio D . Si tratta, in un certo senso, di un trattamento "estensionale" delle intensioni: nulla è detto su come l'estensione di un simbolo primitivo in un dato mondo possibile possa essere calcolata. Quindi, nonostante la semantica modellistica sia incentrata sulla nozione di significato di enunciato come condizioni di verità, tuttavia essa non è in grado di fornire le condizioni di verità di un enunciato che in maniera "ipotetica": vengono date le condizioni di verità di un enunciato a patto che si assumano come note le intensioni dei simboli primitivi extra logici che vi compaiono.

L'impossibilità di rappresentare il significato dei simboli primitivi extra logici è dovuta dunque a un limite inerente all'approccio modellistico in quanto tale. Di conseguenza, molti ricercatori che si collocano all'interno del paradigma modellistico hanno considerato il problema del significato lessicale come estraneo ai compiti della semantica formale. Thomason (1974) ad esempio, a proposito della teoria di Montague, ha affermato:

i problemi della teoria semantica dovrebbero essere distinti da quelli della lessicografia. E' compito della semantica rendere conto dei significati. Uno scopo centrale di questo resoconto è spiegare come tipi diversi di significati si colleghino a differenti categorie sintattiche [...] *Ma non dobbiamo aspettarci che una teoria semantica renda conto di come due espressioni che appartengono alla stessa categoria sintattica differiscano in significato.* "Camminare" e "correre", ad esempio, e "unicorno" e "zebra" certamente hanno un significato diverso, e ci serve un dizionario per sapere quale. Ma costruire un dizionario richiede una conoscenza considerevole del mondo. [...] Forse anche i quadri di un museo di esemplari rappresentativi dovrebbero essere considerati come "terminologia" [...] In ogni caso, la lessicografia deve prendere in prestito concetti da tutte le aree del sapere e della pratica: astronomia, giurisprudenza, cucina, automobilismo, allevamento dei piccioni, e così via. [...] Queste sono faccende di applicazioni, non di teoria. (pp. 48-49)

Nell'ambito dell'approccio modellistico alla semantica del linguaggio naturale questo tipo di problemi è stato parzialmente affrontato mediante l'introduzione di *postulati di significato*. Questa ad esempio è anche la soluzione adottata nella semantica di Montague (si veda ad esempio Montague 1973). I postulati di significato (Carnap 1952) sono formule che si assume valgano in tutte le interpretazioni di una teoria e in tutti i mondi possibili², con la conseguenza di limitare il numero delle interpretazioni ammissibili dei simboli primitivi extra-logici di un linguaggio. Semplici esempi di formule che possono essere utilizzate come postulati di significato in una teoria estensionale sono le seguenti:

$$\begin{aligned}(1) & \forall x (leone(x) \rightarrow mammifero(x)) \\ (2) & \forall x (mammifero(x) \rightarrow \neg uccello(x)).\end{aligned}$$

In entrambi i casi, assumendo che queste formule valgano in tutti i modelli viene ristretto il numero di interpretazioni che possono essere modelli di una teoria. Ad esempio (1) impone che l'interpretazione di *leone* sia un sottoinsieme dell'interpretazione di *mammifero*; (2) impone che l'interpretazione di *mammifero* sia disgiunta dall'interpretazione di *uccello*. Tuttavia questo non è sufficiente per fissare l'interpretazione dei simboli: nessun insieme di postulati di significato può garantire che l'interpretazione dei simboli primitivi sia quella intesa. Una volta imposti i vincoli espressi da (1) e (2), nulla dice che l'interpretazione di *leone* sia l'insieme dei leoni e che l'interpretazione di *mammifero* sia l'insieme dei mammiferi. Ad esempio, (1) e (2) sono pienamente compatibili con un'interpretazione in cui *mammifero* è interpretato sull'insieme dei treni, *leone* sull'insieme degli elettrotreni e *uccello* sull'insieme delle querce. In generale, dato qualsiasi insieme di postulati di significato ricco a piacere, esiste sempre un numero infinito di interpretazioni "non intese" che lo soddisfa. Il metodo dei postulati di significato può essere esteso facilmente alle logiche di tipo intensionale (in questo caso i postulati di significato devono essere imposti come necessariamente validi, devono cioè vincolare le interpretazioni dei simboli in tutti i mondi possibili). Tuttavia i problemi restano immutati.

A questo proposito, Thomason afferma ancora:

² Carnap non faceva riferimento alla nozione di mondo possibile nella accezione tecnica modellistica attuale (che è più recente), ma al concetto, in senso lato analogo, di *descrizione di stato* (cfr. Carnap 1947).

La teoria potrebbe quindi evolvere verso la concretezza aggiungendo postulati di significato o altri espedienti che consentano di distinguere il trattamento semantico di espressioni che appartengono alla stessa categoria sintattica. Ma una tale evoluzione dovrebbe senza dubbio fermarsi assai prima di ottenere qualcosa che assomigli a un dizionario [...] La teoria semantica non si impegna su quali insiemi di entità [...] siano quelle "corrette", o quelle "intese", per l'interpretazione di un dato linguaggio. (p. 50).

Un ulteriore problema posto dalla rappresentazione del lessico mediante postulati di significato consiste nel fatto che i postulati di significato possono esprimere esclusivamente condizioni necessarie e/o sufficienti per l'applicazione di un concetto. Il punto è che per i concetti che corrispondono a parole del linguaggio naturale è estremamente difficile individuare insiemi di condizioni necessarie e di condizioni sufficienti adeguate allo scopo. Di solito, quello che è possibile fare è individuare insiemi di tratti che caratterizzano un concetto nei casi "tipici", e che tuttavia ammettono numerose eccezioni. In filosofia del linguaggio, l'uso di caratteristiche tipiche nella caratterizzazione di termini del lessico del linguaggio naturale è stato posto in evidenza soprattutto con lo sviluppo della *teoria del riferimento diretto*, o *nuova teoria del riferimento* (Kripke 1972; Putnam 1975). Non è questa la sede per addentrarci in questo argomento³; ci limitiamo a ricordare che, in base a questa teoria, alcuni termini del linguaggio naturale si comporterebbero come *designatori rigidi*, sarebbero cioè collegati al loro riferimento in maniera diretta, senza la mediazione di un senso. Esempi di designatori rigidi sarebbero i nomi propri in senso stretto e i termini per *tipi* (o *generi*) *naturali*, ossia per classi di oggetti dati in natura, come ad esempio le specie biologiche, le sostanze chimiche, e così via. Nel caso dei termini per tipi naturali, afferma ad esempio Putnam, il nome sarebbe stato collegato originariamente al suo riferimento attraverso un "battesimo iniziale" e poi tramandato di generazione in generazione senza che necessariamente i parlanti ne conoscessero il "vero significato". Consideriamo ad esempio il termine "tigre", e supponiamo che il tipo naturale tigre sia caratterizzato da alcune proprietà della struttura genetica. Ora, è evidente che la stragrande maggioranza dei parlanti non conosce tale struttura genetica, e quindi, in questo senso, non conosce il significato di tigre (anzi, per molti secoli, nessuno conosceva il significato di tigre, e potrebbe anche darsi il caso che neppure oggi nessuno lo conosca, se ad esempio le nostre teorie biologiche risultassero sbagliate). In origine, il termine "tigre" è stato "agganciato" al suo riferimento battezzando con questo nome alcuni esemplari appartenenti a quella specie. Lo stesso vale per un termine come "acqua". Supponendo vere le teorie chimiche attuali, "acqua" denota la sostanza chimica la cui formula è H_2O , ma per secoli nessun parlante era al corrente di ciò. In questo senso Putnam afferma che i significati dei termini per tipi naturali non si possono collocare "nella testa" dei parlanti. Nel caso di parole come "tigre" e "acqua", prosegue Putnam, i parlanti normalmente identificano il riferimento utilizzando non il vero significato, ma uno *stereotipo*, ossia una caratterizzazione in termini di un insieme di proprietà tipiche che tuttavia non sono né proprietà necessarie, né sufficienti per caratterizzare il concetto. Ad esempio, la tigre stereotipica ha la forma di un grosso gatto, ha le strisce, vive nella giungla, e così via. Tuttavia, possono esistere esseri che hanno queste proprietà ma non sono tigri, così come delle tigri atipiche potrebbero non godere di queste proprietà. Analogamente, lo stereotipo di acqua comprende proprietà come essere un liquido potabile, incolore ed insapore, ma tali proprietà non si applicano sempre e soltanto all'acqua (ad esempio, il vapore è certamente una forma non stereotipica di acqua, ed infatti per molto tempo non è stato riconosciuto come tale).

E' evidente che la teoria del riferimento diretto per i tipi naturali comporta forti assunzioni filosofiche di tipo realistico. Per chi condivida tali assunzioni, e sia inoltre convinto della necessità di caratterizzare la semantica in termini non soggettivi, ciò che è rilevante ai fini semantici è il riferimento reale dei termini per tipo naturale, e non il prototipo che gli è associato dai parlanti. Per un non realista, tuttavia, potrebbe risultare del tutto chimerico parlare di riferimento reale, e potrebbe essere ovvio considerare lo stereotipo come il *significato* di un termine per tipo naturale. A questo punto, i postulati di significato del tipo tradizionalmente utilizzato nella semantica modellistica risulterebbero del tutto inadeguati a catturare la caratterizzazione di un concetto in termini stereotipici. Vi sono poi alcune considerazioni in base alle quali, anche accettando una posizione realista per i tipi naturali, può risultare rilevante la rappresentazione di stereotipi per una teoria semantica. La prima di tali considerazioni ha a che fare con il problema degli atteggiamenti proposizionali, e la prenderemo in considerazione nel seguito. La seconda considerazione riguarda il fatto che gli stereotipi sono cruciali anche per quanto concerne termini che non denotano tipi naturali. Ad esempio, si considerino espressioni lessicali che denotano classi di manufatti, come ad esempio "sedia", o termini che denotano pratiche sociali ed attività, come, riprendendo un esempio di Wittgenstein, la parola "gioco"⁴. Nessuno si sognerebbe di affermare che si tratta di designatori rigidi, che si riferiscono in maniera diretta a una estensione data in natura indipendentemente dalle nostre pratiche sociali e dalle nostre attività cognitive. Tuttavia, una loro caratterizzazione in termini di condizioni necessarie e sufficienti è estremamente problematica, mentre risulta molto più naturale (ed è forse l'unica soluzione possibile) associare loro un insieme di proprietà stereotipiche. Questo non è tuttavia possibile mediante i postulati di significato tradizionali. Inoltre tutto questo mette in crisi la possibilità di una distinzione netta fra informazioni di tipo

³Una importante raccolta di saggi concernenti il dibattito filosofico su questi temi è (Schwartz 1977), per una introduzione si veda (Napoli 1992).

⁴Si veda (Wittgenstein 1956), § 66.

puramente semantico, che riguardano esclusivamente il significato dei termini, e informazioni di tipo fattuale⁵. Anche l'invarianza dei significati da parlante a parlante verrebbe meno: gli stereotipi e i criteri per fissare il riferimento sono diversi da persona a persona, sulla base della cultura e delle informazioni disponibili ai singoli parlanti (Putnam a questo proposito parla di divisione del lavoro linguistico).

Veniamo ora al secondo problema della semantica modellistica che prenderemo in considerazione. La difficoltà "interna" più importante per la semantica dei mondi possibili concerne gli *atteggiamenti proposizionali*. A partire da Russell (1921; 1940), l'espressione *atteggiamenti proposizionali* (*propositional attitudes*) viene usata nella filosofia contemporanea per indicare quegli atteggiamenti di tipo mentale che riguardano 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 α , dove α è un certo enunciato. I contesti linguistici generati da espressioni per atteggiamenti proposizionali sono contesti indiretti, in cui non vale il principio di sostituibilità *salva veritate*: dato un enunciato come "Marco spera che α " non è detto che sostituendo ad α un altro enunciato β con lo stesso valore di verità, il valore di verità dell'enunciato di partenza non cambi. Tuttavia, per i contesti di atteggiamento proposizionale sorgono ulteriori problemi, in quanto in essi neppure espressioni di eguale intensione risultano di norma sostituibili *salva veritate*⁶. Consideriamo ad esempio i contesti legati alla credenza o alla conoscenza (detti anche *contesti epistemici*). E' plausibile assumere che gli enunciati dell'aritmetica siano veri in tutti mondi possibili. Quindi, due enunciati aritmetici veri hanno la stessa intensione (intensione che in questo caso è la funzione costante che ad ogni mondo possibile associa il valore di verità vero).

Quindi, ad esempio, "2+2=4" e " $\frac{3259^2 \cdot 16^{3527}}{2^3}$ è pari" avranno la stessa intensione. Tuttavia, sembra implausibile che

dalla verità di "Marco sa che 2+2=4" segua la verità di "Marco sa che $\frac{3259^2 \cdot 16^{3527}}{2^3}$ è pari".

Questo fa sì che per i contesti epistemici, e, in generale, per i contesti di atteggiamento proposizionale, i risultati ottenuti con la semantica dei mondi possibili risultino estremamente problematici e poco soddisfacenti. Affronteremo nei dettagli questo tema nel cap. 8. Qui ci limitiamo ad anticipare qualche considerazione generale. Una semantica dei mondi possibili per la logica epistemica e per i contesti di credenza è stata proposta da Jikko Hintikka (1969). L'idea intuitiva alla base della semantica di Hintikka è che ad ogni soggetto di credenza B sia associato un insieme di mondi possibili. Ogni membro di questo insieme rappresenta uno stato di cose compatibile con ciò che è creduto da B . Così, B crede un enunciato α se e soltanto se α è vero in tutti i mondi che egli considera possibili (quindi, nella semantica dei mondi possibili il credere si comporta in maniera analoga alla necessità nelle logiche modali). Questo approccio comporta tuttavia il ben noto problema dell'*onniscienza logica*. Date due formule α e β , supponiamo che $\alpha \rightarrow \beta$ sia una formula valida, ossia che β sia conseguenza logica di α . Ciò vuole dire che ogni interpretazione che rende vera α rende vera a sua volta anche β . Quindi, in ogni mondo possibile in cui sia vera α , anche β è vera (infatti i mondi possibili sono corretti e completi, assegnano cioè uno ed un solo valore di verità ad ogni formula chiusa). Quindi, secondo la semantica dei mondi possibili per la logica epistemica, se qualcuno, diciamo Marco, crede che α , allora α deve essere vera in tutti i mondi possibili compatibili con le sue credenze. Ma, se β è una conseguenza logica di α , allora anche β deve essere vera in tutti i mondi possibili compatibili con ciò che crede Marco. Così, se Marco crede che α , allora deve credere anche che β . In altre parole, ogni soggetto epistemico deve credere tutte le conseguenze logiche delle sue credenze. Questa assunzione è certamente troppo forte per essere accettabile: implica, ad esempio, che se qualcuno conosce gli assiomi di Peano, allora debba conoscere tutti i teoremi dell'aritmetica. Strettamente legato a questo è il problema che, se è creduta una verità necessaria, allora devono essere credute *tutte* le verità necessarie: infatti, tutte le verità necessarie condividono la stessa intensione, e sono quindi sostituibili *salva veritate* in tutti i contesti di credenza. Analogamente, due termini con la stessa intensione risultano sostituibili *salva veritate* in ogni contesto di credenza. Questo è ad esempio il caso dei nomi propri se si accetta la teoria del riferimento diretto. In base a tale teoria, i nomi propri sono designatori rigidi, e denotano quindi lo stesso individuo in tutti i mondi possibili. Dati due nomi propri con la stessa estensione, ne consegue che essi devono avere anche uguale intensione, e risultano quindi sostituibili *salva veritate* in tutti i contesti di atteggiamento proposizionale. "Espero" e "Fosforo" sono nomi propri con la stessa estensione (denotano entrambi il pianeta Venere). Se si assume che siano designatori rigidi, allora essi avranno anche la stessa intensione. Quindi, nella semantica dei mondi possibili, per ogni soggetto di credenza B , da " B crede che Espero = Espero", segue che " B crede che Espero = Fosforo", che è certamente poco plausibile. A prescindere dal fatto che i nomi propri siano o meno designatori rigidi, un problema analogo si pone per quei termini che hanno lo stesso significato "per definizione" o per semplice convenzione linguistica, come ad esempio i predicati "greco" ed "elleno". Essi evidentemente hanno la stessa intensione, in quanto denotano lo stesso insieme di oggetti in tutti i mondi possibili. Tuttavia, ad esempio, qualcuno può credere che Pericle era greco senza credere che Pericle fosse elleno, semplicemente perché sa cosa significa "greco", ma

⁵Si veda ad esempio (Marconi 1989).

⁶ Per questa ragione Cresswell (1975) ha introdotto l'espressione *contesti iperintensionali* (*hyperintensional contexts*) per indicare i contesti di atteggiamento proposizionale.

non sa che cosa significa "elleno". Le intensioni forniscono quindi un livello di analisi insufficiente per un trattamento dei contesti di atteggiamento proposizionale.

Più sopra si è accennato al fatto che, anche dal punto di vista di un realista che accetti la teoria del riferimento diretto per i termini che denotano tipi naturali, i contesti di tipo epistemico, e, più in generale, di atteggiamento proposizionale, forniscono buoni argomenti affinché anche gli stereotipi vengano presi in considerazione all'interno della teoria semantica. Si tratta di un problema per molti versi simile a quello posto dai nomi propri. La tesi che i termini per tipi naturali siano designatori rigidi comporta ad esempio che, se la vera essenza dell'acqua è di essere la sostanza chimica la cui formula è H_2O , l'intensione di "acqua" e l'intensione di "la sostanza chimica la cui formula è H_2O " coincidano (in ogni mondo possibile, queste due espressioni devono avere la stessa estensione). Nella semantica dei mondi possibili questo comporta che ad esempio l'enunciato "Dante sapeva che l'acqua è potabile" implichi l'enunciato "Dante sapeva che la sostanza chimica la cui formula è H_2O è potabile". E' evidente che tale implicazione è difficilmente accettabile (almeno rispetto alla lettura più intuitiva del secondo enunciato). Ed è altrettanto evidente che nessuno, per quanto convinto della validità della teoria realista del riferimento diretto, sarebbe disposto a considerarla valida. Il punto è che, anche se si accettano le posizioni realiste della teoria del riferimento diretto rispetto ai termini per tipi naturali, nell'ambito dei contesti epistemici, o, in generale, nei contesti di atteggiamento proposizionale, ciò che è rilevante per l'analisi semantica non è tanto il *significato* di un termine di tipo naturale, quanto piuttosto lo *stereotipo* ad esso associato, cioè i criteri, incompleti ed eventualmente anche in parte scorretti, utilizzati dai parlanti per fissarne il riferimento.

Si noti che le 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 parlanti⁷. Infatti, accettare le forti idealizzazioni imposte dalla semantica dei mondi possibili per gli atteggiamenti proposizionali implica non solo che si stia descrivendo una competenza semantica "idealizzata" (cioè, 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 di credenza idealizzati" (Partee 1982).

In generale, sembra che questi problemi traggano origine dal fatto che la semantica dei mondi possibili fornisca una caratterizzazione del significato di grana troppo "grossolana" per riuscire a rendere conto dei contesti di atteggiamento proposizionale. Sembrerebbe necessario un livello di analisi semantica più fine, in base al quale, ad esempio, due enunciati diversi ma logicamente equivalenti possano risultare distinguibili a livello semantico (mentre, come abbiamo visto, nella semantica a mondi possibili ad essi viene associata la stessa intensione).

E' interessante a questo proposito (anche nell'ottica di comprendere in che misura le intensioni della semantica modellistica forniscano una formalizzazione adeguata del concetto di senso fregeano) cercare di comprendere a che livello di analisi Frege collocasse il concetto di senso, e quanto, in base alla concezione originale fregeana, i sensi fossero adatti a fornire una risposta a questo tipo di problemi. Tuttavia, l'analisi delle opere di Frege rivela a questo proposito una certa ambiguità (cfr. ad esempio Casalegno 1992). Per un verso Frege sembra concepire i sensi come entità più articolate rispetto alle intensioni della semantica modellistica. Ad esempio egli afferma che:

⁷La nozione di *competenza* trae origine dalla teoria linguistica chomskiana, ed è stata elaborata da Chomsky in riferimento a problemi di carattere sintattico piuttosto che di carattere semantico. Secondo Chomsky la competenza linguistica consiste in ciò che un parlante conosce della propria lingua. Si tratta tuttavia di un parlante *idealizzato*, e, in questo senso, la nozione di competenza viene contrapposta alla nozione di *esecuzione*, che concerne invece la pratica linguistica effettiva dei parlanti reali. A differenza della competenza, l'esecuzione è sottoposta a vincoli di natura extra-linguistica. "La teoria linguistica si occupa principalmente di un parlante-ascoltatore ideale, in una comunità linguistica completamente omogenea, il quale conosce perfettamente la sua lingua e non è influenzato da condizioni grammaticalmente irrilevanti quali le limitazioni di memoria, le distrazioni, i cambiamenti di attenzione e di interesse, e gli errori (casuali o caratteristici) nell'applicazione della propria conoscenza della lingua nel corso dell'esecuzione effettiva [...]. Per considerare l'esecuzione linguistica effettiva, dobbiamo considerare l'interazione di vari fattori, e la competenza sottostante del parlante-ascoltatore non è che uno di essi." (Chomsky 1965, pp. 44-45 della trad. it., corsivi nostri). Per quanto idealizzata, tuttavia, la competenza è sempre competenza *di un parlante*. Nel testo quindi contrapponiamo la nozione di "competenza semantica" non a una nozione di esecuzione, ma alla nozione di significato model teorica classica, caratterizzata esclusivamente nei termini delle relazioni fra espressioni linguistiche e loro significati, prescindendo completamente dal ruolo dei parlanti. Vi è tuttavia chi è propenso a considerare la semantica modellistica come una teoria della competenza semantica (Cresswell 1978; si veda anche Chierchia 1992). Cresswell identifica nella semantica modellistica un possibile candidato per una teoria della competenza semantica in quanto ritiene che concetti come quello di verità e di riferimento siano irrinunciabili per caratterizzare il concetto stesso di competenza semantica. Pur concordando su questo punto, sosterremo una posizione in base alla quale una teoria della competenza impone vincoli più restrittivi rispetto alla concezione model teorica di significato (cap. 12).

L'enunciato può essere concepito come una rappresentazione del pensiero, nel senso che al rapporto fra parti e tutto che sussiste fra il pensiero e le sue parti corrisponde, nel complesso, il rapporto che intercorre fra l'enunciato e le sue parti. (Frege 1969, p. 400 della trad. it.).

Sembra quindi che i pensieri (cioè i sensi degli enunciati) abbiano per Frege una struttura che sia in qualche modo isomorfa alla struttura sintattica dell'enunciato, il che non accade appunto per le intensioni model teoretiche, in cui enunciati di forma sintattica totalmente diversa possono avere la stessa intensione. Altrove tuttavia Frege sembra concepire i pensieri in maniera del tutto analoga alle intensioni della semantica dei mondi possibili. Casalegno (1992) cita ad esempio il seguente passo di una lettera di Frege a Husserl del 1906:

In logica bisogna risolversi a considerare enunciati equivalenti come diversi solo nella forma. Enunciati equivalenti hanno [...] qualcosa di comune nel contenuto: chiamo questo qualcosa il *pensiero* da essi espresso. (Frege 1976, p. 82 della trad. it., corsivo nostro).

La semantica intensionale ha sviluppato quindi soprattutto questa seconda accezione del concetto fregeano di senso.

Una prima proposta nella direzione di individuare distinzioni di significato più sottili di quelle a livello intensionale, e che sembra porsi direttamente sulla linea della prima delle due citazioni fregeane sopra riportate, è stata fornita da Carnap (1947) con il concetto di *isomorfismo intensionale*. In breve, perché due espressioni complesse di un certo linguaggio siano intensionalmente isomorfe non è sufficiente che abbiano la stessa intensione, ma devono essere costruite nello stesso modo a partire da elementi primitivi equiintensionali. Ad esempio, due formule α e $\alpha \wedge (\beta \vee \neg \beta)$ hanno la stessa intensione (sono infatti logicamente equivalenti), ma non sono intensionalmente isomorfe, in quanto hanno una struttura diversa (ad esempio nella seconda appare la formula β che non appare nella prima). Di fatto, la teoria dell'isomorfismo intensionale è un modo per inserire degli elementi sintattici nella semantica di un linguaggio. Uno dei problemi di questa strategia consiste tuttavia nel fatto che non funziona quando due enunciati differiscono esclusivamente per l'uso di simboli primitivi diversi, ma con la stessa intensione. Ad esempio se "greco" ed "elleno" sono sinonimi e quindi dotati della stessa intensione, allora "Pericle è greco" e "Pericle è elleno" sono intensionalmente isomorfi, ma qualcuno potrebbe benissimo credere il primo senza credere allo stesso tempo il secondo (se, come si è già detto, sapesse cosa significa "greco" ma non sapesse cosa significa "elleno").

Dopo Carnap, e in parte a partire dalla sua proposta dell'isomorfismo intensionale, sono stati sviluppati diversi tentativi di risolvere il problema degli atteggiamenti proposizionali in ambito in senso lato model teoretico⁸. Su alcune di queste proposte torneremo nella terza parte. In ogni caso, nessuna soluzione globale soddisfacente è stata ancora individuata.

2.2 Verso la Semantica Cognitiva

Le difficoltà e i limiti della semantica modellistica, primi fra tutti quelli relativi agli atteggiamenti proposizionali, hanno fatto sì che, a partire grosso modo dalla metà degli anni settanta, alcuni filosofi del linguaggio ritenessero troppo vincolanti e riduttive le assunzioni fortemente antipsicologiche proprie della tradizione fregeana e model teoretica, e si avvicinassero ad altre concezioni della semantica, di ispirazione, in senso lato, psicologica. Nel caso degli atteggiamenti proposizionali, la motivazione di questo spostamento di interessi è in un certo senso evidente. Già a livello di senso comune, non sembra plausibile che si possa rendere conto del significato di espressioni come "Tizio crede che ..." o "Caio sa che ..." senza tenere conto in qualche modo di che cosa succede "nella testa" di Tizio e di Caio. Per poter inferire "Carlo crede che α " da "Carlo crede che β " sembra plausibile che si debba fare riferimento a un modello di qualche tipo delle capacità inferenziali di Carlo. I problemi incontrati dalla semantica modellistica in questo settore sembrano confermare questa prospettiva. Per quanto riguarda la rappresentazione del significato lessicale, la sostanziale povertà ed inadeguatezza degli strumenti offerti dalla semantica modellistica ha fatto sì che si guardasse con interesse al lavoro svolto in psicologia cognitiva nell'ambito della rappresentazione concettuale⁹. Il problema della rappresentazione di concetti mediante tratti tipici anziché mediante condizioni necessarie e/o sufficienti, rispetto al quale i tradizionali postulati di significato sono totalmente insufficienti, era stato affrontato da psicologi come Eleanor Rosch (si veda ad esempio Rosch 1975). In base alle teorie della Rosch, il significato associato a livello cognitivo a un grande numero di parole può essere ricondotto a un *prototipo*, ossia alla rappresentazione di un esemplare tipico, di una sorta di "campione" della classe di oggetti denotata. Ad esempio, il prototipo associato alla parola "uccello" potrebbe essere un esemplare di passero, ma non un pinguino o uno struzzo, il prototipo associato alla parola "albero" potrebbe essere una quercia, ma non un abete o una palma (anche se, evidentemente, le cose andrebbero diversamente per un lappone o per un nativo di un'isola dei Caraibi). Sono evidenti le analogie fra il prototipo e il concetto di stereotipo in Putnam. Inoltre, la constatazione della sostanziale impossibilità di una distinzione netta fra significato linguistico e conoscenza fattuale

⁸Per una rassegna di queste varie proposte si veda ad esempio (Bäuerle e Cresswell 1988).

⁹Si vedano ad esempio (Medin e Smith 1984); (Tabossi 1985).

nella semantica lessicale (e la stessa dipendenza dei giudizi di tipicità dal bagaglio cognitivo individuale) ha fatto sì che molti ritenessero impossibile una rappresentazione del significato lessicale che prescindesse da tale bagaglio cognitivo e da una rappresentazione della competenza semantica globale dei singoli parlanti.

L'interesse dei filosofi del linguaggio per le teorie semantiche di indirizzo psicologico si è orientato verso quell'ambito di studi che prende il nome di *semantica cognitiva*¹⁰. La semantica cognitiva non si configura come una teoria o come un progetto di ricerca unitario, quanto piuttosto come una serie di ricerche, di proposte e di modelli che condividono un certo numero di assunzioni comuni. Una prima maniera di individuare la semantica cognitiva consiste nell'identificarla con quella parte della scienza cognitiva che concerne la semantica delle lingue naturali. La *scienza cognitiva*, a sua volta, è un ambito interdisciplinare che non si identifica con una teoria unitaria e consolidata, ma si configura piuttosto come un crocevia di ricerche e di interessi che condividono l'assunto della rilevanza degli strumenti e dei modelli computazionali per la comprensione della mente e del pensiero. Emersa nel corso degli anni settanta, vi confluirono ricerche svolte, oltre che in intelligenza artificiale, in discipline quali la psicologia cognitiva, la linguistica, la filosofia del linguaggio e della mente, le neuroscienze, la sociologia, l'antropologia e la biologia¹¹. Nell'introduzione a un volume che raccoglie i saggi presentati a un convegno che per molti versi sancì l'affermarsi della scienza cognitiva, Donald Norman (1981) afferma che l'obiettivo della scienza cognitiva è di comprendere la mente in astratto e nei suoi aspetti generali, sia essa naturale o artificiale, umana o non umana, reale o ipotetica. Caratterizzare in maniera più precisa la scienza cognitiva risulta problematico, proprio per l'estrema varietà dei contributi disciplinari che vi confluiscono. In generale, si può dire che la scienza cognitiva, nella sua forma più "tradizionale", assume che le attività cognitive possano essere ridotte, spiegate o simulate nei termini di computazioni effettuate su rappresentazioni di tipo simbolico. Ad esempio, Norman (1981) afferma che "il tema centrale del libro è il ruolo della computazione e della manipolazione simbolica nella cognizione" (ibid., p.3). Approfondiremo nel prossimo capitolo cosa si debba intendere in questo contesto per rappresentazione e computazione di tipo simbolico. La centralità delle rappresentazioni simboliche per una teoria della mente non è accettata in maniera universale da tutti gli scienziati cognitivi. Soprattutto in anni recenti, molti hanno messo in discussione la natura simbolica e, più in generale, rappresentazionale della mente e dei processi cognitivi¹². Tuttavia, si può affermare che la dottrina simbolico-rappresentazionale costituisce la forma "classica" della scienza cognitiva. Tornando alla semantica cognitiva, in essa i significati vengono ricondotti appunto a rappresentazioni di questo genere: il significato di un'espressione linguistica è spiegato nei termini delle rappresentazioni che i parlanti costruiscono ed elaborano nel corso del processo di comprensione. Possiamo introdurre a questo punto una caratterizzazione più precisa del settore della semantica cognitiva, riportando le seguenti tesi individuate da Diego Marconi (1992a) come caratteristiche del paradigma cognitivo in semantica:

1. la semantica dev'essere una teoria della comprensione, e la comprensione è un'attività mentale;
2. la semantica modellistica, quali che ne siano i meriti, non è una teoria della comprensione;
3. una teoria semantica adeguata deve rispondere alla domanda: che cosa viene *costruito* quando una frase viene compresa?
4. la comunicazione è resa possibile dalla somiglianza o convergenza delle costruzioni mentali, non dall'esistenza di entità astratte, oggettive e pubblicamente accessibili (sensi, intensioni, ecc.);
5. le costruzioni mentali non hanno bisogno a loro volta di interpretazione (in senso modellistico), perché *sono* interpretazioni (Marconi 1992a, pp. 431-432).

Va precisato che tali tesi forniscono non tanto un insieme di condizioni necessarie e sufficienti che caratterizzano il concetto "semantica cognitiva", quanto piuttosto una serie di "tratti stereotipici", che possono ammettere diverse eccezioni rispetto alle posizioni dei vari ricercatori che operano nel settore. In particolare, il punto 5. non è accettato universalmente. Fra coloro che lo condividono, la posizione più estrema è probabilmente quella del solipsismo metodologico sostenuto da Jerry Fodor in (Fodor 1980a), il quale nega ogni rilevanza dei concetti semantici (sia intensionali che estensionali) per la caratterizzazione delle rappresentazioni mentali. Per una posizione di tipo diverso si veda però, ad esempio, (Harman 1987). O anche Woods (1980), che, parlando di "un linguaggio interno, che precede sia logicamente che cronologicamente lo sviluppo del linguaggio esterno, e nei cui termini i significati delle espressioni del linguaggio esterno sono definite" afferma: "questo stesso linguaggio interno richiede una semantica, e sosterrò che senza

¹⁰ Sul rapporto fra semantica modellistica e semantica cognitiva, l'introduzione a (Peters e Saarinen 1982), (Santambrogio e Violi 1986) e (Marconi 1992a) costituiscono utili saggi introduttivi; (Peters e Saarinen 1982) e (Eco et al. 1986) sono importanti raccolte di saggi sull'argomento.

¹¹ Sulla scienza cognitiva si vedano (Bara 1990); (Collins e Smith 1988); (Gardner 1985); (Johnson-Laird 1988); (Tabossi 1988).

¹² Per una panoramica del recente dibattito su questi temi nel settore dell'IA si veda ad esempio (Kirsh 1991).

una qualche comprensione della sua semantica, non si può avere una spiegazione semantica completa del linguaggio esterno" (p. 304-5). Nel seguito prenderemo in esame la linea "logicista" in IA, che risulta certamente rilevante per il problema della semantica delle lingue naturali nel filone, in senso lato, cognitivo. Come vedremo, i logicisti in IA negano esplicitamente il punto 5¹³.

In generale, è pur vero tuttavia che, nelle teorie semantiche di tipo cognitivo, l'enfasi viene posta più sulla comprensione degli enunciati, intesa come costruzione di rappresentazioni interne, che non, ad esempio, sulle loro condizioni di verità. Analogamente, le relazioni inferenziali fra enunciati non vengono di norma ricondotte a nessi di conseguenza logica, ma a elaborazioni computazionali effettuate sulle rappresentazioni. Tali elaborazioni vengono effettuate da *procedure* computazionali, da cui il termine *semantica procedurale*, che caratterizza molti degli approcci alla semantica cognitiva, sviluppati sia in ambito di IA che in ambito più propriamente psicologico. Il termine semantica procedurale viene impiegato con significati diversi. Sulle diverse accezioni di procedura e di semantica procedurale si veda ancora (Marconi 1992a). Qui ci limitiamo a ricordare, per il momento, che, in IA, il termine semantica procedurale è stato diffuso da Winograd (1973), e poi ripreso ed approfondito, ad esempio, da Woods (1980). In ambito psicologico, il modello procedurale più noto e autorevole è probabilmente la teoria dei *modelli mentali* di Philip Johnson-Laird (Johnson-Laird 1982 e 1983; Johnson-Laird et al. 1984). Secondo Johnson-Laird esistono due livelli di rappresentazione semantica. Il primo livello consiste in ciò che lui chiama una rappresentazione "intensionale", cioè una rappresentazione simbolica del tipo, ad esempio, che si può ottenere mediante una rete semantica (cfr. infra, capp. 4-7; si noti che questo uso del termine "intensionale" non ha nulla a che fare con le intensioni della tradizione modellistica formale: qui si ha a che fare con un piano puramente sintattico di rappresentazione). Il secondo livello consiste di una rappresentazione "estensionale" che può essere generata ed aggiornata ricorsivamente a partire dalla rappresentazione simbolica "intensionale". Tale rappresentazione estensionale è un *modello mentale* del mondo, costruito sulla base della memoria, della percezione e dell'immaginazione. Un modello mentale è una rappresentazione analogica, con una struttura che in qualche modo "corrisponde" al mondo reale rappresentato. Così, la rappresentazione mentale di un concetto consiste in parte delle procedure che rendono possibile costruire ed elaborare un modello mentale: "per esempio, la semantica dell'espressione 'a destra di' specifica la direzione da seguire per formare un modello mentale di un'asserzione come 'A è alla destra di B'. Il sistema interpretativo colloca un'entità in un vettore spaziale, e quindi lo percorre nella direzione appropriata rispetto al punto di vista dell'osservatore in maniera da localizzare la seconda entità" (Johnson-Laird et al. 1984, pp. 311-312).

La semantica cognitiva è stata concepita da alcuni suoi sostenitori in contrapposizione esplicita, e con intenti dichiaratamente polemici, rispetto alla semantica modellistica di tipo filosofico. Ad esempio Johnson-Laird (1982) sottolinea i seguenti punti di disaccordo fra semantica modellistica e semantica di tipo psicologico/cognitivo. Innanzi tutto il problema del *lessico*. Abbiamo visto che per la tradizione modellistica il problema del lessico è al di fuori degli scopi della semantica, mentre per la semantica cognitiva essa è parte integrante di una teoria del significato. Inoltre, le posizioni di Putnam sui designatori rigidi postulano che il significato di certe classi di termini del linguaggio non siano "nella testa" dei parlanti - siano quindi, in un certo senso, inconoscibili. Per Johnson-Laird invece il significato di tutti i termini lessicali, compresi quelli per tipi naturali, vanno identificati con entità di tipo mentale. In secondo luogo Johnson-Laird contrappone il concetto di modello mentale al concetto di *mondo possibile*. In semantica modellistica il significato delle varie classi di espressioni dipende da ciò che è vero in un insieme di norma infinito di mondi possibili. Tuttavia gli esperimenti degli psicologi hanno mostrato che gli esseri umani hanno grandi difficoltà a prendere in considerazione anche un piccolo numero di alternative contemporaneamente. Quindi Johnson-Laird ipotizza che gli esseri umani costruiscano un solo modello mentale alla volta, rivedendolo e modificandolo quando si rivelasse inadeguato. D'altro canto i mondi possibili e i modelli della semantica formale sono strutture *complete*, nel senso che assegnano un riferimento ad ogni espressione linguistica. Viceversa, i modelli mentali sono modelli parziali, che assegnano un valore di verità soltanto ad alcuni degli enunciati del linguaggio. Infine, Johnson-Laird sottolinea il problema degli atteggiamenti proposizionali, che, afferma, non può essere affrontato senza tenere conto delle limitazioni psicologiche dei singoli agenti.

Alcune delle obiezioni mosse dagli scienziati cognitivi alla semantica modellistica sono chiaramente inconsistenti e basate su fraintendimenti. Barbara Partee (1979b) esamina ad esempio l'obiezione di parte cognitivista in base alla quale le intensioni (intese come funzioni da mondi possibili ad estensioni) non sarebbero plausibili da un punto di vista psicologico in quanto oggetti infiniti di tipo insiemistico, e quindi non elaborabili da soggetti finiti. Tuttavia questa obiezione si basa sulla confusione fra il fatto di conoscere una funzione, nel senso di conoscere un metodo per calcolarne i valori, e conoscere l'intero decorso dei valori della funzione stessa (che appunto, può essere un oggetto insiemistico infinito). In base a questa obiezione non sarebbe ad esempio possibile conoscere le operazioni aritmetiche, come la somma, perché anch'esse hanno un decorso di valori infinito. Sembra che l'obiezione di Johnson-Laird

¹³ Nel par. 3.2 vedremo che i logicisti richiedono esplicitamente che, affinché le rappresentazioni interne di un sistema computazionale possano essere considerate effettivamente delle *rappresentazioni*, deve essere associata loro una semantica modellistica (McDermott 1978). A ciò si può obiettare che questo non costituisce ancora una condizione sufficiente, in quanto la semantica modellistica, come abbiamo visto, non rende conto del significato delle costanti primitive non logiche.

sull'implausibilità cognitiva dei mondi possibili si fonda anch'essa su di un fraintendimento di questo genere. In realtà modelli mentali e modelli della semantica model teoretica si collocano su piani totalmente diversi. Ad esempio, è possibile costruire un dimostratore automatico di teoremi per una logica modale, ed implementarlo su un calcolatore; dopo di che, si può assegnare una semantica dei mondi possibili al linguaggio utilizzato in maniera da rendere conto delle prestazioni del dimostratore stesso (provando, poniamo, che il dimostratore è corretto rispetto a quella semantica). In questo modo, la semantica dei mondi possibili può fornire un resoconto adeguato del comportamento del dimostratore di teoremi. Ciò però non implica certamente che i mondi possibili, o qualche versione finita di essi, stiano "nella memoria" del calcolatore. In altri termini, quand'anche la semantica modellistica fosse pensabile come un modello psicologicamente adeguato delle prestazioni linguistiche umane, questo non comporterebbe che le intensioni, o i modelli, o i mondi possibili sono qualcosa che sta nella mente dei parlanti. Questi fraintendimenti poggiano probabilmente su accezioni diverse del termine "semantica", accezioni che Woods ad esempio esamina in (Woods 1975 e 1980), e che prenderemo in considerazione nel seguito (par. 4.4). Tutto ciò non toglie ovviamente che il concetto di modello mentale sia interessantissimo dal punto di vista psicologico o per una teoria del significato. Esso si colloca però a livello di formalismo di rappresentazione, a livello sintattico quindi, cui, eventualmente, se lo si ritenesse utile per ragioni di ordine metateorico o di ordine filosofico o di altro genere, si potrebbe forse assegnare una semantica, di tipo modellistico o di qualche altro tipo.

Per quanto riguarda il problema dei rapporti fra semantica modellistica e semantica cognitiva, Partee (1982) ha individuato due possibili atteggiamenti, che ha chiamato rispettivamente *posizione degli obiettivi comuni* (*common goal position*) e *posizione separatista* (*separatist position*). In base alla prima, semantica modellistica e semantica cognitiva possono essere considerate parti di un'unica impresa di ricerca, che può condividere obiettivi e strumenti. Barbara Partee stessa condivide una prospettiva di questo genere. Una posizione degli obiettivi comuni in IA è quella di Woods (1980), che, parlando di semantica procedurale, afferma ad esempio:

Io vedo l'approccio procedurale al problema della semantica non come un'alternativa alla più tradizionale spiegazione model teoretica tarskiana, ma piuttosto come un mezzo per completare tale spiegazione con quello che nella terminologia informatica si chiamerebbe un'estensione "compatibile verso l'alto". Cioè, io vedo la spiegazione tarskiana come un caso particolare di spiegazione procedurale - nella quale le procedure coinvolte sono le definizioni dei quantificatori e dei connettivi logici in quanto procedure per assegnare valori di verità a proposizioni complesse come funzione dei valori di verità dei loro costituenti. (Woods 1980, p. 317)

In base alla posizione separatista invece semantica modellistica e semantica cognitiva assumono punti di vista completamente distinti sul linguaggio, e non esiste quindi una reale possibilità di interazione fra i due progetti. Una posizione di questo genere è ad esempio quella di Cresswell¹⁴ che, da un lato, considera irrinunciabile per la semantica filosofica il riferimento a concetti come quelli di verità e di denotazione, e dall'altro rifiuta alle rappresentazioni mentali qualsiasi ruolo in una teoria filosofica del significato. Una posizione simile è quella di Putnam (1978), che afferma: "la teoria della comprensione del linguaggio e la teoria del riferimento e della verità hanno molto meno a che fare l'una con l'altra di quanto molti filosofi hanno assunto". Una terza posizione, chiamata *abbandonismo* viene identificata da Marco Santambrogio e Patrizia Violi (1986). Mentre sia i separatisti che i teorici degli obiettivi comuni concedono una piena legittimità teorica e scientifica ad entrambe le imprese, della semantica modellistica e di quella cognitiva, gli abbandonisti sostengono che l'intero progetto della semantica modellistica poggia su basi errate e inconsistenti. Una posizione di questo genere è sostenuta da alcuni scienziati cognitivi, come ad esempio Ray Jackendoff (1986).

In ogni caso, la difficoltà di articolare in maniera precisa i rapporti fra semantica modellistica e semantica cognitiva deriva in gran parte dalla natura stessa della seconda (e della scienza cognitiva più in generale), che si presentano come un fenomeno estremamente variegato, sia per gli strumenti utilizzati, sia per la diversa formazione dei ricercatori, sia per gli obiettivi che vengono perseguiti. L'interesse per gli aspetti psicologici, ad esempio, è un fattore estremamente variabile. L'enfasi sull'adeguatezza empirica dal punto di vista psicologico dei modelli proposti è un fattore che varia molto di caso in caso, e talvolta non viene affatto presa in considerazione. La stessa cosa vale per gli aspetti computazionali: sebbene la concezione computazionale della mente sia un'idea guida di tutta la ricerca in scienza cognitiva, non ogni teoria produce modelli formulati in termini algoritmici, oppure implementati (o implementabili) su calcolatore. Il grado di formalizzazione e i tipi di strumenti formali impiegati sono estremamente diversi. Di questo complesso panorama, nel seguito prenderemo in considerazione l'IA di tipo simbolico, e in particolare quella linea di ricerca dell'IA simbolica che va usualmente sotto il nome di linea "logicista", che propugna la centralità degli strumenti logico formali nell'IA. Chiariremo meglio nel prossimo capitolo queste posizioni, limitandoci qui a motivare brevemente questa scelta. Dal nostro punto di vista, il filone logicista in IA è particolarmente interessante per diverse ragioni. Innanzi tutto, si è dimostrato essere uno dei filoni più fecondi della ricerca nelle scienze cognitive, ed anche quello in cui pare emergere qualcosa che più assomiglia ad un paradigma unitario, ad un progetto globale, con una linea di sviluppo progressiva e "cumulativa", e non una serie di modelli, di singoli esempi isolati. Questo ha fatto sì che alla

¹⁴ Si veda ad esempio (Cresswell 1982).

lunga si imponesse come il paradigma dominante nella ricerca della cosiddetta IA simbolica (almeno del punto di vista della sociologia della ricerca). Ciò è indubbiamente legato al fatto di proporre una metodologia generale di ricerca, o almeno di evidenziare l'esigenza di individuarne una. Inoltre, il fatto che la metodologia utilizzata si rifaccia agli strumenti della logica formale può fare sperare in una maggiore continuità con il lavoro svolto nell'ambito della semantica formale di tipo filosofico. D'altro canto, proprio il fatto di fare riferimento agli stessi strumenti formali può fare sorgere il dubbio che ci si trovi di fronte a una sorta di "circolo vizioso": i filosofi, per affrontare i problemi insoluti posti dalla semantica formale di tipo logico, si rivolgono verso i modelli computazionali dell'IA, e contemporaneamente l'IA, per fare fronte alle proprie esigenze di rigore metodologico, utilizza i metodi della logica. Il nostro scopo sarà appunto quello di spezzare almeno in parte tale circolo, ponendo in evidenza i possibili contributi dell'IA logicista alla teoria logica del significato, e i rapporti che intercorrono fra le due discipline. Prima di proseguire, è tuttavia necessario fare un passo indietro, ed esaminare più a fondo gli assunti generali dell'IA di tipo simbolico, quelli dell'IA del filone "logicista" in particolare, il dibattito fra sostenitori della logica e suoi oppositori, e, in generale, il posto in cui si colloca l'approccio logico nell'ambito dell'IA. Questo sarà il tema del capitolo seguente.