FILOSOFIE ANALITICHE / METAFISICA

n. 12

http://www.filosofiaanalitica.blogspot.com
Collana diretta da *Massimiliano Carrara*

Ciro De Florio

LA FORMA DELLA VERITÀ

Logica e filosofia nell'opera di Alfred Tarski

MIMESIS
Filosofie Analitiche / Metafisica

Collana: *Filosofie Analitiche / Metafisica* n. 12
Isbn 9788857518053
www.mimesisedizioni. it
Via Risorgimento, 33 – 20099 Sesto San Giovanni (MI)
Telefono +39 0224861657 / 0224416383
Fax: +39 02 89403935
E-mail: mimesis@mimesisedizioni.it

INDICE

PREFAZIONE

«La verità non è mai semplice»
Oscar Wilde

Il problema filosofico della verità occupa gli uomini da circa duemilacinquecento anni; il problema della verità *tout court* da sempre. La verità, il sapere per *davvero* come stanno le cose ci interessa: ne va molto spesso della nostra sopravvivenza, del nostro successo, della nostra felicità. Ma se è quasi scontato - poco più che buon senso - affermare la centralità della verità, ben altro discorso è chiedersi che cosa sia la verità, come possiamo conoscerla, come distinguerla dalla sua gemella, la falsità. Avere a che fare con *le* verità (della politica, della scienza, delle relazioni interpersonali e così via) è tipico degli uomini, chiedersi come sia fatta *la* verità è affare dei filosofi.

Spesso si crede (e talvolta con fondate ragioni) che nello sviluppo della riflessione filosofica non vi sia alcun progresso: la storia del pensiero altro non sarebbe che un riaffrontare ciclicamente alcune questioni eterne senza possibilità di paragonare, e poi di valutare, il contributo di ogni singolo protagonista. Non è questa la sede per prendere posizione circa lo statuto epistemologico della filosofia; sembra però che talvolta le cose non stiano propriamente così. Vi sono guadagni dello spirito umano che, una volta assicurati, non possono essere più rubricati come 'questioni del passato': bisogna farci i conti, magari attaccandoli ferocemente, ma certo mai ignorandoli. La teoria della verità di Alfred Tarski è, a nostro avviso, uno di questi. Ogni lavoro filosofico che si interessi al tema della verità non può non prendere in considerazione - lo ripetiamo, anche in senso assolutamente negativo - la concezione semantica della verità.

Tarski era di professione un logico e un matematico, non un filosofo. Come vedremo, il suo intento era quello di fornire una definizione rigorosa, precisa e scientificamente rispettabile di una nozione, la verità, centrale per la scienza ma offuscata da vaghezza e confusione concettuale. *Il concetto di verità nei linguaggi formalizzati* ha, così, una doppia natura: da un lato è un momento fondamentale nello sviluppo della logica matematica perché getta le basi della semantica logica (o teoria dei modelli), uno dei settori più importanti nella ricerca logica contemporanea; dall'altro, e ci ricolleghiamo a quanto detto poc'anzi, la sua portata eccede i confini della logica e invade benevolmente i dipartimenti di filosofia.

Le cose però non sono così semplici. È evidente il contributo che l'opera di Tarski ha fornito alla logica: non è solo la fondazione della semantica, come dicevamo, ma anche la base a partire dalla quale sviluppare concezioni alternative (si pensi, a titolo esemplificativo, alle teorie della verità *à la* Kripke o alla semantica dei mondi possibili). Molto più controverso è invece stabilire il contributo dell'opera di Tarski alla filosofia, o meglio, al dibattito filosofico circa la natura della verità. Questo volume cerca di dare una risposta esauriente e motivata a tale questione.

Per fare questo, il punto di partenza non poteva non essere la memoria originale di Tarski ove è presentata la sua celebre definizione. Nella ricchezza del lavoro del logico polacco, però, sono presenti *due* approcci radicalmente differenti tra loro: una definizione *semantica*, che occupa buona parte della costruzione tarskiana e, almeno *in nuce*, una definizione *assiomatica*, ovvero un sistema di principi che regolano l'impiego di un predicato primitivo indefinito "T". Alla base della prima definizione c'è la ben nota idea di definire la verità attraverso l'impiego di un metalinguaggio; alla base della seconda, il linguaggio è unico e la verità viene definita tramite una caratterizzazione assiomatica. Come vedremo si tratta di due strade differenti che riflettono differenti concezioni epistemologiche.

Il volume è, quindi, costituito da due parti, *relativamente* indipendenti: la prima, prende in considerazione la definizione semantica, la seconda quella assiomatica. La domanda che innerverà tutta la prima sezione è se, e in che misura, la teoria di Tarski possa essere considerata una teoria della verità come *corrispondenza*. Il nostro interesse sarà quello di cogliere come l'apparato logico tarskiano possa gettare luce sul problema dell'essenza della verità intesa come una particolare relazione tra linguaggio e mondo. Parallelamente, nella seconda parte si prenderanno in esame le proprietà logiche che ha una teoria formale della verità. La rilevanza filosofica, in questo senso, non riguarderà il tipo di concezione della verità, quanto, piuttosto, la portata epistemica, ontologica e fondazionale che questo concetto ha. Vedremo, cioè, come l'analisi di alcune teorie formali della verità (che rispecchiano le intuizioni di Tarski) metta in crisi le cosiddette posizioni *deflazioniste*, affermando al contrario che la nozione di verità è sostanziale e profondamente informativa.

I due progetti, benché relativamente indipendenti (come testimoniato dalla letteratura attuale che si concentra solitamente sull'uno o sull'altro), hanno in realtà una radice comune: se la verità è, in qualche misura, corrispondenza, possederà una natura specifica e non sarà, contro il deflazionismo, 'metafisicamente sottile'. Il nostro lavoro è quindi a metà tra la riflessione filosofica e quella logica; potremmo dire che si tratta di una filosofia della logica, intesa proprio come un'indagine filosofica sui presupposti e gli esiti di un capitolo essenziale della logica matematica contemporanea.

Continuare a parlare di Tarski come logico (matematico) in opposizione, o alternativa, al suo essere filosofo, è riduttivo e forse, in ultima analisi, inutile. Come dicevamo poche pagine fa, alcune conquiste del genio umano travalicano i confini disciplinari e contribuiscono al mutamento radicale della concezione del mondo e di noi stessi. Per restare nel nostro ambito, solo le opere di Gottlob Frege e di Kurt Gödel hanno la stessa portata rivoluzionaria della definizione tarskiana di verità. Dopo Tarski le cose non sono più come prima. È nostra convinzione che siano più belle, affascinanti e complesse. E la filosofia non può non raccogliere la sfida della complessità e resistere al fascino della bellezza.

Alla bellezza e alla profondità dell'opera tarskiana mi ha introdotto, ormai molti anni fa, il professor Sergio Galvan il quale ha seguito questo progetto di ricerca fin dagli inizi. So quanto gli sia cara la definizione di verità di Tarski, spero che questo lavoro sia il miglior ringraziamento possibile per la sua guida. Con Alessandro Giordani ho discusso tutte (e molte altre) le tematiche di questo volume: pur essendo, contemporaneamente, d'accordo e non, il suo brillante contributo è stato essenziale. Carlo Nicolai ha letto e commentato molte parti di questo lavoro e di ciò (oltre che dell'ospitalità oxoniense) gli sono grato. Dario Palladino ha letto l'intero manoscritto e mi ha aiutato moltissimo nel rendere il testo più fruibile grazie alla sua estesissima competenza logica e alla sua consuetudine alla chiarezza. Vorrei infine esprimere un ringraziamento profondo a Massimiliano Carrara per aver accolto questo volume nella Collana da lui diretta e per i tanti momenti di confronto sui temi più disparati. Ovviamente, la responsabilità di quanto scritto rimane soltanto mia. Dedico questo lavoro alla mia famiglia: a Francesca, la piccola Emma e il piccolissimo Tommaso.

Alessandria, giugno 2013

I PARTE

TEORIE SEMANTICHE DELLA VERITÀ

I
LA VERITÀ DI ALFRED TARSKI: IL CONCETTO DI VERITÀ NEI LINGUAGGI FORMALIZZATI

Praticamente tutti gli studiosi concordano sul ruolo decisivo che *Il concetto di verità nei linguaggi formalizzati*[1] ha giocato per la storia della logica e della filosofia del Novecento. Il giudizio unanime è al di là del-l'effettivo apprezzamento filosofico dell'impresa tarskiana; d'altro canto, proprio perché con il saggio di Tarski ha origine, si può dire, la semantica formale (o semantica logica) è complesso rintracciare una presentazione adeguata dei risultati ivi contenuti. In un certo senso, cioè, ogni manuale di introduzione alla logica fa uso della semantica tarskiana anche se in

---

1 Una nota bibliografica; il testo polacco *Pojecie prawdy w jezykach nauk de dukcyjnych* fu pubblicato nel 1933 ben due anni dopo la sua presentazione ad opera di Lukasiewicz alla Società Scientifica di Varsavia. In realtà, Tarski preparò gran parte del materiale già nel 1929 e lo presentò in alcuni seminari tenuti l'anno dopo presso la sezione di Logica della Società filosofica di Varsavia e la Società Filosofica Polacca di Lvov. Numerose sono state le traduzioni; la più significativa è sicuramente quella tedesca: Der Wahrheitsbegriff in den formalisierten Sprachen, «Studia Philosophica», I, 1935, 261-405; il testo apparve poi in inglese, tradotto da J.H. Woodger nel volume *Logic, Semantics, Metamathematics*. *Papers from 1923 to 1938*, Oxford University Press 1956. Nel seguito indicheremo sempre la traduzione inglese nella seconda edizione di *Logic, semantics and Metamathematics* a cura di J. Corcoran, edito presso Hackett nel 1983. Esiste anche una traduzione italiana a cura di Francesca Rivetti Barbò, *Il concetto di verità nei linguaggi formalizzati*, in *L'antinomia del mentitore nel pensiero contemporaneo*, *da Peirce a Tarski*, Vita e Pensiero, Milano 1961.

forma più o meno modificata[2]. Tuttavia, per i nostri scopi è interessante un'analisi puntuale e precisa della memoria tarskiana nella quale, come sarà evidente alla fine del capitolo, sono presenti sostanzialmente tutti i problemi presenti in letteratura: la concezione semantica della verità, il rapporto tra verità e dimostrabilità, la possibilità di assiomatizzare la nozione stessa di verità. Un'agenda insomma dei temi che hanno animato e animano ancora uno dei più fecondi dibattiti in filosofia della logica[3].

1. *Il concetto di enunciato vero nel linguaggio quotidiano*

Nella prima parte del suo saggio, Tarski prende in esame la possibilità di costruire una definizione corretta di verità nel linguaggio quotidiano. La prospettiva che sembra più allettante è quella di una definizione *semantica* della verità ed egli specifica immediatamente cosa intende per definizione semantica:

(1) «A true sentence is one which says that the state of affairs is so and so, and the state of affairs indeed is so and so».[4]

---

2 Non si tratta di semplici cambiamenti notazionali (anche se, come avremo modo di vedere, l'impiego di una determinata simbologia non è affatto un accessorio per lo sviluppo e la comunicazione delle idee) ma di spostamenti teorici. Citiamo a titolo puramente esemplificativo il fatto che, per Tarski, i linguaggi per i quali costruiremo una definizione rigorosa di verità sono linguaggi *interpretati*, cioè con un significato preciso; al contrario, nelle presentazioni manualistiche della semantica, si definisce una funzione di interpretazione che mette in relazione i segni del linguaggio, intesi come meri elementi sintattici, con un dominio oggettuale.
A ben vedere, però, si tratta di una differenza spesso enfatizzata nella letteratura: in realtà, la definizione di Tarski è costruita per uno specifico linguaggio interpretato - come vedremo, il linguaggio elementare delle classi - ma è facilmente generalizzabile per qualunque linguaggio formalizzato.

3 Galvan 1973 è un'analisi puntuale del saggio di Tarski; come anche Stegmüller 1968. La trattazione di Mancosu et al. 2008 è molto lucida pur nella sua brevità. Altre presentazioni utili della teoria di Tarski sono Kirkham 1992, Soames 1999, Künne 2003. Praticamente ogni testo di teoria dei modelli presenta la semantica tarskiana; possiamo qui ricordare i classici Chang Keisler 1973 e Hodges 1993.

4 Tarski 1983, 155.

Come è facile notare, per Tarski semantica indica primariamente la relazione del linguaggio (parla infatti di enunciati) con il mondo (tramite il riferimento allo stato di cose). Lo stesso Tarski fa risalire questa formulazione al lavoro di Kotarbinski e connette la definizione semantica della verità alla concezione classica o corrispondentista. In nota, infatti, presenta la celebre definizione di Aristotele secondo la quale, «Dire di ciò che è che è e di ciò che non è che non è è vero» come antesignano della concezione semantica della verità. Non discuteremo qui se la concezione tarskiana della verità possa essere definita come una posizione corrispondentista (cfr. capitolo seguente); sembra però davvero difficile affermare, in base a questa evidenza testuale, l'estraneità di Tarski dalla questione *filosofica* della verità.

La prima difficoltà che si ravvede in (1) è data, sostanzialmente, dall'intrinseca vaghezza della terminologia impiegata. Si pensi, a tal proposito, non solo al concetto sfuggente di 'stato di cose' ma anche alla relazione semantica di significazione che mette in relazione l'enunciato vero e lo stato di cose descritto. Per questo motivo, la prima proposta di Tarski, divenuta poi celebre, è il famoso schema T (o T-schema[5]) in base al quale:

(T) x è un enunciato vero se e solo se p

Dove x è una variabile per nomi di enunciati e p una variabile per enunciati. Che si tratti di uno schema è ovvio dal momento che (T) presenta appunto due variabili (di tipo differente) al suo interno che devono essere istanziate per avere esempi concreti di definizione di verità.

Un primo punto da sottolineare è proprio la presenza di una variabile per *nomi* di enunciati. La ragione di ciò è del tutto intuitiva; se, infatti, istanziassimo (T) senza tener conto della procedura di nominalizzazione degli enunciati avremmo qualcosa come

(2) Emma è bionda è un enunciato vero se e solo se Emma è bionda.

Ora, in italiano, (2) è sgrammaticato: la parte sinistra del bicondizionale presenta una confusione tra soggetti e predicati. Naturalmente, una sorta di principio di carità potrebbe permettere di interpretare corretta-

5 È importante non confondere il T-schema con la convenzione *T* di cui parleremo a breve.

mente il significato inteso di (2) ma a rigore il predicato nominale ‘essere un enunciato vero’ deve riferirsi a un soggetto grammaticale che non può, a sua volta, avere la forma di un enunciato. È necessario, dunque, avere a disposizione dei nomi per enunciati del linguaggio. Le proposte di Tarski sono essenzialmente due: *nomi da virgolette* e nomi *strutturali-descrittivi*[6].

Il metodo più semplice è quello di racchiudere l’enunciato che vogliamo nominare tra virgolette e ottenere così la possibilità di menzionare e non usare l’enunciato in questione. Seguendo il nostro esempio, avremmo allora che

(2b) ‘Emma è bionda’ è un enunciato vero se e solo se Emma è bionda.

Il che risulta perfettamente corretto. Un’alternativa è costituita dai nomi strutturali descrittivi, ovvero da descrizioni definite che riproducono la struttura sintattica dell’enunciato in questione. L’idea, anche in questo caso è semplice: immaginiamo di doverci riferire a una determinata parola (per esempio, ‘gatto’) senza pronunciare la parola in questione. Come fare? Basta dire al nostro interlocutore di prendere in considerazione quella parola di cinque lettere formata dalla sequenza di una ‘g’, una ‘a’, una ‘t’, un’altra ‘t’ e una ‘o’. Con questo sistema possiamo ‘codificare’ ogni espressione del nostro linguaggio e quindi nominare gli enunciati che ci servono.

Chiarito questo punto, Tarski passa a esaminare in che senso (T), cioè il T-schema, possa essere considerato una definizione di verità[7]. Il problema cruciale riguarda la sua generalizzazione; è chiaro infatti che una definizione della verità deve prendere in considerazione qualsivoglia enunciato e non solo alcune istanze del T-schema. Tarski presenta due tentativi di generalizzazione di (T) che falliscono per la medesima ragione. Siano infatti:

---

6 Sono infelici traduzioni italiane per *quotation-mark names* e *structural-descriptive names*.

7 In effetti Tarski alterna i concetti di definizione e di spiegazione utilizzandoli talvolta come sinonimi talvolta con sfumature di significato filosoficamente interessanti. Nel caso della generalizzazione del T-schema, per esempio, parla di spiegazione; probabilmente, sapendo che il T-schema è insufficiente per i suoi scopi, riserva il termine definizione a ciò che soddisfa i requisiti di rigore e scientificità richiesti mentre spiegazione diventa sinonimo di chiarificazione e nulla più.

(3) Per tutti i *p*, ‘*p*’ è un enunciato vero se e solo se *p*

dove la variabile per nomi di enunciati è sostituita da una variabile enunciativa e

(4) Per tutti gli *x*, *x* è un enunciato vero se e solo se per un certo *p*, *x* è identico a ‘*p*’ e *p*

dove si mantengono i due tipi di variabili e si quantifica universalmente sulle variabili per nomi di enunciati. Il fallimento di (3) e (4) dipende sostanzialmente per una fallacia nell’uso dell’operatore di virgolettatura. I nomi da virgolette devono essere considerati come parole di un linguaggio non ulteriormente scomponibili in particelle sintattiche più semplici. Tarski è chiarissimo su questo punto:

> Quotation-mark names may be treated like single words of a language, and thus like syntactically simple expressions. [...] [T]hey can possess no independent meaning. Every quotation-mark name is then a constant individual name of a definite expression and in fact a name of the same nature as the proper name of a man[8].

Quindi, come Tarski efficacemente sottolinea, è scorretto quantificare su ‘*p*’ poiché non si tratta di una variabile ma di un semplice nome proprio di un enunciato. Analogamente in (4), dove il quantificatore esistenziale a destra del bicondizionale dovrebbe vincolare proprio la *p* in ‘*p*’. Al contrario, il nome ‘*p*’ potrebbe denotare una delle lettere dell’alfabeto e quindi si avrebbe come istanza di (3) qualcosa come:

(3b) ‘L’iniziale di Paolo’ è un enunciato vero se e solo piove

E, data la totale arbitrarietà del riferimento, anche la sua contraddittoria:

(3c) ‘L’iniziale di Paolo’ è un enunciato vero se e solo non piove

L’impossibilità di generalizzare il T-schema sembra, quindi, derivare direttamente dalla natura sintattica semplice dei nomi per enunciati. Naturalmente è possibile tentare un’altra interpretazione dei nomi da virgo-

8 Tarski (1983, 159).

lette e considerarli cioè come termini funzionali. In questo caso, allora, l'espressione 'p' è costituita da un segno funzionale '…' il cui argomento è una variabile enunciativa e il cui valore è il nome dell'enunciato. Se, per esempio, q è una variabile enunciativa qualunque, 'q' sarà l'applicazione della funzione di citazione a q ottenendo appunto il termine 'q', cioè il nome di q. Da questo punto di vista, diventa possibile quantificare su una variabile all'interno di un termine funzionale[9]. Rimane però il problema circa la natura della funzione di citazione impiegata; Tarski è scettico riguardo alla funzione di citazione a causa della non estensionalità di quest'ultima. Infatti, è ovvio che proposizioni equivalenti potrebbero non avere nomi equivalenti; siano infatti $2 + 2 = 4$ e $3 + 3 = 6$ le nostre proposizioni in esame. Naturalmente vale che $(2 + 2 = 4) \leftrightarrow (3 + 3 = 6)$ ma i rispettivi nomi degli enunciati sono differenti: '$2 + 2 = 4$' $\neq$ '$3 + 3 = 6$'. Il comportamento della funzione di citazione non incontra, quindi, i *desiderata* di Tarski non solo per la natura non vero-funzionale ma soprattutto perché può portare alla formazione di alcune antinomie.

Ed eccoci giunti al cuore della prima parte del lavoro di Tarski: il paradosso del mentitore. In ben due punti della memoria, il logico polacco prende in esame la possibilità di costruire istanze di (T) che conducono al paradosso. Infatti:

(4) L'enunciato alla linea (4) non è vero.

Poiché (4) è un enunciato ben formato, è possibile applicare il T-schema e ottenere, quindi

(5) 'L'enunciato alla linea (4) non è vero' è vero se e solo se l'enunciato alla linea (4) non è vero.

Ora, che cosa intendiamo con la descrizione definita 'L'enunciato alla linea (4)'? È il nome di un enunciato che dice di se stesso che non è vero. Ma allora è possibile stabilire la seguente identità:

(6) L'enunciato alla linea (4) = 'L'enunciato alla linea (4) non è vero'

Ma, da (5) e (6), per sostituzione otteniamo

---

9 Come, ad esempio, $\forall x(f(x) = g(x))$.

(7) L'enunciato alla linea (4) è vero se e solo se l'enunciato alla linea (4) non è vero.

(7) è chiaramente contraddittorio. La conclusione di Tarski[10] è piuttosto secca: il paradosso deriva dall'impossibilità di tenere insieme tre condizioni tipiche dei linguaggi naturali e cioè:

I. La validità delle classiche leggi logiche
II. La possibilità di effettuare le sostituzioni di carattere empirico
III. Il linguaggio nel quale è stata riformulata l'antinomia possiede sia le risorse espressive per riferirsi agli enunciati stessi del linguaggio sia il predicato 'vero' applicato ai nomi degli enunciati.

Il terzo requisito è sicuramente il più significativo: indica che il linguaggio naturale è un linguaggio universale (cioè è in grado, di principio, di parlare di qualunque cosa) e, possedendo le risorse per predicare verità e falsità dei suoi stessi enunciati, risulta *semanticamente chiuso*. La chiusura semantica del linguaggio è considerata da Tarski come il tratto che porta alla formazione delle antinomie e che, quindi, mina alla base la possibilità stessa di una definizione rigorosa, scientificamente accettabile, della nozione di verità.

---

10 Tarski si riferisce a (6) come a una identità stabilita 'empiricamente'; molti commentatori hanno riportato fedelmente l'aggettivo originale. Tuttavia, dal punto di vista filosofico sembra una scelta poco felice: in che senso la stipulazione dell'identità tra il nome di enunciato e un segno è empirica? Sembra piuttosto che si tratti di una convenzione contingente, arbitraria. Tarski, però, riutilizza lo stesso aggettivo nella nota 1 a pagina 165 dove si legge:

> The antinomy of heterological words […] is simpler than the antinomy of the liar in so far as no empirical premises analogous to [6] appears in its formulation. (Tarski 1983, 165)

In effetti, tra le cause del paradosso è citata proprio la possibilità di formulare e accettare premesse come (6) e Tarski, che aveva in mente la trattazione del problema in un contesto formale, sa bene che una simile assunzione è bloccata dall'adozione di una precisa distinzione tra linguaggio oggetto, metalinguaggio e nomi (metalinguistici) degli elementi del linguaggio oggetto.

2. *La costruzione del linguaggio formalizzato*

Nella seconda parte del suo saggio, Tarski costruisce in maniera rigorosa un linguaggio formalizzato[11] in modo da assicurarsi una solida base sulla quale edificare, nella parte seguente, la definizione di verità. La notazione utilizzata da Tarski è sicuramente datata (basti pensare all'impiego dello stile 'polacco') mentre l'impostazione generale è impeccabile; anzi, risulta interessante proprio riproporre in maniera più fedele possibile all'originale il percorso tarskiano in modo da evidenziarne la portata rivoluzionaria. Ovviamente, in sede di discussione critica non saremo più obbligati a seguire le scomode convenzioni degli albori della semantica logica.

La presentazione del linguaggio formalizzato ha per Tarski un duplice scopo: da un lato, come abbiamo appena detto, è propedeutica alla definizione stessa di verità; dall'altro, risolve la difficoltà emersa con il tentativo di definizione della verità nei linguaggi ordinari, assicurando, cioè, la *correttezza formale* della definizione. Cosa si intende per correttezza formale? È, in poche parole, l'assicurazione logica che dalla definizione tarskiana non possano sorgere antinomie come quella del mentitore. Ebbene, l'impiego del linguaggio formalizzato assicura il requisito minimale della correttezza formale grazie all'attenta distinzione in *linguaggio oggetto* e *metalinguaggio*. Infatti, l'enunciato del mentitore è tale grazie all'autoreferenzialità: dice di se stesso che è falso. Ma, la possibilità di applicare termini semantici (come verità e falsità) a tutti gli enunciati del linguaggio è proprio la caratteristica della chiusura semantica che abbiamo accennato alla fine del paragrafo precedente. Distinguendo attentamente tra linguaggio oggetto e metalinguaggio, Tarski impone delle precise condizioni circa l'applicabilità dei termini semantici; in particolare,

---

11 Adoperiamo coscientemente l'aggettivo 'formalizzato' e non 'formale'; infatti, il linguaggio presentato da Tarski non è a rigore un linguaggio formale dal momento che è interpretato. Anzi, per lui, il problema della definizione della verità può applicarsi sensatamente solo in contesti di linguaggi interpretati. Nell'uso comune è prevalsa l'abitudine di indicare con l'aggettivo 'formale' sia i linguaggi non interpretati (come per esempio, il calcolo dei predicati) sia le teorie formali (come l'aritmetica di Peano). 'Formalizzato' indica quindi un linguaggio perfettamente rigoroso, simbolizzato, che non presenta fenomeni di equivocità e varianza semantica nello spirito dei linguaggi scientifici proposti dal neopositivismo. Per una panoramica dei differenti significati relativi all'uso del termine "formale" rimandiamo ad Agazzi (1994).

mentre il linguaggio oggetto non contiene termini semantici, questi potrebbero essere presenti nel metalinguaggio. Quindi, nel metalinguaggio è possibile affermare falsità o verità di enunciati appartenenti solo al linguaggio oggetto, bloccando così il fenomeno dell'auto-predicazione che è alla base del paradosso del mentitore.

Iniziamo allora la presentazione del linguaggio oggetto (LO) seguendo il più possibile lo spirito tarskiano. Tarski fornisce un esempio concreto di linguaggio formalizzato: il linguaggio delle classi al primo ordine.

Linguaggio oggetto

Alfabeto
a) Variabili individuali: $x_1, x_2, x_3, \ldots$
b) Costanti logiche: $\neg, \vee, \forall$
c) Costante non logica a due posti: $\subseteq$

Termini
a) Le variabili individuali sono termini
b) Niente altro è un termine

Formule
a) per ogni i, j, $x_i \subseteq x_j$ è una formula
b) Se α e β sono formule e x_i è una variabile allora: $\neg\alpha$ è formula,
$\alpha \vee \beta$ è formula, $\forall x_i \alpha$ è formula
c) Niente altro è una formula

La teoria delle classi sarà poi costituita da assiomi logici, assiomi specifici e regole di inferenza. Adottando uno sistema di derivazione naturale è possibile togliere la classe degli assiomi logici a fronte di un incremento delle regole di inferenza.

Metalinguaggio

La struttura del metalinguaggio di Tarski è molto più complessa rispetto al linguaggio oggetto; innanzitutto, nel metalinguaggio avremo variabili, costanti logiche e costanti non logiche che, in un certo senso, corrispondono agli elementi del linguaggio oggetto; inoltre, il metalinguaggio deve essere in grado di 'nominare' gli oggetti linguistici del linguaggio oggetto. Tutto questo sarà più chiaro in seguito ma fin d'ora è facile rendersi conto dell'importanza dei nomi: nel T-schema sono presenti variabili sia per enunciati (o meglio per traduzioni metalinguistiche di enunciati del linguaggio oggetto) che variabili per nomi di enunciati. Del resto, anche nella trattazione del linguaggio naturale Tarski ha discusso due procedure per nominare gli enunciati e, come abbiamo visto, le ha scartate a causa di un ineliminabile aspetto intensionale.

Alfabeto
a) Costanti logiche: non, vel, om
b) Costante relazionale: *I*
c) Costanti non logiche ausiliarie: $\in$

d) Variabili per classi: a, b, …
e) Variabili per sequenze di classi: $\sigma, \tau, \dots$
f) Variabili per numeri naturali: k, l, m, n, …
g) Variabili per espressioni qualsiasi del linguaggio: $\alpha, \beta, \dots$

È necessario qualche commento e qualche precisazione. Il significato intuitivo delle costanti logiche è lo stesso della negazione (non), disgiunzione (vel) e quantificazione universale (om) del linguaggio oggetto. Analogamente anche per la relazione di inclusione. Per quanto riguarda le costanti non logiche ausiliarie si tratta di una serie di costanti insiemistiche (come il segno di appartenenza $\in$) che permettono a Tarski di definire nel metalinguaggio alcuni concetti importanti per la definizione di verità. Fra tutti, come vedremo a breve, è essenziale l'idea di successione di oggetti (nel caso specifico, classi).

Nomi
a) Nomi per variabili individuali: $v_1, v_2, v_3, \dots$
b) neg è il nome di $\neg$
c) disg è il nome di $\vee$
d) all è il nome di $\forall$
e) in è il nome di $\subseteq$

Regole di formazione di nomi complessi
Siano e_i ed e_j nomi di espressioni, allora $e_i{}^\frown e_j$ è il nome dell'espressione ottenuta dall'espressione e_i seguita da e_j. Prendiamo in esame a titolo esemplificativo la formula del linguaggio oggetto $x_1 \subseteq x_2$; il suo nome (metalinguistica) sarà pertanto: $(\text{in}{}^\frown v_1){}^\frown v_2$.

Come si può notare, la distinzione tra linguaggio oggetto e metalinguaggio porta a una tripartizione del livelli: il linguaggio oggetto, il metalinguaggio e, in quest'ultimo, i nomi per le espressioni del linguaggio oggetto.

Come si ricorderà, il T-schema presenta la seguente struttura:

(T) x è un enunciato vero se e solo se p

già allora avevamo messo in evidenza che la variabile x si riferisce a nomi di enunciati mentre la variabile p sta per enunciati del metalinguaggio che siano la traduzione dell'enunciato nominato da x. Ora abbiamo a disposizione le risorse tecniche per poter rendere più chiara questa idea. Infatti, si prenda ad esempio l'enunciato $\forall x_1 (x_1 \subseteq x_1)$ che è vero nell'interpretazione intesa della teoria delle classi e che intuitivamente dice che ogni classe è inclusa in se stessa. Il nome dell'enunciato sarà, in base a quanto detto in precedenza, $(\text{all} \frown v_1) \frown (\text{in} \frown v_1) \frown v_1$; d'altro canto la sua traduzione metalinguisitica è: (om a)(a *I* a). Quindi l'istanza del T-schema diventerà:

(T*) $(\text{all} \frown v_1) \frown (\text{in} \frown v_1) \frown v_1$ è vero se e solo se (om a)(a *I* a)

dove sono perfettamente rintracciabili le componenti illustrate in precedenza.

Il sistema messo a punto da Tarski è del tutto rigoroso anche se estremamente scomodo. In effetti, egli introduce delle abbreviazioni per quanto riguarda i nomi degli enunciati che abbiamo preferito tralasciare al fine di non complicare un quadro già non propriamente maneggevole. Una possibile semplificazione, in realtà, è data dall'assumere che il metalinguaggio contenga come parte propria il linguaggio oggetto e che quindi, dal punto di vista grafico, i due siano indistinguibili. Lo stesso Tarski prende in esame l'alternativa:

> To simplify the further developments and avoid possible misunderstanding we shall suppose the metalanguage to be so constructed that the language we are studying forms a fragment of it; every expression of the language is at the same time an expression of the metalanguage, but not vice versa. This enable us in certain cases [...] to speak simply of

> the expressions of the language itself, instead of expressions of the metalanguage which have the same meaning[12].

Nel nostro caso, dunque è possibile tralasciare la traduzione metalinguistica e affermare che:

(T**) (all⌢v₁)⌢(in⌢v₁)⌢v₁ è vero se e solo se $\forall x_1 (x_1 \subseteq x_1)$

La distinzione di metalinguaggio e linguaggio oggetto è, quindi, il primo passo per la definizione della verità; rimane però molta strada da compiere e i passi decisivi verranno mossi nella sezione successiva dove, usando le parole dello stesso Tarski, si tratta «il problema chiave di questo articolo – la costruzione della definizione di enunciato vero»[13].

3. *Il concetto di enunciato vero nel linguaggio del calcolo delle classi*

Come abbiamo visto in precedenza, il primo problema che Tarski fronteggia è costituito dal requisito di correttezza formale che una nozione scientificamente accettabile di verità impone alla sua procedura di definizione. L'insorgenza delle antinomie porta il logico polacco ad abbandonare il linguaggio naturale come luogo dove definire il concetto di verità; da qui, la costruzione del linguaggio formalizzato e la distinzione in linguaggio oggetto e metalinguaggio su cui ci siamo soffermati nel paragrafo precedente.

Tuttavia, una definizione di verità non deve essere solo corretta formalmente (o, alternativamente, immune da antinomie): è necessario che il concetto di verità che definiamo si accordi con le nostre intuizioni preteoriche. Detto altrimenti, se la nostra teoria dovesse ascrivere la verità a enunciati che, in via del tutto intuitiva, considereremmo falsi, non soddisferebbe il requisito di adeguatezza materiale. Ma come rendere più rigoroso il requisito dell'adeguatezza materiale? Per Tarski, una definizione di verità materialmente adeguata deve rispondere alla convenzione T:

> A formally correct definition of the symbol 'Tr', formulated in the metalanguage, will be called an adequate definition of truth if it has the following consequence[s]:

---

12 Tarski 1983, 247.
13 Tarski 1983, 186.

> (α) all sentences which are obtained from the expression '$x \in \mathrm{Tr}$ if and only if p' by substituting for the symbol 'x' a structural-descriptive name of any sentence of the language in question and for the symbol 'p' the expression which forms the translation of this sentence into the metalanguage[14].

Tarski aggiunge un'altra condizione (β) in cui si stabilisce che l'insieme delle proposizioni vere deve essere contenuto nell'insieme di tutti gli enunciati S. Si tratta di una specificazione non essenziale, come ammesso dallo stesso Tarski e quindi ci concentreremo sul primo punto della convenzione T.

In base a quanto dichiarato, quindi, il requisito di adeguatezza materiale è rappresentato dalla possibilità di ottenere tutti i T-schemi, ovvero tutte le istanze di

(T) x è vero se e solo se p

Come nota giustamente Tarski il problema di costruire una definizione di verità materialmente adeguata si riduce a una semplice procedura di elencazione nel caso in cui il linguaggio in esame possieda solo un numero finito di enunciati. Ipotizziamo, infatti, che il nostro linguaggio oggetto possieda solo tre enunciati:

p_1 : Emma è bionda
p_2 : Emma è bruna
p_3 : Emma è rossa

A questi tre enunciati corrispondono altrettanti nomi:

x_1 : 'Emma è bionda'
x_2 : 'Emma è bruna'
x_3 : 'Emma è rossa'

In questo caso, la nostra definizione di verità procederà nel modo seguente:

(T-fin) x è vero sse ($x = x_1$ e p_1) vel ($x = x_2$ e p_2) vel ($x = x_3$ e p_3)

---

14 Tarski 1983, 188.

e risulterà perfettamente adeguata secondo la convenzione T. Infatti, per rendersene conto, è sufficiente sostituire a *x* un nome qualunque degli enunciati in questione e, logicamente, uno dei disgiunti risulterà vero (esaurendo essi tutti i casi possibili) comportando così anche la verità della disgiunzione.

Il problema, naturalmente, sorge nel momento in cui ci si rende conto che i linguaggi normalmente possiedono un numero infinito di potenziali enunciati. Basta infatti che nel linguaggio sia disponibile un sistema che permetta di concatenare gli enunciati (come i connettivi) e immediatamente il metodo appena esposto non diventa più praticabile.

Ciò conduce direttamente verso un'altra grande intuizione di Tarski: la definizione della verità deve possedere un carattere *ricorsivo*. Deve cioè essere definita per un caso base, di complessità minima e per ricorsione deve riuscire a generare tutte le possibili combinazioni, quand'anche queste fossero infinite. Seguendo l'esempio di prima, possiamo immaginare un linguaggio molto povero, costituito dai tre enunciati più un connettivo, la congiunzione. La base della nostra definizione sarà del tutto simile a (T-fin); inoltre ammetteremo una clausola ricorsiva che (x = 'x_i & x_j' e p_i è vero e p_j è vero) in base alla quale la verità di una congiunzione è definita a partire dalla verità dei due congiunti. Il fatto che la definizione di Tarski sia ricorsiva è di capitale importanza per la sua rilevanza filosofica e sarà oggetto di discussione nel prossimo capitolo.

3.1 *Soddisfazione*

Tuttavia, anche la procedura per ricorsione non è sufficiente per gli scopi tarskiani. Infatti, in molti casi, enunciati complessi sono costruiti a partire da espressioni che non sono a loro volta enunciati. Nel caso della congiunzione, per esempio, i componenti sono enunciati, con un preciso valore di verità ma cosa accade se consideriamo l'espressione '*x* è biondo e *y* è moro'? Non possiamo definire la verità di questa frase partendo dai suoi componenti perché, essi stessi, non sono enunciati. Infatti, *x* è biondo non è un enunciato ma uno pseudo-enunciato, o per utilizzare la terminologia di Tarski una funzione enunciativa. La presenza della variabile individuale *x* non permette di ascrivere un valore di verità all'espressione: se *x* viene sostituito con il nome di un individuo biondo, allora l'enunciato risulterà vero; falso altrimenti. Ma in presenza di una o più variabili, l'espressione non può descrivere uno stato di cose determinato e quindi non può essere sensatamente dichiarata vera o falsa.

Si può, pertanto, dire che la soluzione tarskiana fino ad ora prospettata è sufficiente per caratterizzare la definizione di verità per linguaggi costituiti solo da enunciati, come ad esempio, il linguaggio della logica proposizionale ma non riesce a trattare casi più complessi comprendenti quantificatori e variabili come accade nella logica dei predicati.

L'intuizione di Tarski è, allora, quella di rinvenire una proprietà che obbedisca a tre condizioni particolari[15]:

i) Deve essere posseduta sia dagli enunciati aperti (cioè dalle funzioni enunciative) che dagli enunciati veri e propri.
ii) Deve essere ricorsiva anch'essa.
iii) Deve essere possibile definire la verità in base al possesso o meno di quella determinata proprietà.

Il candidato ideale per Tarski è il concetto di *soddisfazione*. La soddisfazione è una proprietà relazionale che vale tra (sequenze di) oggetti da un lato ed enunciati aperti dall'altro. Vedremo a breve il perché del riferimento alle sequenze; per adesso, concentriamoci sul caso più semplice, quello tra oggetti individuali ed enunciati aperti. Si consideri il seguente enunciato aperto:

(8) x è rosso

Quando, intuitivamente, (8) risulterà vero? Quando alla x sostituiamo un oggetto, o meglio il nome di un oggetto, che gode della proprietà di essere rosso. Se per esempio, mettiamo al posto di x 'il sangue' otterremo una proposizione vera; quindi, si può dire che l'oggetto *o* (cioè il sangue) soddisfa l'enunciato aperto (8). Convenendo di indicare con a, b, c oggetti del dominio, con a, b, c i nomi degli oggetti e con $\alpha(x)$ un enunciato di complessità qualsiasi in cui vi sia un'occorrenza libera della variabile x, possiamo caratterizzare la relazione di soddisfazione come segue:

(9) $\mathrm{Sod}(a,\alpha(x))$ se e solo se $\alpha(\mathrm{a})$ è vero

La relazione di soddisfazione mette in collegamento linguaggio e mondo e si qualifica come un concetto squisitamente semantico. Questo

15 Kirkham 1992, 152.

aspetto sarà discusso e analizzato in seguito vista la sua rilevanza nella valutazione filosofica del lavoro di Tarski.

Nel linguaggio, tuttavia, vi possono essere occorrenze multiple di variabili, come nel caso delle relazioni:

(10) x ama y

Nel caso di (10) saranno due gli oggetti che soddisferanno l'enunciato aperto. Ma non solo. È essenziale anche l'ordine di presentazione dei due oggetti dal momento che la relazione 'amare' non è simmetrica. Quindi, la coppia ordinata <Romeo, Giulietta> soddisferà (10). Tarski utilizzerà un concetto derivato dalla logica delle relazioni per generalizzare casi come (9) e (10): le sequenze di oggetti. Una sequenza di oggetti è molto simile a un insieme con due differenze fondamentali: nella sequenza l'ordine conta e un oggetto può occorrere più di una volta. Per cui mentre

{il Colosseo, il numero 3, Socrate}

e

{il numero 3, Socrate, il Colosseo}

sono lo stesso identico insieme, al contrario le sequenze

<il Colosseo, il numero 3, Socrate>

e

<il numero 3, Socrate, il Colosseo>

sono due sequenze differenti, mutando l'ordine degli oggetti. Così è possibile avere una sequenza così formata:

<il Colosseo, il numero 3, il Colosseo, il numero 3, il Colosseo, …>

Poiché non vi sono particolari impedimenti, è possibile assumere le sequenze infinite di oggetti per cui un enunciato aperto come (10) sarà soddisfatto da una sequenza infinita se il primo oggetto ama il secondo oggetto, come nel caso di una sequenza così composta: <Romeo, Giulietta, il numero 3, il Colosseo, …>

Vediamo all'opera il concetto di soddisfazione nel caso specifico del calcolo delle classi. Una sequenza soddisfa una funzione enunciativa se e solo se vale la condizione espressa. Naturalmente, proprio perché la relazione di soddisfazione deve essere definita nel metalinguaggio, dovremo badare a indicare il *nome* della funzione enunciativa e la *traduzione metalinguistica* della condizione.

Sia σ una sequenza infinita di classi[16]. Sia i_σ l'i-esimo elemento della σ sequenza. Assumiamo per semplicità che le variabili α, β, … stiano per nomi di espressioni qualunque. Allora:

a) σ soddisfa la funzione enunciativa $(in^\frown v_1)^\frown v_2$ se e solo se $1_\sigma\, I\, 2_\sigma$

b) σ soddisfa la funzione enunciativa $(neg^\frown \alpha)$ se e solo se σ non soddisfa α

c) σ soddisfa la funzione enunciativa $(disg^\frown \alpha)^\frown \beta$ se e solo se σ soddisfa α oppure σ soddisfa β

d) σ soddisfa la funzione enunciativa $(all^\frown v_1)^\frown \alpha$ se e solo se per ogni sequenza τ che differisce da σ al massimo per ciò che riguarda v_1, τ soddisfa $(\alpha^\frown v_1)$.

La clausola teoricamente più significativa è senz'altro l'ultima, quella riguardante cioè i quantificatori. Prendiamo in esempio il seguente enunciato quantificato:

(11) $\forall x_4 \mathrm{R} x_4$

il cui significato inteso sia: ogni cosa è rotonda. In base alla definizione di soddisfazione tarskiana, diremo che una sequenza σ deve soddisfare (11) senza il quantificatore (e cioè, $\mathrm{R}x_4$) e, questa funzione enunciativa deve essere soddisfatta da ogni sequenza che è esattamente come σ eccetto il fatto che possiede un oggetto differente al quarto posto. Infatti, sia σ la seguente sequenza:

---

16 Tarski costruisce la sua definizione di verità per il linguaggio delle classi; pertanto, le sequenze di oggetti che soddisferanno le funzioni enunciative e renderanno veri gli enunciati di tale linguaggio saranno sequenze di classi. Questo aspetto è rilevante per le ricadute ontologiche della semantica tarskiana e sarà discusso in seguito.

<Mario, il numero 3, la libertà, il cerchio, Luigi, …>

È chiaro che σ soddisfa Rx_4 dal momento che il quarto oggetto della sequenza (secondo la nostra terminologia, 4_σ) è un oggetto che gode della proprietà di essere rotondo. Ebbene, la condizione sul quantificatore universale dice proprio che ogni altra sequenza deve soddisfare Rx_4 con un oggetto differente al suo posto; prendiamo per esempio la sequenza τ:

<Mario, il numero 3, la libertà, il Sole, Luigi, …>

Anche τ soddisfa Rx_4, dal momento che il Sole è rotondo. Se questo vale per tutte le sequenze possibili allora vuol dire che ogni oggetto (al quarto posto) è rotondo: proprio il significato inteso del nostro enunciato quantificato.

3.2 *Definizione di verità*

A questo punto possiamo compiere l'ultimo passo e dare la definizione di verità. Abbiamo visto come la proprietà di soddisfazione riguarda funzioni enunciative, cioè espressioni che presentano variabili e che quindi non possono essere definite vere o false. Un enunciato può essere considerato un caso limite di una funzione enunciativa. Che cosa accade, quindi, se proviamo a utilizzare la proprietà della soddisfazione in riferimento a una formula chiusa, cioè un enunciato?

È utile notare che, nei casi precedentemente analizzati, la soddisfazione o meno di una funzione enunciativa dipende esclusivamente dalle variabili libere in essa presenti. Ma, allora, nel caso limite la soddisfazione di una formula da parte di una sequenza non dipende dalle proprietà dei termini della sequenza. Rimangono solo due possibilità: o tutte le sequenze soddisfano l'enunciato o nessuna. Nel primo caso, diremo allora che l'enunciato è vero, nel secondo che è falso. Quindi, un po' più formalmente:

(12) x è vero se e solo se (om σ)(σ soddisfa x)

Dove, ricordiamolo, al posto di x è necessario porre i nomi metateorici dell'enunciato in oggetto.

Si tratta di una definizione di verità materialmente adeguata? Per rispondere a questa domanda è necessario andare a vedere se dalla definizione sia possibile derivare la lista dei T-schemi, in ottemperanza alla convenzione T. Si può mostrare che la definizione risponde al requisito della convenzione T e, pertanto, la definizione tarskiana è materialmente adeguata.

4. *Linguaggi di ordine finito*

Dopo aver esposto per linee generali la costruzione di una definizione di verità scientificamente accettabile, Tarski estende il risultato a linguaggi di complessità crescente. Tuttavia, appare immediatamente una difficoltà riguardante il concetto stesso di soddisfazione, cardine per la definizione della verità. Il linguaggio delle classi, infatti, è costituito da una sola specie di variabili, segnatamente variabili individuali che stanno per oggetti del dominio, cioè, nel caso in esame, per classi. Nel caso di linguaggi più complessi, possiamo avere a che fare con più tipologie di variabili: variabili per predicati monadici (o proprietà), diadici (relazioni a due posti), n-adici (relazioni a n-posti). Quindi, il concetto di soddisfazione utilizzato in precedenza si rivela insufficiente dal momento che le sequenze che abbiamo preso in esame godono di una sorta di *omogeneità ontologica*. Infatti, nel caso del calcolo delle classi, gli oggetti presenti nelle sequenze appartengono tutti alla medesima categoria e cioè quella degli individui[17]. Ma se vogliamo estendere la relazione di soddisfazione a linguaggi più complessi, diventa indispensabile modificare la struttura stessa della relazione e porre delle distinzioni a livello semantico che rispecchino, in un certo modo, le differenze linguistiche. In altre parole, nella misura in cui sono presenti specie diverse di variabili, dovremo presentare differenti tipi di entità che possono soddisfare le rispettive funzioni proposizionali.

Qual è dunque una possibile tassonomia delle variabili, o meglio delle espressioni, che possiamo trovare in un linguaggio di qualsivoglia complessità? Per rispondere a questa domanda Tarski introduce il concetto di *categoria semantica*, di cui riconosce la paternità a Husserl ma che uti-

17 Non è rilevante, in questa sede, soffermarsi sul fatto che le classi possono essere considerate, da un punto di vista ontologico, oggetti astratti non individuali o pluralità.

lizza nell'accezione indicata da Lesniewski. La teoria delle categorie semantiche si avvicina, per certi versi, alla stratificazione dei concetti operata da Russell e Whitehead nei *Principia Mathematica* ma, con le parole dello stesso Tarski:

> [...][I]t is scarcely possible to imagine a scientific language in which the sentences have a clear intuitive meaning but the structure of which cannot be brought into harmony with the above theory[18].

Un'analisi precisa della nozione di categoria semantica è però fuori dagli intenti di Tarski; per questo motivo, egli enuncia solamente una sorta di principio di identità per categorie semantiche: due espressioni sono della medesima categoria semantica se e solo nessuna funzione enunciativa che contiene una di queste espressioni cessa di essere tale se la prima espressione è sostituita dall'altra. Prendiamo ad esempio le espressioni x e y e chiediamoci se appartengono alla medesima categoria semantica. Sia $\alpha(x)$ una funzione enunciativa; sostituiamo[19] in $\alpha(x)$ x con y e otteniamo $\alpha(y)$ che è a sua volta una funzione enunciativa. In base al criterio proposto da Tarski diremo allora che x e y appartengono alla medesima categoria semantica[20]. Al contrario, nelle espressioni P^1x e Q^2xy, P^1 e Q^2 non appartengono alla stessa categoria semantica. Sostituire Q^2 per P^1 in P^1x porterebbe all'espressione Q^2x che risulta chiaramente malformata dal momento che Q^2 è un predicato diadico.

A loro volta le categorie semantiche possono essere raggruppate in differenti *ordini*. Tarski parte dall'assunzione che i nomi di individui e le variabili individuali possiedono ordine 1. Quindi un'espressione sarà di ordine n+1 se l'espressione con il tipo più alto in essa occorrente avrà ordine n. Pertanto, le variabili individuali x e y saranno di ordine 1 (per definizione). Le espressione P^1x e Q^2xy saranno entrambe di ordine 2, dal momento che le espressioni di ordine più elevato che occorrono in esse sono di ordine 1. In base a tutto ciò, tutte le espressioni che appartengono a una categoria semantica possiedono lo stesso ordine mentre non vale, come abbiamo appena visto, il viceversa. Infatti, P^1x e Q^2xy hanno lo stesso ordine ma appartengono a due categorie semantiche differenti.

---

18 Tarski 1983, 215.

19 La sostituzione deve ovviamente rispettare i criteri di correttezza formale; deve cioè essere legittima.

20 Per essere rigorosi, dovremmo dire che la classe delle sostituzioni possibili che ottemperano al criterio tarskiano costituisce una categoria semantica.

Infine, Tarski offre un terzo criterio tassonomico dei linguaggi e cioè il *tipo*. Diremo che due funzioni (enunciative) possiedono lo stesso tipo semantico se e soltanto se il numero delle variabili libere di ogni categoria semantica nelle due funzioni è lo stesso. In questo senso, P^1x e T^1z appartengono al medesimo tipo semantico, dal momento che le variabili individuali (x e z) sono della stessa categoria semantica e possono essere messe in corrispondenza biunivoca.

Con i tre concetti (di categoria semantica, di ordine e di tipo) appena definiti Tarski offre la seguente classificazione dei linguaggi:

> (1) languages in which all the variables belong to one and the same semantical category; (2) languages in which the number of categories in which the variables are included is greater than 1 but finite; (3) languages in which the variables belong to infinitely many different categories but the order of these variables does not exceed a previously given natural number n; and finally (4) languages which contain variables of arbitrarily high order[21].

I linguaggi del primo tipo sono linguaggi in cui le variabili appartengono ad un'unica categoria semantica; il calcolo delle classi presentato da Tarski ne è un esempio ma pensiamo, anche, al calcolo dei predicati del primo ordine. Tutti i tipi rimanenti sono linguaggi di ordine superiore, con la differenza che nel secondo caso, le variabili appartengono a un numero finito di categorie semantiche mentre nel terzo le categorie semantiche ammesse sono infinite a patto che l'ordine delle variabili sia finito. Infine, il quarto tipo di linguaggio presenta variabili appartenenti a categorie semantiche di ordine arbitrariamente grande. I primi tre tipi di linguaggio sono detti linguaggi di ordine finito, l'ultimo di ordine infinito.

Tarski prosegue l'esposizione presentando alcuni tratti della costruzione della definizione di verità per le tipologie di linguaggio prese in esame. Prenderemo in considerazione solo gli aspetti rilevanti per la nostra trattazione per arrivare alla presentazione (anch'essa per linee generali) del risultato noto come teorema di indefinibilità della verità.

Come esempio di linguaggio di tipo 2, Tarski tratta il linguaggio delle relazioni diadiche. L'estensione rispetto al linguaggio delle classi è piuttosto semplice.

---

21 Tarski 1983, 220.

Alfabeto (Logica delle relazioni diadiche)
a) Variabili individuali: $x_1, x_2, x_3, \ldots$
b) Variabili predicative relazionali: $X_1, X_2, X_3, \ldots$
c) Costanti logiche: $\neg$, $\vee$, $\forall$

In questo caso non è necessario introdurre alcuna costante non logica dal momento che $X_1x_1x_2$ è una formula ben formata[22]. In maniera del tutto analoga, V_1, V_2, V_3, … sono i nomi metalinguistici delle variabili predicative. Per questo, $(\mathrm{V}_1{}^\frown \mathrm{v}_1){}^\frown \mathrm{v}_2$ è il nome di $X_1x_1x_2$.

La definizione di soddisfazione per le variabili predicative ricalca quella per le variabili individuali; nel caso di una funzione proposizionale come

(13) X_1ab

diremo che una sequenza σ soddisfa X_1ab se e solo il primo elemento è la relazione che vale tra a e b. Quindi, in maniera analoga, avremo sequenze di attributi diadici come <amare, uccidere, …>. In questo senso, è sufficiente scegliere la sequenza (o le sequenze) adatte alla categoria semantica delle variabili presenti nella funzione proposizionale e procedere come prima.

Tuttavia, le cose non sono così semplici. Infatti, in un linguaggio come questo potrebbero occorrere in una stessa funzione enunciativa sia variabili individuale sia variabili predicative; come abbiamo già accennato in precedenza una relazione di soddisfazione di una formula che contiene variabili di categorie semantiche differenti non può a sua volta essere univoca. Si prenda ad esempio la funzione proposizionale

(14) $X_1x_1x_2$

In essa sono presenti variabili del primo ordine e una variabile del secondo ordine. È ovvio che una sequenza di oggetti non potrà soddisfare tutte le variabili, dal momento che o gli oggetti della sequenza sono individui oppure sono attributi. Ma in un caso e nell'altro si esclude una ca-

22 Teoricamente, bisognerebbe introdurre due segni per i due differenti tipi di quantificatori differenti, il primo per le variabili individuali, il secondo per le variabili predicative; tuttavia, è possibile fare a meno di questa distinzione senza perdere nulla quanto a precisione e con il vantaggio di non appesantire troppo il vocabolario di partenza.

tegoria di variabili dalla relazione di soddisfazione. Come procedere, dunque?

Tarski mette a punto due metodi che chiama rispettivamente *delle sequenze multi-riga* e *dell'unificazione semantica*.

Il primo è sicuramente il più intuitivo e consiste nel ridefinire la relazione di soddisfazione come una relazione che vale tra tre termini: il nome di una funzione proposizionale, una sequenza di individui e una sequenza di attributi (in questo caso diadici). Possiamo introdurre delle variabili per sequenze di quest'ultimo tipo come Σ, Λ, e avremo che:

(15) σ, Σ soddisfano $(V_1 \frown v_1) \frown v_2$ se e solo se la relazione 1_Σ vale tra 1_σ e 2_σ

dove, lo ricordiamo, con 1_Σ indichiamo il primo elemento della sequenza Σ che, nel nostro caso, è una relazione diadica e con 1_σ e 2_σ indichiamo, rispettivamente, il primo e il secondo elemento della sequenza σ, cioè due individui[23]. Questo metodo può poi essere generalizzato non considerando σ e Σ come due sequenze distinte ma come *righe* della medesima sequenza che assume, quindi, l'aspetto di una matrice a due righe.

Il secondo metodo esposto da Tarski fa uso di alcuni concetti che appartengono alla teoria degli insiemi e che vengono normalmente adoperati negli studi di logica matematica. L'intuizione alla base del procedimento è, in questo caso, piuttosto differente che nelle sequenze multi-riga. Mentre in precedenza, infatti, si aumentava la complessità ontologica delle sequenze a disposizione per far fronte a una relazione di soddisfazione che comprendesse anche variabili per relazioni, qui l'idea è per certi versi speculare; si tratta, in poche parole, di uniformare gli elementi delle varie sequenze in modo da non dover utilizzare due categorie differenti. Come procedere nel nostro caso? È sufficiente trasformare le variabili individuali in variabili di relazione a due posti; così facendo avremmo una sola categoria di variabili (quelle di ordine superiore) e, di conseguenza, una sola tipologia di sequenze per la relazione di soddisfazione. Questo può essere fatto mettendo in relazione biunivoca ogni individuo *a* con una coppia ordinata *a** costituita da due termini che sono identici con *a* (e quindi, per proprietà dell'identità, anche identici tra loro). Un esempio concreto: l'individuo *a* è correlato con la coppia ordinata $a^* = \langle b,c \rangle$

23 Ricordiamo anche che $(V_1 \frown x_1) \frown x_2$ è il nome metalinguistico di $X_1x_1x_2$.

tale che $b = a$ e $c = a$. Quindi la relazione di soddisfazione ritorna ad essere una relazione tra nomi di funzioni proposizionali e sequenze.

(16) Σ soddisfa $(V_1 \frown v_1) \frown v_2$ se e solo se la relazione 1_Σ vale tra 1_σ e 2_σ dove 1_σ è correlato con un 1^*_Σ e 2_σ è correlato con 2^*_Σ tali che $1^*_\Sigma = \langle n,m \rangle$ dove $n = 1_\sigma$ e $m = 1_\sigma$ e tali che $2^*_\Sigma = \langle k,l \rangle$ dove $k = 1_\sigma$ e $l = 1_\sigma$.

Quest'ultima condizione risulta piuttosto complessa ed è bene spendere qualche parola di chiarimento[24]. In base al metodo dell'unificazione semantica dobbiamo cercare di ridurre la categoria degli individui a quella delle relazioni diadiche. Ora, nel caso in esame ci sono due individui, che nel linguaggio oggetto sono rappresentati dalle variabili x_1 e x_2 i cui nomi metateorici sono v_1 e v_2. Grazie alla procedura di correlazione esposta sopra, 1_σ cioè l'individuo che 'va al posto di' x_1 viene messo in relazione con una coppia ordinata, cioè un particolare attributo diadico che abbiamo chiamato 1^*_Σ costituito a sua volta da elementi identici a 1_σ e quindi identici tra loro[25].

Tarski riconosce che il metodo dell'unificazione semantica presenta aspetti di artificiosità che mancano, invece, nella prima strategia presentata. Tuttavia, solo l'unificazione semantica è in grado di essere applicata a linguaggi del terzo tipo, cioè a linguaggi che presentano un numero di categorie semantiche infinito nonostante il loro ordine rimanga finito. Linguaggi siffatti si differenziano dalla logica delle relazioni diadiche dal momento che presentano una serie infinita di variabili di ordine superio-

---

24 Naturalmente, potrebbero sorgere confusioni tra gli indici di variabili appartenenti a categorie differenti. Per ovviare a questo Tarski propone di assegnare alle variabili individuali solo indici dispari e alle variabili relazionali solo indici pari.

25 Come è facile notare, Tarski considera gli attributi, in questo caso le relazioni diadiche, in maniera puramente estensionale, cioè come insiemi di coppie ordinate; questo non ha alcuna rilevanza sulla dimensione logica della definizione ma è molto interessante in sede di discussione filosofica. Come avremo modo di vedere, questo fatto suggerisce almeno due spunti: innanzitutto l'atteggiamento filosofico di Tarski sembra essere scettico circa le intensioni; in secondo luogo, rimane da chiarire se la categoria fondamentale di enti che la semantica tarskiana deve assumere non sia in fondo costituita dagli individui dal momento che le coppie ordinate sono, appunto, in ultima analisi, particolari insiemi di individui. Ci riserviamo, comunque, di discutere questi aspetti nel prossimo capitolo.

re. Non ci sono cioè solo variabili per relazioni a due posti, ma a tre, quattro e così via. Il nostro alfabeto si complica ulteriormente; infatti:

Alfabeto (Logica delle relazioni n-adiche)
a) Variabili individuali: $x_1, x_2, x_3, \ldots$
b) Variabili predicative: $X_1^1, X_2^1, X_3^1, \ldots, X_1^2, X_2^2, \ldots$
c) Costanti logiche: $\neg, \vee, \forall$

Dove con X_k^n indichiamo la *k*-esima variabile predicativa a *n* posti. $X_k^n(x_1,...,x_n)$ sarà quindi una formula ben formata il cui nome, metateorico, sarà $(V_k^n \frown v_1) \frown v_2 \frown \ldots \frown v_n$.

Vediamo ora perché in linguaggi di questo tipo il metodo delle sequenze multi-riga fallisce:

> In the case we are investigating the number *n* [of sequences of various categories] is indefinitely large and the metalanguage we are using – like all other actually existing formalized languages – provides no means for dealing with the relation of mutual dependence between objects which belong to infinitely many distinct categories[26].

La ragione è quindi molto semplice: nel metalinguaggio avremmo bisogno di una matrice di sequenze con un numero infinito di righe per comprendere tutte le possibili categorie semantiche delle variabili di ordine superiore. Ma questo è al di là delle capacità espressive del linguaggio e quindi l'unico metodo a disposizione rimane quello dell'unificazione semantica.

Ma come avviene, nel concreto, l'unificazione semantica per linguaggi di questo tipo? È sufficiente correlare ogni relazione a *n* posti R tra individui, una classe R* che consiste di *n* sequenze di individui, ovvero la classe di tutte le sequenze che soddisfano la seguente condizione: la relazione R vale tra gli individui $a_1,\ldots,a_n$. In parole più semplici, ogni relazione è, estensionalmente, un insieme di *n*-ple ordinate. Un esempio chiarirà tutto: assumiamo di avere a che fare con la relazione 'essere più alto di', che chiameremo per comodità A. Ora, A vale tra due individui a e b se e solo se a è più alto di b. Quale sarà l'estensione di A? Sarà costituita dall'insieme di tutte le coppie ordinate $\langle x,y \rangle$ in cui *x* è più alto di *y*. Lo stesso procedimento può essere esteso a relazioni più complesse. Ma

26 Tarski 1983, 232.

ogni sequenza di individui può essere rappresentata come una particolare relazione diadica tra individui e numeri naturali e quindi, indipendentemente dal numero di elementi nella sequenza, ogni sequenza apparterrà alla medesima categoria semantica. Quindi ogni funzione proposizionale che contiene variabili di ordine superiore che indicano relazioni tra n-termini con n grande a piacere può essere trasformata in una funzione che contiene esclusivamente variabili per relazioni diadiche, e quindi appartenenti alla stessa categoria semantica.

È importante considerare un ultimo aspetto del metodo dell'unificazione semantica: la categoria tramite la quale compiamo l'unificazione deve essere almeno di ordine uguale al massimo ordine delle variabili del linguaggio in esame[27]. Se, come nel nostro caso, abbiamo una relazione R che è del secondo ordine, la categoria unificante dovrà essere al minimo del secondo ordine e, infatti, si tratta di una classe di sequenze di individui. Questa condizione rivestirà un ruolo cruciale per i risultati di limitazione che esporremo nel paragrafo presente.

In conclusione, quindi, la definizione tarskiana di verità, costruita inizialmente per un linguaggio abbastanza povero, quale il calcolo delle classi, può essere estesa a linguaggi considerevolmente più espressivi. In particolare, sono a disposizione due metodi per adattare il concetto di soddisfazione alle pluralità di categorie semantiche presenti in questi linguaggi: il metodo delle sequenze multi-riga e il metodo dell'unificazione semantica. Mentre il primo metodo, benché più naturale, è applicabile solamente a linguaggi del secondo tipo, cioè con un numero finito di categorie semantiche, è necessario il ricorso all'unificazione semantica per i linguaggi del terzo tipo, ovvero linguaggi che presentano una serie di variabili predicative di complessità qualsiasi.

Cosa accade, però, se passiamo all'ultima tipologia di linguaggi presentati, quelli di ordine infinito? L'ultima sezione del lavoro di Tarski è dedicata proprio allo studio di questi linguaggi e alla presentazione dei risultati che riguardano l'indefinibilità della verità.

---

27 Su questo punto vedi Tarski 1983, 244.

5. *Linguaggi di ordine infinito*

Nell'ultima parte del suo saggio, Tarski presenta come esempio di linguaggio di ordine infinito il calcolo generale delle classi[28]. Si tratta della generalizzazione del linguaggio del primo ordine usato come caso prototipico nella definizione della verità. Qui, però, non c'è nessun limite all'ampiezza delle categorie semantiche in gioco: non solo, infatti, il numero delle categorie semantiche è infinito ma anche l'ordine delle variabili è infinito. Per cui, data una variabile di ordine n è sempre possibile costruirne una di ordine n+1. L'alfabeto presenta una certa somiglianza strutturale con i linguaggi di ordine finito (del terzo tipo):

Alfabeto (calcolo generale della classi)
a) Variabili predicative a un posto: $X_1^1, X_2^1, X_3^1, \ldots$
b) Variabili predicative a due posti: $X_1^2, X_2^2, X_3^2, \ldots$
c) Variabili predicative a n posti: $X_1^n, X_2^n, X_3^n, \ldots$
d) Costanti logiche: ¬, ∨, ∀

Abbiamo diversificato la classe delle variabili per semplificare la presentazione anche se si tratta di un'unica categoria di segni. In effetti, gli individui sono rappresentati dalle variabili di primo ordine (X_1^1, X_2^1, X_3^1, …); come esempio di formula ben formata abbiamo, quindi,

(17) $X_k^{n+1} X_l^n$

il cui significato inteso è: l'entità descritta dalla variabile X_k^{n+1} si predica dell'entità descritta dalla variabile X_l^n . Questo permette, poi, due

---

28 Tarski spende qualche decina di righe per enfatizzare la semplicità e insieme la potenza del calcolo generale delle classi. In particolare, afferma che esso

> suffices for the formulation of every idea which can be expressed in the whole of mathematical logic (Tarski, 1983, 242).

La teoria generale delle classi è una teoria degli insiemi; è materia di sicuro interesse filologico ed esegetico indagare l'effettiva consapevolezza della distinzione tra linguaggio e teoria nel Tarski degli anni '30. In particolare, ciò risulta dalla comparazione che lo stesso Tarski fa tra il calcolo generale della classi e la teoria dei tipi presentata nei *Principia* di Russell e Whitehead.

interpretazioni: l'una, estensionale, che legge (17) come l'elemento X_l^n appartiene all'insieme (o alla classe) X_k^{n+1}; l'altra, intensionale, secondo la quale X_l^n ha la proprietà X_k^{n+1}. Secondo le convenzioni simboliche adottate da Tarski, il nome metateorico di (17) è $\mathrm{V}_k^{n+1} \frown \mathrm{V}_l^n$.

Gli assiomi specifici del calcolo generale delle classi consistono, poi, di tre principi insiemistici: il principio di comprensione, di estensionalità e dell'infinito. Ci soffermeremo un istante sulla formulazione che Tarski offre di questi assiomi perché ci aiuteranno (soprattutto il primo) a comprendere un passaggio cruciale nella dimostrazione del teorema di indefinibilità della verità. Tarski chiama 'pseudodefinizioni' le istanze del seguente assioma[29]:

(18) $\exists X_k^{n+1} \forall Y_l^n \left(\left(X_k^{n+1} Y_l^n \wedge \alpha \right) \vee \left(\neg X_k^{n+1} Y_l^n \wedge \neg\alpha \right) \right)$

Di per sé il significato del principio rimane piuttosto oscuro; per gettare un po' di luce è bene assumere che le variabili in questione siano *X* per la *k*-esima variabile di ordine n+1, cioè X_k^{n+1} e *y* per la *l*-esima variabile di ordine n e cioè Y_l^n. Il principio diventa allora:

(19) $\exists X \forall y \left(\left(Xy \wedge \alpha \right) \vee \left(\neg Xy \wedge \neg\alpha \right) \right)$

Ora, poiché da $\left(Xy \wedge \alpha \right)$ seguono sia $Xy \rightarrow \alpha$ che $\alpha \rightarrow Xy$, abbiamo che $\left(Xy \wedge \alpha \right) \rightarrow \left(Xy \leftrightarrow \alpha \right)$. D'altro canto anche da $\left(\neg Xy \wedge \neg\alpha \right)$ seguono $Xy \rightarrow \alpha$ e $\alpha \rightarrow Xy$; cioè $\left(\neg Xy \wedge \neg\alpha \right) \rightarrow (Xy \leftrightarrow \alpha)$. Ma allora abbiamo che $\left(\left(Xy \wedge \alpha \right) \vee \left(\neg Xy \wedge \neg\alpha \right) \right) \rightarrow \left(Xy \leftrightarrow \alpha \right)$; quindi il principio di comprensione può essere formulato in maniera logicamente equivalente come:

---

29 La simbologia originale di Tarski non è, come si è potuto constatare, affatto maneggevole. Tuttavia, abbiamo ritenuto importante non discostarci troppo dalla formulazione originale per mettere in luce alcuni punti teoricamente significativi. Nella formulazione del principio di comprensione però sorvoliamo sulla distinzione tra nomi metalinguistici dei segni e i segni stessi, riservandoci di chiarire qualora necessario il livello linguistico in cui ci troviamo. Del resto, utilizzare i vari nomi presentati in precedenza comporterebbe una sequenza di segni davvero lunga e di fatto inutile; lo stesso Tarski introduce delle ulteriori regole di accorciamento per i nomi di elementi del linguaggio oggetto, cfr. Tarski 1983, 175-176.

(20) $\exists X \forall y (Xy \leftrightarrow \alpha)$

secondo l'uso consolidato; Tarski aggiunge poi la nota restrizione circa la variabile X che non deve occorrere libera nella formula α. Si noti che senza il quantificatore esistenziale di ordine superiore, (18) è un principio di definizione.

La legge (o assioma di estensionalità) ha invece questa forma:

(21) $\forall X^{n+2} \forall Y^{n+1} \forall Z^{n+1} \left(\left(\neg \exists x \left(Yx \wedge \neg Zx \vee \neg Yx \wedge Zx \right) \right) \rightarrow \left(XY \rightarrow XZ \right) \right)$

Date due classi Y e Z, se non esiste alcun elemento che Y possiede e Z no (o viceversa) allora Y e Z hanno le stesse proprietà e quindi sono uguali.

Infine, è necessario per lo sviluppo delle varie teorie matematiche, un assioma dell'infinito che dichiari l'esistenza di un numero infinito di individui:

$$\exists R \left(\exists X \left(R(X) \wedge \forall X \left(\begin{array}{l} R(X) \rightarrow \exists Y \left(R(Y) \wedge \forall x (Yx \rightarrow Xx) \wedge \right. \\ \left. \exists x \left(Xx \wedge \neg Yx \right) \right. \end{array} \right) \right) \right)$$

Con questi principi specifici del calcolo generale delle classi e con un insieme di regole di derivazione (quali il *modus ponens* e la generalizzazione universale) si conclude la presentazione del linguaggio di ordine infinito.

Perché in linguaggi di questo tipo è impossibile definire il concetto di verità? Il motivo è concettualmente semplice. Come si ricorderà, avendo variabili appartenenti a più categorie semantiche è necessario adottare il metodo dell'unificazione semantica al fine di stabilire la relazione di soddisfazione. Tuttavia, un requisito indispensabile per l'applicazione del metodo prevede che la categoria unificante sia almeno pari all'ordine massimo delle variabili presenti nel linguaggio. Ma proprio qui sta il nocciolo della questione. Nei linguaggi di ordine infinito non esiste un ordine massimo delle variabili dal momento che per definizione, data una variabile di ordine n ne esiste sempre una di ordine n+1. Quindi non è possibile trovare nel metalinguaggio una categoria unificante di ordine superiore e, di conseguenza, non è possibile stabilire la relazione di soddisfazione necessaria alla definizione di verità.

Si tratta di un fatto tanto chiaro quanto sorprendente; in effetti, data la definizione di soddisfazione e i suoi ampliamenti teorici, attraverso l'introduzione delle categorie semantiche e del metodo dell'unificazione semantica, risulta che per un linguaggio di ordine infinito è impossibile definire la verità. In effetti, i linguaggi di ordine infinito sono, per così dire, strutturalmente aperti, in modo che non è possibile la costruzione di un metalinguaggio adeguato alla descrizione del predicato di verità.

L'interrogativo che Tarski si pone[30] immediatamente è, però, fondamentale: ci si chiede se l'indefinibilità della nozione di verità per linguaggi di ordine infinito abbia una causa tutto sommato contingente, (derivante, magari, dai metodi impiegati) o sia invece un tratto strutturale dei concetti coinvolti in queste ricerche. In altre parole, si tratta di un insuccesso momentaneo o di un vero e proprio risultato di impossibilità, ove si sancisce un limite fondamentale, di principio? Il problema assume quindi la seguente forma:

> [I]t is a question of *whether on the basis of the metatheory of the language we are considering the construction of a correct definition of truth in the sense of convention* T *is in principle possible*[31] .

La risposta a questa domanda è negativa ed è contenuta nell'ultimo dei molti tesori presenti nel saggio di Tarski: il teorema di indefinibilità della verità.

L'idea di fondo del teorema è piuttosto semplice e si basa su due punti fondamentali:

i) è possibile codificare aritmeticamente le espressioni del metalinguaggio

ii) è possibile esprimere relazioni aritmetiche nel calcolo generale delle classi

Il primo punto riguarda una procedura, detta comunemente aritmetizzazione, che si attribuisce solitamente al genio di Kurt Gödel. In base all'aritmetizzazione si stabilisce una corrispondenza biunivoca tra le espressioni di un linguaggio (cioè la sua sintassi) e i numeri naturali in

30 Tarski 1983, 246.
31 Tarski 1983, 246.

modo che a ogni espressione corrisponda un codice, cioè un numero e viceversa[32].

Il secondo punto, invece, concerne il fatto che in una teoria delle classi sufficientemente potente è possibile immergere l'aritmetica; in base alla ricostruzione compiuta da Russell e Whitehead (sfruttando le intuizioni di Frege) nei *Principia*, un numero naturale può essere considerato come la classe di tutte le classi equinumerose dove l'equinumerosità viene definita come corrispondenza biunivoca degli elementi. In altre parole, che cos'è il numero 5? È la classe di tutte le classi composte da cinque oggetti, i cui elementi, cioè, possono essere messi in corrispondenza biunivoca. I numeri sono, allora, oggetti del terzo ordine, ovvero classi di classi di individui.

Questi due risultati insieme conducono al risultato che ci interessa. Infatti, se il metalinguaggio può essere completamente aritmetizzato, cioè espresso nel linguaggio dell'aritmetica, e se, a sua volta, l'aritmetica può essere espressa nel linguaggio del calcolo delle classi, segue che il linguaggio oggetto, cioè il calcolo delle classi è in grado di 'parlare' del suo metalinguaggio. Come si ricorderà, presupposto per la trattazione tarskiana del predicato di verità è la distinzione tra linguaggio oggetto e metalinguaggio, con il secondo essenzialmente più ricco del primo. Ora, in base ai punti i) e ii) questa distinzione cade, dal momento che, *via* aritmetica, il linguaggi oggetto comprende il metalinguaggio. Riappare, cioè, proprio quell'aspetto di universalità del linguaggio che era stata la sorgente delle antinomie nel tentativo di definizione della verità per il linguaggio naturale. Diventa possibile, cioè, ricostruire l'antinomia del mentitore nel metalinguaggio, ponendo nel linguaggio un enunciato x tale che l'enunciato ad esso correlato nel metalinguaggio dice che x non è

---

32 Anche in questo caso il materiale per l'esegesi del testo tarskiano non manca; l'aritmetizzazione fu un risultato che Gödel e Tarski scoprirono indipendentemente o vi furono delle influenze di uno sull'altro? In una nota al testo, Tarski chiarisce la sua posizione e pur riservando parole di grandissima ammirazione per il lavoro di Gödel afferma la piena autonomia della sua scoperta. D'altro canto, Mancosu è chiaro nel ribadire che il teorema di indefinibilità della verità è presentato da Tarski *dopo* aver letto il lavoro di Gödel (Mancosu et al. 2009, 441); ci sono anzi indizi testuali circa la priorità di Gödel su Tarski non solo riguardo la procedura di aritmetizzazione quanto circa lo stesso risultato di indefinibilità della verità. Cfr., a tal proposito, Murawski 1998, Wolenski, Simons 1989, Feferman 2004, Grattan-Guinness 1979.

vero[33]. Ma come avviene tutto ciò nel concreto? Ci soffermeremo solo sul secondo dei punti enunciati, e cioè sulla possibilità di interpretare nel linguaggio delle classi gli enunciati aritmetici sorvolando sulla procedura di aritmetizzazione che, oltre a essere solo citata nel lavoro di Tarski, ci sottrarrebbe troppo tempo.

Dato un numero (naturale) k è possibile costruire una funzione enunciativa α_k che presenta un'unica variabile libera e che asserisce che la classe (il cui nome è rappresentato dal simbolo α_k) è identica al numero k. In termini un po' più colloquiali: dato un qualsiasi numero è possibile costruire una classe che ha esattamente quel numero di individui. Di che tipo sarà la variabile libera in α_k? Si tratterà di una variabile per numeri, ovvero una variabile del terzo ordine; noi la indicheremo sia con la simbologia classica X_1^3 sia con l'abbreviazione n. A titolo esemplificativo prendiamo in esame il caso di α_1 (cioè la classe costituita da un solo elemento):

$$\alpha_1 =_{def.} \forall X_1^2(X_1^3 X_1^2 \wedge \exists X_1^1 \forall X_1^2 \forall X_2^2(X_1^2 X_1^1 \wedge (\neg(X_1^2 X_2^1) \vee \neg(X_2^2 X_1^1) \vee (X_2^2 X_1^1)) \vee$$
$$\neg X_1^3 X_1^2 \wedge \forall X_1^1 \exists X_2^1 \exists X_2^2(\neg(X_1^2 X_1^1) \vee X_1^2 X_2^1 \vee X_2^2 X_1^1 \wedge \neg(X_2^2 X_2^1))))$$

Per chiarire un po' la formulazione, si osservi che occorrono tre categorie di variabili:

$X_1^1, X_2^1, X_3^1, \dots$	sono variabili del primo ordine, per individui e verranno indicate con x, y, z,…
$X_1^2, X_2^2, X_3^2, \dots$	sono variabili del secondo ordine, per classi di individui e verranno indicate con X, Y, Z,…
$X_1^3, X_2^3, X_3^3, \dots$	sono variabili del terzo ordine, per classi di classi di individui e verranno indicate con R, S, …

Secondo queste convenzioni, avremo che:

$$\alpha_1 =_{def.} \forall X[[R(X) \wedge \exists x \exists y \forall Y(Xx \wedge (\neg Xy \vee \neg Yx \vee Yy))] \vee$$
$$[\neg R(X) \wedge \forall x \exists y \exists Y(\neg Xx \vee Xy \wedge Yx \wedge \neg Yy)]]$$

come è facile vedere, si tratta di un'istanza del principio di comprensione senza, naturalmente, la chiusura esistenziale; in base a quanto mostrato in precedenza possiamo riscrivere il tutto come:

---

33 Cfr. Tarski 1983, 248 e, prima, 164.

$$\alpha_1 =_{def.} \forall X \big(R(X) \leftrightarrow \exists x \exists y \forall Y \left(Xx \wedge \left(Xy \rightarrow (Yx \rightarrow Yy) \right) \right) \big)$$

Il significato intuitivo è piuttosto semplice: una classe possiede un solo individuo se e soltanto se possiede un individuo e se ne possiede un altro, questo coincide con il primo.

La forma logica di questa definizione è quindi:

$$\alpha_1 =_{def.} \forall X \left(R(X) \leftrightarrow \beta(X) \right)$$

dove l'unica variabile libera è la *R*, che corrisponde a una variabile del terzo ordine (X^3) e dove $\beta(X)$ rappresenta la condizione espressa dalla stringa di tre quantificatori. Con ciò abbiamo mostrato come il linguaggio delle classi è in grado di rappresentare concetti aritmetici come i numeri cardinali.

Supponiamo, a questo punto, di aver definito la classe Tr di enunciati veri nel metalinguaggio. In base alla procedura di aritmetizzazione a Tr corrisponde una classe di numeri naturali, cioè i codici degli enunciati presenti in Tr; è possibile quindi definire una sequenza ϕ di oggetti, in questo caso numeri naturali, in cui sono presenti i codici delle proposizioni in esame. Si consideri l'espressione:

(23) $\exists X_1^3 \left(\alpha_n \wedge \phi_n \right) \notin \mathrm{Tr}$

In (23) si asserisce l'esistenza di una classe di *n* elementi e che il numero corrispondente, cioè l'ennesimo elemento della sequenza ϕ, non appartiene alla classe Tr. Si ricordi che i numeri sono per Tarski classi di classi di individui, secondo l'impostazione di Russell. Chiaramente (23) è una funzione proposizionale del metalinguaggio che contiene come unica variabile n. Ora, poiché (23) è formulata nel metalinguaggio segue, per aritmetizzazione, che esiste una funzione di carattere aritmetico che costituisce la 'traduzione' di (23); chiamiamo $\psi(n)$ la versione aritmetizzata di $\exists X_1^3 \left(\alpha_n \wedge \phi_n \right) \notin \mathrm{Tr}$. Poiché si tratta di una traduzione fedele, avremo il seguente risultato:

per ogni *n*, $\exists X_1^3 \left(\alpha_n \wedge \phi_n \right) \notin \mathrm{Tr}$ se e solo se $\psi(n)$

La funzione $\psi(n)$ è un determinato codice numerico e poiché il linguaggio dell'aritmetica può essere trattato nel calcolo generale delle classi (in base alla seconda condizione enunciata in precedenza), alla funzione $\psi(n)$ di carattere aritmetico corrisponderà un termine della sequenza ϕ, per esempio il termine con indice *k*. Possiamo quindi porre che $\psi(n) = \phi_k$. Ma se sostituiamo l'indice *k* per la variabile *n* otteniamo:

(24) $\exists X_1^3 \left(\alpha_k \wedge \phi_k\right) \notin \mathrm{Tr}$ se e solo se $\psi(k)$

A questo punto, $\exists X_1^3 \left(\alpha_k \wedge \phi_k\right)$ è un enunciato dal momento che non possiede più la variabile libera *n*. A che cosa corrisponde, dal punto di vista intuitivo, $\exists X_1^3 \left(\alpha_k \wedge \phi_k\right)$? Un attimo di riflessione ci fa comprendere che altro non è che la traduzione logico-aritmetica della proposizione del mentitore. Poiché, però, $\exists X_1^3 \left(\alpha_k \wedge \phi_k\right)$ è un enunciato dovrà seguire le condizioni imposte dal T-schema che ricordiamo presenta la seguente forma generale:

(T) *x* è un enunciato vero se e solo se *p*

Ora, al posto di *x* è necessario porre un nome dell'enunciato in questione e cioè, nel nostro caso, $\exists X_1^3 \left(\alpha_k \wedge \phi_k\right)$. La variabile *p* va invece sostituita con una traduzione metalinguistica dell'enunciato; ma, dal momento che il metalinguaggio può essere completamente aritmetizzato, segue che la traduzione di $\exists X_1^3 \left(\alpha_k \wedge \phi_k\right)$ è proprio $\psi(k)$. Otteniamo allora che:

(25) $\exists X_1^3 \left(\alpha_k \wedge \phi_k\right) \in \mathrm{Tr}$ se e solo se $\psi(k)$.

Ma, come si vede, (24) e (25) sono in contraddizione. Pertanto ecco dimostrato il teorema di indefinibilità della verità: nella misura in cui il linguaggio oggetto diventa in grado di parlare del metalinguaggio (grazie ai requisiti prima citati) si ritorna nella situazione di chiusura semantica, fonte delle antinomie.

Il commento di Tarski al teorema è di notevole lucidità e incisività.

> The metalanguage in which we carry out the investigation contains exclusively structural-descriptive terms, such as names of expressions of the language, structural properties of these expressions, […] What we call metatheory is, fundamentally, the *morphology of language* – a science of the form of expressions[34].

---

34 Tarski 1983, 251.

Qui Tarski prende in esame la natura del metalinguaggio (oggi noi diremo della metateoria) nel quale formulare la definizione di verità. Il metalinguaggio ha una natura essenzialmente logica nel senso che ha a che fare con la dimensione strutturale del linguaggio («Science of the form of expressions»). È tuttavia improprio affermare che Tarski pensasse di aver ridotto la dimensione semantica a quella sintattica anche per il semplice motivo che, in quelle pagine, egli sta trattando proprio i concetti semantici fondamentali. Piuttosto, la definizione tarskiana di verità mostra come un concetto semantico intuitivo (e cioè quello di verità insieme ai suoi correlati: denotazione, riferimento e così via) può essere studiato in maniera rigorosa impiegando le risorse espressive di un metalinguaggio quale quello presentato da lui.

La possibilità di definire la verità attraverso l'impiego del metalinguaggio non è però omogenea. Quando però si passa a linguaggi di ordine infinito emerge un limite strutturale: è impossibile costruire una definizione corretta di verità per tali linguaggi. Infatti:

> [A]ll concepts and all grammatical forms of the metalanguage found an interpretation in the language and hence we were able to show conclusively that the semantics of the language could not be established as a part of its morphology[35].

Ma davvero il teorema di Tarski e, in generale, la sua definizione della verità, intendono la semantica come una controparte sintattica del linguaggio in esame? Discuteremo questo punto in seguito, nelle interpretazioni filosofiche della concezione tarskiana della verità.

Il risultato di indefinibilità sembra essere dunque un limite strutturale per la definizione di verità; non appena il linguaggio assume una certa potenza e generalità diventa di principio impossibile definire, nel metalinguaggio, la classe delle proposizioni vere Tr. Ma esiste un'altra strada (già intravista da Tarski) la cui analisi occuperà la seconda parte di questo volume. Si tratta di introdurre il predicato T come un primitivo e fornire degli assiomi che ne regolino l'applicazione. In embrione, Tarski enuncia ciò che nella letteratura seguente è divenuto lo studio delle teorie assiomatiche della verità. Nelle pagine successive discuteremo invece la portata filosofica della concezione semantica della verità e le critiche che, da vari versanti, che sono state mosse all'opera di Tarski.

---

35 Tarski 1983, 254.

II
LA RILEVANZA FILOSOFICA DELLA CONCEZIONE TARSKIANA DI VERITÀ

Quasi tutti gli studiosi concordano che è difficile sovrastimare l'importanza del *Concetto di verità nei linguaggi formalizzati* di Tarski per lo sviluppo della logica matematica e della semantica formale. Ma non è solo all'interno di questi ambiti disciplinari che l'opera del logico polacco è cruciale; anche per il sapere filosofico, il contributo di Tarski segna una cesura; dal 1935 (anno di pubblicazione della versione tedesca[1]) tutte le discussioni circa la natura della verità devono in qualche modo fare i conti con la cosiddetta concezione semantica. Tuttavia, se l'importanza del *Concetto di verità* nella storia della logica del Novecento non è sostanzialmente messa in discussione ed è, se mai, interessante ricostruire come in effetti sia avvenuta la diffusione della semantica tarskiana nella comunità degli studiosi di logica, fino a costituire parte del canone indiscusso della disciplina, per quanto riguarda l'impatto sul mondo filosofico le cose non sono così semplici.

A tal proposito si possono delineare due linee di ricerca in un certo senso complementari ma ben distinguibili: la prima comprende questioni eminentemente esegetiche che si interrogano sulle effettive intenzioni di Tarski. Fino a che punto il logico polacco voleva contribuire al dibattito filosofico circa la nozione di verità[2]? E ammettendo la non estraneità a simili questioni concettuali (in opposizione alle ragioni logico-formali del suo lavoro), possiamo dire quali fossero le intenzione di Tarski nei riguardi della sua teoria della verità tarskiana in relazione alle classiche concezioni filosofiche della verità? Naturalmente, accanto a simili problemi piuttosto generali, la ricostruzione storica si specifica nell'analisi di

---

1 Per una rassegna bibliografica, vedi nota 1 del capitolo precedente.

2 Significativo è già il titolo di un saggio di J. Wolenski, *Alfred Tarski as a Philosopher* (cfr. Wolenski 1993).

questioni più circoscritte, quale per esempio la concezione di Tarski dell'universo oggettuale di riferimento e il suo cambiamento di prospettiva avvenuto tra i primi lavori della metà degli anni Trenta e gli anni Cinquanta.[3] Dedicheremo alle questioni di ordine esegetico poco spazio preferendo concentrarci sull'altro ordine di problemi suscitati dalla riflessione sull'opera di Tarski; d'altro canto, come avremo modo di vedere, per i nostri scopi, le evidenze testuali a nostra disposizione sono del tutto sufficienti per una risposta argomentata ad alcune domande preliminari. Ciò che invece occuperà il resto del capitolo è l'analisi della seconda linea di ricerca che concerne, al contrario, la rilevanza filosofica della concezione tarskiana di verità, *al di là* delle effettive intenzioni o posizioni del logico polacco. In particolare, ci interessa una questione assolutamente centrale nel dibattito attuale: vedremo, cioè, se e in quale misura la teoria semantica della verità può essere considerata una teoria corrispondentista e in che relazione sta con alcune intuizioni di base che caratterizzano le teorie filosofiche corrispondentista.

Nel prossimo capitolo, invece, prenderemo in considerazione alcune critiche mosse alla teoria tarskiana. Come spesso capita, le questioni filosofiche non presentano confini netti e, anzi, particolari posizioni assunte in un ambito si riflettono in altri; per questa ragione le questioni che ci interessano non possono essere trattate in maniera completamente disgiunta ma solo con una serie di richiami interni. Si noti infine che tutti gli ambiti indagati hanno a che vedere con la domanda di fondo che innerva tutto il nostro lavoro: la rilevanza filosofica della concezione semantica della verità. Infatti, qualora si riuscisse a mostrare convincentemente che la proposta tarskiana è assimilabile a una posizione corrispondentista e che, pertanto, non è affatto banale e non equivale a una riduzione della dimensione semantica a quella sintattica, avremo mostrato *a fortiori* la sua rilevanza filosofica.

1. *Le intenzioni di Tarski*

Prima di iniziare con le questioni teoriche che abbiamo appena citato, è bene indugiare un po' sul giudizio che lo stesso Tarski riservò al suo

---

3 Ci stiamo riferendo al cosiddetto 'problema del modello a dominio fisso' correlato strettamente con le nozioni di verità *tout court* o verità *in una struttura*. A questo proposito si vedano: Hodges 1985; Etchemendy 1990; Gomez-Torrente 1996, 1998; Bays 2001; Mancosu 2010.

lavoro in ottica filosofica. Sebbene la concezione semantica della verità sia contenuta ed esposta nell'ampio saggio *Il concetto di verità nei linguaggi formalizzati*, Tarski scrisse due altri lavori dedicati completamente a questo tema ma destinati alla divulgazione: si tratta de *The Semantic Conception of Truth and the Foundations of Semantics* e *Truth and Proof*.[4] In tutti questi saggi, Tarski è completamente chiaro circa le intenzioni anche filosofiche del suo contributo: la concezione semantica della verità è in linea con la teoria della verità classica, ovvero, la verità intesa come corrispondenza.

> We shall understand by semantics the totality of considerations concerning those concept which, roughly speaking, express certain connections between the expressions of a language and the objects and states of affairs referred to by these expressions […] The concept of truth also – and this is not commonly recognized – is to be included here, at least in his classical interpretation, according to which 'true' signifies the same as 'corresponding with reality'[5].

Del resto anche *Il concetto di verità* si apre con una caratterizzazione informale del predicato 'essere vero' che Tarski mutua da Kotarbinski ma che specifica poi, in nota, come risalente allo stesso Aristotele:

> Dire di ciò che è che non è, o di ciò che non è che è, è falso, mentre dire di ciò che è che è e di ciò che non è che non è, è vero[6].

Tarski sembra apprezzare molto questo passo di Aristotele, tanto che risulta presente nei tre contributi citati. D'altro canto, Tarski associa a questa formulazione antica, alcune definizioni più moderne:

> The truth of a sentence consists in its agreement with (or correspondence to) reality[7].

O anche:

---

4 Cfr. Tarski 1944 e Tarski 1969. In ciò che segue citeremo sempre le edizioni originali dei saggi; le traduzioni italiane sono in Linsky 1969 e Tarski 1972.

5 Tarski 1983, 401.

6 Aristotele, *Metafisica*, 1011, 25ss.

7 Tarski 1944, 342.

> A sentence is true if it designates an existing state of affairs[8].

Tarski non menziona mai la provenienza precisa di queste formulazioni, probabilmente si tratta di elaborazioni e sintesi che egli aveva studiato dai suoi maestri. Che il concetto di verità sia, per Tarski, intrinsecamente connesso all'idea di una corrispondenza tra un polo linguistico e un polo extra-linguistico lo si può notare analizzando alcune sue affermazioni circa il carattere delle nozioni semantiche:

> Semantics is a discipline which, speaking loosely, deals with certain relations between expressions of a language and the objects (or 'states of affairs') 'referred to' by those expressions[9].

Proprio la mancanza di precisione di nozioni quali 'stato di cose', 'corrispondenza' e così via spinse Tarski alla ricerca di una formulazione scientificamente accettabile della teoria della corrispondenza. Notiamo, innanzitutto, che le nozioni semantiche hanno un carattere tipicamente *relazionale* (per esempio, la relazione di *denotazione* che vale tra un segno del linguaggio e un oggetto del dominio, per cui diremo che il segno *x denota* l'oggetto *o*); al contrario, almeno in apparenza, il concetto di verità descrive una proprietà: la proprietà di *essere vero*. Alcuni enunciati sono veri, cioè possiedono questa proprietà, mentre altri no. Del resto, anche nell'elaborazione della semantica Tarski spesso fa uso della seguente scrittura:

$$x \in \mathrm{Tr}$$

dove con x si intende il nome di un determinato enunciato e con Tr la classe degli enunciati veri. Ma tale classe non è altro che l'estensione della proprietà 'essere vero'. Ebbene, lo stesso Tarski chiarisce che il concetto di verità è semantico in un senso secondario (ma non per questo meno importante) rispetto altri; i concetti semantici si differenziano in due categorie: quelli che esprimono le connessioni tra linguaggio e mondo e quelli che descrivono tali relazioni, caratterizzando certe classi di espressioni attraverso tali relazioni. È evidente che il concetto di verità appartiene a quest'ultima categoria; la nozione di verità caratterizza la

---

8 Tarski 1944, 343.
9 Tarski 1944, 345. Ma vedi anche Tarski 1983, 252; Tarski 1983, 401-403.

classe degli enunciati veri tramite il riferimento a una nozione semantica definita in maniera rigorosa quale è la relazione di soddisfazione.

In conclusione, si può dire che le evidenze testuali non lasciano spazio a dubbi ragionevoli. Tarski considerava la sua teoria perfettamente inserita nella tradizione classica delle teorie della verità come corrispondenza. Tanto più che quando egli accenna ad altre teorie della verità, i riferimenti sono quasi sempre verso teorie pragmatiste (vero è ciò che è utile) e coerentiste (vero è ciò che è coerente con un insieme di assunzioni di fondo). Tuttavia è riscontrabile in lui uno scetticismo di fondo circa il sapere filosofico in quanto tale; interessante è, per esempio, uno dei commenti conclusivi a Tarski 1944:

> The history of science shows many instances of concepts which were judged metaphysical […] before their meaning was made precise; however, once they received a rigorous, formal definition, the distrust in them evaporated[10].

Contro un radicale anti-metafisicismo, Tarski si rivela più possibilista condividendo, almeno sembra, una sorta di tesi di continuità tra metafisica e scienza. Molti concetti prima appartenenti all'ambito metafisico si sono man mano chiariti (distillati, potremmo dire) fino a essere parte integrante della visione scientifica del mondo. Lo stesso potrebbe darsi, continua Tarski, con il concetto di verità alla cui emancipazione dal dominio dei filosofi ha contribuito lo stesso Tarski con la fondazione della semantica scientifica[11].

Non ci interessa addentrarci ulteriormente nella ricostruzione storica del pensiero tarskiano; è però interessante notare ancora alcuni aspetti circa la 'percezione' di Tarski del contesto filosofico a lui contemporaneo. Illuminante in questo senso è il resoconto di Jan Wolenski sulla genealogia che da Tarski risale fino a Brentano[12]. Perché, si chiede Wolenski, la definizione formale della semantica è nata proprio in Polonia, quando sia nella Vienna dei Circolisti sia nella Germania di Hilbert erano fiorenti

---

10 Cfr. Tarski 1944, 364.

11 Una simile impostazione risente di una certa eco popperiana; il che è plausibile considerando i rapporti intercorsi tra Popper e Tarski. Anche cronologicamente, il primo incontro fra i due avvenne a Praga nel 1935, quindi nove anni prima dello scritto di Tarski.

12 Wolenski, 2002; Rojszczak 2002.

scuole di logica? I fattori che determinarono la presenza di un *humus* filosofico adatto perché il genio di Tarski potesse svilupparsi nella forma che è passata alla storia sono cinque:

1. La dottrina dell'intenzionalità
2. Una particolare concezione del linguaggio e della logica
3. L'ammissione di metodi radicalmente non finitari
4. Un approccio esclusivamente estensionale
5. L'attenzione a definizioni precise

È chiaro che non tutti questi fattori presentano lo stesso 'peso' teorico. In particolare, gli ultimi due (che lo stesso Wolenski dichiara essere questioni più particolari) possono essere messi in relazione. Una definizione precisa, secondo gli standard metodologici che Tarski spesso richiama[13], deve soddisfare determinate caratteristiche formali; tra queste spicca il rifiuto di entità intensionali a favore di caratterizzazioni puramente estensionali. Per riprendere una questione già discussa: definire il concetto di verità equivale a definire in maniera rigorosa la classe degli enunciati veri e non la proprietà di 'essere vero' che appartiene proprio a quella sfera semantica intuitiva che abbisogna di una rigorizzazione.

Gli altri fattori che costituirono la base su cui sorgerà il lavoro di Tarski sono di ben altra portata e aiutano a spiegare (almeno parzialmente) la mancanza di una semantica logica presso altri autori. Innanzitutto la concezione del linguaggio (e della logica) che si affermò in Polonia è simile, per certi versi, a ciò che Jaakko Hintikka[14] chiama *Logic as a Calculus* in contrapposizione a un'altra impostazione filosofica generale - *Logic as a Language* - secondo cui il linguaggio è visto come un *medium universale*. Cosa significa tutto ciò? In estrema sintesi, secondo Hintikka, in base alla concezione del linguaggio come *medium* universale, ogni accesso al mondo avviene in una prospettiva linguistica: siamo immersi nel linguaggio che diviene, appunto, il tramite di ogni rapporto conoscitivo. Immediata conseguenza di questo è l'impossibilità *di principio* di trascendere il linguaggio e di interrogarci su di esso. Conseguentemente diventa priva di senso la domanda sui rapporti tra linguaggio e mondo

13 Cfr. Tarski 1983, 155 e anche il molto discusso passaggio sul fisicalismo in Tarski 1983, 406.

14 Vedi Hintikka 1988; Kusch 1989. Ma anche vanHeijenoort 1967 e Cocchiarella 1988 che ripresentano, pur con differenze, il tema.

che, come si ricorderà, è alla base delle intuizioni semantiche. In questa sorta di pan-linguismo spicca la figura di Wittgenstein che nel *Tractatus* scrive:

> *The limits of my language* mean the limits of my world[15].

Hintikka accomuna in questa concezione generale del linguaggio anche Frege e il primo Russell. Non è questa la sede per una disamina critica sull'effettiva consonanza tra le posizioni di Frege e Russell e la tesi del linguaggio come *medium* universale. Di certo, e forse è sufficiente per la tesi di Wolenski, in Frege non ha senso chiedersi la rispondenza del linguaggio con il mondo a patto che si precisi che con 'linguaggio' stiamo intendendo l'*Ideografia*, ovvero la lingua del pensiero puro.

Ma una conseguenza cruciale per lo sviluppo della semantica è che, secondo questo approccio, è bloccato ogni tentativo di reinterpretazione del linguaggio; la possibilità di reinterpretare il linguaggio è però alla base della semantica logica, perlomeno nella fase matura, cioè dagli anni Cinquanta. Se si considera, al contrario, il linguaggio come un calcolo, è perfettamente lecito interrogarsi su di esso, adottando un punto di vista metalinguistico. Ciò calza perfettamente con i rilievi dello stesso Tarski a proposito della chiusura universale dei linguaggi naturali. L'unica via di uscita dall'antinomia del mentitore è costituita dalla differenziazione del linguaggio (comunemente inteso) in linguaggio oggetto e metalinguaggio. La possibilità di variare gli universi di riferimento e di uscire così dall'unica struttura ammessa da Frege e da Wittgenstein (nel *Tractatus*) è presente già in Löwenheim[16] il cui lavoro è stato, a ragione, definito come l'inizio della teoria dei modelli.

Ci si potrebbe allora domandare per quale motivo, all'interno della scuola hilbertiana non sia fiorita una trattazione della semantica come quella tarskiana. Tanto più che, a livello del tutto intuitivo, Hilbert si pose problemi di natura strettamente semantica quale, per esempio, la comple-

---

15 Wittgenstein 1922, 5.6.
16 Cfr. Badesa 2009.

tezza del calcolo funzionale ristretto[17]. Tuttavia, lo sviluppo della semantica richiede (perlomeno) due requisiti non presenti nella scuola hilbertiana e nei Circolisti; si tratta dell'ammissione di metodi di dimostrazione radicalmente non finitari e di una sorta di 'sensibilità' filosofica nei confronti delle questioni semantiche. Nell'analisi di Wolenski, questi requisiti corrispondono agli ultimi due fattori che prenderemo in esame.

Benché discepolo di Lesniewski, Tarski non ereditò dal suo maestro il suo nominalismo estremo, almeno per quanto riguarda la cosiddetta 'pratica matematica'. Per quanto riguarda l'orientamento filosofico generale, invece, Tarski stesso dichiarava esplicitamente che:

> I am a nominalist. This is a very deep conviction of mine. It is so deep, indeed, that even after my third reincarnation, I will still be a nominalist[18].

Pur tralasciando i non trascurabili tratti del carattere di Tarski, incline a prese di posizioni decise e a effetto, rimane la questione esegetica di cosa intendesse Tarski con l'etichetta 'nominalismo'. È noto, per esempio, che intorno agli anni Trenta con 'realismo' si intendeva quasi sempre una posizione di realismo gnoseologico, secondo cui il soggetto conoscente è distinto, in certa maniera, dall'oggetto conosciuto. Quindi è lecito chiedersi quale accezione di nominalismo fosse adottata da Tarski. A titolo esemplificativo[19] si consideri la questione dei *truth-bearers* affrontata in controluce da Tarski. Sebbene egli respinga la nozione di proposizione, in quanto mancante della chiarezza necessaria per le sue analisi, egli specifica (in nota[20]) che è meglio intendere gli enunciati non come iscrizioni, o *token*, quanto piuttosto, come classi di enunciati, o *type*. Del resto, l'ammissione dei *types* è tipica anche in una prospettiva radical-

17 Per calcolo funzionale ristretto oggi si intende, sostanzialmente, il calcolo dei predicati del primo ordine. La completezza di cui si parla è di carattere semantico perché ci si chiedeva se ogni formula valida espressa in quel linguaggio potesse essere dimostrata a partire dagli assiomi (logici) e dalle regole di derivazione. In Hilbert Akermann 1929 il problema è lasciato aperto. Bisognerà aspettare la tesi di dottorato di Kurt Gödel per una risposta positiva al problema. Cfr. Gödel 1930.

18 Vedi Feferman, Feferman 2004, 52. Tarski affermò quanto sopra durante un simposio che si tenne quando egli era intorno ai settant'anni.

19 La questione dei *truth-bearers* è infatti molto più complessa. Vedi Rojszczak 2002 per una panoramica esauriente.

20 Tarski 1944, 370n.

mente finitista. Ciò che è significativo, per l'interpretazione delle concezioni filosofiche di Tarski, è il suo deciso impegno nei confronti del nominalismo pur ammettendo metodi, procedure e concetti radicalmente infinitari come si evince non solo dalle ricerche sulla verità ma anche, e forse maggiormente, dal lavoro in teoria degli insiemi.

In altre parole, e per tornare alla questione della pratica matematica, gli interessi di Tarski per la teoria degli insiemi mal si conciliano con l'ammissione dei metodi finitari perseguiti nel programma hilbertiano. È vero, del resto, che l'intento di Tarski non è primariamente quello di fornire una giustificazione del sapere matematico, come avviene invece per i programmi di fondazione classici. In questo senso, egli assume come perfettamente legittimo il discorso su insiemi infiniti o su sequenze infinite di oggetti, tanto che, nello specificare le caratteristiche determinanti del metalinguaggio, include in esso una sorta di sintassi allargata, comprendente, quindi, nozioni insiemistiche.[21]

Infine, ciò che favorì la nascita della semantica e che aiuta a chiarire un po' il patrimonio filosofico ereditato da Tarski è la teoria dell'intenzionalità. Qui sta sicuramente il punto più discusso e interessante della ricostruzione di Wolenski. La teoria dell'intenzionalità approdò nella Polonia dei primi anni del Novecento grazie all'influsso di Twardowski, allievo di Brentano. Come è noto, Twardowski introdusse la distinzione tra *contenuto* e *oggetto* di presentazione, difendendo la tesi secondo cui benché tutte le presentazioni abbiano oggetti cui sono dirette (secondo l'assunto fondamentale della teoria dell'intenzionalità), gli oggetti intenzionali sono reali nella maggior parte dei casi. In ciò Twardowski si differenzia dalla posizione di Brentano secondo cui gli oggetti intenzionali sono parti di fenomeni mentali[22].

La distinzione di Twardowski, nata in contesto fenomenologico non poteva essere assunta *qua talis* dalla comunità logica polacca a causa delle profonde convinzioni antipsicologiste di questi ultimi. Il processo derivante fu allora una sorta di spostamento di queste categorie, dall'ambito del *mentale* a quello *linguistico*, trasformando così aspetti intenzionali in aspetti semantici. I *nomi* diventano, allora, le controparti linguistiche delle presentazioni che, come abbiamo appena detto, rimanevano

21 Si tratta di un punto molto interessante per capire i rapporti tra sintassi e semantica; per questo motivo ci riserviamo di discuterlo in seguito.

22 Una discussione approfondita sulle teorie fenomenologiche dell'intenzionalità è totalmente fuori dai nostri intenti.

troppo ancorate alla dimensione psichica degli atti mentali. Il contenuto divenne pertanto il *significato* mentre l'oggetto di presentazione il *riferimento*. Così, le espressioni linguistiche ereditarono le proprietà semantiche dal carattere intenzionale degli atti mentali[23].

In conclusione, quindi, il sostrato filosofico di Tarski affonda le sue radici in una certa interpretazione della dottrina dell'intenzionalità che, unita a un forte antipsicologismo, diede origine alle categorie semantiche di contenuto e riferimento essenziali per lo sviluppo della semantica scientifica. Ma ciò ovviamente non bastava. Era necessario un apparato tecnico formale in grado di modellare matematicamente le nozioni semantiche: la teoria degli insiemi, perfettamente accettata una volta accantonate le preoccupazioni finitiste della scuola hilbertiana.

Tutto ciò sembra suggerire un 'primato del semantico' nella filosofia polacca e soprattutto nell'opera di Alfred Tarski. Questa è la conclusione di Wolenski che dichiara esplicitamente:

> [T]he concept of truth exceeds the arithmetical hierarchy of objects definable by arithmetical predicates. […] for any language sufficient for arithmetic, its semantics transcend its syntax. […] [S]emantics is prior to syntax[24].

Con ciò abbiamo però abbandonato la ricerca storica circa le opinioni dello stesso Tarski riguardo la sua teoria e, in generale, riguardo la semantica per sconfinare in una interpretazione squisitamente filosofica della semantica tarskiana. E questo ci conduce direttamente alla seconda parte di questo capitolo in cui ci chiederemo se la teoria di Tarski sia davvero corrispondentista.

2. *Quale corrispondenza?*

Prima di affrontare la questione se la concezione tarskiana della verità[25] possa essere annoverata tra le teorie corrispondentiste è necessario chiarire almeno in via preliminare cosa intenderemo in questa sede con tale denominazione.

---

23 Wolenski 2002, 20.

24 Wolenski 2002, 20, 22.

25 In ciò che segue non ci interesserà più l'effettiva rispondenza tra la proposta filosofica del corrispondentismo e le effettive opinioni di Tarski.

Si sottolinea spesso come l'intuizione della corrispondenza sia alla base del concetto comune di verità e costituisca, al tempo stesso, il nucleo delle più antiche teorie filosofiche della verità. In questo senso, è abusatissima la citazione del famoso passo di *Metafisica* 1011, 25 in cui Aristotele caratterizza una nozione corrispondentista di verità. Tuttavia, a un esame appena più attento, si osserva come a partire da questa radice comune, si dipani una serie infinita di problemi e questioni filosofiche di primaria importanza. Si può pertanto dire che lo scopo filosofico di una teoria della verità corrispondentista sia la spiegazione filosofica degli elementi che giocano un ruolo chiave nelle intuizioni di fondo circa la natura della verità.

Per i nostri scopi è sufficiente delineare alcuni tratti della verità come corrispondenza prima di entrare *in medias res*, ovvero nella comparazione della proposta tarskiana entro il quadro concettuale della teoria della corrispondenza. In via del tutto preliminare, l'intuizione di fondo della teoria corrispondentista può essere caratterizzata così:

(Corr-Base) *t* è vero se e solo se un elemento extra-linguistico lo rende tale

In (Corr-Base) *t* sta per qualsiasi truth-bearer cioè un'entità che può istanziare la proprietà di 'essere vero' o appartenere all'insieme degli oggetti veri, a seconda che si adotti un punto di vista intensionale o estensionale. A destra del bi-condizionale abbiamo invece due poli che dovranno essere esplicitati in analisi ulteriori:

1. Elemento extra-linguistico (polo ontologico)
2. Elemento relazionale (relazione di dipendenza)

Ovvero, l'elemento extra-linguistico è ciò che si chiama normalmente truth-maker[26]: il pezzo di realtà che rende vero *t*. Ma non è sufficiente indicare la polarità linguistica e quella ontologica; queste devono risultare relate in modo tale che la verità di *t* dipenda da come stanno le cose nel mondo. Si noti che, senza specificare ulteriormente i termini della questione, potremmo avere una situazione in cui l'elemento extra-lingui-

26 Come sarà chiaro a breve, per i nostri fini non è richiesta un'approfondita analisi del concetto di truth-maker e della relazione connessa di truth-making. Nella letteratura sterminata sul tema, rimandiamo al classico Armstrong (2004) per una panoramica delle posizioni più diffuse.

stico coincide con il mondo nella sua interezza che diventa quindi *truth-maker* di ogni proposizione vera. Tuttavia, una simile teoria della verità, che ricorda per certi versi alcune posizioni di filosofia rinascimentale riprese, seppur con distinzioni, nell'idealismo trascendentale tedesco, difficilmente verrà considerata corrispondentista. Non basta che la verità dipenda da come stanno le cose 'fuori dalla finestra' per essere arruolati nelle fila dell'esercito della corrispondenza; è necessario, infatti, aggiungere alcuni vincoli che andiamo a esaminare.

(Corr-weak) *t* è vera se e solo c'è un fatto *p* e *t* è correlato con *p*

In (Corr-weak) si è specificato il polo ontologico: ciò che rende veri i portatori di verità sono i fatti; ma cosa sono i fatti? Inutile dire che un'ontologia dei fatti esula dagli intenti di questo lavoro; assumiamo che i fatti siano stati di cose attuali, dove con stato di cose si intende un complesso di oggetti in relazione[27]. Talvolta ci si riferisce a (Corr-weak) come alla teoria della *corrispondenza-come-correlazione*[28]; secondo questa impostazione, il mondo è costituito da fatti che rendono veri gli enunciati (o le proposizioni o le credenze o qualsiasi portatore di verità); non si indaga però sulla natura della relazione che connette da un lato entità linguistiche e dall'altro pezzi di realtà: ciò che è lasciato indeterminato, in altre parole, è proprio il polo relazionale.

Wolenski e Simons connettono la corrispondenza-come-correlazione alla definizione aristotelica

> Dire di ciò che è che non è, o di ciò che non è che è, è falso, mentre dire di ciò che è che è e di ciò che non è che non è, è vero

sottolineando come sia sufficiente l'«è» (o rispettivamente il «non è») per garantire la connessione tra i due poli. A sua volta, poi, la copula può essere intesa in senso più o meno forte, cioè come un'asserzione esistenziale o di predicazione. Esiste però almeno un altro passaggio in Aristotele che suggerisce una posizione più complessa circa la corrispondenza:

27 Una chiarificazione: per complesso di oggetti in relazione contempliamo anche il caso minimale in cui un individuo *a* gode della proprietà *P*, come nel caso di 'Emma è bionda'.

28 Cfr. Wolenski, Simons 1989, 393ss; Kirkham 1992, 119ss.

> [S]arà nel vero chi ritiene essere separate le cose che effettivamente sono separate ed essere unite le cose che effettivamente sono unite; sarà, invece, nel falso, colui che ritiene che le cose stiano in modo contrario a come effettivamente stanno[29].

In questo caso, è in gioco un'altra intuizione: i portatori di verità sono resi veri dai fatti (o da altri truth-makers) in virtù del loro essere strutturalmente simili e di 'rispecchiare' la realtà sul piano linguistico. Secondo la corrispondenza-come-congruenza, quindi, non basta una semplice correlazione tra mondo e linguaggio ma è necessario che vi sia una sorta di similarità tra i due:

(Corr-strong) *t* è vera se e solo c'è un fatto p e t ≈ p

Dove con t ≈ p indichiamo una non meglio specificata relazione di congruenza; è ovviamente compito di chi sostiene una simile impostazione chiarire ed esplicitare questa nozione.

Quindi le questioni di base di ogni proposta corrispondentista possono essere sintetizzate nei punti seguenti:

- Natura dei portatori di verità
- Natura dei truth-makers (fatti, oggetti, eventi, situazioni, …)
- Natura della relazione tra linguaggio e mondo

(Corr-strong) ha il vantaggio di risultare maggiormente esplicativa rispetto alla sua controparte (Corr-weak) e come tale necessita di una serie di giustificazioni ulteriori. Ciò che è interessante per i nostri scopi è indagare se sia possibile un'analisi filosofica della teoria tarskiana senza considerare questioni cruciali come quelle appena enunciate. In altre parole, la rilevanza filosofica della teoria di Tarski è relativamente indipendente da un quadro ontologico di fondo? Come vedremo alla fine, rispondere a questa domanda equivale, per certi versi, proprio a chiarire l'importanza filosofica della teoria in esame.

Passiamo ora a chiarire un po' la relazione di congruenza che è presente in (Corr-Strong) e che, in qualche modo, dovrebbe essere il collante tra i due poli (quello linguistico e quello ontologico) prima descritti. Tra le innumerevoli caratterizzazioni proposte, tre tipologie sono particolar-

---

29 Metafisica, 1051b 3ss.

mente interessanti: *isomorfismo*, *connessione causale* e *relazione semantica*.

Seguendo Russell (1919) e alcuni passi di Wittgenstein (1922) si può affermare che una *t* è vera perché esiste un fatto che la rende tale e che c'è una sorta di somiglianza strutturale tra l'enunciato e lo stato di cose descritto. Il mondo è organizzato in modo analogo al linguaggio e alla relazione di predicazione linguistica che vale, per esempio, tra un nome e un predicato corrisponde un nesso di predicazione ontologica (o istanziazione, o partecipazione) tra un individuo e una proprietà. Interpretare la connessione posta dalla teoria della corrispondenza come una relazione di isomorfismo è un'assunzione estremamente forte e altrettanto problematica. Come mette bene in luce Kolár:

> To wit, there is no reason for the truth-makers, if any, to mirror exactly the structure of the truth-bearers[30].

Nella sua semplicità, il rilievo coglie perfettamente nel segno. Senza ulteriori giustificazioni, l'assunzione di isomorfismo è troppo impegnativa sia dal punto di vista ontologico che da quello linguistico. A questo proposito, poi, di quale linguaggio si sta parlando? Sembra che ogni riferimento alle lingue naturali sia fuori questione; rimarrebbero i linguaggi artificiali, come la logica dei predicati. Ma anche in questo caso, presupporre un isomorfismo in grado di spiegare la corrispondenza degli enunciati veri con la realtà è un'opzione ontologica piuttosto impegnativa.

Di carattere quasi opposto è l'idea di collegare linguaggio e mondo facendo leva su una connessione di tipo *causale*. Gli enunciati veri corrispondono al mondo perché sono in qualche modo dipendenti dall'azione causale di quest'ultimo. Anche in questo caso però i problemi che scaturiscono sono moltissimi. Innanzitutto, come nel caso dell'isomorfismo, non è affatto chiaro di quale tipo di relazione causale si tratti. Se assumiamo un contesto naturalistico, il rapporto di causa dovrà configurarsi come esclusivamente empirico. Come avremo modo di vedere, questo è uno dei punti più discussi da Hartry Field nel suo saggio sulla teoria della verità di Tarski (1972). Ovviamente, si possono postulare altri tipi di relazione causale ma in assenza di una teoria di riferimento sufficientemente solida, anche la proposta della connessione causale sembra portare più problemi di quelli che è in grado di risolvere.

---

30 Kolar 1999, 68.

Infine, la proposta che ci sembra più promettente anche se non scevra da problemi è quella di interpretare la relazione di congruenza come una relazione *semantica*. In un certo senso, gli enunciati significano determinate cose, hanno un significato che è lo stato di cose descritto dall'enunciato stesso: «Snow is white» significa che la neve è bianca. Parlando di semantica ci troviamo esattamente al punto in cui Tarski inizia il suo saggio sulla verità; come avevamo detto in precedenza (cfr. capitolo 1), la nozione di verità che Tarski ha in mente è una nozione semantica e corrispondentista e adesso possiamo chiarire meglio i termini della questione. La corrispondenza è basata sull'assunto minimale per cui la verità dipende da qualcosa di extra-linguistico mentre la specificazione semantica sta ad indicare, seppur molto sommariamente, il tipo di relazione che connette il polo linguistico da quello ontologico.

Si noti, per inciso, che in tutte le formulazioni prese in esame sopra rimane costante un punto di vista di realismo minimale: la verità dipende dal mondo, o meglio da come è fatto il mondo. Il che si accorda benissimo con le intuizioni del senso comune e di buona parte del sapere scientifico. Da questo punto di vista si può dire che la teoria corrispondentista è la teoria della verità del senso comune; lo stesso senso comune che ispira il "robusto" senso della realtà del sapere scientifico Ovviamente, adottare una teoria della corrispondenza pone alcuni problemi quando il dominio inteso delle nostre teorie non riguarda, almeno in prima battuta, oggetti del mondo di tutti i giorni. «7 + 5 = 12» e «Piove o non piove» sono proposizioni sicuramente vere; ma quale tratto nel mondo le rende tali? La questione cruciale può essere espressa come un dilemma; delle due l'una: o si ammette che la nostra teoria della verità è plausibile e bisogna fare i conti con un ontologia che ammetta regioni particolari quali quelle degli enti matematici oppure dobbiamo far saltare la nostra teoria della verità (e quindi la semantica) e parlare di più concetti di verità, come verità matematica, verità empirica, verità storica e così via.[31] Non ci interessa proseguire la discussione che ci porterebbe davvero troppo lontano; è però importante tenere a mente non solo che la nozione di verità si applica anche a proposizioni il cui dominio oggettuale di riferimento è più rarefatto rispetto a quello di tutti i giorni ma un uso del concetto di

---

31 Una variante di questo dilemma è rintracciabile nel celeberrimo articolo di Benacerraf, *Mathematical Truth*. Lì oltre al problema semantico e ontologico veniva inserita la questione dell'accesso conoscitivo agli enti matematici su uno sfondo di teoria causale della conoscenza. Cfr. Benacerraf 1973.

verità che Tarski aveva in mente era proprio quello speso all'interno del dibattito delle ricerche di metalogica; in un contesto, cioè, in cui gli esempi di sopra sono del tutto comuni.

In base alla caratterizzazione semantica è possibile analizzare ulteriormente (Corr-Strong). Infatti:

(Corr-strong) t è vera se e solo c'è un fatto p e $t \approx p$

Ma abbiamo appena visto che $t \approx p$ viene spiegato come

(Corr-sem) t è vera se e solo c'è un fatto p e t significa p

Ora, dire che «c'è un fatto *p*» e dire semplicemente «*p*» può essere considerata la stessa cosa a patto di non ammettere letture possibiliste.[32] Bisogna cioè assumere che l'operatore «esiste un fatto che» sia dispensabile; per questo

(Corr-sem2) t è vera se e solo p e t significa p

Al posto di *t*, che è un segno generico per un truth-bearer possiamo introdurre una variabile enunciativa (i cui valori sono cioè enunciati) e riscrivere il tutto come:

(Corr-sem3) x è vera se e solo p e x significa p

A questo punto rimane da analizzare la relazione semantica «*x* significa *p*». Ebbene, in maniera del tutto intuitiva *x* significa *p* quando la comprensione di *x* implica la comprensione di *p*, come accade nei consueti esempi di traduzione nelle lingue naturali. Ma allora è chiara la similarità con il T-schema di Tarski. Infatti:

(T) x è vero se e solo se p

Dove, come si ricorderà, *x* è una variabile per nomi di enunciati e *p* è una variabile per enunciati del metalinguaggio che siano gli equivalenti (metalinguistici) dell'enunciato in questione. Ma cosa vuol dire che *p* è un enunciato del metalinguaggio che rappresenta il contenuto di *x* se non

32 Ciò significa che, dichiarando *p*, dichiariamo anche l'attualità di *p*.

che esiste una relazione di sinonimia tra di essi? Si può quindi affermare che nel T-schema tarskiano è presente una teoria della corrispondenza nella sua versione forte? Vi sono autori che hanno abbracciato senza esitazione l'apparentamento tra il T-schema e una teoria corrispondentista della verità[33]. Altri sono molto più scettici; ciò che ci preme non è una tassonomia delle varie posizioni al riguardo quanto piuttosto una valutazione degli argomenti in quanto tali. È tempo di affrontare nel dettaglio la questione se e come la teoria tarskiana possa dirsi corrispondentista.

3. *Due definizioni di verità*

Il primo problema che sorge nell'interpretare la teoria tarskiana in ottica corrispondentista è il seguente: il T-schema non è la definizione di verità di Tarski. Si tratta probabilmente di uno dei più frequenti *misunderstanding* in letteratura ma la posizione del logico polacco è chiara. La convenzione T afferma che una teoria soddisfacente della verità deve soddisfare il requisito di adeguatezza materiale (oltre che di correttezza formale). Tale requisito consiste nella possibilità di derivare tutti i T-schemi; ma questa lista (infinita) non è una definizione della verità. Men che meno lo è una sola istanza, quale l'abusata

(1) «La neve è bianca» è vero se e solo se la neve è bianca.

Tarski è esplicito al riguardo in almeno due punti della sua memoria. *In primis*, quando presenta i fondamenti della sua teoria e introduce il T-schema; poi, nelle considerazioni finali dove prende in esame la possibilità di una teoria della verità costituita dall'insieme di tutte le istanze del T-schema (teoria che possiamo indicare come *TB*).

Le istanze del T-schema possono al massimo costituire una definizione parziale della verità; al limite, la loro congiunzione infinita, potrebbe costituire la somma delle teorie parziali e quindi una teoria totale. Tarski è scettico circa l'impiego di un dispositivo logico del genere ma ancora più importante a questo proposito è proprio l'accenno alla teoria *TB*. Quand'anche si possedesse una lista infinita dei T-enunciati relativizzati a

33 Si pensi, per esempio, a Popper (cfr. Popper 1963, 223). Altri studiosi favorevoli a un'interpretazione corrispondentista della teoria tarskiana sono, tra gli altri: Niiniluoto 1994; 2003; Sher 1998, Moreno 2001.

un certo linguaggio L, tale teoria risulterebbe estremamente debole e non in grado di dimostrare le chiusure universali di principi logici di base quali il principio di non contraddizione e del terzo escluso.[34] Naturalmente è interessante andare a vedere in che misura il T-schema racchiuda parte (o tutte) delle intuizioni corrispondentiste ma non si può dimenticare che la definizione di verità tarskiana è completamente differente. Infatti,

(2) Un enunciato *p* è vero sse *p* è soddisfatto da tutte le sequenze

Quale definizione è più adatta a un raffronto filosofico con la teoria della verità corrispondentista? E poi, possiamo tralasciare una definizione a vantaggio di un'altra?

3.1 *T-schemi e corrispondentismo*

Inizieremo con l'osservare che il T-schema può essere interpretato secondo due ottiche differenti: in senso *linguistico* e in senso *realistico*.

Secondo la prima accezione, si parte dal constatare che

(T) *x* è vero se e solo se *p*

è un enunciato del metalinguaggio. Come è noto, *x* sta per il nome di un enunciato e *p* sta per la traduzione metateorica dell'enunciato in questione. Se assumiamo che il nostro linguaggio oggetto sia l'italiano e il metalinguaggio l'inglese avremo un T-schema di questo tipo:

(3) 'La neve è bianca' is true if and only if snow is white

In (3) vengono fornite le condizioni di verità dell'enunciato nominato attraverso il dispositivo della virgolettatura; il lato destro del bicondizionale non è il polo della realtà, o del mondo. Questa lettura del T-schema si accorda bene con una teoria generale della verità come dispositivo linguistico di devirgolettatura secondo una particolare concezione generale di tipo deflazionista.

---

34 Questi temi verranno affrontati in profondità nella seconda parte del volume.

Anche Patterson sembra convinto di una posizione simile quando afferma esplicitamente che il bicondizionale non può essere preso come esempio di teoria corrispondentista dal momento che non afferma affatto una relazione:

> The simplest criticism is that what this T-biconditional, like any T-biconditional, does it to state a condition exactly under which a sentence bears the property of truth. [...] [I]t doesn't predicate a relation of the sentence it mentions and some other objects[35].

La stessa forma grammaticale del bicondizionale porterebbe a escludere che esso possa indicare una relazione e quindi *a fortiori* nemmeno una relazione tra linguaggio e mondo.

La critica a questa posizione è altrettanto semplice e diretta: il bicondizionale esprime una relazione. In particolare si tratta di una relazione diadica complessa tra un enunciato *s* e un fatto *p*. La 'forma logica' di un bicondizionale è in realtà quella di una relazione R che vale tra *s* e *p*: R(*s*,*p*). Naturalmente ci potranno essere, conclude la replica, dei problemi riguardanti le entità implicate, in particolare i fatti, o gli stati di cose, ma ciò non toglie che una relazione del genere sia almeno possibile e che non si possa demandare ad altra sede la discussione filosofica sui *truth-makers*.

Del resto, c'è un'altra possibile lettura realista del T-schema; esso esprime sì una condizione circa l'applicazione del predicato 'esser vero' ma con un riferimento a una predicazione oggettuale (o realistica). Infatti:

(1) 'La neve è bianca' è vero se e solo se la neve è bianca

Può essere formalizzato mettendo in luce proprio i due differenti tipi di predicazione:

(4) VERO('La neve è bianca') se e solo se BIANCO(neve)

In (4) sono presenti due tipi di predicati differenti: un predicato linguistico VERO che si applica a un individuo 'sintattico' cioè all'enunciato 'La neve è bianca' e un predicato molto più comune, cioè l'esser bianco, che si predica della neve. Quindi, seguendo Sher:

---

35 Patterson 2003, 427.

> The left side of a T-biconditional consists of a linguistic predication, its right side of an objectual or "worldly" predication[36].

Il fatto che siano presenti due tipi di predicazione, l'uno linguistico e l'altro oggettuale (o realistico) chiarisce la funzione del T-schema che è proprio quella di connettere due dimensioni della relazione di predicazione. Ma qual è l'interpretazione più plausibile del T-schema?

La lettura disquotazionale risponde a una precisa istanza deflazionistica; dal momento che la nozione di verità non possiede una 'natura' o, per usare le parole di Stewart Shapiro, è 'metafisicamente sottile', rimangono al predicato vero solo funzioni connesse con il suo impiego linguistico che precludono, pertanto, da ogni riferimento semantico in senso proprio. È proprio del deflazionismo, l'intendere 'vero' come un dispositivo linguistico utile per riferimenti indiretti e per compiere generalizzazioni.[37] Tra le molte critiche che si possono muovere a una concezione deflazionistica ci interessa qui sottolineare i rilievi mossi contro un'interpretazione linguistica (o deflazionistica) del T-schema. Una ragione significativa per cui un simile atteggiamento teorico non è plausibile è che non tutte le istanze del T-schema vedono il predicato T come dispositivo disquotazionale. Si ricordi che in

(T) x è vero se e solo se p

p equivale a una traduzione metalinguistica dell'enunciato nominato da x. Ora, mentre il classico esempio

(1) 'La neve è bianca' è vero se e solo se la neve è bianca

può indurre a formulare giudizi circa la banalità e l'uso disquotazionale del predicato 'vero', lo stesso non si può dire per

(5) 'La neve è bianca' è vero se e solo se la neve riflette la luce di tutte le frequenze in maniera diffusa

---

36 Sher 1999, 135.

37 Non approfondiamo ulteriormente la tematica del deflazionismo in connessione alla teoria della verità tarskiana perché la tratteremo in seguito.

In (5) è evidente che il predicato 'essere vero' non compie la (piuttosto banale) operazione di de-virgolettatura; anzi, conoscere (5) sembra aumentare in maniera significativa il nostro bagaglio di informazioni. Naturalmente il deflazionista può replicare e sostenere che (5) rimane comunque compatibile con un'interpretazione decitazionale della verità. In ogni caso, non ci impegniamo a sostenere che *solo* una lettura del T-schema sia quella legittima ma ci interessa indagare quale delle due prima esposte sia la più plausibile in ottica corrispondentista. E la scelta non può che ricadere sulla seconda opzione, ovvero sull'interpretazione *oggettuale*.

3.2 *Metalinguaggio e sinonimia*

Tuttavia, anche sposando un'interpretazione oggettuale del T-schema sorgono alcune difficoltà; in particolare, ne prenderemo in esame due: la questione del *metalinguaggio* e il problema della *sinonimia*. Abbiamo detto più volte che il T-schema è formulato in un linguaggio particolare (il metalinguaggio) che risulta essenzialmente più ricco del linguaggio oggetto: contiene infatti i termini semantici, quale appunto 'vero'. Proprio la gerarchizzazione dei livelli linguistici impedisce quei fenomeni di autopredicazione e di chiusura semantica responsabili dell'antinomia del mentitore. Ma anche il metalinguaggio è un linguaggio e quindi ci si può sensatamente domandare se sia interpretato o no. La prospettiva di Tarski è chiarissima: il metalinguaggio, esattamente come il linguaggio oggetto, è interpretato; se non ci fosse un'interpretazione, cioè se fosse presente la sola dimensione sintattica, lo stesso problema della verità sarebbe insensato. Vediamo ora le conseguenze di questo fatto per la nostra discussione sul T-schema e prendiamo in considerazione

(4) VERO('La neve è bianca') se e solo se BIANCO(neve)

Ora, in (4), è chiaro che il lato destro del bicondizionale rappresenta il polo oggettivo o realistico. Ma in che senso? Una risposta possibile è che il nesso di predicazione BIANCO(neve) rappresenta il fatto (o lo stato di cose) in base al quale l'individuo 'neve' gode della proprietà 'essere bianco'. Dire questo, però, equivale ad assumere già una forma di teoria della verità come corrispondenza, solo spostata a livello del metalinguaggio. E la semplice presentazione del T-schema non aiuta a chiarire la natura di questa relazione fra linguaggio e mondo, quand'anche questo nesso sia declinato tra un particolare linguaggio, e cioè il metalinguaggio, e il mondo. Viene, insomma, riproposto il problema: un enunciato è vero quando corrisponde a qualcosa nella realtà ma, in base al T-schema, diremo che un enunciato è vero se e solo se vale la traduzione metalinguistica dell'enunciato stesso; tuttavia, quest'ultima condizione presuppone *già* la teoria della corrispondenza.

Questo rilievo ricorda per certi versi una classica critica alla teoria della corrispondenza: non è possibile, secondo alcuni avversari del corrispondentismo, osservare il nesso tra linguaggio e mondo (o tra pensiero e realtà) dal momento che non si può uscire dal linguaggio (o dal pensiero) e assumere, così, un terzo punto di vista neutrale[38]. In questo senso, il problema dell'interpretazione del metalinguaggio ricalca abbastanza bene questo tipo di obiezione.

Detto altrimenti: l'interpretazione oggettuale (o realistica) di

(4) VERO('La neve è bianca') se e solo se BIANCO(neve)

risulta essere esplicativa del lato sinistro del bicondizionale a patto che BIANCO(neve) sia naturalmente interpretato sul mondo, secondo un'ottica corrispondentista.

Wolenski e Simons riconducono questa riflessione a una critica di Brentano nei confronti della teoria della corrispondenza. Il problema è, come dicevamo in precedenza, sempre quello del 'terzo occhio', ovvero di un punto di vista che si collochi al di là del linguaggio e del mondo in modo da poter osservare la relazione di corrispondenza. I due autori rispondono all'accusa difendendo l'interpretazione classica di Tarski:

> There is no problem about comparing sentences with reality, provided we step up to a metalanguage. The correspondence of sentences of L_n with reality is elucidated in L_{n+1}. [...] This disposes of Brentano's [...] objection, which probably represents the point of view of natural, i.e. universal and hence semantically closed languages. But the semantical antinomies show the need for caution and revision[39].

---

38 Non ci interessa indagare nel dettaglio della critica alla teoria della verità come corrispondenza. Si noti, in ogni caso, che la validità delle obiezioni mosse a questa concezione della verità dipendono, tra l'altro, *anche* dalla precisione e accuratezza con cui è stata specificata la teoria corrispondentista.

39 Wolenski Simons 1989, 424.

A questo punto, però, sorge un'altra questione intimamente correlata con la precedente. Nel T-schema, *p* sta per un enunciato del metalinguaggio che è la traduzione metalinguistica dell'enunciato nominato da *x*. Il concetto di traduzione sembra richiamare però il concetto semantico per eccellenza, quello di significato. Eppure Tarski è chiarissimo circa l'assunzione di altre nozioni semantiche nella definizione della verità:

> In this construction I shall not make use of any semantical concept if I am not able to previously reduce it to other concepts[40].

Ma richiedere che *p* sia la traduzione metalinguistica dell'enunciato nominato da *x* sembra violare proprio il vincolo che Tarski si è autoimposto. Ci sono due possibili risposte: la prima è molto semplice e consiste nel reinterpretare 'caritatevolmente' la frase di Tarski. In base a questo approccio, le nozioni semantiche cui Tarski fa riferimento sono altre e, in particolare, le relazioni di 'designare', 'denotare' e così via. Insomma, la nozione di sinonimia non rientrerebbe nei casi problematici e può essere, pertanto, utilizzata liberamente. Che questa ipotesi non convinca è abbastanza chiaro. Innanzitutto, si tratta di un'opzione esegetica e in quanto tale dovrebbe confrontarsi con altri passi di Tarski in cui vi è esplicito riferimento alla dimensione semantica; in secondo luogo, è davvero poco plausibile che una nozione cardine come quella di 'significato' non sia considerata tra le nozioni semantiche rilevanti. Molto più interessante, è invece, la risposta offerta da Panu Raatikainen che sottolinea, giustamente, come all'attenzione non sia tanto la nozione di significato quanto quella di 'identità di significato' o 'sinonimia' che è alla base della funzione di traduzione.

> I submit that it is possible to view translation in this context, as a purely synctactic mapping between two languages, without assuming any relations between either language and the external objects. Translation, so viewed, is not a semantical concept in Tarski's sense[41].

Quale può essere, dunque, la correlazione sintattica che, almeno nella lettura di Raatikainen, costituisce la nozione di traduzione impiegata da Tarski? Si tratta di una relazione a due posti che mette in connessione nomi di enunciati del linguaggio oggetto con elementi del metalinguag-

---

40 Tarski 1983, 153.
41 Raatikainen 2007, 113.

gio definita mediante una lista. Un esempio può chiarire; assumiamo l'italiano come linguaggio oggetto, le virgolette semplici come dispositivo per nominare enunciati e l'inglese come metalinguaggio. La relazione di sinonimia si presenterebbe quindi così:

Trans(x,y) $\leftrightarrow$ (x = 'La neve è bianca' $\wedge$ y = Snow is white) $\vee$ (x = 'L'erba è verde' $\wedge$ y = Grass is green) $\vee$ (x = 'Emma è bionda' $\wedge$ y = Emma is blonde) $\vee$ …

In questo senso, non c'è nessun riferimento al concetto di significato ma solo alla disponibilità di liste come quella appena esibita. Naturalmente la proposta di Raatikainen[42] solleva, o meglio rimanda, ad altre questioni di carattere più fondamentali. In particolare due sono i punti cruciali: *in primis*, ci si può chiedere se definizioni esplicite caratterizzate come liste siano filosoficamente informative (e si tratta di un aspetto della pervasiva critica di Hartry Field); in secondo luogo, anche accettando questa proposta, rimane un problema di sfondo: quello che vale per i T-schemi e per la funzione di traduzione, ovvero la riducibilità di una nozione intuitivamente semantica (come quella di traduzione o sinonimia) a una struttura puramente sintattica (quale la disgiunzione appena illustrata) non potrebbe valere *in generale* nella trattazione tarskiana della verità? In altre parole, la concezione *semantica* non è forse una concezione *sintattica* della verità? Discuteremo in seguito sia la critica di Field sia la relazione tra sintassi e semantica; ora ci concentreremo sulla definizione vera e propria di verità e sulla sua possibile interpretazione in ottica corrispondentista.

3.3 *Soddisfazione e corrispondenza*

Ben diverso è il discorso sull'interpretazione corrispondentista della concezione di Tarski quando si prende in considerazione l'autentica definizione di verità:

(6) *x* è vero se e solo se (om σ)(σ soddisfa *x*)

---

42 Che peraltro è differente almeno in un punto importante: la nostra funzione di traduzione Trans(x,y) concerne enunciati mentre l'idea di Raatikainen è quella di fornire relazioni di denotazione per nomi e per predicati.

Cioè, un enunciato è vero se e solo se tutte le sequenze soddisfano *x*. Tarski precisa che la definizione della verità è un caso limite della relazione di soddisfazione tra sequenze di oggetti e funzioni proposizionali.[43] Una lettura corrispondentista di (6) può stabilire la seguente proporzione:

Soddisfazione : corrispondenza = sequenze : truth-makers

La relazione di soddisfazione sta alla relazione di corrispondenza tra linguaggio e realtà come le sequenze stanno ai truth-makers (fatti, stati di cose o altro). Se posta in questi termini, l'interpretazione si espone a una serie di obiezioni piuttosto serie. La più significativa riguarda proprio una delle intuizioni di fondo della teoria corrispondentista e cioè che enunciati differenti sono resi veri da differenti truth-makers. È chiaro che sarebbe necessario un criterio di identità per enunciati (o proposizioni) che non faccia riferimento agli stati di cose o ai fatti descritti; non si può pertanto dire che le proposizioni A e B sono differenti, e quindi possiedono differenti truth-makers, se la giustificazione di A ≠ B dipende proprio dal fatto che il truth-maker di A è diverso al truth-maker di B. Ma resta il fatto che, intuitivamente, proposizioni come

(7) La neve è bianca

e

(8) I neutrini non hanno una massa costante

debbano in qualche modo essere rese vere da pezzi di realtà completamente differente. Eppure, nella teoria tarskiana se (7) è vera allora è soddisfatta da tutte le sequenze di oggetti; esattamente come (8). Ma in questo caso il truth-maker di (7) dovrebbe essere identico al truth-maker di (8) il che risulta assolutamente controintuitivo. Tra coloro che in qualche modo concordano circa questa obiezione ricordiamo Haack (1978) e Grayling (1998).

Una possibile risposta consiste nel considerare la generalizzazione universale su tutte le sequenze un *escamotage* tecnico di Tarski, conseguenza della sua impostazione formale, più che un'effettiva tesi filosofica. Per chiarire questo, prendiamo in considerazione una proposizione

---

43 Cfr. la presentazione svolta nel primo capitolo.

quantificata esistenzialmente come $\exists x_1 Px_1$, dove Px_1 voglia dire "x_1 è il maestro di Platone". Secondo la clausola di soddisfazione di Tarski, una sequenza σ soddisfa $\exists x_1 Px_1$ se e solo se esiste almeno una sequenza σ_1 che concorda con σ su tutti gli oggetti tranne che per l'oggetto designato dalla variabile x_1, che soddisfa Px_1. Quali sono allora le uniche sequenze di oggetti rilevanti per la soddisfazione di Px_1? Saranno quelle che hanno al primo posto l'oggetto 'Socrate'; tutte le altre non hanno alcuna influenza sulla verità o meno del nostro enunciato. Quindi è plausibile considerare l'oggetto 'Socrate' come il truth-maker dell'enunciato $\exists x_1 Px_1$ anche se quest'ultimo è soddisfatto da ogni sequenza di oggetti. Con le parole di Raatikainen:

> Given an arbitrary sequence σ, we are, so to say, allowed to switch its relevant member […] to a relevant object […] and produce a sequence σ_1 which does the real work[44].

La proposta appena presa in esame mette in luce un tratto fondamentale della questione. Da un lato è certamente corretto dire che la definizione di verità tarskiana è un 'caso limite' delle clausole sulla soddisfazione e come tale presenta aspetti controintuitivi, d'altro canto sembra che riflessioni più feconde derivino proprio dal considerare la nozione di soddisfazione come il candidato più probabile per interpretare la teoria di Tarski come una posizione corrispondentista.

In generale, ogni teoria della corrispondenza deve fronteggiare una sorta di dilemma: da un lato, spiegare la relazione di corrispondenza assumendo una forma di isomorfismo tra stati di cose (o fatti) e enunciati (o proposizioni) sembra del tutto ingiustificato dal punto di vista ontologico[45]; se però si rimane neutrali circa la relazione di isomorfismo (o somiglianza) allora viene meno proprio la capacità di spiegare il nesso tra linguaggio e mondo.[46] Vediamo quindi nel dettaglio se e in che modo la relazione di soddisfazione e il suo caso limite, ovvero la definizione di verità, può essere interpretata come una forma di corrispondenza.

Ripetiamo per comodità le clausole tarskiane sulla soddisfazione, aiutandoci con una simbologia più maneggevole dell'originale:

---

44 Raatikainen 2007, 115.

45 E, forse, in contrasto con una serie di evidenze circa il funzionamento delle lingue naturali.

46 Cfr, Kolàr 1999, 68.

α è soddisfatta dalla sequenza σ se e solo se:
1. $\alpha \equiv P(x_1,...,x_n)$ e il primo elemento di σ, il secondo elemento di σ, …, l'ennesimo elemento di σ stanno nella relazione *P*
2. $\alpha \equiv \neg\beta$ e σ non soddisfa β
3. $\alpha \equiv \beta \vee \gamma$ e σ soddisfa β oppure σ soddisfa γ
4. $\alpha \equiv \forall x_n \beta$ e ogni sequenza τ che differisce da σ al massimo all'en nesimo posto soddisfa β.

Infine,

5. α è un enunciato vero sse α è soddisfatto da tutte le sequenze σ

L'idea che alcuni studiosi[47] condividono circa la relazione tra corrispondenza e soddisfazione è che sia errato interpretare la proposta di Tarski come una teoria della corrispondenza *tout court* ma che la definizione semantica costituisca una sorta di *terzo polo* tra linguaggio e mondo. Cerchiamo di chiarire. Come abbiamo visto in precedenza, le posizioni corrispondentiste identificano due poli dalla cui relazione (più o meno specificata) è possibile definire la nozione di verità. In questo caso, la teoria di Tarski è il *medium* tra linguaggio e mondo. Scendendo un poco nel dettaglio, da un lato abbiamo il portatore di verità che assumiamo essere un enunciato di un linguaggio già interpretato. Dall'altro, abbiamo la realtà, dove un 'pezzo' di mondo rende vero (o meno) il portatore di verità. La dimensione semantica è ciò che rela questi due poli. Ma in che modo, il terzo regno della semantica risulta interessante per i nostri scopi? Detto in altri termini, non si tratta di una futile ripetizione? A nostro avviso, questa interpretazione riesce a risolvere almeno parzialmente proprio il dilemma prima presentato. Il seguente schema può aiutare a visualizzare la situazione:

Polo linguistico	Polo semantico	Polo oggettivo
Truth-bearers	Sequenze + soddisfazione; strutture insiemistiche	Truth-makers

47 Per esempio Kolàr 1999, Niiniluoto 2004 e per certi versi anche Sher 1999.

Illustriamo le componenti dello schema: a sinistra, come già accennato abbiamo il linguaggio, caratterizzato dalla presenza dei truth-bearers ovvero da enti che hanno la proprietà di essere veri; a destra il polo oggettivo, o mondo, in cui vi sono truth-makers. Si noti che, a questo livello, non ci interessa chiarire la natura, l'individuazione e l'esistenza stessa dei truth-makers. Al centro, c'è la semantica che può apparire sia nella versione 'arcaica' delle sequenze e della relazione di soddisfazione, sia nella versione più moderna, tipica della teoria dei modelli, che utilizza strutture insiemistiche.

Uno dei problemi fondamentali della teoria della corrispondenza è proprio quello di 'unire' entità così differenti, quali enunciati e fatti o stati di cose in maniera coerente con alcune intuizioni di fondo. Infatti:

(9) Londra è a ovest di Berlino

e

(10) Berlino è a est di Londra

sono due enunciati almeno sintatticamente differenti ma cui corrisponde un unico stato di cose e cioè il fatto che Berlino e Londra sono in una determinata relazione spaziale. Ebbene, la semantica tarskiana, tramite la relazione di soddisfazione, costituisce un ponte tra le due regioni. Infatti sia (9) che (10), in quanto enunciati veri, sono soddisfatti da tutte le sequenze. Qui però, interviene una nota tecnica e le considerazioni di Raatikainen si fanno particolarmente pertinenti. In ottica tarskiana, un enunciato vero è sì soddisfatto da tutte le sequenze ma questo accade perché, non possedendo variabili, l'enunciato non discrimina tra sequenze che lo soddisferebbero e sequenze che non lo soddisferebbero. Infatti, come mette in luce bene Kirkham:

> Since a definition of truth as satisfaction by some sequence works perfectly well, why did Tarski define truth as satisfaction by all sequences? The answer is that, as a pratical matter, the two definitions amount to exactly the same thing. […][T]he condition that must be met for S to satisfy $\forall x_4 Rx_4$ is exactly the same condition that any sequence must meet to satisfy this sentence, namely, everything must be R. […] So if one sequence satisfies the sentence, they all do. […] The same point […] applies to all other genuine sentences: to be satisfied by one sequence is to be satisfied by them all. Thus it matters not whether we

> define truth as satisfaction by some sequence or as satisfaction by all sequences[48].

Ora, se da un punto di vista tecnico, dire che un enunciato è soddisfatto da una sequenza è equivalente ad affermare che quell'enunciato è soddisfatto da tutte le sequenze, da un punto di vista filosofico le cose cambiano e di molto. Infatti, l'identificazione (almeno parziale) tra truth-makers e sequenze si scontra con quanto appena detto. Non è affatto irrilevante affermare che un enunciato è reso vero da tutti i truth-maker se esiste un truth-maker che lo rende vero. Possiamo concludere, alla luce di queste considerazioni, che la teoria semantica di Tarski è sì una teoria corrispondentista ma che differisce per alcuni tratti essenziali da ciò che normalmente si ha in mente quando si parla di corrispondentismo e di truth-maker. Infatti, tornando al caso di Berlino e Londra, diremo che ci sono sì due sequenze (o infinite) che soddisfano gli enunciati in questione ma, intuitivamente, si tratta dello stesso fatto che li rende veri. Quello che resta da fare è allora cercare di capire i motivi per i quali la teoria semantica è una teoria corrispondentista e quale forma di corrispondenza sia qui in gioco. Ed è ciò che ci proponiamo di fare nel prossimo paragrafo.

4. *Logica, semantica e corrispondenza*

Dopo aver esplorato alcune possibili linee di analisi, eccoci giunti al nocciolo della questione; prima di chiarire quanto espresso nell'ultimo paragrafo circa l'interpretazione filosofica della teoria di Tarski è bene riassumere brevemente il percorso compiuto.

Nella trattazione tarskiana compaiono, come abbiamo visto, due definizioni di verità: la prima, parziale, espressa nel T-schema, la seconda, più specifica, che si basa sui concetti di soddisfazione e sequenza. Il T-schema sembra essere aperto a due letture: l'una *oggettuale* ove si collegano due tipi di predicazione, quella linguistica e quella ontologica e che, quindi, si situa in ottica corrispondentista; l'altra di carattere *deflazionistico*, che legge il predicato di verità come dispositivo di de-virgolettatu-

48 Kirkham 1992, 156-157.

ra. Sebbene entrambe le interpretazioni non siano immuni da obiezioni, non è scontata la loro reciproca esclusione[49].

Del resto, ciò sembra accordarsi anche con i numerosi rilievi di Tarski a proposito dell'incompletezza del T-schema come definizione di verità; proprio perché si tratta di una definizione al massimo parziale, essa è soggetta a varie interpretazioni, egualmente legittime.

Ma che dire della definizione vera e propria, quella ottenuta cioè tramite il concetto di soddisfazione da parte di sequenze di oggetti? Sembra fuori di dubbio che la relazione di soddisfazione sia una relazione linguaggio-mondo, cioè una nozione semantica nel senso proprio del termine. E infatti, questa è l'unica concessione che Tarski fa nella costruzione della sua definizione: la relazione di soddisfazione è il *primum* da cui parte per edificare la sua teoria. Ora, il problema è il seguente: se la relazione di soddisfazione è semantica, in che senso possiamo intendere la relazione fra sequenze e truth-makers? Per esempio, le sequenze di oggetti non possiedono alcuna specificazione ontologica, per cui non sappiamo a quali categorie appartengono i membri di tali sequenze. In questo senso, può essere utile rimandare alla nostra analisi del capitolo precedente sui linguaggi di ordine finito.

Come si ricorderà, Tarski elabora due strategie per adattare il concetto di soddisfazione a linguaggi in cui siano presenti variabili di più ordini: il metodo delle sequenze multi-riga e quello dell'unificazione semantica. Al di là dei dettagli tecnici, si evince che l'atteggiamento di Tarski è di un'effettiva *indifferenza ontologica*. Non è rilevante la 'natura' degli oggetti che devono soddisfare le funzioni enunciative, tanto che nella seconda strategia (l'unica disponibile quando si passa a linguaggi di complessità ancora maggiore) gli individui possono essere 'ridotti' a coppie ordinate secondo il metodo che abbiamo illustrato in precedenza. Ma ciò significa che le sequenze di oggetti, che Tarski chiamerà a breve modelli,

---

49 Per esempio, Gila Sher (1999) è abbastanza chiara a tal proposito:

> I would like point out […] that the disquotational reading of the T-biconditional does not conflict with their correspondence reading. We switch from one to the other by a change in gestalt, so to speak. […] Using the medieval distinction between *suppositio materialis* and *suppositio formalis*, we can say that the first reading [la disquotazionale, ndA] assumes the *suppositio formalis* mode, the second – the *suppositio materialis* mode.

non costituiscono il nucleo dell'ontologia della semantica. Questo ci porta a concludere che al di là delle varianti tecniche nelle procedure messe a punto da Tarski, l'intuizione ontologica di fondo, realista, non muta.

Jan Wolenski, nel difendere la concezione semantica della verità come una teoria filosoficamente rilevante, è molto chiaro su questo punto:

> There is a temptation to regard [the sequences of objects] as something like facts. This is, I believe, a completely wrong interpretation. Sequences are not facts. If A is a sentence, no sequence is distinguished from its point of view. We achieve nothing if we say that A corresponds with all sequences. There is no relation of correspondence between sentences and portions of reality[50].

Abbiamo, con questo, reciso ogni ponte tra la teoria di Tarski e le teorie filosofiche della verità, in particolare con le teorie corrispondentiste? Alcuni autori sono pronti senz'altro a compiere quest'ultimo passo. In effetti, lo sviluppo della teoria dei modelli può essere letto come una naturale evoluzione di questo atteggiamento. Le strutture model-teoretiche non sono altro che strutture insiemistiche o algebriche, appartenenti cioè ad altre teorie, senza nessuna immediata ricaduta di carattere ontologico.

Eppure vi sono almeno due ragioni per considerare questa posizione non del tutto convincente: innanzitutto, per le numerose e non fraintendibili dichiarazioni dello stesso Tarski che ha sempre considerato la sua teoria della verità ispirata in un certo senso alla concezione classica; in secondo luogo, perché sebbene non immediatamente riconducibile a una teoria della corrispondenza, la teoria tarskiana non solo è filosoficamente feconda ma risulta, in un certo senso, *normativa* nei confronti di altre teorie della verità. Vediamo allora, di chiarire meglio quanto detto finora proponendo una linea interpretativa che tenga conto delle difficoltà emerse e riesca a ricomprenderle in una sintesi più ampia.

Lo schema riportato in precedenza mostra chiaramente una tripartizione dei livelli; qualcosa di simile è presente in Kuipers[51] ed è ripreso da Niiniluoto[52]. Vengono identificati una regione *C*, costituita da oggetti e sistemi del mondo, una regione *I* di strutture insiemistiche e una regione *II* di enunciati del linguaggio. Il ponte tra il linguaggio e il mondo è allo-

50 Wolenski 1999, 62.
51 Kuipers 2000.
52 Niiniluoto 2004, 63.

ra fornito proprio da quella «terra di mezzo» di entità insiemistiche. Ora, non ci sembra del tutto azzardato tentare un paragone con la nostra impostazione a patto di evidenziare alcuni punti importanti. Che la semantica tarskiana sia una sorta di *medium* tra linguaggio e mondo è la tesi centrale; con ciò naturalmente si lasciano sullo sfondo una serie di problemi affatto banali. Solo qualche esempio: da un punto di vista ontologico rimane da chiarire la natura delle costruzioni semantiche siano esse sequenze infinite, siano strutture insiemistiche; la divisione di Niiniluoto e Kuipers è convincente in casi standard ma richiede qualche approfondimento in situazioni più complesse eppure altrettanto comuni. Si prenda, a titolo esemplificativo, un asserto β di teoria degli insiemi, che essendo una parte di linguaggio dovrebbe abitare la regione *II*. Che β voglia descrivere una qualche relazione astratta lo si può evincere dal significato inteso di β stesso. La semantica (tarskiana) fornirà quindi il modello rigoroso delle condizioni di verità di β. E qui iniziano i problemi. Infatti, sia al livello *I* (quello semantico) sia al livello *C* (quello ontologico) abbiamo strutture insiemistiche. Sembrerebbe che il materiale di collegamento tra le due regioni sia parte costitutiva di una delle due. Non ci interessa approfondire in questa sede l'ontologia dei modelli e la relativa indipendenza delle strutture semantiche; è importante però ritenere il cuore dell'intuizione che abbiamo discusso finora e cioè che la semantica tarskiana si basa su una sorta di *corrispondenza indiretta* tra linguaggio e mondo.

La dichiarazione di Tarski che all'inizio della sua memoria indicava la sua concezione della verità come semantica e, contemporaneamente, appartenente all'alveo delle classiche teorie corrispondentiste trova allora plausibilità. Infatti, la caratterizzazione semantica indica proprio che la verità è qualcosa che ha a che fare con la connessione linguaggio / mondo e proprio per questo motivo si qualifica come realista, anche da un punto di vista minimale. Non è, pertanto, un caso che le concezioni della verità rivali citate da Tarski siano quasi sempre quelle coerentiste (dove il polo ontologico è, in un certo senso, eliminato) e quelle utilitariste (dove addirittura viene considerata preponderante la funzione pragmatica del linguaggio a discapito di quella semantico referenziale).

È necessaria una precisazione. In un passo famoso di Tarski (1944), l'autore rivendica una sorta di neutralità filosofica della sua concezione semantica. Dice Tarski:

> We may accept the semantic conception of truth without giving up any epistemological attitude we may have had: we may remain naïve realists, critical realists or idealists, empiricist or metaphysicians – whate-

> ver we were before. The semantic conception is completely neutral toward all these issues. (Tarski 1944, 362)

Ora, nell'interpretazione di questo passo è bene adoperare alcune precauzioni: innanzitutto, come abbiamo già avuto modo di constatare, le denominazioni di posizioni filosofiche non sono affatto standard e anzi spesso risultano fonte di ambiguità. Proprio nel passaggio in esame desta qualche perplessità il riferimento a posizioni realiste (sia ingenue che critiche), idealiste e a non meglio identificati orientamenti metafisici. Diremmo, oggi, che idealismo, realismo, empirismo sono differenti punti di vista metafisici, magari gerarchizzati al loro interno (per cui ha perfettamente senso dire che un empirista è un realista, a seconda, naturalmente, del tipo di entità cui ci si sta riferendo). Per tornare a noi, dunque, sembra più convincente una lettura di questo passo che ascriva alla teoria di Tarski un impegno filosofico non diretto, ovvero non identificabile immediatamente con una corrente filosofica classica. Ciò però non vuol dire che, al di là delle intenzioni di Tarski, non sia invece legittimo argomentare a favore di una interpretazione corrispondentista, benché sui generis, della concezione semantica della verità.

Una delle grandi intuizioni di Tarski sta nell'aver affrontato il problema della verità in maniera squisitamente *logica*, intendendo con ciò il porre l'attenzione al polo linguistico della relazione di corrispondenza. Tarski parte dalla struttura linguistica e in ultima analisi, logica, dell'enunciato: questa è infatti la forma stessa della definizione ricorsiva di verità. La possibilità di *costruire* le condizioni di verità di enunciati di complessità qualsiasi a partire dalle clausole di base è alla radice della *composizionalità*. Cerchiamo di specificare meglio quanto detto.

La relazione di corrispondenza - ovviamente, secondo le teorie della verità corrispondentiste - connette un truth-bearer vero, e cioè un portatore di verità che ha la proprietà di essere vero, e un truth-maker, e cioè ciò in virtù del quale il truth-bearer possiede quella determinata proprietà. Ora, possiamo rintracciare in tale relazione due aspetti fondamentali: un aspetto logico e un aspetto ontologico.

Quest'ultimo - che possiamo chiamare anche *materiale* - riguarda l'essenza della relazione di corrispondenza e dei suoi relati, stati di cose e proposizioni, per esempio. A sua volta, e semplificando un po', la relazione di corrispondenza può essere assunta come ontologicamente primitiva o derivata da altre relazioni più basilari. Nel primo caso, sarà compito dello studioso indagare le conseguenze di questa assunzione, in particolare per la struttura stessa dei truth-makers. Nel secondo caso, invece, la sfida è fornire una spiegazione della corrispondenza sulla base di altre relazioni più immediate; un tentativo, in tal senso, può essere la distinzione - già citata ampiamente in precedenza - di *corrispondenza come congruenza* e *corrispondenza come correlazione*. Dal punto di vista ontologico, l'analisi tarskiana non è rilevante, o perlomeno non lo è in maniera così significativa. E molte interpretazioni della teoria di Tarski alla luce dell'aspetto materiale, ontologico, della nozione di corrispondenza risultano a nostro avviso deludenti proprio per questa ragione: perché la concezione semantica della verità non è, almeno in prima battuta, una caratterizzazione ontologica della relazione di corrispondenza[53]. Per capire la natura della semantica dobbiamo indagare, al contrario, l'aspetto logico della relazione di corrispondenza; ed è in questo senso che il contributo di Tarski è stato davvero epocale.

Iniziamo da una considerazione del tutto generale: caratterizzare l'aspetto logico della relazione di corrispondenza significa in un certo senso stabilire le possibili forme che questa particolare relazione potrà assumere poi attualmente. Riprenderemo questo concetto di 'delimitazione' delle possibilità per chiarire ancora la nostra interpretazione della teoria tarskiana. Quali sono, quindi, i punti fondamentali di questo approccio? Sono quattro: la *distinzione linguaggio / metalinguaggio*; la *struttura ricorsiva*; la *limitazione* e, infine, la *rilevanza ontologica schematica*.

La distinzione linguaggio / metalinguaggio suggerisce che per indicare le condizioni minimali di *signficatività* del predicato di verità è necessario eliminare i linguaggi semanticamente chiusi.[54] Detto in altri termini, la caratterizzazione della relazione di soddisfazione non può avvenire al

53 Si potrebbe estendere la discussione anche all'aspetto *epistemico* ove il portatore di verità non sarebbe solamente più una proposizione (o un enunciato) ma un giudizio ovvero una proposizione affermata o creduta vera da un certo soggetto conoscente. Un'indagine in tale direzione è interessante per mettere a confronto le teorie contemporanee della corrispondenza (si pensi, per esempio, alla versione di Russell) con la teoria classica adeguazionista presente, pur in molteplici varianti, nella tradizione aristotelico-tomista.

54 Questa è ovviamente l'opinione di Tarski. Delle tre diagnosi del paradosso del mentitore, egli indica nella chiusura semantica il motivo della contraddizione. Si può naturalmente mantenere, in un certo senso, la chiusura del linguaggio e ammettere fenomeni di riflessività del predicato di verità, senza per questo cadere nei paradossi. L'esempio classico sono le teorie della verità *à la* Kripke.

medesimo livello linguistico dei nomi dei *relata*; è necessario, quindi, assumere una prospettiva meta-teorica in cui predicare la verità a oggetti appartenenti a un livello linguistico inferiore.

La struttura ricorsiva della definizione di verità mette in luce il secondo aspetto logico; un punto fondamentale della strategia di Tarski è concentrarsi sulla *complessità logica* della proposizione (o enunciato) vera. Proprio perché abbiamo un'intuizione composizionale circa la struttura della verità, dobbiamo costruire una procedura di generazione delle proposizioni vere di un linguaggio. Chiaramente, questa procedura deve assumere come data una *base* (la clausola base,appunto) per poi procedere con il meccanismo di ricorsione. Ciò è estremamente significativo per il nostro discorso: significa che le condizioni di verità di una proposizione sono strettamente legate alla sua struttura logica e rende ragione dei ripetuti richiami di Tarski alla necessità, al fine di risolvere il problema della definizione della verità, di impiegare linguaggi completamente formalizzati, linguaggi, cioè, ove la struttura logica fosse esibita completamente in modo da poter applicare la procedura di ricorsione[55].

Il terzo punto riguarda proprio il teorema di Tarski, ovvero l'impossibilità di definire *tout court* la verità. Detto altrimenti, nessun linguaggio, minimamente significativo, è in grado di definire il proprio predicato di verità. La portata filosofica di questo teorema di limitazione va oltre i nostri scopi; qui ci basta sottolineare che se la nozione di corrispondenza vale tra mondo e linguaggio, il teorema di Tarski indica un limite preciso alla possibilità di caratterizza questa particolare relazione. La ragione di questa limitazione non è, però, contingente, quanto invece strutturale: è la forma stessa del predicato di verità (inteso tarskianamente) che comporta la sua indefinibilità generale. Possiamo allora dire che *la semantica tarskiana stabilisce lo schema per la relazione tra linguaggio e realtà relativamente a ogni teoria*.

Infine, come quarto punto, ci si può chiedere se anche dal punto di vista logico esiste una componente ontologica alla teoria semantica di Tarski. Questa è ridotta, a nostro avviso, a una sorta di ontologia formale intesa in questo senso: per Tarski, ciò che rende veri gli enunciati sono oggetti (o sequenze di oggetti)[56]. Ma si tratta di oggetti puramente schematici, arbitrari: non è in nessun modo rilevante la loro natura; ciò che

55 È noto che molti autori, capofila Davidson, hanno cercato di estendere l'approccio tarskiano anche alle lingue storico-naturali.

56 E, ancora una volta, questo lo qualifica come realista.

interessa è che *vi siano*. Ora, il senso di esistenza è qui inteso in maniera particolarmente debole: gli oggetti devono essere presenti nel dominio del modello della teoria (o del linguaggio) per il quale stiamo costruendo la definizione di verità. Sarà eventualmente un discorso squisitamente ontologico a dover prendere posizione circa l'esistenza attuale (o meno) degli elementi presenti nel modello della teoria. Se volessimo coniare uno slogan, potremmo dire che la rilevanza ontologica della teoria di Tarski sta nell'aver sostituito il mondo con i modelli. Dopo Tarski le cose non sono più vere o false *sic et simpliciter* ma vere o false in un modello. E questo ci sembra mostri molto bene l'aspetto logico della definizione tarskiana e la sua connessione con l'ontologia: passare dal mondo ai modelli non è in alcun modo una *riduzione*; anzi, si tratta di estendere la nozione di corrispondenza, o tarskianamente di soddisfazione, fino a comprendere gli *oggetti in quanto tali*, purché dati in una interpretazione (modello) del linguaggio in esame.

III
UNA CRITICA ALLA CONCEZIONE TARSKIANA DI VERITÀ

Nonostante l'incredibile 'successo' riscosso dalla concezione semantica della verità, sia dal punto di vista logico-formale che dalle ricadute di carattere filosofico, non sono mancati autori che hanno messo in discussione la validità di alcuni aspetti o, talvolta, dell'intera proposta tarskiana. Se è vero che i teoremi di logica non sono negoziabili è altrettanto vero che i teoremi si basano su assunzioni di carattere logico ed extra-logico e queste sono passibili di critica (si pensi, ad esempio, alla posizione costruttiva che ritiene per lo più illegittima la logica classica bivalente) ed è, soprattutto in ottica filosofica, l'interpretazione dei risultati formali che anima il dibattito.

In ciò che segue prenderemo in esame solo una critica mossa alla concezione della verità di Tarski e cioè la potente accusa mossa da Hartry Field. La critica di Field è estremamente interessante perché prende in considerazione la *rilevanza filosofica* della definizione tarskiana di verità. In estrema sintesi: secondo Field, la definizione di Tarski non è conciliabile con una posizione fisicalista ma poiché sia Field sia Tarski (almeno secondo Field) sono fisicalisti da ciò segue l'impraticabilità della concezione semantica della verità.

Discutere solamente la posizione di Field richiede però una qualche giustificazione. Infatti, come dicevamo, sono molti gli autori che hanno attaccato la cogenza della teoria semantica della verità: basti pensare a Black, Etchemendy, Haack, Putnam, solo per citarne alcuni. Tuttavia, se si vanno a vedere le varie obiezioni, si scopre per lo più che si tratta di travisamenti dell'intento tarskiano originario o, alternativamente, attacchi alla teoria della corrispondenza in quanto tale. Non è nostro interesse difendere la teoria corrispondentista, quanto piuttosto esaminare - come abbiamo fatto nel capitolo precedente - la rilevanza filosofica della teoria della verità tarskiana. Per questo motivo, il lavoro di Hartry Field è un *case study* esemplare: mette in luce proprio il problema della connessione tra una struttura logica e le implicazioni filosofiche che l'adozione di tale struttura comporta.

1. *Field, Tarski e il fisicalismo*

Il lavoro di Field, *Tarski's theory of truth*, apparso nel 1972 sul *Journal of Philosophy*, può a ragione essere considerato uno dei più significativi contributi al dibattito filosofico circa la concezione semantica della verità. Il *paper* di Field non è di semplice lettura dal momento che sono condensate almeno tre istanze al suo interno: un'istanza *esegetica*, in cui l'autore cerca di delineare quali fossero gli intenti di Tarski, un'istanza *concettuale specifica*, che prevede una valutazione critica della proposta tarskiana e infine un'istanza *concettuale generale*, in cui si discute l'effettiva compatibilità della semantica tarskiana con i limiti imposti da una metafisica di stampo fisicalista. Queste tre tematiche, come avremo modo di vedere, sono intrinsecamente connesse nell'argomentazione di Field ed è bene, nella nostra esposizione, chiarirne i confini; inoltre, questi nuclei concettuali costituiranno i principali bersagli su cui dirigere altrettanti attacchi all'interpretazione di Field.

In estrema sintesi l'accusa che Field muove a Tarski è di non aver rispettato i patti che lo stesso Tarski aveva preso all'inizio della trattazione del concetto di verità. Come si ricorderà, l'articolo di Tarski

> is almost wholly devoted to a single problem – *the definition of truth*. [...] In this construction I shall not make use of any semantical concept if I am not able previously to reduce it to other concept[1].

Punto cruciale degli intenti di Tarski è allora la costruzione di una definizione di verità che non faccia uso di nozioni semantiche primitive. L'unica possibilità rimane allora la riduzione degli eventuali concetti semantici ad altri concetti che, possiamo pensare, risulteranno più chiari rispetto quelli indagati.

Ebbene, la tesi di Field è, di fatto, la negazione del proposito tarskiano:

> My contrary claim will be that Tarski succeeded in reducing the notion of truth *to certain other semantic notions*; but that he did not in any way

1 Tarski 1983, 152-153.

> explicate these other notions, so that his results ought to make the word 'true' acceptable only to someone who already regarded these other semantic notions as acceptable[2].

La strategia argomentativa di Field si basa sui seguenti punti:

i) Innanzitutto, sviluppa due teorie della verità di stampo tarskiano (che egli chiama T1 e T2) il cui proposito è esibire il riferimento, prima esplicito e poi implicito, a nozioni semantiche primitive.
ii) Mostra, poi, come ci siano due modi per interpretare la teoria tarskiana: o si fa uso di nozioni semantiche primitive oppure risulta del tutto non esplicativa dei concetti di verità e soddisfazione.
iii) Ma, poiché vogliamo (ed è plausibile che lo volesse anche Tarski) una teoria della verità esplicativa, dobbiamo scegliere il primo disgiunto, cioè quello che prevede l'assunzione di nozioni semantiche primitive.
iv) Tuttavia, questa strada non è percorribile perché ciò sarebbe in patente conflitto con un'impostazione fisicalista che Tarski avrebbe assunto nel lavoro sulla verità

Segue la conclusione. Si noti che Field non considera la teoria di Tarski come banale o, peggio, errata. Anzi, è convinto che i suoi risultati

> are extremely important, and have applications not only to mathematics but also to linguistics and to more directly "philosophical" problems about realism and objectivity. I think, however, that the real value of Tarski's discoveries for linguistics and philosophy is widely misunderstood, and I hope to eradicate the most central misunderstanding by clarifying and defending the claim that Tarski merely reduced truth to other semantic notions[3].

Come si vedrà in seguito, la fecondità della teoria tarskiana per Field può emergere solo se accompagnata da una serie di teorie ausiliarie del riferimento, compatibili con una cornice naturalista e fisicalista.

---

2 Field 1972, 347.
3 Field 1972, 348.

1.1. *Due teorie della verità*

Presentiamo ora le teorie T1 e T2 con lo scopo di mettere in luce le assunzioni implicite nella semantica tarskiana. T1 è la teoria di Tarski *come avrebbe dovuto essere presentata*, secondo Field.[4]

Alfabeto di T1

(a) costanti logiche primitive: $\neg$, $\vee$, $\forall$
(b) costanti non logiche primitive:
   1. costanti individuali: $c_1, c_2, c_3, \ldots$
   2. costanti funzionali monadiche: $f_1, f_2, f_3, \ldots$
   3. costanti predicative monadiche: $P_1, P_2, P_3, \ldots$
(c) variabili individuali: $x_1, x_2, x_3, \ldots$
(d) segni ausiliari: (,)

Definizione induttiva di termine

Base: tutte le costanti individuali e le variabili sono termini.

Passo: se t è un termine, allora $f_i(t)$ è un termine. Nient'altro è un termine.

Definizione induttiva di formula

Base: se t è un termine, allora $P_i(t)$ è una formula

Passo: se α è formula e β è formula e x_i è una variabile allora, $\neg\alpha$, $\alpha \vee \beta$, $\forall x_i \alpha$ sono formule. Nient'altro è una formula.
Definizione di enunciato

α è un enunciato se e solo se α è una formula senza variabili libere.

4 Si possono muovere obiezioni di principio a questo modo di procedere; senza dubbio, Field sta compiendo qui una sorta di 'ingegneria concettuale', discostandosi, quindi, dal dato testuale. Tuttavia, poiché questa discussione è primariamente centrata sulla rilevanza filosofica della teoria tarskiana possiamo accettare questo modo di procedere.

T1 è di fatto equivalente al calcolo dei predicati del primo ordine[5]; Field passa quindi alla presentazione della semantica di T1 e segue la traccia tarskiana introducendo il concetto di soddisfazione. Come è noto, data una sequenza di oggetti $s = < s_1, s_2, ... >$, la denotazione della variabile x_k rispetto all'assegnamento s, sarà s_k, ovvero il *k*-esimo elemento della sequenza. Fino a qui nulla di particolare. Ma poiché la definizione di soddisfazione (e, poi, di verità) comprende anche le costanti, bisogna definire la relazione di denotazione per i nomi, i predicati e le funzioni. Si chiede infatti Field

> But what is the denotation relative to s of c_k?[6]

Evidentemente, gli oggetti assegnati alle variabili sono del tutto irrilevanti per comprendere la relazione di denotazione delle costanti. E infatti queste ultime avranno delle denotazioni fisse (costanti, appunto). Così avremo che c_k denota rispetto a un certo assegnamento s un certo individuo. Una cosa analoga accade per quanto riguarda i predicati: per sapere la denotazione di P_k dovremo andare a vedere la sua estensione, ovvero a quali oggetti si *applica* il predicato P_k. Analogamente, la funzione f_k sarà *soddisfatta* (o *saturata*) da determinate coppie ordinate di elementi.[7]

Clausole di denotazione per termini singolari

(a) variabili: x_k denota rispetto all'assegnamento s (denota_s) s_k (ovvero il k-esimo elemento della sequenza s)
(b) costanti individuali: c_k denota_s ciò che c_k denota
(c) costanti funzionali: $f_k(t)$ denota_s y se e solo se c'è uno z tale che t denota_s z e f_k è saturata da $<y,z>$

Clausole di soddisfazione

(a) $P_k t$ è soddisfatto da una sequenza s (vero_s) se e solo se
    1. C'è un oggetto b tale che t *denota*$_s$ b

5 O meglio a una porzione di esso, dal momento che sono presenti solo predicati e funzioni monadiche.
6 Field 1972, 349.
7 Ricordiamo che nella presentazione di Field, predicati e le funzioni sono esclusivamente monadici; non è difficile estendere l'argomento ai casi poliadici.

2. P_k si *applica* a b
(b) $\neg\alpha$ è $vero_s$ se e solo se α non è $vero_s$.
(c) $\alpha \wedge \beta$ è $vero_s$ se e solo se α è $vero_s$ e β è $vero_s$
(d) $\forall x_k \alpha$ è $vero_s$ se e solo se per ogni sequenza s* che differisce da s al massimo al k-esimo posto, α è $vero_{s^*}$.

(Clausola di verità) Un enunciato è vero sse è $vero_s$ per tutte le s.

Ora, come si può vedere facilmente dalla presentazione di Field, la definizione di verità dipende dalla definizione di soddisfazione, secondo l'intuizione di Tarski. Tuttavia, il caso base, cioè quello delle formule atomiche, presenta due occorrenze di nozioni semantiche: il concetto di denotazione (rispetto a un determinato assegnamento) e quello di applicazione di un predicato. Del resto, anche nelle clausole di denotazione compaiono queste nozioni semantiche con in aggiunta anche la relazione di saturazione di una costante funzionale. Ma, ed ecco il punto cruciale dell'argomento, la teoria di Tarski riduce una nozione semantica (quella di verità) ad altre tre nozioni semantiche (e, segnatamente, quelle di denotazione, di applicazione e di saturazione). Con le parole dello stesso Field:

> It explains what it is for a sentence to be true in terms of certain semantic features of the primitive components of the sentence: in terms of what it is for a name to denote something, what it is for a predicate to apply to something, and what it is for a function symbol to be fulfilled by some pair of things[8].

Field propone di racchiudere queste tre nozioni semantiche primitive sotto un'unica denominazione, quella di *denotazione primitiva*.[9] Il problema è allora chiarissimo: la teoria T1 non riesce a ridurre realmente la nozione di verità ad altri concetti non semantici.

Ma T1 non è affatto la teoria presentata da Tarski! Proprio per questo motivo Field passa a illustrare la teoria T2 che, nei suoi intenti, dovrebbe essere la versione fedele al dettato tarskiano. La costruzione del linguaggio di T2 è del tutto identica a quella della controparte T1, le differenze

---

8 Field 1972, 350.

9 Seguiremo anche noi la convenzione di Field. Talvolta useremo il termine 'riferimento primitivo', dal momento che, normalmente, il concetto di riferimento è più ampio di quello di denotazione.

sono, ovviamente, nell'apparato semantico. L'idea di Tarski è, come abbiamo visto in precedenza, quella di tradurre ogni espressione del linguaggio oggetto nel metalinguaggio, quindi utilizzare le espressioni del metalinguaggio per formulare le condizioni di soddisfazione. Per semplicità grafica modificheremo leggermente la simbologia di Field, usando il grassetto per indicare le espressioni del metalinguaggio, mentre tralasceremo di indicare i nomi metalinguistici delle componenti del linguaggio oggetto.

Clausole di denotazione per termini singolari

(d) variabili: x_k denota rispetto all'assegnamento s ($denota_s$) s_k (ovvero il k-esimo elemento della sequenza s)
(e) costanti individuali: c_k $denota_s$ c_k
(f) costanti funzionali: $f_k(t)$ $denota_s$ $f_k(t)$ dove t $denota_s$ *t*

Clausole di soddisfazione

(e) $P_k t$ è soddisfatto da una sequenza s ($vero_s$) se e solo se $P_k t$, dove t $denota_s$ *t*.

Le altre clausole sono perfettamente identiche e quindi possiamo tralasciarle. Secondo Field, l'impiego del dispositivo di traduzione da linguaggio oggetto a metalinguaggio sembra eliminare il ricorso alla nozione generale di riferimento primitivo che è ingrediente essenziale in T1. Riservandoci di discutere in sede critica la ricostruzione di Field, continuiamo con il prossimo punto dell'argomento.

1.2. *Riduzionismo*

Abbiamo visto che, in base ai propositi tarskiani, è necessario adottare una teoria della verità come T2, una teoria cioè che non impieghi altre nozioni semantiche per definire il concetto di verità. Ci si potrebbe chiedere per quale motivo sia così indispensabile una definizione della verità che risulti a-semantica. La risposta di Field è semplice: perché altrimenti avremmo una teoria che non è in accordo con i presupposti fisicalisti. Quindi, la definizione di verità deve essere data utilizzando una semantica compatibile con una descrizione del mondo fisicalista. La ragione per cui Field ascrive a Tarski una simile posizione filosofica deve essere ricercata nel passaggio molto discusso di un articolo *The establishment of scientific semantics*, presentato al Congresso Internazionale della Filosofia Scientifica che si tenne a Parigi nel 1935.[10]

> It would then be difficult to bring this method [cioè studiare la verità introducendo un termine primitivo nel sistema assiomatico] into harmony with the postulates of the unity of science and of physicalism (since the concepts of semantics would be neither logical nor physical concepts)[11].

Che cosa intenda Field con 'fisicalismo' è espresso nella terza sezione del suo articolo:

> The doctrine that chemical facts, biological facts, psychological facts and semantical facts, are all explicable (in principle) in terms of physical facts[12].

La visione fisicalista del mondo cerca di fornire una spiegazione di ogni evento in termini esclusivamente fisici; quali esempi di impostazioni differenti, Field cita il 'vitalismo', ovvero la teoria secondo la quale vi sono fatti biologici irriducibili e il 'cartesianesimo' in base a cui vi sono fatti mentali irriducibili. Accanto a queste metafisiche, ci sarebbe anche il semanticismo, per cui alcuni fatti semantici sono primitivi e non riducibili a null'altro. Quali antidoti, per Field, contro queste impostazioni filosofiche? Essenzialmente vi sono due strategie: l'una consiste nello spiegare i termini di una teoria biologica in termini non biologici; l'altra nel bollare la credenza filosofica come mitologica.[13] Quindi, prosegue Field, poiché il contesto filosofico di Tarski era di tipo riduzionista e fisicalista, la sua teoria della verità non deve fare riferimento a nessun primitivo semantico, pena una qualche forma di semanticismo.

---

10 Cfr. Tarski 1983, 401-408. L'articolo venne pubblicato prima in polacco con il titolo di *O ugruntowaniu naukowej semantyki* in Przeglad Filozoficzny, 39, 1936, pp. 50-57 e poi in tedesco come Grundlegung der wissenschaftlichen Semantik, vol. 3., (Actualités Scientifiques et Industrielles, 390), Paris, 136, pp. 1-8.

11 Tarski 1983, 406.

12 Field 1972, 357.

13 Field 1972, 359.

La teoria T2 sembrerebbe quindi incontrare i *desiderata* di Tarski letto alla luce di Field. Infatti il criterio di una definizione della verità deve rispondere a una condizione simile a questa:

(M) $\forall\alpha(\alpha \text{ è vero } \equiv \mathrm{B}(\alpha))$

dove $\mathrm{B}(\alpha)$ è una formula ben formata che non contiene termini semantici[14]. Ora $\mathrm{B}(\alpha)$ deve risultare in qualche modo informativa. Detto altrimenti, comprendere $\mathrm{B}(\alpha)$ deve aiutarci a comprendere cosa voglia dire che α è vero. D'altro canto, l'equivalenza estensionale tra α è vero e $\mathrm{B}(\alpha)$ non è certo sufficiente; è necessario un legame più forte che indichi la *rilevanza* di $\mathrm{B}(\alpha)$ per il concetto di verità. Ma, ed è l'altro corno di un ipotetico dilemma, Field non può accettare un nesso di equivalenza intensionale: *in primis*, perché tale relazione è di difficilissima caratterizzazione e poi, soprattutto, perché comporterebbe proprio una violazione del criterio di riduzione fisicalista. Il problema rimane aperto, alla ricerca di una condizione più forte di quella estensionale ma non come quella intensionale. Pertanto, secondo Field, la riduzione tarskiana del concetto di verità non è affatto esplicativa. Questo è l'ultimo passaggio del suo argomento: nonostante la teoria T2 (l'originale di Tarski) non faccia riferimento a nozioni semantiche primitive, è del tutto inadeguata perché il tipo di riduzione (fisicalista) di Tarski non è affatto *esplicativo*. Qui è cruciale come intendere il concetto di *spiegazione* in gioco. In filosofia della scienza, segnatamente nella teoria della spiegazione, l'oggetto principe della spiegazione è un fenomeno. In questo caso, invece, abbiamo un concetto (che, seguendo Tarski, viene *definito* e non spiegato). Ora, Field si può giustamente chiedere se la definizione risulti esplicativa nel senso di spiegare (e cioè dipanare) il contenuto del *definiendum*. In questa accezione, è esplicativa una definizione tale che la comprensione del *definiens* migliora la nostra comprensione del *definiendum*. Vediamo, quindi, perché la definizione offerta da T2 non è affatto esplicativa (e cioè epistemicamente vantaggiosa) agli occhi di Field.

---

14 Field riconosce che nulla di simile a (M) è stato espresso da Tarski; dedica pertanto una lunga nota del suo *paper* a discutere le relazioni tra (M) e la convenzione T. Non ci soffermeremo ulteriormente su questo punto.

1.3 *Dalla chimica alla semantica*

L'esempio che Field porta è piuttosto semplice: consideriamo il concetto di valenza di un elemento chimico. Che cos'è la valenza? È un numero intero che indica il tipo di combinazioni chimiche che quell'elemento può intrattenere[15]. Ebbene, per ridurre il concetto di valenza dobbiamo costruire uno schema simile a quello enunciato in precedenza per il concetto di verità:

$$\forall E \forall n\big(E \text{ ha valenza } n \equiv B(E,n)\big)$$

dove la variabile E varia sugli elementi conosciuti e n sui numeri interi. Poiché la nostra riduzione deve essere di carattere fisicalistico è chiaro che in $B(E,n)$ non devono comparire concetti chimici. Un buon esempio di riduzione fisicalistica del concetto di valenza può essere quello in cui si fa riferimento a determinate proprietà fisiche degli atomi, in particolare alla costituzione dei loro orbitali. In questo caso, il concetto di valenza sarà spiegato mediante il riferimento ad altri concetti (elettroni di legame, orbitali e così via) appartenenti all'ambito della fisica e quindi accettabili in base agli standard epistemologici del fisicalismo.

Secondo Field è possibile, però, un'altra riduzione (estensionale) del concetto di valenza che proceda in questo modo:

$$\forall E \forall n\big(E \text{ ha valenza } n \equiv E \text{ è potassio } \& \ n = +1 \vee E \text{ è zolfo } \& \ n = -2 \vee \ldots\big)$$

La disgiunzione a destra del bicondizionale è in grado di esaurire tutti i casi possibili (poiché gli elementi conosciuti sono in numero finito) e quindi risulterà estensionalmente equivalente con la riduzione fornita in precedenza utilizzando la fisica atomica.[16]

Abbandonando la metafora, Field accusa Tarski di offrire una riduzione della semantica in termini fisicalistici del tutto inadeguata esattamente come la definizione della valenza in base a una lista delle valenze degli

15 Field 1972, 362-364.

16 Come qualche commentatore ha notato, *en passant*, Field considera termini come 'sodio', 'calcio', 'ferro', appartenenti all'ambito della fisica. Questo sembra perlomeno dubbio. In ogni caso, si può riformulare l'esempio, sostituendo al posto dei nomi degli elementi una proprietà fisica che li identifichi, come il numero atomico.

elementi chimici. Infatti la riduzione implicita di Tarski dovrebbe assomigliare, nel caso delle costanti individuali, a qualcosa come:

$$\forall c \forall a \left(c \text{ denota } a \equiv \left(c = 'c_1' \wedge a = \mathsf{c}_1 \right) \vee \left(c = 'c_2' \wedge a = \mathsf{c}_2 \right) \vee \ldots \right)$$

Qui abbiamo due variabili speciali: *c* sta per nomi (o meglio, nomi metalinguistici di termini del linguaggio oggetto) e *a* sta per oggetti in una sequenza. Quando, allora avremo che *c* denota *a*? Quando *c* è uguale, per esempio, al nome c_4 e *a* coincide con l'oggetto c_4. In questo senso, diventa cruciale, il caso *base* delle definizioni induttive di denotazione e soddisfazione.

Secondo Field, una specificazione della denotazione come questa è del tutto insufficiente, si tratta di una «bogus reduction» per utilizzare la terminologia di McDowell[17]. Il vantaggio di T2, allora è del tutto apparente: è vero che non fa riferimento a nozioni semantiche primitive ma utilizza un sistema di caratterizzazione della denotazione primitiva per elencazione che, benché estensionalmente corretta, non è affatto esplicativa del concetto in esame. La conclusione di Field è quindi netta: la teoria della verità di Tarski non riesce a soddisfare i requisiti enunciati dal suo stesso autore.

2. *Obiezioni a Field*

Come si ricorderà dalla presentazione del pervasivo e influente argomento di Field, ci sono tre componenti nel suo lavoro che rimandano rispettivamente a: una ricostruzione della definizione di Tarski, una valutazione filosofica della concezione semantica e infine una discussione più generale sull'ammissibilità della semantica in un contesto fisicalista. Ebbene, è possibile attaccare tutti gli assi portanti del ragionamento di Field eccependo cioè sia sulla fedeltà della sua ricostruzione, sia sull'interpretazione filosofica (specifica e generale) dell'opera di Tarski. In ciò che segue, percorreremo le tre vie e vedremo che, nell'imbastire una 'difesa' della concezione semantica, ne metteremo in luce alcuni aspetti filosoficamente rilevanti.

17 McDowell 1978, 118.

2.1. *Fedeltà a Tarski*

È bene spendere qualche parola per sgombrare il campo da alcuni possibili equivoci. Nella presentazione della memoria tarskiana non è ovviamente obbligatorio seguire pedissequamente la simbologia originale;[18] tuttavia, a volte, *details matter*. Soprattutto quando, come questo è il caso, si sta vagliando in ottica filosofica (o comunque metateorica) la plausibilità di una teoria della verità. L'oggetto del contendere è in questo caso il riferimento alle nozioni semantiche primitive che Tarski dichiara di eliminare nella sua definizione di verità e che, secondo Field, sono invece o introdotte implicitamente tramite la nozione generale di riferimento primitivo e non incontrano i gusti fisicalisti oppure vengono 'ridotte' tramite una lista assolutamente non esplicativa.

Nella sua memoria, Tarski non parla quasi mai di 'denotazione' o 'riferimento'; la menzione più significativa è sicuramente questa:

> To say that the name *x* denotes a given object a is the same as stipulate that the object *a* (or every sequence of which *a* is the corresponding term) satisfies a sentential function of a particular type[19].

Il passo è tratto da una lunga nota a piè pagina nella quale Tarski chiarisce l'importanza del concetto di soddisfazione per l'analisi del linguaggio e la semantica, in generale. Scrive infatti Tarski:

> The concept just defined is of the greatest importance for investigations into the semantics of language. With its help the meaning of a whole series of concepts in this field can easily be defined, e.g. the concepts of denotation, definability, and truth, with the last of which we are especially concerned here[20].

Anche la posizione all'interno del testo è particolarmente significativa: siamo infatti tra la definizione 22, in cui vengono enunciate le famose clausole di soddisfazione, e la definizione 23, cioè la caratterizzazione della verità come soddisfazione da parte di tutte le sequenze. Siamo, pertanto, nel cuore della costruzione della definizione di verità per il lin-

---

18 Nel primo capitolo, dove siamo stati il più fedeli possibile all'originale, abbiamo sperimentato direttamente la poca maneggevolezza del linguaggio di Tarski.

19 Tarski 1983, 194n.

20 Tarski 1983, 193-194.

guaggio delle classi, l'esempio scelto da Tarski. Quindi, ricostruire, come fa Field, la semantica tarskiana introducendo i concetti di denotazione, applicazione e saturazione non è certo una mossa fedele al dettato originale. Tanto è vero che lo stesso Tarski spiega in nota che queste nozioni semantiche possono essere derivate dalla nozione di soddisfazione e, con ciò, ridotte.

Se poi analizziamo con attenzione il testo notiamo che Tarski non fornisce una teoria schematica della verità, ma un esempio specifico, costruito sul linguaggio delle classi. Di questo egli è ovviamente consapevole:

> For an extensive group of formalized languages it is possible to give a method by which a correct definition of truth can be constructed for each of them. The general abstract description of this method and of the languages to which it is applicable would be troublesome and not at all perspicuous. I prefer therefore to introduce the reader to this method in another way[21].

Benché alcuni commentatori[22] abbiano rimarcato le differenze tra la presentazione originale di Tarski e la ricostruzione di Field, sembra che quest'ultima citazione non sia stata messa adeguatamente in luce. Qui Tarski è perfettamente conscio che non sta presentando un metodo generale (o diremmo noi oggi una teoria schematica) per la definizione di verità. Anzi, la versione astratta sembra essere più difficoltosa di un esempio concreto, come quello per il linguaggio delle classi. Tutto questo per giustificare la totale estraneità al Tarski storico alla questione delle nozioni semantiche primitive.

Ciò che è importante considerare è che Tarski lavora con linguaggi interpretati. Ciò significa che, dato un linguaggio come quello della teoria delle classi, il significato della (unica) costante non logica, $\subseteq$, è dato dall'interpretazione intesa. Quindi la prima clausola di soddisfazione, così discussa da Field, può essere resa come:

(Sod I) '$X \subseteq Y$' è soddisfatto dalla sequenza σ se e solo se l'elemento X della sequenza è incluso nell'elemento Y della sequenza.

21 Tarski 1983, 167-168.
22 Come per esempio Gila Sher, cfr. Sher 1999.

In Sod I, ciò che è compreso tra le virgolette è un nome di enunciato aperto, mentre l'italiano è utilizzato come metalinguaggio. Il significato della costante predicativa $\subseteq$ è dato, perché noi sappiamo che $\subseteq$ vuol dire inclusione fra classi. Non è necessario specificarlo perché è incluso nel bagaglio di *informazione semantica* data nell'interpretazione del linguaggio.

Questo vuol dire che se la nostra teoria (o, per usare una terminologia tarskiana, il nostro linguaggio oggetto) prevede due costanti predicative, avremo bisogno di due clausole atomiche di soddisfazione e così via. Ancora una volta: la definizione fornita da Tarski di verità non è schematica ma è istanziata in un esempio particolare e presuppone il fatto che i linguaggi di riferimento siano interpretati. Tuttavia, ciò non vuol dire che il contributo tarskiano sia una *specifica* teoria della verità per una teoria T. Infatti, come dicevamo a chiusura del capitolo precedente, la concezione tarskiana della verità fornisce una struttura per la definizione formale del predicato vero. E tale struttura può essere poi istanziata dalla teoria che si desidera.

Una posizione simile alla nostra è sostenuta da Luis Fernandez Moreno; ciò che gioca un ruolo chiave nella costruzione di Tarski è il concetto di 'sinonimia' o di 'identità di significato' tra le espressioni del linguaggio oggetto e le espressioni del metalinguaggio. Ed è ciò che noi avevamo espresso con il riferimento all'interpretazione intesa del linguaggio oggetto.

> What in any case may be objected to Tarski is not to have recognized an exception to the thesis that the rest of semantic concepts are definable in terms of the concept of satisfaction and hence are reducible to it, since that claim is not right – and moreover, it *cannot be right* – with respect to the concept of identity of meaning, synonymy or translation and therefore to the concept of *meaning* itself, since this concept is involved in the characterization of the interpreted languages for which Tarski proposed to define the semantic concepts[23].

Ora, detto questo, non abbiamo scagionato Tarski dall'accusa di Field. Infatti, quand'anche la memoria originale presenti le cose in maniera differente, resta il fatto che effettivamente è possibile generalizzare la semantica tarskiana per qualsiasi teoria formalizzata. Ed è ciò che si fa nel-

---

23 Moreno 2001, 143-144.

le presentazione di semantica logica dei manuali. Diremo per esempio che

$\mathcal{M} \vDash Pt$ se e solo se $I_{\mathcal{M}}(P)$ si applica a $I_{\mathcal{M}}(t)$ o alternativamente $I_{\mathcal{M}}(t) \in I_{\mathcal{M}}(P)$

Qui P è una lettera *schematica* per qualsiasi predicato a un posto mentre compare la funzione di interpretazione che permette di connettere elementi del linguaggio (siano costanti o variabili) con elementi del modello (individui, classi e così via). Assumendo questa versione di semantica tarskiana, possiamo riproporre il dilemma di Field.

Se analizziamo bene la clausola di verità appena esposta, notiamo che sono all'opera implicitamente due teorie: una teoria del *riferimento* e una teoria della *predicazione*. Infatti, la prima opera nella costituzione della funzione di interpretazione (come si connette il linguaggio con il mondo?) mentre la seconda deve spiegare la relazione che sussiste tra il denotato di t e il denotato di P. Per questo motivo abbiamo esibito due caratterizzazioni possibili (l'una intensionale, l'altra estensionale): nella prima infatti bisogna chiarire la natura della relazione di applicazione tra una proprietà e un individuo; nella seconda, invece, si assume l'appartenenza insiemistica come modello matematico della relazione di predicazione. In entrambi i casi, senza un contributo esplicativo in un senso o nell'altro, non possiamo dire di aver chiarito le condizioni di verità di una proposizione come Pt. Quindi la critica di Field, benché forse imprecisa se riferita alla versione originale di Tarski, risulterebbe puntuale nel caso della generalizzazione della teoria tarskiana.

Il punto fondamentale è questo: la clausola atomica è esplicativa a patto che si riesca a dare un qualche significato alle relazioni di denotazione e predicazione. Ma queste nozioni sono *presupposte* nel metalinguaggio e quindi la definizione, legittima, è esplicativa. Come avevamo detto in precedenza, la teoria di Tarski stabilisce le condizioni minimali (o strutturali o logiche) perché si dia corrispondenza tra linguaggio e mondo. In conclusione: la teoria di Tarski non presuppone le nozioni semantiche fondamentali cui abbiamo fatto riferimento. Infatti queste possono essere ridotte alla nozione di soddisfazione, come Field stesso ammette. Tarski era convinto che la relazione di soddisfazione fosse primitiva e su questo si può ovviamente dissentire. Ma non crediamo che ciò sia sufficiente per avvalorare la critica di Field. La teoria semantica della verità è una caratterizzazione soprattutto logica (e con ciò intendiamo strutturale, formale) della verità e pertanto presuppone alcune nozioni semantiche chiave come quelle di riferimento, denotazione, soddisfazione o in generale, quella di significato[24].

2.2 *Banalità*

La seconda strategia di difesa di Tarski si basa su un assunto non del tutto esplicito nell'articolo di Field ma molto interessante dal punto di vista filosofico.

Secondo Field, come si ricorderà, l'unico tipo di riduzione dei concetti semantici disponibile a Tarski è di stampo fisicalista[25] ma le definizioni di riferimento primitivo che egli offre non sono assolutamente informative, essendo nient'altro che elenchi di disgiunzioni. A tal proposito, Field cita l'esempio della valenza chimica e conclude che la riduzione di Tarski è strutturalmente simile e pertanto non riesce nel suo compito. Ora, tralasciando questioni esegetiche di cui abbiamo discusso in precedenza, si possono comunque muovere due rilievi: innanzitutto, una definizione per elencazione può risultare non particolarmente soddisfacente dal punto di vista filosofico ma non ha nulla di scorretto; in secondo luogo, quand'anche il caso atomico non risulti informativo, non segue, *ipso facto*, che anche gli altri casi non siano informativi e che quindi *tutta* la definizione sia banale.

Per quanto riguarda il primo punto, Raatikainen si esprime così:

---

24 Un altro esempio che mi è stato suggerito in conversazione da Alessandro Giordani riguarda la teoria dei modelli. Normalmente, la funzione di interpretazione I associa costanti non logiche del linguaggio a particolari elementi di una struttura. Si assume cioè che le costanti logiche (connettivi, quantificatori) abbiano un'interpretazione intesa chiara. Ma è davvero così? La sterminata letteratura sulla natura delle costanti logiche suggerirebbe proprio l'opposto eppure, nella teoria dei modelli, si assume un determinato punto di vista. Ciò sarebbe problematico solo nel momento in cui si stabilisca che la teoria dei modelli copra l'intero ambito della semantica logica (posizione che probabilmente è sostenuta da alcuni studiosi) esattamente come lo sarebbe il dichiarare la concezione semantica di Tarski una teoria completa della verità.

25 La discussione di questo assunto costituirà la terza e ultima critica a Field.

> A list-like characterization of primitive denotation such as above may strike one as disappointingly shallow philosophically, […] but there is logically speaking nothing in principle wrong in Tarski's approach[26].

Abbiamo visto, inoltre, come, nel caso della valenza chimica, le due definizioni risultano estensionalmente equivalenti. Il problema delle caratterizzazioni di concetti tramite liste è, da un punto di vista formale, un altro: possono essere utilizzate solo se il numero degli elementi in esame è *finito*. Ma nel *Concetto di verità* non è la prima volta che si incontrano definizioni del genere: durante la costruzione della definizione di verità, Tarski prende in considerazione linguaggi con un numero finito di enunciati e afferma che il T-schema sarebbe in grado, solamente in quei casi, di fornire una definizione formalmente corretta e materialmente adeguata di verità. Infatti, siano s_1 e s_2 gli unici enunciati del nostro linguaggio e p_1 e p_2 le corrispettive traduzioni metalinguistiche. Avremo allora che:

$$\forall x\left(x \text{ è vero } \equiv \left(x = s_1 \,\&\, p_1\right) \vee \left(x = s_2 \,\&\, p_2\right)\right)$$

In questo modo la disgiunzione esaurisce tutti i casi possibili; e lo stesso vale per le clausole di soddisfazione per i predicati primitivi del linguaggio. Naturalmente, la definizione vale nella misura in cui si assume la traduzione metalinguistica del predicato espresso nel linguaggio oggetto, come abbiamo discusso in precedenza. Ciò che è rilevante ora riguarda la fecondità filosofica di una definizione di questo tipo; Field sembra accusare Tarski del fatto che una definizione di questo tipo non offra nessuna spiegazione del perché la relazione di soddisfazioni funzioni in quel modo; detto in altri termini, per quale motivo il simbolo $\subseteq$ è tradotto metalinguisticamente con la relazione di inclusione? La ragione è che si tratta della teoria della verità per la teoria delle classi in cui il simbolo $\subseteq$ indica proprio la relazione di inclusione. Ripetendo ancora una volta il punto: Tarski non offre una teoria del significato ma una teoria della verità, presupponendo una teoria (o se non altro, la nozione) del significato. E tale presupposizione si basa sul fatto, minimale, che *almeno* il metalinguaggio risulti interpretato, sia dotato cioè di un significato.

Il secondo punto riguarda un'assunzione implicita nell'argomento di Field. Egli conclude che la definizione di soddisfazione (e quindi quella strettamente relata di verità) di Tarski è banale perché la base, cioè il caso atomico, è banale. Anche ammettendo la verità della premessa (e cioè che

---

26 Raatikainen 2008, 255.

le caratterizzazioni concettuali tramite disgiunzioni non siano informative) non segue immediatamente la conclusione. O meglio, seguirebbe a patto che non vi sia alcun incremento informativo nelle clausole riguardanti i connettivi e i quantificatori. Perché mai

(Sod¬) σ soddisfa '$\neg\beta$' se e solo se σ *non* soddisfa 'β'

o anche

(Sod∨) σ soddisfa '$\beta \vee \gamma$' sse σ soddisfa 'β' *oppure* 'γ'

devono risultare banali? In fondo, ciò che sta a destra del bicondizionale chiarisce il significato di ciò che si trova a sinistra a patto naturalmente di comprendere il significato del metalinguaggio stesso. Normalmente le critiche che vengono mosse all'informatività di clausole come (Sod¬) e (Sod∨) si basano sulla circolarità delle definizioni: si spiega il significato del connettivo logico '∨' grazie al connettivo del linguaggio naturale 'oppure' e questo viene considerato illegittimo. Non ci interessa approfondire qui il tema dell'informatività della logica con i suoi collegamenti con la teoria della verità[27]; ciò che però ci preme è notare come, in assenza di argomenti positivi, la semplice assunzione di banalità del caso atomico (ancora tutta da dimostrare) non è sufficiente per dichiarare la non informatività di tutta la definizione induttiva.

2.3 *Quale fisicalismo?*

La terza (e ultima) linea di attacco all'argomento di Field tratta direttamente la questione del fisicalismo che gioca un ruolo cruciale per la tenuta di tutta la discussione. Ricordiamo, in estrema sintesi, che il progetto tarskiano non mantiene le sue promesse, per Field, perché non riesce a rendere conto della riduzione fisicalista di nozioni semantiche primitive.

---

27 Questa tematica è affrontata con notevole profondità di analisi e originalità da Gila Sher (cfr, a tal proposito, Sher 1991, 1999, 1999b, 2004).

La maggior parte dei commentatori[28] concorda con l'ascrivere a Tarski simpatie fisicaliste; ciò che però è rilevante è se la concezione semantica della verità di Tarski abbia lo scopo di ridurre le nozioni semantiche a nozioni puramente fisiche. La posizione di Frost-Arnold è diametralmente opposta:

> [M]y central contention […] is that Tarski's work on the concept of truth is *not* motivated by a philosophical desire to reduce the concept of truth to purely physical concepts[29].

Quindi il problema non riguarda tanto le convinzioni filosofiche di Tarski (anche se queste naturalmente giocheranno un qualche ruolo) ma la dipendenza concettuale del lavoro sulla verità da un'impostazione metafisica fisicalista. La presentazione del fisicalismo compiuta da Field lascia qualche perplessità; innanzitutto, il riferimento continuo alla fisica, cioè alle teorie fisiche, rende il fisicalismo una posizione filosofica indicizzata temporalmente. Nel Settecento, un fisicalista sarebbe stato del tutto legittimato a credere all'esistenza del flogisto, salvo poi vederlo scomparire dalla propria ontologia in seguito al progresso della termodinamica. Si tratta solo di un aspetto di una serie di problemi circa la connessione tra le ontologie implicite delle teorie fisiche e il fisicalismo. Un altro passaggio che pare un po' affrettato riguarda le strategie di eliminazione dei termini che Field presenta:

> But there is another possible strategy, which is to argue that the biological terms are illegitimate. The second strategy seems reasonable to adopt in dealing with the following predicate of (reincarnationist) biology: '*x* ha the same soul as *y*'. A physicalist would never try to find physical or chemical facts that underlie reincarnation; rather, he would reject reincarnation as a myth[30].

Perché mai un concetto come quello di 'valenza' può essere ridotto fisicalisticamente mentre un concetto come quello di 'anima' deve essere

---

28 Per esempio, McDowell 1978, Soames 1984, Stalnaker 1987. Kirkham 1993 presenta invece una posizione leggermente più complessa. In ciò che segue, muoveremo, invece, da una recente analisi di Frost-Arnold 2004. Interessanti per la questione sono anche Wolenski 2002, Mancosu 2008 e Wolenski, Simons 1989.

29 Frost-Arnold 2004, 267.

30 Field 1972, 359

bollato come appartenente alla mitologia? Certamente, di valenza ne parla la chimica mentre anima non compare nel vocabolario di nessuna scienza. Ma che dire di casi appena più complessi, o borderline? Si pensi a ‘mente’, ‘vita’, ‘caso’ e così via, quale infallibile guida ci permetterà di capire subito se si tratta di questioni meritorie di riduzione o destinate alla pattumiera della mitologia?

Passando poi al tema vero e proprio, la posizione di Field è giustificata in base al passaggio

> It would then be difficult to bring this method [cioè studiare la verità introducendo un termine primitivo nel sistema assiomatico] into harmony with the postulates of the unity of science and of physicalism (since the concepts of semantics would be neither logical nor physical concepts)[31].

Frost-Arnold dedica buona parte del suo *paper* a relativizzare l’importanza di questa citazione; le ragioni addotte sono essenzialmente di tre ordini.

Innanzitutto, la teoria semantica di Tarski era in odore di metafisica, in un contesto storico in cui la metafisica era bollata come inutile nonsenso. Ne è testimonianza la serie di critiche riportate in appendice a Tarski (1944) e il famoso *disclaimer* in cui Tarski dichiara la neutralità filosofica della sua concezione rispetto alle principali opzioni filosofiche.[32] Il congresso del 1935 a Parigi accoglieva il *gotha* dei fautori della filosofia scientifica e la presentazione di Tarski avrebbe sicuramente mostrato il fianco a una serie di obiezioni in questa direzione. Tra l’altro bisogna aggiungere che, mentre una valutazione seria del lavoro di Tarski non può prescindere dall’aspetto tecnico-formale, il contributo di Tarski al Congresso era assolutamente espositivo. Il riferimento al fisicalismo (l’unico in tutta la produzione filosofica di Tarski) è allora spiegabile come una sorta di *captatio benevolentiae* nei confronti di un uditorio pervaso da sentimenti profondamente anti-metafisici.

Inoltre, nella versione polacca del testo, il punto in questione risulta leggermente differente:

> It would be difficult to bring […] scientific semantics into harmony with postulates of the unity of science and of physicalism, which are

---

31 Tarski 1983, 406.
32 Vedi nota 88.

> propagated by a great number of philosopher from the so called Vienna Circle[33].

La presenza della specificazione non è ovviamente una prova definitiva a favore della tesi di Frost-Arnold ma di sicuro ridimensiona la portata dell'esegesi di Field.

Ancora più significativo a nostro avviso risulta l'esame attento delle condizioni poste da Tarski sul metalinguaggio. Esso deve contenere: i) termini della logica; ii) i termini corrispondenti del linguaggio oggetto; iii) i nomi di espressioni del linguaggio oggetto. Non c'è nessun riferimento a vincoli di carattere ontologico. Anzi, la costituzione del metalinguaggio è data nei termini di un linguaggio di ordine superiore (in grado cioè di parlare del linguaggio oggetto), lasciando aperta la questione circa l'effettiva riduzione della semantica alla sintassi (o morfologia del linguaggio secondo la terminologia di Tarski).

La posizione di Frost-Arnold è, in questo senso, chiarissima:

> And it is clear, given the definition of satisfaction, that Tarski is reducing semantic terms to set-theoretical (and other logical) terms[34].

La semantica è calcolo logico più teoria degli insiemi e quindi la riduzione dei termini semantici c'è stata ma non in senso fisicalista. Questa tematica, cruciale per interpretare la portata filosofica della definizione tarskiana, verrà discussa in seguito approfonditamente. Qui ci basta solo considerare come la riduzione fisicalista non sia l'unica possibile.

Nell'interessantissimo Wolenski Simons (1989) la questione del fisicalismo è presa in considerazione in un passaggio illuminante:

> Field remarks that Tarski's construction does not fulfill the postulates of physicalism. This is correct, since set theory is, to put mildly, difficult to build on a physicalistic basis. However, Field's objections cannot be arguments against the semantics conception of truth as such. They are convincing only for adherents of physicalism. Since physicalism is not self-evident, we may conclude that although Tarski fails to reduce semantics to physicalisticaly interpreted syntax, his definitions do reduce semantics to a well-established vocabulary. (Wolenski Simons 1989: 414)

---

33 Frost-Arnold 2004, 271.
34 Frost-Arnold 2004, 273.

Si potrebbe forzare un po' la mano in tal senso e ribaltare l'argomento: proprio perché è così complesso accordare una posizione ontologica fisicalista con la semantica tarskiana e, del resto, quest'ultima si è rivelata uno strumento efficace e affidabile, ciò potrebbe valere come argomento contro il fisicalismo, al di là delle personali convinzioni di Tarski e della temperie culturale nella quale egli si trovò a operare.[35]

Concludendo la questione circa il fisicalismo, in quanto teoria formale (o perlomeno formalizzabile e assiomatizzabile) la teoria della verità di Tarski dovrebbe essere compatibile con un'ontologia fisicalista nella misura in cui lo sono altre teorie formali quali, per esempio, la teoria degli insiemi. La risposta a questa domanda dovrebbe provenire da un fisicalista. Storicamente vi sono stati autori di simpatie fisicaliste (o comunque naturaliste) che hanno comunque accettato teorie formali profondamente infinitarie come la teoria degli insiemi di Zermelo-Fränkel (si veda per esempio Quine).

Il punto però è un altro, a nostro avviso. Ed è se una certa interpretazione filosofica della teoria tarskiana sia o meno compatibile con una posizione fisicalista. Ciò equivale a chiedersi se una teoria generale (cioè logica) della verità come corrispondenza possa essere accettata da un fisicalista. Probabilmente la riposta dipende da quanto il fisicalista è disposto ad ammettere in fatto di *infinito*. La teoria tarskiana è radicalmente *infinitaria* (non solo per la definizione stessa di verità come soddisfazione da parte di tutte le sequenze ma anche per la stessa clausola di verità del quantificatore universale). Ora, se il fisicalista è coerente dovrebbe preferire maggiormente una definizione finitaria, allontanandosi da Tarski, ma sono disponibile, ovviamente, numerose altre interpretazioni. In ogni caso, l'argomento di Field non sembra minare la legittimità e la portata della teoria di Tarski al di là delle scelte ontologiche di ciascuno.

---

35 In effetti Tarski era probabilmente un fisicalista e nominalista. Feferman (2004, 52) riporta questa sua illuminate ammissione:

> However, I represent this very crude, naïve kind of anti-Platonism, one thing which I could describe as materialism, or nominalism with some materialistic taint, and it is very difficult for a man to live his whole life with this philosophical attitude, especially if he is a mathematician, especially if for some reason he has a hobby which is called set theory.

II PARTE

TEORIE ASSIOMATICHE DELLA VERITÀ

IV
TEORIE ASSIOMATICHE DELLA VERITÀ E DEFLAZIONISMO

1. *Approccio semantico vs approccio assiomatico*

La concezione semantica di Tarski non è l'unico esito della trattazione logico-formale della nozione di verità. È, infatti, possibile adottare un'altra strategia: al posto di definire, in un metalinguaggio, il predicato 'essere vero' si può introdurre questo predicato nel nostro linguaggio come un segno primitivo e fornire degli assiomi, cioè dei principi, che ne regolano la funzione e l'uso. In altre parole, è possibile assiomatizzare il concetto di verità. È importante notare subito alcuni punti fondamentali: in primo luogo, l'approccio assiomatico (così ci esprimiamo per indicare la posizione che stiamo presentando differenziandola, quindi, dall'approccio semantico, esaminato nella prima parte del volume) non significa, *ipso facto*, la formalizzazione della teoria tarskiana della verità. Come è noto (cfr. capitolo 1, prima parte) Tarski non fornisce una trattazione completamente formalizzata della sua teoria della verità: non è specificato, per esempio, il calcolo logico assunto nel metalinguaggio. Naturalmente si può formalizzare la concezione tarskiana, assiomatizzandola completamente. Nell'approccio semantico, elaboriamo, come abbiamo visto, una definizione del predicato di verità all'interno di un metalinguaggio. Negli approcci assiomatici, invece, si introduce il predicato T al linguaggio della teoria di base ℓ_B e si stipulano degli assiomi che governino l'uso di T. L'aggiunta del predicato costituisce un'estensione linguistica essenziale (essendo noto il fatto che T non è definibile all'interno di ℓ_B). Quindi c'è una perfetta simmetria tra LO e ML (linguaggio oggetto e metalinguaggio) e ℓ_B + T (dove con T si intendono anche gli assiomi che riguardano T).

Tarski è all'origine di entrambi gli approcci. Come specificheremo meglio in seguito, dopo aver esposto la concezione semantica della verità, Tarski prende in considerazione anche l'impostazione assiomatica come metodologia per l'analisi della verità. E, pur con qualche sfumatura, il suo atteggiamento nei confronti dell'assiomatizzazione del predicato di verità è scettico. Nel prossimo capitolo vedremo nel dettaglio le ragioni della perplessità tarskiana; qui è importante solo considerare che, proprio questo scetticismo, mostra che concezione semantica e assiomatica rispondono a intuizioni differenti e che le preferenze di Tarski sono, ovviamente, orientate verso la prima a scapito della seconda.

Una teoria assiomatica della verità è quindi una teoria formalizzata che consta di tre parti: una parte *logica*, comprendente cioè il linguaggio e le regole di deduzione nel quale la teoria è espressa; una parte *matematica* o teoria di *base* che comprende una serie di assiomi che specificano il tipo di oggetti di cui si predica la verità (per esempio enunciati aritmetici); infine la vera e propria parte specifica che consta di assiomi che contengono il predicato di verità T. Le cose diventeranno più chiare non appena ci accingiamo a presentare 'concretamente' alcune teorie assiomatiche della verità.

La più importante classificazione delle teorie assiomatiche della verità riguarda la possibilità di predicare la verità a enunciati che contengano, a loro volta, occorrenze del predicato di verità. La riflessione del predicato di verità è alla base, come si ricorderà, della formazione dell'antinomia del mentitore; quindi, le scelte sono due: o non si ammette la predicazione iterata della verità, oppure si pongono in essere altre limitazioni che pur permettendo formule come $T(T(p))$, è vero che è vero che p, non portano a contraddizione. La prima opzione è quella di avere teorie assiomatiche *tipate*, cioè suddivise in *tipi*, la seconda è quella di avere teorie *non tipate*.

Le teorie assiomatiche tipate rispondono all'intuizione originale di Tarski circa la suddivisione del linguaggio in linguaggio oggetto e metalinguaggio; la strategia per evitare i fenomeni di autopredicazione è simile a quella impiegata da Russell e Whitehead nella teoria dei tipi: la predicazione (o la relazione di appartenenza) può sussistere solo tra oggetti che sono di tipo differente. Al contrario, le teorie assiomatiche non tipate nascono dall'intuizione di Saul Kripke e dalla sua teoria della verità. Non ci occuperemo in questa sede delle teorie non tipate, sia per ragioni di spazio che di opportunità contenutistica, essendo il pensiero tarskiano (con un approccio tipato) alla base del nostro lavoro.

Le teorie assiomatiche della verità, dicevamo, non definiscono in maniera esplicita il predicato di verità: questo viene assunto come un segno primitivo, non definito. Ci si può chiedere se e in quale misura l'assiomatizzazione comporti, comunque, una qualche forma di definizione. È la domanda che sta alla base della proposta formalista di Hilbert e della sua scuola: l'assiomatizzazione fornisce - risponde il formalismo - una definizione *implicita* dei concetti in gioco. Ovvero, non si viene a dire, per tornare al nostro caso, che un enunciato è vero se e solo se è soddisfatto da tutte le sequenze (o, secondo teorie non logiche della verità, se e solo se descrive uno stato di cose esistente) ma si chiarisce la natura di questa nozione come tutto ciò che obbedisce agli assiomi circa il predicato di verità.

Lo stesso discorso può essere fatto per tutti i concetti che vengono assiomatizzati; si prenda il caso emblematico di numero naturale. La strategia di Frege, come quella di Dedekind, mira a una definizione (data in termini logici o logico-strutturalisti) di numero naturale. Diverso è, invece, l'approccio della definizione implicita: numero è tutto ciò che obbedisce agli assiomi di Peano. Così come retta è tutto ciò che obbedisce agli assiomi della retta e così via. Tipico dell'approccio formalista è, pertanto, la de-sostanzializzazione dei concetti matematici: non esiste una natura propria del numero, colta attraverso l'intuizione e codificata negli assiomi ma sono gli assiomi a determinare l'ontologia degli oggetti che 'soddisfano' le condizioni lì enunciate.

Al di là di queste considerazioni, ciò che è più importante per i nostri scopi riguarda, però, i rapporti tra approcci semantici ed assiomatici: sono antitetici o complementari? E a quali condizioni? Come Halbach mette bene in luce i metodi possono essere complementari:

> [T]he axiomatic approach is not opposed to definitional approaches. In fact, both approaches complement one another. For instance, one may start by formulating truth-theoretic principles like Tarski's T-sentences for a theory like Peano arithmetic and then show that a suitable truth predicate for the language of arithmetic can be defined in set theory by defining a model for Peano arithmetic expanded by these truth- theoretic principles. Hence one knows that these truth-theoretic principles are consistent, at least if set theory is to be trusted. (Halbach 2011, 6)

Del resto, molte teorie formali della verità hanno una struttura simile alle teorie aritmetiche (e si possono stabilire precise relazioni di potenza deduttiva tra queste). Proprio come indaghiamo interessanti proprietà dell'aritmetica di Peano facendo riferimento ai modelli (e alle loro proprietà) che rendono veri gli assiomi di Peano così si può fare per le teorie della verità (per esempio, proprio per la teoria della verità dell'aritmetica di Peano).

Il punto fondamentale non riguarda tanto gli approcci quanto la specifica definizione di T. Come abbiamo detto, nella memoria originale di Tarski la nozione di verità è definita (tramite da definizione ricorsiva del predicato di soddisfazione). È possibile assiomatizzare la teoria tarskiana della verità aumentando il linguaggio di una teoria di base (L_B) una serie di assiomi specifici che riguardano il predicato indefinito T e che sono perfettamente analoghi alle classiche clausole tarskiane. Da questo punto di vista non c'è particolare differenza se non che in un caso T è definito tramite un altro predicato e nell'altro viene assunto come primitivo. La differenza fondamentale tra i due approcci si nota nel caso in cui venga variato l'insieme degli assiomi che governano il predicato di verità. Avremo quindi tante teorie assiomatiche della verità che cercano di catturare ciascuna una certa concezione (pre-formale) di verità; in molti casi, l'idea sottostante e la seguente 'traduzione' assiomatica è differente dalla concezione tarskiana e i due approcci risultano così antitetici quanto per contenuto, più che per forma[1].

Come è facile vedere da questi brevi cenni, ci sono numerose questioni concettuali al fondo della trattazione logica (assiomatica, in particolare) della verità. Ma il tema è ancora più interessante perché l'approccio assiomatico è stato adottato da molti esponenti di una particolare concezione della verità che costituirà il nostro contraltare teorico: il *deflazionismo*.

2. *Deflazionismo*

Tra le cose più difficili nella discussione filosofica sulla verità c'è sicuramente la caratterizzazione di una delle posizioni più diffuse nel dibattito contemporaneo: il deflazionismo. Questo è dovuto essenzialmente al fatto che non esiste un riferimento preciso cui ispirare le varie proposte deflazioniste. Non c'è, in altre parole, un 'Aristotele' dei deflazionisti; si tratta, piuttosto, di un atteggiamento filosofico o di un orientamento generale che si declina, poi, in proposte che talvolta divergono anche note-

---

1 Il tutto sarà chiarissimo a breve quando analizzeremo il caso del deflazionismo: una particolare concezione della verità - una concezione deflazionista - viene tradotta da una opportuna scelta di assiomi - i T-enunciati.

volmente. Talvolta si fa risalire l'origine del deflazionismo a Frege citando, normalmente, un noto passo de *Il pensiero*.

> It is worthy of notice that the sentence 'I smell the scent of violets' has the same content as the sentence 'it is true that I smell the scent of violets'. So it seems, then, that nothing is added to the thought by my ascribing to it the property of truth. (Frege, 1918)

Non è questa la sede per tentare un'esegesi fregeana; tuttavia, da una semplice considerazione circa la teoria di Frege sulle proposizioni, sembra quantomeno improbabile ascrivergli un orientamento deflazionista. Se mai, l'unico punto di tangenza tra Frege e i deflazionisti può stare nell'indefinibilità della verità, cioè nella sua semplicità e irriducibilità. Tuttavia, dall'impostazione filosofica di Frege si può dedurre che per lui la verità sia sostanziale e per questo è meglio collocabile, a nostro avviso, nella schiera dei primitivisti più che dei deflazionisti.

Più vicino alla sensibilità deflazionista è la posizione ridondandista, secondo la quale, il predicato di verità è semplicemente ridondante, apportatore di nessun incremento informativo e pertanto dispensabile. In effetti, come vedremo a breve, i deflazionisti concordano sulla parte negativa e cioè sul fatto che la verità non abbia una natura ma non si spingono ad affermare che la nozione di verità sia, *tout court*, eliminabile dal linguaggio.

In ogni caso, non è nostro interesse ricostruire un panorama dettagliato delle varie correnti deflazioniste attraverso il loro sviluppo storico e teorico. Inoltre, il fatto che sia complesso cogliere in maniera precisa la posizione deflazionista, proprio perché a ben vedere non si tratta di una posizione ma di un ventaglio teorico, non implica che non si possa ragionare in astratto e indicare alcune tesi fondamentali circa le quali tutti gli studiosi di orientamento deflazionista dovrebbero concordare.

I tesi: la verità non ha una natura propria

La prima tesi è quella fondamentale e non è un caso che sia espressa come una negazione. La verità, secondo il deflazionismo, non è una proprietà reale che determinati enunciati (o proposizioni, credenze...) possiedono o non possiedono. Non c'è nulla - nel mondo - che corrisponda, in qualche senso da precisare, al predicato 'esser vero'. La I tesi qualifica le posizioni deflazioniste come antitetiche a tutte quegli approcci che cercano invece di indicare la natura della verità. Da questo punto di vista, il deflazionismo si pone a un altro livello rispetto alle varie teorie della verità che si sono avvicendate durante l'arco della storia della filosofia. Infatti, sia teorie realiste (corrispondentiste) sia teorie antirealiste (costruttiviste) sia teorie pragmatiste cercano di afferrare la natura autentica della verità, identificandola di volta in volta, nella 'corrispondenza' con i fatti, nella 'costruibilità' nel pensiero, nell'"utilità" rispetto a un determinato fine, nella 'coerenza' con un sistema concettuale di riferimento e così via. Tutte queste teorie vengono accomunate, in letteratura, dall'essere teorie *sostanzialiste*. Per esse, la verità ha una natura, un'essenza, e i nostri sforzi filosofici sono tesi a chiarirla. In opposizione alle teorie sostanzialiste (che tra loro, come si vede, non sono affatto in accordo), ci sono le teorie deflazioniste. La I tesi specifica negativamente il carattere di questa opposizione. La verità non ha una natura, è, per usare una felice espressione di Stewart Shapiro, metafisicamente sottile. Qualcuno potrebbe obiettare che il non avere una natura è, in un certo senso, possederne una. Si tratta, in realtà, di poco più che un sofisma perché l'intento deflazionista è molto chiaro su questo punto: dire che la verità non ha una natura non vuol dire che alla verità non possiamo coerentemente applicare una serie di predicati informativi ma che la nozione di verità non può essere spiegata (o anche ridotta) in base a tratti costitutivi del mondo o dell'accesso epistemico ad esso. Il deflazionismo, insomma, riduce le pretese conoscitive della filosofia, sottraendo la nozione di verità a una possibile spiegazione filosofica; e non è certo un caso che uno dei più agguerriti deflazionisti, Hartry Field, sia anche propugnatore del programma finzionalista in filosofia della matematica.

II tesi: il contenuto della nozione di verità è espresso dai bicondizionali

Dopo la caratterizzazione negativa della prima tesi, è necessario 'dire qualcosa' su come si possa intendere, deflazionisticamente, la verità. Una delle opinioni più diffuse è che il contenuto informativo della nozione di verità sia completamente caratterizzato dalla serie dei bicondizionali di Tarski cioè da tutte le istanze dello schema seguente:

(Dec) "p" è vero se e solo se p

La verità è quindi un mezzo di decitazione, o devirgolettatura. Permette, cioè di passare da un enunciato al nome di questo enunciato e viceversa. Il *locus classicus* di questa posizione è espresso da Quine:

> Quotation marks make all the difference between talking about words and talking about snow. The quotation is a name of a sentence that contains a name, namely 'snow', of snow. By calling the sentence true, we call snow white. The truth predicate is a device for disquotation. (Quine 1970)

La verità è allora un dispositivo di decitazione e tutto quello che c'è da sapere sulla verità è consegnato alle equivalenze di Tarski. In questo senso, le proprietà della verità sono di natura essenzialmente logico-linguistica; un parallelo che viene spesso evocato è tra la nozione di verità e quella di connettivo proposizionale. Esattamente come - di solito - non si ascrive nessuna rilevanza ontologica al connettivo "&" ma esclusivamente una funzione logica così, allo stesso modo, si comporta la verità. Nel prossimo capitolo vedremo alcune teorie decitazionali, ovvero sistemi assiomatici che rispondono all'intuizione decitazionale. È chiaro che tra le priorità del decitazionismo vi sarà anche la questione circa la natura dei bicondizionali: sono proposizioni necessarie o contingenti? Sono analitiche? Sono conoscibili a priori o a posteriori? Comprendere la nozione di verità coincide infatti con il comprendere il significato dei bicondizionali; ma i condizionali sono in numero infinito e quindi difficilmente trattabili da un soggetto epistemico con capacità limitate.

III tesi: la nozione di verità permette riferimenti indiretti e generalizzazioni.

Se, in ottica deflazionista, il concetto di verità non comporta alcun impegno metafisico (I tesi) e se il contenuto di questa nozione è veicolato dal T-schema (o, alternativamente, dalle istanze del T-schema) in base alla II tesi, rimane da spiegare perché in tutti i linguaggi naturali più importanti sia diffusa una parola il cui significato può essere rubricato come "essere vero". Del resto, se davvero avessero ragione i sostenitori di una posizione ridondantista, e quindi se il predicato di verità fosse eliminabile in ogni situazione linguistica, ci si potrebbe chiedere ragionevolmente perché sia sopravvissuto così tanto. La III tesi dichiara che le uniche funzioni espressive nelle quali il predicato di verità gioca un ruolo di fatto insostituibile sono quelle dei *riferimenti indiretti* e delle *generalizzazioni*; la *raison d'être* della verità è quindi il permettere la costruzione di simili strutture linguistiche necessarie nella composizione di una lingua.

Abbiamo detto che, per il deflazionista, l'enunciato «la neve è bianca» e l'enunciato «'la neve è bianca' è vero» possiedono il medesimo *valore epistemico*; pertanto, essi possono essere sostituiti senza alcuna modificazione del contesto generale. Ma non sempre è possibile questa manovra. Ipotizziamo di volerci riferire a un determinato enunciato che però non conosciamo: immaginiamo, per esempio, di aver avuto una conversazione ieri sera con il nostro amico Alfredo e di essere stati perfettamente d'accordo con lui circa l'opinione che abbiamo sulla nostra comune conoscente Maria. Tuttavia, non ci ricordiamo i dettagli delle nostre asserzioni (se, per esempio, stavamo parlando dei capelli di Maria o del suo carattere). Potremmo allora affermare qualcosa come:

(RI) Quello che ha detto Alfredo su Maria è vero

Qui non abbiamo a disposizione l'enunciato preciso cui ci vogliamo riferire, e quindi non possiamo applicare lo schema di decitazione ed eliminare il predicato di verità. Anzi, in questo caso, il predicato di verità svolge una funzione essenziale. Se non fosse a disposizione avremmo una frase grammaticalmente malformata. Pertanto, anche la più sobria posizione deflazionista, deve ammettere l'insostituibilità del predicato di verità nei contesti i cui il riferimento agli enunciati è opaco.

Ma non è questa la parte della III tesi più significativa, per i nostri scopi. La verità è infatti insostituibile anche in un'altra situazione linguistica che, filosoficamente parlando, è ancora più interessante. Se, per esempio, riponiamo una grande fiducia nelle competenze calcistiche del nostro amico Arturo, possiamo dire:

(gen) Tutto quello che dice Arturo sul calcio è vero

Questo enunciato è una generalizzazione e anche in questo caso non possiamo fare a meno del predicato di verità. La ragione è analoga al caso discusso precedentemente: in queste situazioni non abbiamo a disposizione tutti i nomi degli enunciati che vogliamo affermare. Infatti, potremmo provare a parafrasare (gen) come:

(gen*) Se Arturo dice 'p' sul calcio allora p
se Arturo dice 'q' sul calcio allora q
se Arturo dice 'r' sul calcio allora r
...

Tuttavia (gen*) non possiede, in generale, la medesima estensione di (gen). Infatti, nel secondo caso potremmo non sapere che cosa dice effettivamente Arturo e non avere, di conseguenza, i nomi (‘p’, ‘q’, ‘r’, ...) degli enunciati.

Le generalizzazioni sono filosoficamente significative perché normalmente descrivono stati di cose strutturali, come leggi o principi. Un tipico esempio di generalizzazione che ci impegnerà a lungo nelle pagine seguenti è:

(corr) Tutti i teoremi di T sono veri

(corr) esprime il fatto che la teoria **T** in questione è corretta: se un enunciato è ottenibile a partire dagli assiomi di **T** per mezzo dell’applicazione delle regole di deduzione presenti nel calcolo di T, allora diremo che si tratta di un enunciato vero.

In conclusione le tesi I-III definiscono, con qualche tollerabile imprecisione, le condizioni minimali che una posizione deflazionista sul concetto di verità deve soddisfare. Benché il deflazionismo non sia una posizione intuitiva - istintivamente pensiamo che la verità sia ‘qualcosa’ e anzi, spesso, qualcosa di importante che vale la pena conoscere - tuttavia, come avremo modo di vedere, non è affatto semplice confutarlo.

3. *Critiche al deflazionismo: l’argomento di conservatività*

Non indicheremo in questa sede le più importanti accuse mosse alla posizione deflazionista. Paul Horwich, nel suo famoso *Verità*, elenca circa quaranta critiche alla sua teoria e altrettante controbiezioni. In maniera del tutto generale, gli argomenti contro il deflazionismo possono essere suddivisi in due grandi famiglie che, parafrasando Carnap, chiameremo critiche *esterne* e critiche *interne*. Nel caso delle critiche esterne si mettono, normalmente, in discussione una o più tesi che caratterizzano la posizione deflazionista cercando poi di mostrare come tali assunzioni siano meno plausibili rispetto alle controparti di una concezione sostanzialista della verità. Si può affermare, per esempio, che una concezione deflazionista della verità non rende conto a sufficienza dell’importanza di questa nozione nell’attività esplicativa. È bene, nel formulare critiche esterne, non incorrere nell’errore metodologico di accusare il nostro avversario proprio della posizione che egli sostiene; così, non è particolarmente efficace attaccare il deflazionista sottolineando il fatto che la sua teoria della verità non tiene conto della profonda e radicata intuizione corrispondentista, secondo la quale la verità è in ultima analisi una relazione di corrispondenza fra linguaggio e mondo. Non è efficace, dicevamo, perché è proprio il nocciolo della posizione deflazionista negare recisamente ogni compromissione metafisica, e quindi anche l'intuizione corrispondentista. Come si può notare, le critiche esterne, oppongono interi sistemi concettuali differenti, in un confronto olistico, generale. In ciò che segue non ci interessano le critiche esterne al deflazionismo ma, al contrario, riteniamo essere più incisivo un argomento che, per la sua peculiarità, può appartenere alle critiche interne.

Secondo quest'ultimo approccio, si assumono tutte le premesse deflazioniste e si mostra come, in determinati casi, queste non consentono di trarre la conclusione filosofica desiderata. Nel nostro caso, quindi, partiremo dalle tesi prima esposte e faremo vedere che, alla fine dell'argomentazione, una delle tesi - segnatamente la più importante, la I - dovrà essere rifiutata. A differenza delle critiche esterne, quelle interne hanno il vantaggio di non opporre sistemi concettuali completamente differenti ma agiscono, se ci concediamo una metafora militare, come guastatori dietro le linee nemiche. Se una critica interna ha successo si può affermare, con Crispin Wright[2], che il deflazionismo è una posizione *internamente instabile*.

Una delle più significative critiche (interne) al deflazionismo è costituita dal cosiddetto *argomento di conservatività*. In breve, il deflazionista è costretto ad ammettere che alcune generalizzazioni importanti devono essere derivabili dalla sua teoria della verità. In particolare la generalizzazione che dichiara che tutti i teoremi della teoria di base (del linguaggio oggetto, in terminologia tarskiana) sono veri, cioè che la teoria è corretta. Ma questo enunciato non è affatto innocuo: la teoria della verità che riesce a derivarlo, può immediatamente derivare anche la consistenza della teoria di base e quindi cessa di essere conservativa. La presenza del predicato di verità e di assiomi specifici che ne regolano l'uso e la natura, diventa così causa di incremento informativo essenziale. Ma, conclude l'argomento, come è possibile che la verità sia una proprietà metafisicamente 'sottile', senza una natura specifica se, al contrario, gioca un ruolo così significativo nel contesto formale?

---

2 Cfr. Wright 1994, 1998.

L'argomento di conservatività, qui appena delineato, è un ottimo *case study* per la discussione filosofica e logica sulla nozione di verità: compendia, infatti, tematiche circa il significato e la portata dell'approccio assiomatico con la discussione sulla natura della verità presente nell'opposizione deflazionismo/sostanzialismo.

La seconda parte del volume è organizzata quindi nella maniera seguente: il prossimo (quinto) capitolo tratterà le teorie decitazionali, ovvero le teorie assiomatiche della verità che assumono come principi essenziali proprio i bi-condizionali di Tarski. Analizzeremo la portata logica di questi sistemi e la loro rilevanza filosofica per la discussione sul concetto di verità; in particolare sarà interessante soffermarci sulla questione delle generalizzazioni (intese come espressioni finitarie di congiunzioni infinite) che - come abbiamo già avuto modo di vedere - costituiscono un punto cruciale per la determinazione della natura della verità.

Il sesto capitolo prende in esame le teorie assiomatiche composizionali, quelle cioè più prossime all'intuizione tarskiana. In effetti, queste teorie altro non sono che la presentazione in forma assiomatica delle clausole tarskiane per la definizione di verità. Le teorie composizionali sono più potenti rispetto alle decitazionali; in particolare, la teoria della verità (composizionale, tarskiana) per l'aritmetica di Peano risulta non conservativa rispetto all'aritmetica ed è in grado di dimostrarne la consistenza. A questo punto si pongono due questioni di carattere logico: la prima è cercare di individuare con precisione la portata logica della teoria della verità. E a questo quesito risponde la dimostrazione in base alla quale faremo vedere che la teoria della verità per l'aritmetica di Peano può essere derivata a partire da un particolare sottosistema di aritmetica del secondo ordine (*ACA*). In secondo luogo, ci si può chiedere se esistono sotto-teorie interessanti della teoria della verità: anche in questo caso la risposta è affermativa. Prenderemo in esame un sistema di assiomi tarskiani che presenta, però, l'induzione aritmetica ristretta alle formule aritmetiche, non contenenti cioè il predicato T. Dimostreremo infine che questa teoria risulta conservativa rispetto alla sua base aritmetica. L'analisi tecnica di queste teorie metterà in luce alcuni aspetti filosofici interessanti riguardanti il nesso tra predicato di verità, concetto di insieme e natura del principio di induzione matematica.

Ma il 'succo' filosofico più prezioso è distillato nel settimo capitolo, dove presenteremo l'argomento di conservatività e sfrutteremo tutti i risultati formali precedentemente dimostrati. Discuteremo criticamente la portata e l'efficacia di questo argomento mostrandone le presupposizioni concettuali e gli esiti filosofici: in particolare, ci concentreremo su due critiche (ad opera di Neil Tennant e Hartry Field) che mettono in dubbio - muovendo da punti di vista molto differenti - la conclusione ottenuta.

Infine una breve conclusione cercherà di sintetizzare sia i risultati della prima parte che quelli della seconda fornendo un quadro il più possibile unitario sul problema logico e filosofico della verità. Data la natura tecnica e formale dei concetti in gioco sono necessarie alcune conoscenze di carattere logico-matematico; di seguito un breve riepilogo delle nozioni principali utilizzate poi nel resto del capitolo.

4. *Nozioni sintattiche preliminari*

In ciò che segue forniremo alcune nozioni formali necessarie per comprendere adeguatamente la discussione sugli aspetti logici e filosofici delle teorie assiomatiche della verità. Si tratta di aspetti noti per gli addetti ai lavori ma non scevri di qualche complicazione tecnica. La nostra presentazione si concentrerà su tre punti fondamentali: innanzitutto vedremo la struttura generale di una teoria formale della verità, composta, cioè, di una teoria di base e di una parte specifica; vedremo poi come sia un requisito irrinunciabile di una teoria assiomatica della verità la possibilità di riferirsi, cioè di nominare, gli oggetti dei quali di predica la verità e mostreremo come questa esigenza sia soddisfatta perfettamente dalla procedura di aritmetizzazione messa a punto da Gödel; infine, illustreremo una teoria 'ingenua' della verità, che ripropone, cioè, la chiusura semantica del linguaggio in cui è formulata e consente la derivazione dell'antinomia del mentitore. Questa sezione è ovviamente dispensabile per chi ha confidenza con le nozioni in gioco.

4.1. *Struttura di una teoria formale della verità*

Una teoria formale della verità[3] è una teoria assiomatica che ha come oggetto privilegiato il predicato T il cui significato intuitivo è "essere vero". La struttura di una teoria assiomatica della verità non è banale. Si

---

3 Cfr. Halbach 2011 per una puntuale e accurata ricostruzione delle più importanti teorie assiomatiche della verità. Seguiremo il testo di Halbach in molti punti della nostra trattazione.

possono infatti riconoscere tre componenti fondamentali: una *logica*, cioè il calcolo nel quale è espressa la teoria, una *teoria base* e una *teoria specifica*.

La logica utilizzata è nella stragrande maggioranza dei casi il calcolo dei predicati del primo ordine con identità; discuteremo brevemente in seguito il ricorso a logiche più potenti, segnatamente la logica del secondo ordine (con la semantica adeguata) e linguaggi che consentono l'espressione di formule infinitamente lunghe.

La verità - secondo un'assunzione piuttosto condivisa - si predica di determinati oggetti: possono essere enunciati, formule, proposizioni e così via. Nel contesto di una teoria formale, tali oggetti devono essere specificati in maniera del tutto rigorosa; se, per ipotesi, vogliamo ascrivere il predicato «essere vero» a una determinata classe di enunciati, la nostra teoria di base dovrà rendere conto della struttura degli enunciati, delle regole di formazione e così via. La teoria di base è a sua volta assiomatizzata e, ovviamente non include il predicato di verità; la scelta delle teorie di base è ampia. Tuttavia in letteratura ci si è orientati su alcune scelte precise. Tendenzialmente si assume una teoria matematica dalle proprietà metateoriche note, come per esempio, l'aritmetica ricorsiva primitiva (PRA), l'aritmetica di Peano al primo ordine (PA) o la teoria degli insiemi di Zermelo-Fränkel (ZF)[4]. In ciò che segue, salvo diversamente espresso, assumiamo proprio l'aritmetica di Peano come teoria di base per le teorie assiomatiche della verità; i suoi assiomi sono:

1. $\forall x\big(s(x) \neq 0\big)$
2. $\forall x \forall y\big(s(x) = s(y) \rightarrow x = y\big)$
3. $\forall x\big(x + 0 = x\big)$
4. $\forall x \forall y\big(x + s(y) = s\big(x + y\big)\big)$
5. $\forall x\big(x \cdot 0 = 0\big)$
6. $\forall x \forall y\big(x \cdot s(y) = x \cdot y + x\big)$
7. $\big[\alpha(0) \wedge \forall x\big(\alpha(x) \rightarrow \alpha(s(x))\big)\big] \rightarrow \forall x \alpha(x)$

7. merita un po' di discussione. Innanzitutto non è un assioma ma uno schema di assiomi, dal momento che contiene la variabile metateorica α; si tratta del principio di induzione matematica che afferma: data una pro-

4 Per una presentazione delle teorie assiomatiche e delle loro proprietà più importanti rimandiamo, tra i molti testi classici, a Hatcher 1968, Shoenfield 1967, Rogers 1971.

prietà α, questa vale per tutti i numeri se vengono soddisfatti due requisiti: la proprietà, se vale di un numero deve vale per il suo successore e deve valere per lo zero.

La potenza del principio di induzione dipende, naturalmente, dalla complessità logica della formula induttiva. Se $\alpha \in L(PA)$ la formula possiede una complessità aritmetica qualsiasi. Un caso più interessante si riscontra quando la formula induttiva presenta occorrenze del predicato T. Si parla, allora, di induzione *non ristretta,* e si indica la corrispondente teoria aritmetica di base con il simbolo PA^T. È importante sottolineare che PA^T non è ancora una teoria assiomatica della verità: mancano, infatti, i principi specifici riguardanti il predicato di verità.

La teoria specifica consiste in una serie (a volte anche infinita) di assiomi che regolano il significato del predicato T. È ovviamente questa la parte di teoria della verità propriamente detta e la presentazione di varie assiomatizzazioni del predicato di verità sarà il cuore della nostra trattazione.

4.2. *Aritmetizzazione*

Come dicevamo in precedenza, ogni teoria formale della verità deve essere in grado di riferirsi in maniera precisa e rigorosa agli oggetti (della teoria) di cui si predica la verità. In altri termini, si richiede una procedura per nominare i *truth-bearers* e, come abbiamo visto già Tarski era perfettamente consapevole del problema. La soluzione tarskiana, però, non è del tutto soddisfacente; molto più agile è il sistema messo a punto da Gödel (da cui il nome di gödelizzazione) in cui viene definita, in primo luogo, una funzione che associa ad ogni segno del linguaggio della teoria T un numero naturale e, in secondo luogo, vengono definite funzioni e predicati sintattici fondamentali che mimano, all'interno del linguaggio stesso, l'organizzazione sintattica del linguaggio. Vediamo allora i punti fondamentali della codifica numerica della sintassi.[5]

Se ε è un espressione del linguaggio della nostra teoria, indichiamo con $\ulcorner \varepsilon \urcorner$ il suo gödeliano. $\ulcorner \varepsilon \urcorner$ è quindi un numero naturale k che è il codice dell'espressione sintattica ε. In PA non abbiamo dei segni per indicare i numeri nella maniera usuale, cioè tramite cifre arabe. È possi-

---

5 Per una presentazione della procedura di aritmetizzazione, rimandiamo a Galvan 1992 e Murawski 2005.

bile, però, ovviare a questo inconveniente utilizzando in maniera iterata la funzione (di) successore. Il numero naturale n sarà quindi scritto in *PA* come $0^{\prime \ldots n\, volte \ldots \prime}$; per questioni di comodità possiamo abbreviare la scrittura precedente con $\overline{n}$ che indica, appunto, il numerale di n, ovvero il nome del numero n nel linguaggio formale di *PA*. Nel caso precedente, quindi, scriveremo $\overline{\ulcorner \varepsilon \urcorner}$ per indicare il nome del numero (cioè il numerale) che corrisponde all'espressione ε. Qualora non fosse esplicitato il contesto teorico nel quale ci troviamo, possiamo usare $\ulcorner \varepsilon \urcorner$ senza preoccuparci se vi siano le risorse linguistiche per esprimere il gödeliano[6].

È chiaro che la funzione di aritmetizzazione dovrà essere definita per i segni primitivi del nostro linguaggio. Nel caso di *PA*, una possibile aritmetizzazione è la seguente:

$\ulcorner \neg \urcorner = 3$	$\ulcorner \leftrightarrow \urcorner = 11$	$\ulcorner + \urcorner = 19$	$\ulcorner x_i \urcorner = 2i$
$\ulcorner \vee \urcorner = 5$	$\ulcorner \exists \urcorner = 13$	$\ulcorner \cdot \urcorner = 21$	
$\ulcorner \wedge \urcorner = 7$	$\ulcorner \forall \urcorner = 15$	$\ulcorner 0 \urcorner = 23$	
$\ulcorner \rightarrow \urcorner = 9$	$\ulcorner s \urcorner = 17$	$\ulcorner = \urcorner = 25$	

Alcune considerazioni. I simboli primitivi di *PA* hanno un gödeliano dispari mentre le variabili, possibilmente infinite, sono pari, il che garantisce l'impossibilità di sovrapposizioni. Le parentesi possono essere incluse o meno. Se, come nel nostro caso non sono presenti, è necessario riscrivere le formule di *PA* in modo tale che la leggibilità sia univoca, per esempio attraverso la riduzione a forme normali prenesse.

---

6 Quando si presenta un linguaggio formale, come per esempio, il calcolo dei predicati del primo ordine, si specificano delle condizioni per la formazione dei termini e delle formule del linguaggio in esame. Tale operazione, normalmente, si svolge all'interno di un metalinguaggio, semi-formale, in cui si esibiscono le clausole induttive. A rigore, infatti, bisognerebbe parlare di metateoria sintattica (e, rispettivamente, di metateoria semantica). Il procedimento dell'aritmetizzazione permette di immergere la metateoria sintattica all'interno della teoria o, se si preferisce, di inglobare il linguaggio oggetto nel metalinguaggio. Come è stato detto in precedenza ciò consente la chiusura 'semantica' del linguaggio in questione insieme alla possibilità di riproporre l'antinomia del mentitore; di qui il fondamentale risultato di indefinibilità della verità.

Una volta stabiliti i gödeliani primitivi bisogna mettere a punto un sistema che ci permetta di concatenare i codici numerici in modo da aritmetizzare espressioni di qualsiasi complessità sintattica. Prendiamo, ad esempio, $\ulcorner x = y \urcorner$. Il codice di x è 2 (assumendo che le variabili del nostro linguaggio siano numerate e che x sia in questo caso la prima variabile); il codice di y è 4 (perché, per assunzione, y è la seconda variabile); il codice di = è 25. Ora, il problema di unire questi numeri è che il risultato deve essere univoco. Il codice di $\ulcorner x = y \urcorner$ deve indicare sempre e solo l'espressione $x = y$ e non un'altra espressione sintattica. Come garantire tutto ciò?

Una delle soluzioni consiste nello sfruttare il teorema fondamentale dell'aritmetica in base al quale ogni numero naturale possiede una e una sola scomposizione in fattori primi la cui forma è $p_1^{k_1} \cdot p_2^{k_2} \cdot, ..., \cdot p_j^{k_i}$ dove $p_1, ..., p_j$ è la sequenza iniziale dei j numeri primi e $k_1, ..., k_i$ i relativi esponenti. Nel caso precedente, avremo allora che:

$$\ulcorner x = y \urcorner = p_1^{k_1} \cdot p_2^{k_2} \cdot p_3^{k_3} = 2^{\ulcorner = \urcorner} \cdot 3^{\ulcorner x \urcorner} \cdot 5^{\ulcorner y \urcorner} = 2^{25} \cdot 3^2 \cdot 5^4$$

Tramite la procedura di aritmetizzazione è possibile definire all'interno della teoria (nel nostro caso, *PA*) alcuni predicati che mimano le proprietà sintattiche di espressioni di *PA*. Per esempio, è possibile codificare il predicato «essere un termine di *PA*».

4.2.1. Termini

La definizione è induttiva e si compone, come al solito, di una base e di un passo.

Base. Nella base abbiamo due clausole: i) ogni variabile individuale è un termine; ii) lo zero è un termine. Quindi, per la prima clausola, avremo che x è il codice di una variabile se x è uguale a $2n$, con n diverso da 0.

$Var(x) \rightarrow Term(x)$, dove $Var(x) \leftrightarrow x = 2n, n \neq 0$

Per la seconda clausola, lo zero è un termine.

$x = \ulcorner 0 \urcorner \rightarrow Term(x)$

Passo. Le clausole del passo sono tre: il successore di un termine è un termine; l'addizione di due termini è un termine; la moltiplicazione di due termini è un termine. Siano t e r segni per termini qualsiasi. Abbiamo che:

$$x = \ulcorner s(t) \urcorner \vee x = \ulcorner t + r \urcorner \vee x = \ulcorner t \cdot r \urcorner \rightarrow Term(x)$$

Dove, ricordiamo, $\ulcorner t + r \urcorner = 2^{\ulcorner + \urcorner} \cdot 3^{\ulcorner t \urcorner} \cdot 5^{\ulcorner r \urcorner}$.

4.2.2. Formule

Base. In maniera del tutto analoga è possibile definire il predicato $FormAt(x)$ che vale di x se e solo se x è il codice di una formula atomica di PA.

$$FormAt(x) \leftrightarrow x = \ulcorner t = r \urcorner$$

Passo. Quindi, il passo della definizione induttiva di formula:

$$FormMol(x) \leftrightarrow$$
$$x = \ulcorner \neg\alpha \urcorner \vee x = \ulcorner \alpha \wedge \beta \urcorner \vee x = \ulcorner \alpha \vee \beta \urcorner \vee x = \ulcorner \alpha \rightarrow \beta \urcorner \vee x = \ulcorner \forall x_i \alpha \urcorner$$

Da cui, $Form(x) \leftrightarrow FormAt(x) \vee FormMol(x)$

4.2.3. Funzioni

Tramite la funzione di aritmetizzazione è possibile poi definire delle funzioni sintattiche che descrivono la struttura delle formule.

Per esempio, se x è il codice di una formula α, si può definire una funzione $neg(x) = y$ dove y è il codice della negazione di x. In maniera del tutto equivalente, possiamo avere:

1. $conj(x, y) = z$
dove z è il codice della congiunzione delle formule codificate da x e y.

2. $disj(x, y) = z$
dove z è il codice della disgiunzione delle formule codificate da x e y.

3. $all(v, x) = y$

dove y è il codice della quantificazione universale della formula codificata da x rispetto alla variabile v.

4. $ex(v,x) = y$
dove y è il codice della quantificazione esistenziale della formula codificata da x rispetto alla variabile v.

L'utilità di queste funzioni è apprezzabile mediante un semplice esempio. Prendiamo in esame la clausola tarskiana di verità che riguarda la congiunzione e il cui significato intuitivo è: una congiunzione è vera se e solo se lo sono anche i due congiunti. Possiamo esprimente il tutto nella maniera seguente:

Per α e β qualsiasi, $T\left(\ulcorner \alpha \wedge \beta \urcorner\right) \leftrightarrow T(\ulcorner \alpha \urcorner) \wedge T(\ulcorner \beta \urcorner)$

Tuttavia, questa formulazione non è del tutto formale perché la quantificazione universale, cioè il fatto che la relazione valga per qualunque formula, è compiuta a livello metateorico, indicando appunto l'arbitrarietà di α e β.

Se vogliamo invece portare a livello teorico la quantificazione dobbiamo servirci delle funzioni sintattiche aritmetizzate appena esposte. Sia $T(PA)$ la teoria della verità composizionale tarskiana per l'aritmetica di Peano, la clausola della congiunzione sarà:

$$T(PA) \vdash \forall x \forall y \forall z \left(\left(conj(x,y) = z \right) \to Tz \leftrightarrow \left(Tx \wedge Ty \right) \right)$$

Nel corso della trattazione useremo a seconda degli scopi e del contesto la versione formalizzata o quella semi-formale in modo da non compromettere né il rigore né la leggibilità. Talvolta può essere utile, poi, impiegare alcune abbreviazioni:
$\forall t$ significa in realtà: $\forall x\left(Term(t) \to ...\right)$
$\forall \alpha$ significa in realtà: $\forall x\left(Form(x) \to ...\right)$

4.2.4. Funzione val e operazione di sostituzione

Vi sono poi ancora due funzioni importanti che è bene introdurre in via preliminare. La prima è la funzione val che assegna a un termine il suo valore e che è definita nel modo seguente:

Base.
$val(\ulcorner 0 \urcorner) = 0$

Passo.
$val(\ulcorner s(t) \urcorner) = s\left(val(\ulcorner t \urcorner)\right)$
$val(\ulcorner t + r \urcorner) = val(\ulcorner t \urcorner) + val(\ulcorner r \urcorner)$
$val(\ulcorner t \cdot r \urcorner) = val(\ulcorner t \urcorner) \cdot val(\ulcorner r \urcorner)$

L'altra funzione significativa è quella di *sostituzione*. L'operazione sintattica di sostituzione coinvolge tre oggetti linguistici: la formula dove viene effettuata la sostituzione, la variabile (o il termine) che viene sostituito e la variabile (o il termine) che viene introdotto. Scriviamo quindi $so(\ulcorner \alpha \urcorner, \ulcorner x \urcorner, \ulcorner t \urcorner)$ per indicare la sostituzione nella formula α della variabile x con il termine t. In maniera del tutto equivalente ai casi precedenti, $so(x, y, z) = w$ indica che w è il codice della formula codificata da x in cui la variabile codificata da y è stata sostituita con il termine codificato da z[7].

Attraverso la funzione di sostituzione è possibile, poi, quantificare su variabili individuali che si trovano all'interno di gödeliani. Infatti, l'espressione $\ulcorner \alpha(x) \urcorner$ corrisponde a un numero naturale e non presenta l'occorrenza libera della x. Come fare per poter operare attraverso la quantificazione *all'interno* di un codice? Con la funzione di sostituzione. Pertanto, $so(\ulcorner \alpha(y) \urcorner, \ulcorner y \urcorner, num(x))$ rappresenta il gödeliano della sostituzione in $\ulcorner \alpha(y) \urcorner$ della variabile y con il codice dell'x-esimo numerale (ed è proprio questo il significato della funzione *num*). Convenzionalmente, $so(\ulcorner \alpha(y) \urcorner, \ulcorner y \urcorner, num(x))$ sta per $\ulcorner \alpha(\dot{x}) \urcorner$.[8] Ma quest'ultima espressione contiene una variabile libera (la x) e quindi può essere quantificata. $\exists x\left(\ulcorner \alpha(\dot{x}) \urcorner\right)$ indica proprio che esiste un numerale - sia per esempio k - che, sostituito al posto di x, rende vera la formula α.

Altre funzioni di fondamentale importanza per le teorie assiomatiche della verità (come il predicato di dimostrabilità Pr_T) verranno introdotte nel seguito della trattazione.

---

7 Per una trattazione completa del predicato di sostituzione, vedi Galvan 1992, 136.

8 Sull'utilizzo delle variabili puntate, vedi Smorinsky 1977.

5. *Teoria naïve della verità*

Chiudiamo questa sezione introduttiva presentando una semplicissima teoria assiomatica della verità. Come si ricorderà [dal primo capitolo], una delle due condizioni che Tarski impone alla teoria della verità soddisfacente è l'adeguatezza materiale. Una teoria deve, cioè, essere in grado di derivare tutti i T-enunciati, ovvero le istanze del T-schema. Nel sistema tarskiano, per T-schema si intende:

(T) x è vero se e solo se p

dove x è il nome (metalinguistico) di un enunciato del linguaggio oggetto e p è la traduzione metalinguistica dell'enunciato del linguaggio oggetto. Ma, grazie alla procedura di aritmetizzazione, il metalinguaggio contiene, come parte propria, il linguaggio oggetto. Per cui, possiamo formalizzare il T-schema nella seguente maniera:

(T) $T\left(\overline{\ulcorner\alpha\urcorner}\right) \leftrightarrow \alpha$

dove $\overline{\ulcorner\alpha\urcorner}$ è il nome dell'enunciato α e T è il predicato di 'essere vero'. Una teoria assiomatica della verità S sarà, quindi, materialmente adeguata se, per ogni α

$$S \vdash T\left(\overline{\ulcorner\alpha\urcorner}\right) \leftrightarrow \alpha$$

ovvero se S è in grado di derivare tutti i T-bicondizionali. Naturalmente, ed è questo il caso delle teorie decitazionali come vedremo tra poco, la teoria può essere costituita *proprio* dalla serie dei bicondizionali. Così facendo, il criterio di adeguatezza materiale è immediatamente soddisfatto.

Se non poniamo nessuna limitazione su α otteniamo fenomeni di re-iterazione del predicato di verità; α, infatti, potrebbe contenere a sua volta il predicato T come nel caso seguente:

$$S \vdash T\left(\overline{\ulcorner T\left(\overline{\ulcorner\beta\urcorner}\right)\urcorner}\right) \leftrightarrow T\left(\overline{\ulcorner\beta\urcorner}\right)$$

È vero che è vero β se e solo se è vero β. Si tratta di teorie della verità *non tipate,* dove è ammessa, cioè, la re-iterazione del predicato di verità.

Tarski, come abbiamo visto, non prende in considerazione le teorie non tipate.

Una teoria della verità S in grado di rappresentare la sua sintassi e senza limitazioni circa la complessità delle formule nel T-schema è inconsistente. È possibile, infatti, riprodurre al suo interno il paradosso del mentitore. S permette la derivazione del lemma di diagonalizzazione:

(DL) Sia $\alpha(x)$ una formula appartenente al linguaggio di S. Allora esiste una formula β tale che:

$$S \vdash \beta \leftrightarrow \alpha(\ulcorner \beta \urcorner)$$

Sia $\alpha(x) \equiv \neg Tx$; avremo, pertanto, $S \vdash \lambda \leftrightarrow \neg T(\ulcorner \lambda \urcorner)$. Come è facile vedere, l'ultima espressione è la formalizzazione dell'enunciato del mentitore: λ dice di sé stessa che non è vera. Ma, poiché, S deriva le T-istanze, abbiamo che $S \vdash T(\ulcorner \lambda \urcorner) \leftrightarrow \lambda$; combinando i due risultati: $S \vdash T(\ulcorner \lambda \urcorner) \leftrightarrow \neg T(\ulcorner \lambda \urcorner)$. Contraddizione.

V
TEORIE DECITAZIONALI

1. *Origini tarskiane*

Come abbiamo detto nel capitolo precedente, il lavoro di Tarski costituisce il punto di inizio per la tradizione semantica nell'analisi della verità; tuttavia, nelle pagine conclusive, Tarski menziona la possibilità di un approccio differente nella trattazione logica della nozione di verità. Scrive:

> If the class of all provable sentences of the metatheory is consistent and if we add to the metatheory the symbol 'Tr' as a new primitive sign, and all theorems which are described in conditions (A) and (B) of the convention T as new axioms, then the class of provable sentences in the metatheory enlarged in this way will also be consistent[1].

La proposta di Tarski è quindi quella di introdurre il simbolo per il predicato di verità T (Tr nella grafia tarskiana) come primitivo e tutti i T-bicondizionali (ovvero le istanze della convenzione T, cui fanno riferimento le condizioni indicate come (A) e (B)) come assiomi. Il teorema afferma che la teoria *S* insieme ai bicondizionali è consistente se la teoria *S* di partenza è consistente. Si tratta di un risultato di conservatività e per comprenderlo appieno è bene introdurre la nozione di *estensione conservativa* di una teoria.

1.1 *Estensioni conservative*

Siano *S* e *S*' due teorie assiomatizzate. Diremo che *S*' è un'estensione conservativa di *S* se e solo se per ogni enunciato α appartenente al linguaggio di *S*, se α è un teorema di *S*', α è un teorema di *S*. In simboli:

---

1 Tarski 1983, 256.

S’ estensione conservativa S $\Leftrightarrow \forall \alpha \in L(S), S' \vdash \alpha \Rightarrow S \vdash \alpha$

L’idea di fondo è che un’estensione conservativa di una teoria non apporta nessun contenuto informativo in più. Se infatti la teoria ‘nuova’, cioè *S*’, riesce a dimostrare teoremi che si possono ottenere già a partire dalla teoria di partenza, cioè *S*, risulta che *S*’ non fa progredire nella conoscenza di nuovi fatti descrivibili attraverso il linguaggio della teoria *S*. È ovviamente fondamentale la condizione per la quale α appartenga al linguaggio della teoria di base; se così non fosse ogni teoria che possieda più segni primitivi di un’altra diventerebbe *ipso facto* un’estensione non conservativa dal momento che la prima teoria non ha, per definizione, le risorse linguistiche per derivare tutti gli enunciati della sua estensione.

Tarski, tuttavia, non parla di estensioni conservative ma di consistenza. In generale vale che la conservatività implica la consistenza mentre non vale il converso. Vi possono essere estensioni consistenti ma non conservative. Mostriamo che se *S* è una estensione conservativa di *T* allora se *T* è consistente lo sarà anche *S*.

Se *S* è un’estensione conservativa di *T*, allora $\forall \alpha \in L(T)$ abbiamo che $S \vdash \alpha \Rightarrow T \vdash \alpha$. Ma poiché per ipotesi *T* è consistente avremo che $T \nvdash \bot$ e quindi $S \nvdash \bot$ e cioè *S* è consistente.

Una tecnica per dimostrare che *T* è un’estensione conservativa di *S* è quella di far vedere che ogni modello di *S* può essere espanso a un modello di *T*. Infatti:

Sia $\forall \alpha (T \vdash \alpha \Rightarrow S \vdash \alpha)$ il nostro demonstrandum. Per teorema di correttezza abbiamo che $\forall \alpha (T \Vdash \alpha \Rightarrow S \Vdash \alpha)$. Assumiamo l’antecedente e per definizione di conseguenza logica abbiamo $\mathcal{M} \vDash T \Rightarrow \mathcal{M} \vDash \alpha$. In base alla nostra ipotesi, cioè la possibilità di espandere ogni modello di *S* in un modello di *T*, possiamo scrivere: $\forall \mathcal{M} (\mathcal{M} \vDash S \Rightarrow \mathcal{M} \vDash T)$. Tramite la regola di eliminazione del quantificatore e per concatenazione si ottiene $\forall \mathcal{M} (\mathcal{M} \vDash S \Rightarrow \mathcal{M} \vDash \alpha)$ ovvero $S \Vdash \alpha$ e, data la completezza della logica, $S \vdash \alpha$. Naturalmente questo tipo di argomento può essere svolto solo nella logica del primo ordine a causa dell’utilizzo rilevante del teorema di completezza.

Nonostante la teoria assiomatica della verità proposta da Tarski non ponga problemi circa la sua consistenza, la sua ammissibilità nel progetto generale tarskiano è quantomeno dubbia. Due sono infatti le fonti di forte perplessità che l'autore esplicita nei confronti della teoria in questione e si tratta di aspetti interconnessi. Il primo riguarda la *debolezza deduttiva* di una teoria simile, il secondo l'aspetto *arbitrario*, *contingente* e 'accidentale' degli assiomi. Nel seguito cercheremo di capire se e in che misura i limiti intravisti da Tarski valgono per le teorie assiomatiche della verità più studiate e se effettivamente questi possono giocare un ruolo nella discussione filosofica sulle teorie assiomatiche della verità.

2. *Teoria minimale della verità*

Prima di procedere con lo studio sistematico delle teorie decitazionali, presentiamo una teoria, *MT* (*Minimal Truth*), che è interessante perché costituisce - per così dire - il limite inferiore del discorso assiomatico circa la verità. Paul Horwich chiama minimalismo la sua versione di deflazionismo (cfr. Horwich 1990). Non sappiamo se *MT* sia in grado di codificare alcune sue intuizioni; di certo, Horwich non prevede un T-schema ristretto ed è costretto ad ammettere restrizioni 'pragmatiche' circa i bicondizionali ammissibili senza contraddizione. Un'altra questione riguarda poi l'interpretazione intesa di *MT*; la teoria di Horwich infatti assume le proposizioni - e non gli enunciati - come portatori di verità. In ogni caso, è bene notare che pur nella sua 'trasparenza' *MT* presenta già dei fenomeni di non conservatività, come vedremo a breve. La teoria minimale della verità non è una teoria matematica, dal momento che i bicondizionali vengono aggiunti alla logica pura e non a una teoria di base come *PA*. Il linguaggio $\ell(MT)$ della teoria minimale della verità è, pertanto, costituito da:

- Il linguaggio di base, $\ell(C)$, e cioè la logica dei predicati del primo ordine con identità
- Un operatore proposizionale monadico: *T*
- Un operatore di formazione per termini: $< ... >$

Le regole di formazione di $\ell(MT)$ sono:

Termini.
a) Tutti i termini di $\ell(C)$ sono termini di $\ell(MT)$.

b) Se α è una formula chiusa di $\ell(C)$, allora $<\alpha>$ è un termine di $\ell(MT)$.
Formule.
a) Tutte le formule di $\ell(C)$ sono formule di $\ell(MT)$
b) Se $<\alpha>$ è un termine, allora $T(<\alpha>)$ è una formula di $\ell(MT)$.

Il significato intuitivo di *T* è, ovviamente, quello del predicato di verità. $<...>$ è una funzione che fa corrispondere a ogni enunciato del linguaggio di base (cioè di $\ell(C)$) il nome dell'enunciato in questione.

Assiomi.
Gli assiomi specifici di *MT* sono tutte le istanze del T-schema:

$T(<\alpha>) \leftrightarrow \alpha$, per ogni $\alpha \in L(C)$

Teorema 1. *MT* non è conservativa rispetto alla logica pura.

Dimostrazione.

$MT \vdash T(<\alpha>) \leftrightarrow \alpha$	T-schema
$MT \vdash \neg T(<\alpha>) \leftrightarrow \neg\alpha$	logica
$MT \vdash T(<\neg\alpha>) \leftrightarrow \neg\alpha$	T-schema
$MT \vdash T(<\neg\alpha>) \leftrightarrow \neg T(<\alpha>)$	logica
$MT \vdash <\alpha> = <\neg\alpha>$	Ipotesi per assurdo
$MT, <\alpha> = <\neg\alpha> \vdash T(<\alpha>) \leftrightarrow \neg T(<\alpha>)$	logica
$MT \vdash <\alpha> \neq <\neg\alpha>$	logica
$MT \vdash \exists x \exists y (x \neq y)$	logica

La teoria *MT* risulta quindi non conservativa rispetto alla logica pura. Ciò significa che anche un dispositivo teorico molto debole (quale quello di *MT*) permette di derivare l'esistenza di due oggetti differenti, ampliando quindi il potere informativo della teoria di base, ovvero della logica dei predicati del primo ordine.

Teorema 2. *MT* è conservativa su $T \Leftrightarrow T \vdash \exists x \exists y (x \neq y)$.

Dimostrazione.
Ad $\Rightarrow$) . Assumiamo che *MT* sia conservativa su T. Quindi, per definizione di conservatività,

$MT \vdash \exists x \exists y(x \neq y) \Rightarrow T \vdash \exists x \exists y(x \neq y)$. Ma, come abbiamo visto dal teorema 1, $MT \vdash \exists x \exists y(x \neq y)$. Da cui $T \vdash \exists x \exists y(x \neq y)$. ■

Ad $\Leftarrow$). Assumiamo che $T \vdash \exists x \exists y(x \neq y)$. Per mostrare la conservatività di *MT* su T basta far vedere che ogni modello di T può essere espanso a un modello di *MT*. Sia $\mathcal{M} \vDash T$ e quindi $\mathcal{M} \vDash \exists x \exists y(x \neq y)$. Il dominio di $\mathcal{M}$ deve contenere (almeno) due elementi, *a* e *b*. Definiamo un'espansione di $\mathcal{M}$, $\mathcal{M}^*$ come segue:

I. $\Im^*(T) = \{a\}$
II. $\alpha \in L(T) \wedge \mathcal{M} \vDash \alpha \Rightarrow \Im^*(<\alpha>) = a$
III. $\alpha \in L(T) \wedge \mathcal{M} \nvDash \alpha \Rightarrow \Im^*(<\alpha>) = b$

Il dominio di $\mathcal{M}^*$ è lo stesso di $\mathcal{M}$.

Le condizioni sulla struttura del modello sono chiare: I) stabilisce che l'estensione del predicato T è l'insieme che ha per elemento a. II) e III) stabiliscono l'interpretazione dei nomi degli enunciati. In breve, ogni enunciato vero è mappato su *a* mentre ogni enunciato falso su *b*. Vogliamo ora far vedere che valgono anche le inverse delle implicazioni in II e III.

Ad II) Assumiamo $\Im^*(<\alpha>) = a$ e $\mathcal{M} \nvDash \alpha$ come ipotesi per assurdo. Se vale $\mathcal{M} \nvDash \alpha$ per III, avremo che $\Im^*(<\alpha>) = b$; ma allora, per identità, $a = b$ contro la definizione di $\mathcal{M}^*$. ■

Ad III) Assumiamo $\Im^*(<\alpha>) = b$ e $\mathcal{M} \vDash \alpha$ come ipotesi per assurdo. Anche in questo caso, per II, avremo che $\Im^*(<\alpha>) = a$ e $a = b$ contro la definizione di $\mathcal{M}^*$. ■

Per cui abbiamo che:
$\mathcal{M} \vDash \alpha \Leftrightarrow \Im^*(<\alpha>) = a$
$\mathcal{M} \nvDash \alpha \Leftrightarrow \Im^*(<\alpha>) = b$
ma, per definizione di $\mathcal{M}^*$ (cioè l'espansione di $\mathcal{M}$) abbiamo che $\mathcal{M}^* \vDash \alpha \Leftrightarrow \Im^*(<\alpha>) = a$

Ora,

$\mathcal{M}^* \vDash T(<\alpha>) \Leftrightarrow \Im^*(<\alpha>) \in \Im^*(T)$	def. di $\vDash$
$\mathcal{M}^* \vDash T(<\alpha>) \Leftrightarrow \Im^*(<\alpha>) \in \{a\}$	def. di $\Im^*(T)$
$\mathcal{M}^* \vDash T(<\alpha>) \Leftrightarrow \Im^*(<\alpha>) = a$	teoria degli insiemi
$\mathcal{M}^* \vDash T(<\alpha>) \Leftrightarrow \mathcal{M}^* \vDash \alpha$	logica

$\mathcal{M}^* \vDash T(<\alpha>) \leftrightarrow \alpha$ def. di $\vDash$

$\mathcal{M}^*$ rende veri i T-enunciati, quindi $\mathcal{M}^* \vDash MT$.■

3. *Teorie decitazionali classiche*

L'idea di fondo del decitazionalismo è, come abbiamo visto nel capitolo precedente, esplicitare la II tesi deflazionista secondo la quale il contenuto della nozione di verità è esaurito completamente nell'elenco dei bicondizionali di Tarski. Le teorie assiomatiche decitazionali traducono formalmente questa intuizione e consistono, solitamente, di una serie (infinita) di assiomi che corrispondono ai bicondizionali. In generale, quindi, le teorie decitazionali non sono teorie finitamente assiomatizzabili: del resto, per scrivere gli assiomi di tali teorie si ricorre a uno schema di assiomi, cioè lo schema T, in maniera analoga a quanto si fa con l'assioma di induzione nel caso dell'aritmetica di Peano del primo ordine. Le teorie decitazionali sono banalmente materialmente adeguate, in senso tarskiano. Come si ricorderà, infatti, il criterio di adeguatezza materiale è dato dalla possibilità di ottenere tutti i bicondizionali; la derivazione di questi è, nel caso di queste teorie, immediata dal momento che essi stessi costituiscono la parte specifica della teoria della verità. Le teorie che prendiamo in considerazione sono *TB* e *UTB*[2]. In questo caso, la teoria di base è costituita dall'aritmetica di Peano e per questo motivo i nomi degli enunciati appartengono al linguaggio stesso della teoria di base, in virtù della procedura di aritmetizzazione.

La teoria *TB* è costituita da $PA \cup T(\overline{\ulcorner \alpha \urcorner}) \leftrightarrow \alpha$ per ogni $\alpha \in L(PA)$. Poiché il linguaggio di *TB* include il predicato di verità *T*, il principio di induzione di *PA* non è ristretto. Ciò significa che è possibile definire proprietà induttive di complessità aritmetica qualsiasi, che contengano, anche, il predicato di verità. La teoria *TB* non è, ovviamente, finitamente assiomatizzabile non solo per la presenza nella teoria di base di uno schema di assiomi ma anche per il fatto che i suoi assiomi specifici sono infiniti essendo tutte le istanze del T-schema.

---

2 Ci rifacciamo alla denominazione di Halbach, *TB* sta per *Tarskian Biconditionals* e *UTB* per *Uniform Tarskian Biconditionals*. Talvolta si trova anche l'indicazione *DT* per *Disquotational Theory*.

Teorema 3. La teoria *TB* è consistente.

Ciò significa che la teoria *TB* possiede un modello e, per di più, un modello basato sulla struttura standard dei numeri naturali. Sia $\mathcal{M} = \langle \mathbb{N}, \mathcal{E} \rangle$ un modello dove $\mathbb{N}$ è il supporto per l'interpretazione della parte aritmetica della teoria mentre $\mathcal{E}$ è l'insieme dei codici degli enunciati che servono come interpretazione del predicato di verità. $\mathcal{E}$, in altri termini, è l'estensione del predicato di verità:

$$\mathcal{M} = \left\langle \mathbb{N}, \left\{ \ulcorner \alpha \urcorner \middle| \alpha \in L(PA) \wedge \mathbb{N} \vDash \alpha \right\} \right\rangle$$

Dimostriamo quindi che, per un α qualsiasi, $\mathcal{M} \vDash T\left(\overline{\ulcorner \alpha \urcorner}\right) \leftrightarrow \alpha$.

Dimostrazione. Partiamo dall'assunzione e cioè $\mathcal{M} \vDash T(\overline{\ulcorner \alpha \urcorner})$. Avremo che $\mathcal{M} \vDash T(\overline{\ulcorner \alpha \urcorner}) \Leftrightarrow \Im_{\mathbb{N}}\left(\ulcorner \alpha \urcorner\right) \in \Im_{\mathbb{N}}\left(T\right)$; in base alla definizione di $\mathcal{M}$, $\mathcal{M} \vDash T(\overline{\ulcorner \alpha \urcorner}) \Leftrightarrow \alpha \in \mathcal{E}$. D'altro canto, vale che $\alpha \in \mathcal{E} \Leftrightarrow \mathbb{N} \vDash \alpha$ per definizione. Quindi $\mathcal{M} \vDash T(\overline{\ulcorner \alpha \urcorner}) \Leftrightarrow \mathbb{N} \vDash \alpha$. Poiché $\alpha \in L(PA)$, cioè non contiene occorrenze del predicato di verità, abbiamo che $\mathbb{N} \vDash \alpha \Leftrightarrow \mathcal{M} \vDash \alpha$. Ma allora $\mathcal{M} \vDash T(\overline{\ulcorner \alpha \urcorner}) \Leftrightarrow \mathcal{M} \vDash \alpha$ e quindi, $\mathcal{M} \vDash T(\overline{\ulcorner \alpha \urcorner}) \leftrightarrow \alpha$. ■

Nella teoria *TB* si prendono in considerazione solo formule chiuse, ovvero enunciati. Se si ammettono anche formule aperte, si passa alla teoria *UTB*. L'idea di fondo è che affermare che $\alpha(x)$ è vera significa che, dato il termine chiuso *t*, l'enunciato $\alpha(t)$ è vero se e solo se $\alpha(t)$. In simboli:

(U1) $\forall t\left(T\left(so\left(\ulcorner \alpha \urcorner, \ulcorner x \urcorner, \ulcorner t \urcorner\right)\right) \leftrightarrow \alpha\left(val\left(\ulcorner t \urcorner\right)\right)\right)$

Si noti la differenza tra questo assioma e il seguente:

(U1*) $\forall x\left(T\left(\ulcorner \alpha(\dot{x}) \urcorner\right) \leftrightarrow \alpha(x)\right)$

In quest'ultimo caso, $\dot{x}$ varia sui numerali come si evince facilmente richiamandone la definizione estesa: $so\left(\ulcorner \alpha(y) \urcorner, \ulcorner y \urcorner, num(x)\right)$. Nel caso precedente, invece, il *range* del quantificatore è su tutti i termini, e quindi anche su termini che contengono altri simboli oltre allo zero e al segno di successione. In (U1), tuttavia, si fa riferimento alla sostituzione di una sola variabile per un solo termine. È possibile generalizzare a una stringa, finita, di termini a patto di estendere la funzione *so*. Abbiamo infatti che

$so(\ulcorner\alpha\urcorner,\ulcorner x_1\urcorner,\ulcorner t_1\urcorner)$ può essere estesa a una sequenza finita di x_n (e, rispettivamente, di t_n) in questo modo:

$$so(\ulcorner\alpha\urcorner,\ulcorner x_1\urcorner,...,\ulcorner x_n\urcorner,\ulcorner t_1\urcorner,...,\ulcorner t_n\urcorner)$$

Complicando un poco l'espressione grafica, avremo così:

$$\text{(U2)}\ \forall t_1,...,t_n \begin{pmatrix} T\left(so(\ulcorner\alpha\urcorner,\ulcorner x_1\urcorner,...,\ulcorner x_n\urcorner,\ulcorner t_1\urcorner,...,\ulcorner t_n\urcorner)\right) \leftrightarrow \\ \alpha\left(val(\ulcorner t_1\urcorner),...,val(\ulcorner t_n\urcorner)\right) \end{pmatrix}$$

Il che si può semplificare come:

$$\text{(U3)}\ \forall t_1,...,t_n\left(T\left(\ulcorner\alpha(\dot{t}_1,...,\dot{t}_n)\urcorner\right) \leftrightarrow \alpha\left(val(\ulcorner t_1\urcorner),...,val(\ulcorner t_n\urcorner)\right)\right)$$

dove $\dot{t}$ sta per l'esito della sostituzione all'interno del codice di $\alpha(x)$ della variabile x con il termine chiuso t.

Teorema 4. Le teorie decitazionali *TB* e *UTB* sono conservative rispetto a *PA*.

Dimostrazione. La dimostrazione[3] è strutturalmente simile per *TB* e per *UTB*; tuttavia, per una maggiore leggibilità, dimostriamo il risultato per *TB*, senza perdita di generalità. L'idea è piuttosto semplice; bisogna mostrare come trasformare una *TB*-dimostrazione in una *PA*-dimostrazione di una formula aritmetica. Ogni dimostrazione è, per il teorema di finitezza, un oggetto sintattico finito. Ne segue che le T-istanze impiegate sono sempre in numero finito n. Quindi in ogni dimostrazione di *TB* possiamo avere: o assiomi di *PA*, o istanze del principio di induzione o, per qualche $i \leq n$ un'istanza del T-schema come:

$$T\left(\overline{\ulcorner\alpha_i\urcorner}\right) \leftrightarrow \alpha_i$$

È necessario allora definire un predicato di verità parziale in cui non compaia il simbolo T.

Sia $\tau(x) =_{def.} (x = \ulcorner\alpha_1\urcorner \wedge \alpha_1) \vee ... \vee (x = \ulcorner\alpha_n\urcorner \wedge \alpha_n)$. L'interpretazione intuitiva di $\tau(x)$ può essere la seguente: x è vero* se e solo se x = «la

3 Vedi Halbach (1999), (2011).

neve è bianca» e la neve è bianca o x = «l'erba è verde» e l'erba è verde o ... e così via. Quello che rimane da fare, a questo punto, è mostrare che, per un i generico, vale:

$$PA \vdash \tau\left(\overline{\ulcorner \alpha_i \urcorner}\right) \leftrightarrow \alpha_i$$

Ad $\rightarrow$). Assumiamo $PA \vdash \tau(\overline{\ulcorner \alpha_i \urcorner})$. Per definizione di $\tau(x)$, si ha che $PA \vdash \tau(\overline{\ulcorner \alpha_i \urcorner}) \leftrightarrow (\overline{\ulcorner \alpha_i \urcorner} = \overline{\ulcorner \alpha_1 \urcorner} \wedge \alpha_1) \vee ... \vee (\overline{\ulcorner \alpha_i \urcorner} = \overline{\ulcorner \alpha_n \urcorner} \wedge \alpha_n)$; quindi $PA, \tau(\overline{\ulcorner \alpha_i \urcorner}) \vdash (\overline{\ulcorner \alpha_i \urcorner} = \overline{\ulcorner \alpha_1 \urcorner} \wedge \alpha_1) \vee ... \vee (\overline{\ulcorner \alpha_i \urcorner} = \overline{\ulcorner \alpha_n \urcorner} \wedge \alpha_n)$. Se $\alpha_i = \alpha_1$, $PA, \tau(\overline{\ulcorner \alpha_1 \urcorner}) \vdash (\overline{\ulcorner \alpha_1 \urcorner} = \overline{\ulcorner \alpha_1 \urcorner} \wedge \alpha_1) \vee ... \vee (\overline{\ulcorner \alpha_1 \urcorner} = \overline{\ulcorner \alpha_n \urcorner} \wedge \alpha_n)$.

Il prossimo passo è quello di eliminare, uno ad uno, i disgiunti. Come fare? Ora, se α_1 e α_2 sono due oggetti differenti, $PA \vdash \overline{\ulcorner \alpha_1 \urcorner} \neq \overline{\ulcorner \alpha_2 \urcorner}$; in maniera analoga, per ogni $j \neq 1$, $PA \vdash \overline{\ulcorner \alpha_1 \urcorner} \neq \overline{\ulcorner \alpha_j \urcorner}$ e quindi $PA \vdash \neg\left(\overline{\ulcorner \alpha_1 \urcorner} = \overline{\ulcorner \alpha_j \urcorner} \wedge \alpha_j\right)$. Ora, j è un indice qualsiasi e poniamo che $j = n$. Avremo allora che $PA \vdash \neg\left(\overline{\ulcorner \alpha_1 \urcorner} = \overline{\ulcorner \alpha_n \urcorner} \wedge \alpha_n\right)$.

Ma così abbiamo eliminato l'ultimo disgiunto e otteniamo, pertanto, che $PA, \tau(\overline{\ulcorner \alpha_1 \urcorner}) \vdash (\overline{\ulcorner \alpha_1 \urcorner} = \overline{\ulcorner \alpha_1 \urcorner} \wedge \alpha_1) \vee ... \vee (\overline{\ulcorner \alpha_1 \urcorner} = \overline{\ulcorner \alpha_{n-1} \urcorner} \wedge \alpha_{n-1})$. Possiamo ripetere il procedimento n-1 volte e arrivare a $PA, \tau(\overline{\ulcorner \alpha_1 \urcorner}) \vdash (\overline{\ulcorner \alpha_1 \urcorner} = \overline{\ulcorner \alpha_1 \urcorner} \wedge \alpha_1)$ da cui segue banalmente $PA \vdash \tau(\overline{\ulcorner \alpha_1 \urcorner}) \rightarrow \alpha_1$. Ma come si ricorderà, la nostra assunzione generale era $\alpha_i = \alpha_1$; ora, lo stesso ragionamento può essere compiuto per qualunque α ottenendo così, $PA \vdash \tau(\overline{\ulcorner \alpha_i \urcorner}) \rightarrow \alpha_i$. ■

Ad $\leftarrow$). Assumiamo $PA, \alpha_1 \vdash \alpha_1$. Per logica, abbiamo che $PA, \alpha_1 \vdash \overline{\ulcorner \alpha_1 \urcorner} = \overline{\ulcorner \alpha_1 \urcorner} \wedge \alpha_1$ e per introduzione della disgiunzione, $PA, \alpha_1 \vdash (\overline{\ulcorner \alpha_1 \urcorner} = \overline{\ulcorner \alpha_1 \urcorner} \wedge \alpha_1) \vee ... \vee (\overline{\ulcorner \alpha_1 \urcorner} = \overline{\ulcorner \alpha_n \urcorner} \wedge \alpha_n)$. Ma, per definizione di $\tau(x)$, $PA, \alpha_1 \vdash \tau(\overline{\ulcorner \alpha_1 \urcorner})$ e quindi $PA \vdash \alpha_1 \rightarrow \tau(\overline{\ulcorner \alpha_1 \urcorner})$. Anche in questo caso, il ragionamento è ripetibile per tutte le α, per cui $PA \vdash \alpha_i \rightarrow \tau(\overline{\ulcorner \alpha_i \urcorner})$ ■

3.1. Commento

I risultati di conservatività sembrano incontrare particolarmente bene i gusti dei deflazionisti. Una teoria conservativa non aggiunge nulla di nuovo alla teoria di base e, nel nostro caso, la nozione di verità non comporta, in forza della conservatività, nessun incremento informativo. Certo, qualsiasi teoria della verità risulta non conservativa sulla logica pura, dimostrando l'esistenza di almeno due oggetti distinti; tutto sommato questo non dovrebbe essere così problematico per il deflazionista: in fondo anche il calcolo dei predicati del primo ordine comporta un impegno ontologico minimale (e cioè l'esistenza di almeno un oggetto nel dominio) e che una teoria della verità discrimini tra *due* oggetti è del tutto coerente con il fatto che, anche in ottica deflazionista, si riesca a distinguere tra il *vero* e il *falso*.

Quali sono dunque i problemi con le teorie decitazionali? Tarski ne individua due connessi, rispettivamente, alla debolezza deduttiva e all'arbitrarietà della scelta iniziale degli assiomi. Ci occuperemo prima (e per lo più) di quest'ultimo perché riguarda il problema circa la natura della definizione della verità.

4. *Definizioni implicite ed esplicite*

Tra le varie critiche mosse alle teorie decitazionali (*TB* o *UTB*), è interessante quella mossa da Jeffrey Ketland:

> A number of writing about Tarski's theory have claimed that Convention T (in effect, the infinite list of T-sentences: i.e. the theories TB and MT [and UTB] "fixes the extension of", or implicitly defines, "true"[4].

Tra i meriti delle teorie decitazionali ci sarebbe dunque la possibilità di fissare l'estensione del predicato T, ovvero di definirlo implicitamente. Tuttavia:

> Convention T does not implicitly define "true". [...] Assume that TB implicitly defines T. Then by the Beth Definability Theorem there exists an explicit definition of T in TB. [...] But this contradicts Tarski's Indefinability Theorem.

Vediamo come si struttura l'argomento di Ketland. Assumiamo che TB (cioè una teoria decitazionale, deflazionisticamente accettabile) sia in grado di definire implicitamente il predicato di verità T. In base al teorema di Beth, se una teoria definisce implicitamente un predicato, allora lo definisce esplicitamente. Ciò vuol dire che $TB \vdash \forall x\big(Tx \leftrightarrow \alpha(x)\big)$. Poiché *TB* ha le risorse sintattiche per il lemma di diagonalizzazione, possiamo dire che, per ogni formula β vale $TB \vdash \alpha(\ulcorner \beta \urcorner) \leftrightarrow \beta$. Ma per conservatività, $PA \vdash \alpha(\ulcorner \beta \urcorner) \leftrightarrow \beta$. Questo equivale a dire che in *PA* è possibile

---

4 Ketland 1999, 84.

definire un predicato di verità per *PA*, contro il teorema di Tarski. Ketland conclude che il deflazionista non può affermare che la teoria decitazionale offra una definizione implicita del predicato di verità.

L'argomento, però, non funziona come è stato messo in luce molto bene da Timothy Bays (Bays 2009). Per rendersene conto bisogna analizzare la portata del teorema di Beth che è uno dei punti fondamentali dell'argomento di Ketland.

5. *Teorema di Beth*

Per dimostrare il teorema di Beth è necessario fare uso di un risultato ottenuto da Craig[5].

Teorema 5 (Craig). Siano α e β due formule appartenenti al calcolo dei predicati del primo ordine con identità. Valga, inoltre che $\alpha \Vdash \beta$; allora esiste una formula γ (detta formula interpolante) tale che $\alpha \Vdash \gamma$ e $\gamma \Vdash \beta$.

Siano R e R' due simboli relazionali a n posti non appartenenti al linguaggio L.

$X(R)$ sia un insieme di enunciati del linguaggio $L \cup \{R\}$
$X(R')$ sia un insieme di enunciati del linguaggio $L \cup \{R'\}$

$X(R')$ è ottenuto sostituendo in $X(R)$ ogni occorrenza di R con R'.

5.1. Definizione implicita

Ora diciamo che $X(R)$ *definisce implicitamente* R se e solo se:

(i) $$X(R) \cup X(R') \Vdash \forall x_1,\ldots,x_n \left(R(x_1,\ldots,x_n) \leftrightarrow R'(x_1,\ldots,x_n)\right)$$

Detto altrimenti: se un insieme di formule definisce implicitamente un predicato, questo è estensionalmente indiscernibile da un altro predicato che presenta le stesse proprietà; si ricordi infatti la costruzione di $X(R')$.

5 La dimostrazione che segue è adattata da Chang Keisler (1973).

Ciò vuol dire che se una teoria definisse implicitamente un predicato di verità, per esempio, questo sarebbe essenzialmente unico.

La definizione implicita può essere espressa in una maniera analoga, dicendo cioè che

(ii) $X(R)$ *definisce implicitamente* R se e solo se c'è un solo modo di espandere un modello $\mathcal{M}$ di X a un modello $\mathcal{M}^*$ in modo che $\mathcal{M}^* \vDash X(R)$.

Un attimo di riflessione mostra come le due definizioni siano in realtà equivalenti. Che cosa vuol dire, infatti, che esiste un solo modo di espandere un modello in modo che renda vero $X(R)$? Significa che in tutte le possibili espansioni, il «significato inteso» di R rimane lo stesso, cioè R ha sempre la stessa estensione, come dichiarato dalla prima espressione.

5.2. Definizione esplicita

(iii) $X(R)$ *definisce esplicitamente* R se e solo se esiste una formula $\phi(x_1,...,x_n)$ del linguaggio L, tale che

$$X(R) \Vdash \forall x_1,...,x_n \left(R(x_1,...,x_n) \leftrightarrow \phi(x_1,...,x_n)\right)$$

Teorema 6 (Beth). Se $X(R)$ definisce esplicitamente R allora la definisce implicitamente e viceversa.

Dimostrazione. Il teorema si compone di due versi:

(1° verso) *DefExp* $\rightarrow$ *DefImp*
(2° verso) *DefImp* $\rightarrow$ *DefExp*

1° verso. Si tratta della direzione semplice. Assumiamo che $X(R)$ non definisca implicitamente R. Quindi ci sono due modelli $\mathcal{M}^1 = \langle \mathfrak{A}, K \rangle$ e $\mathcal{M}^2 = \langle \mathfrak{A}, K' \rangle$ di $X(R)$ tali che $K \neq K'$. Ma allora anche l'estensione di $\phi(x_1,...,x_n)$ sarà differente in $\mathcal{M}^1$ e $\mathcal{M}^2$ e quindi $X(R)$ non definisce esplicitamente R. ■

2° verso. Assumiamo che $X(R)$ definisca R implicitamente. Aggiungiamo una sequenza di n nuove costanti $c_1,...,c_n$ a L. Avremo quindi, per particolarizzazione e sostituzione che:

$$X(R) \cup X(R') \Vdash R(c_1,...,c_n) \leftrightarrow R'(c_1,...,c_n)$$

Per il teorema di compattezza, esistono sottoinsiemi finiti $\Delta \subseteq X(R)$ e $\Delta' \subseteq X(R')$ tali che:

$$\Delta \cup \Delta' \Vdash R(c_1,...,c_n) \leftrightarrow R'(c_1,...,c_n)$$

Sia, a questo punto, $\psi(R)$ la congiunzione di tutte le formule in cui occorre *R* che appartengono o a Δ o a Δ'; cioè:

$$\psi(R) = \bigwedge\{\alpha(R) | \alpha(R) \in \Delta \vee \alpha(R) \in \Delta'\}$$

In maniera equivalente $\psi(R')$ sia la congiunzione di tutte le formule in cui occorre *R’* che appartengono o a Δ o a Δ':

$$\psi(R') = \bigwedge\{\alpha(R') | \alpha(R') \in \Delta \vee \alpha(R') \in \Delta'\}$$

Avremo pertanto:

$$\psi(R) \wedge \psi(R') \Vdash R(c_1,...,c_n) \leftrightarrow R'(c_1,...,c_n)$$

per trasformazioni logiche

$$\psi(R) \wedge R(c_1,...,c_n) \Vdash \psi(R') \rightarrow R'(c_1,...,c_n)$$

A questo punto usiamo il teorema di interpolazione di Craig; esiste, pertanto una formula $\theta(c_1,...,c_n)$ tale che:

(1) $\psi(R) \wedge R(c_1,...,c_n) \Vdash \theta(c_1,...,c_n)$
(2) $\theta(c_1,...,c_n) \Vdash \psi(R') \rightarrow R'(c_1,...,c_n)$

Prendiamo in considerazione la (2). Dal momento che $X(R)$ definisce implicitamente *R*, per ipotesi, un modello $\mathcal{M} = \langle \mathfrak{A}, K' \rangle$ su $L \cup \{R', c_1,...,c_n\}$ è anche un modello su $L \cup \{R, c_1,...,c_n\}$ quando, naturalmente, interpretiamo *R* su K’. Per cui:

(2*) $\forall \mathcal{M}(\mathcal{M} \vDash \theta(c_1,...,c_n) \Rightarrow \mathcal{M} \vDash \psi(R') \rightarrow R'(c_1,...,c_n))$

per definizione di conseguenza logica. Ma l'ipotesi di definizione implicita comporta che:

(3) $\forall \mathcal{M} \forall \alpha \left(\mathcal{M} \vDash \alpha(R', c_1, ..., c_n) \Leftrightarrow \mathcal{M} \vDash \alpha(R, c_1, ..., c_n) \right)$

Quindi, mettendo insieme (2*) e (3) si ottiene che

(4) $\forall \mathcal{M} \left(\mathcal{M} \vDash \theta(c_1, ..., c_n) \Rightarrow \mathcal{M} \vDash \psi(R) \to R(c_1, ..., c_n) \right)$

ovvero

(4*) $\theta(c_1, ..., c_n) \Vdash \psi(R) \to R(c_1, ..., c_n)$, per definizione di conseguenza logica. Quindi, riassumendo, abbiamo:

(1) $\psi(R) \wedge R(c_1, ..., c_n) \Vdash \theta(c_1, ..., c_n)$
(4*) $\theta(c_1, ..., c_n) \Vdash \psi(R) \to R(c_1, ..., c_n)$

Quindi:

$\theta(c_1, ..., c_n), \psi(R) \Vdash R(c_1, ..., c_n)$ da (4*)
$\psi(R) \Vdash \theta(c_1, ..., c_n) \to R(c_1, ..., c_n)$ logica
$\psi(R) \Vdash R(c_1, ..., c_n) \to \theta(c_1, ..., c_n)$ da (1)

per cui

$\psi(R) \Vdash R(c_1, ..., c_n) \leftrightarrow \theta(c_1, ..., c_n)$ logica

Ora, $c_1, ..., c_n$ non occorrono in $\psi(R)$ che, come si ricorderà, è stato ottenuto da $X(R)$. Possiamo quindi sostituire le costanti con le variabili e quantificare universalmente:

$\psi(R) \Vdash \forall x_1, ..., x_n \left(R(x_1, ..., x_n) \leftrightarrow \theta(x_1, ..., x_n) \right)$

Ma poiché $\psi(R) \subseteq X(R)$ avremo che

$X(R) \Vdash \forall x_1, ..., x_n \left(R(x_1, ..., x_n) \leftrightarrow \theta(x_1, ..., x_n) \right)$ ■

6. *Discussione critica*

L'argomento di Ketland può essere riassunto come segue: una teoria deflazionisticamente accettabile (come *TB*) dovrebbe essere in grado di definire implicitamente il predicato *T*. Ma, continua Ketland, se la teoria definisse implicitamente *T*, allora lo definirebbe anche esplicitamente, in forza del teorema di Beth. Ora, poiché *TB* non può definire esplicitamente il predicato di verità (in ragione del teorema di Tarski) segue che *TB* non può neanche definire implicitamente *T*.

Il problema con l'argomento di Ketland è che con "definizione implicita" nel senso di Beth si richiede qualcosa di estremamente forte e cioè fissare l'estensione di *T* su tutti i modelli di *PA*, inclusi i modelli non standard.

Come dice bene Bays:

> They would have to determine the application of T, not just to every sentence in L, but to every object that any model of *PA thinks* is a sentence in L (Bays 2009, 8).

Ora, è possibile che il deflazionista rimoduli il suo concetto di definizione implicita in modo che non si applichi il teorema di Beth? Sempre secondo Bays:

> It seems to me, however, that this is not what deflationists have in mind when they claim that the T-schema 'fixes the extension' of the truth predicate. As far as I can see, they are simply claiming that the T-schema fixes the application of T to every *genuine* sentence in L, and there is nothing in Ketland's argument which tells against this more restricted claim. In fact, I think that this restricted claim is pretty clearly correct (Bays 1999, 7).

Per renderci conto del rilievo di Bays circa la "genuinità" degli enunciati è bene notare questi fatti di natura tecnica. *TB* fissa l'estensione di T rispetto al modello standard di *PA*. Sia $\mathcal{M}$ un modello basato sul modello standard e così composto: $\langle \mathbb{N}, T \rangle$ dove $\mathbb{N}$ è il dominio dei numeri naturali e T è l'estensione del predicato T. Ora, vale che $\langle \mathbb{N}, T \rangle \vDash TB$. Assumiamo che n sia il codice numerico di un enunciato $\phi \in L(PA)$.Quindi, avremo che $\langle \mathbb{N}, T \rangle \vDash T(\ulcorner \phi \urcorner) \leftrightarrow \phi$ cioè $\langle \mathbb{N}, T \rangle \vDash T(\ulcorner \phi \urcorner) \Leftrightarrow \langle \mathbb{N}, T \rangle \vDash \phi$; poiché ϕ appartiene al linguaggio di *PA*, non contiene occorrenze del predicato *T*, quindi, in virtù della definizione di relazione modello, abbiamo che $\Im(\ulcorner \phi \urcorner) \in \Im(T) \Leftrightarrow \mathbb{N} \vDash \phi$. In altre parole, n, e cioè il numero della formula ϕ sarà nell'estensione di *T* se e solo se $\mathbb{N} \vDash \phi$. Poiché ogni enunciato di L(*PA*) ha un valore di verità determinato in $\mathbb{N}$, e poiché la funzione di aritmetizzazione è fissata sulla parte standard si può dire che il T-schema fissa l'estensione di *T* per ogni enunciato.

Inoltre, *TB* fissa l'estensione di *T* sulla parte standard dei modelli non standard di *PA*. Anche questo è del tutto chiaro non appena si consideri che la parte standard dei modelli non standard è, per l'appunto, il segmento standard della struttura dei numeri naturali e, in questo segmento, è possibile riprodurre l'argomentazione appena esposta.

Quindi abbiamo che *TB* fissa l'estensione di *T* sul modello standard e sulla parte standard dei modelli non standard; il motivo per cui *TB* non riesce a definire implicitamente *T* nel senso richiesto dal teorema di Beth è che non è in grado di fissare l'estensione di *T* sulla parte non standard dei modelli non standard. Sia $\mathbb{N}^*$ un modello non standard di *PA* e *m* un elemento non standard tale che $\mathbb{N}^* \vDash Sent(m)$. Il problema è che - secondo il modello non standard $\mathbb{N}^*$, *m* è il codice di un enunciato ma ciò è impossibile dal momento che la funzione di aritmetizzazione è fissata solo per la parte standard. Quindi il ragionamento di prima non può essere riproposto per il semplice fatto che non c'è nessun enunciato ϕ tale che *m* sia il suo codice.

Bays mette in luce un punto interessante, in nota. Possiamo immaginare il T-schema come una procedura (una sorta di "ricetta") che determina l'applicazione di *T* a particolari codici di enunciati del linguaggio. Nel caso dei modelli non standard, il modello dichiara che *m* è il codice di un enunciato (il che è evidentemente falso) e di conseguenza il T-schema non è in grado di applicare T a quello pseudo-enunciato. Il problema sta però, secondo Bays, nell'estensione del predicato aritmetizzato *Sent*(*x*) e non nel T-schema in quanto tale.

In realtà è facile rendersi conto che nessuna teoria della verità che possieda una teoria aritmetica del primo ordine come teoria di base riesce a definire implicitamente il predicato di verità nel senso del teorema di Beth. Infatti, se così fosse, la teoria in questione sarebbe in grado di definire il predicato di verità anche esplicitamente, contro il risultato di limitazione di Tarski.

A questo punti il deflazionista deve chiarire cosa intende con "definizione implicita". Abbiamo visto due alternative: l'una intende "definizione implicita" come fissare l'estensione di *T* su tutti i modelli (standard e non) di *PA* e questo è impossibile per i teoremi di Beth e Tarski (come messo in luce da Ketland); l'altra possibilità è quella - almeno implicitamente suggerita dalla riflessione di Bays - di considerare “definizione implicita” il fissare l’estensione di T rispetto al modello standard della teoria in questione, nel nostro caso, *PA*.

L’idea che una trattazione assiomatica del predicato di verità non sia in grado di fissare in maniera assoluta l’estensione, pena il ricadere nella situazione appena illustrata, era nota già a Tarski, il quale discutendo proprio la plausibilità di un’assiomatizzazione indicava nel carattere «arbitrario» e «contingente» della scelta degli assiomi uno dei limiti costitutivi di questo approccio. Scrive infatti, Tarski:

> Thus it seems natural to require that the axioms of the theory of truth, together with the original axioms of the metatheory, should constitute a categorical system. It can be shown that this postulate coincides in the present case with another postulate, according to which the axiom system of the theory of truth should unambigously determine the extension of the symbol ‘*Tr*’ which occurs in it, and in the following sense: if we introduce into the metatheory, alongside this symbol, another primitive sign, e.g. the symbol “ *Tr'*” and set up analogous axioms for it, then the statement «*Tr* = *Tr'*» must be provable[6].

Il riferimento di Tarski al teorema secondo cui T = T’ coincide con ciò che è stabilito nella definizione implicita impiegata nel teorema di Beth. Se le cose, infatti, stessero così, ciò verrebbe a significare che l’estensione del predicato di verità sarebbe fissata su *tutti* i modelli come sta a testimoniare la proprietà metateorica di categoricità evocata da Tarski. Che cosa significa mostrare che un sistema di assiomi è categorico? Significa far vedere che i suoi modelli sono tra loro isomorfi, cioè che possiede un solo modello a meno di isomorfismi. Ma dire questo equivale ad affermare che il sistema non presenta modelli non standard, o patologici, cioè proprio quei modelli sui quali la teoria decitazionale *TB* non è in grado di fissare l’estensione del predicato T.

Se la teoria in questione fosse categorica, ovvero in grado di definire implicitamente il predicato T in senso forte, e la logica nella quale è formulata completa, allora ciò sarebbe in contrasto con il teorema di Tarski di indefinibilità della verità. Questo argomento che sfrutta, come abbiamo visto, il teorema di Beth era stato proposto già da Tarski che infatti dice:

> But this postulate [cioè che T = T’, ndA] cannot be satisfied. For it is not difficult to prove that in the contrary case the concept of truth could

[6] Tarski 1983, 258.

> be defined exclusively by means of terms belonging to the morphology of language, which would be in palpable contradiction to theorem I[7].

L'aspetto curioso è che la memoria di Tarski *precede* il lavoro di Beth (e quello di Craig) sulle definizioni implicite ed esplicite. Al di là di questa notevole anticipazione tarskiana ciò che è interessante è proprio lo scetticismo del padre della concezione semantica circa l'approccio assiomatico: uno scetticismo, come abbiamo visto, che è motivato dall'impossibilità di compiere una scelta degli assiomi non del tutto esente da tratti di arbitrarietà. E questo fenomeno è ben rappresentato, formalmente, dall'impossibilità di caratterizzare in maniera univoca (cioè, a meno di isomorfismi) l'estensione del predicato *T*.

7. *Debolezza deduttiva*

L'altra fonte di perplessità sulle teorie assiomatiche deriva dalla estrema debolezza deduttiva di *TB* e *UTB*. Anche su questo punto, Tarski è molto chiaro:

> A theory of truth founded on them [cioè sugli assiomi derivati dal T-schema] would be a highly incomplete system, which would lack the most important and most fruitful general theorems[8].

In breve, *TB* non riesce a dimostrare la generalizzazione di principi logici fondamentali del tutto intuitivi: $TB \nvdash \forall\alpha\big(T(\ulcorner\alpha\urcorner) \vee T(\ulcorner\neg\alpha\urcorner)\big)$. Qualche precisazione a tal proposito. *TB* è perfettamente in grado di mostrare che, per un'α qualunque, vale $TB \vdash T(\ulcorner\alpha\urcorner) \vee T(\ulcorner\neg\alpha\urcorner)$. Ciò che *TB* non riesce a fare è importare il quantificatore a livello teorico. Si ricordi che la formulazione appena data è un'abbreviazione per $TB \nvdash \forall x\big(Sent(x) \rightarrow T(x) \vee T(neg(x))\big)$.

Teorema 7. $TB \nvdash \forall\alpha\big(T(\ulcorner\alpha\urcorner) \vee T(\ulcorner\neg\alpha\urcorner)\big)$

Dimostrazione. Assumiamo per assurdo che $\forall\alpha(T(\ulcorner\alpha\urcorner) \vee T(\ulcorner\neg\alpha\urcorner))$ sia dimostrabile in *TB* e che *D* sia una dimostrazione in *TB*. *D* è per ipotesi un oggetto sintattico finito. *D* conterrà solo un numero finito di T-bicondizionali:

---

7 Tarski 1983, 258.

8 Tarski 1983, 257.

$$T(\ulcorner\alpha_1\urcorner) \leftrightarrow \alpha_1$$
$$T(\ulcorner\alpha_2\urcorner) \leftrightarrow \alpha_2$$
$$\ldots$$
$$T(\ulcorner\alpha_n\urcorner) \leftrightarrow \alpha_n$$

Costruiamo a questo punto un modello $\mathcal{M} = \langle \mathbb{N}, \mathcal{E} \rangle$ che rende veri tutti i bicondizionali in D. In particolare $\mathcal{E}$ è l'estensione di T e quindi sarà l'insieme costituito da quegli enunciati β tali che o β o $\neg\beta$ è nell'insieme $\{\alpha_1, ..., \alpha_n\}$. Quindi $\mathcal{M}$ rende veri gli assiomi di *PA* e la sequenza dei bicondizionali che abbiamo indicato e non gli altri. Ora, sia θ un enunciato tale che né θ né la sua negazione appartengono a $\{\alpha_1, ..., \alpha_n\}$. Quindi $\mathcal{M} \nvDash T(\ulcorner\theta\urcorner) \leftrightarrow \theta$. Tuttavia, in base all'ipotesi per assurdo si ha che $TB \vdash \forall\alpha(T(\ulcorner\alpha\urcorner) \vee T(\ulcorner\neg\alpha\urcorner))$ cioè, per la correttezza del calcolo, $TB \Vdash \forall\alpha(T(\ulcorner\alpha\urcorner) \vee T(\ulcorner\neg\alpha\urcorner))$. Per quanto detto prima, c'è un modello $\mathcal{M}$ tale che $\mathcal{M} \vDash TB$ e $\mathcal{M} \nvDash T(\ulcorner\theta\urcorner) \vee T(\ulcorner\neg\theta\urcorner)$. Per definizione si ha $TB \nVdash \forall\alpha(T(\ulcorner\alpha\urcorner) \vee T(\ulcorner\neg\alpha\urcorner))$, cioè $TB \nvdash \forall\alpha(T(\ulcorner\alpha\urcorner) \vee T(\ulcorner\neg\alpha\urcorner))$.

La debolezza deduttiva di *TB* ha a che fare direttamente con il concetto di generalizzazione. Si può affermare che le teorie decitazionali sono teorie 'concrete' nel senso che possiedono un assioma per ogni enunciato vero. Il che risulta, per altro, molto poco intuitivo. Nella concezione assiomatica (sia classica che moderna) l'idea di assioma è legata a quella di principio generale, cioè di una espressione di determinate proprietà che valgono di classi di individui. Introdurre uno specifico assioma per ogni 'oggetto' di verità sarebbe come elencare, caso per caso, le istanze dell'addizione. Nessuno si sognerebbe di formulare un sistema assiomatico aritmetico in cui gli assiomi comprendono la lista infinita: 1+1=2, 1+2=3, 1+3=4, ...

Una teoria della verità autentica dovrebbe, al contrario, riuscire a stabilire principi strutturali circa la nozione che intende trattare (in questo caso, la verità). Non contraddizione e terzo escluso, così come le generalizzazioni di tutte le leggi logiche, sono esempi chiarissimi di generalizzazioni logicamente vere che si assume essere teoremi della nostra teoria della verità.

8. *Congiunzioni infinite*

Come abbiamo avuto modo di vedere in precedenza, uno degli scopi fondamentali del predicato di verità, deflazionisticamente inteso, è quello di permettere l'espressione di generalizzazioni.[9] Nei classici contesti di proferimento, infatti, il predicato di verità è eliminabile: sempre mantenendo l'ottica deflazionsita, risulta perfettamente uguale dire che «La neve è bianca» e dire che «'La neve è bianca' è vero». Le cose cambiano sensibilmente quando sono in gioco enunciati generali, dove l'eliminazione del predicato di verità non è possibile.

> We may affirm the single sentence by just uttering it, unaided by quotation or by the truth predicate; but if we want to affirm some infinite lot of sentences, then the truth predicate has its use[10].

L'esempio classico è il seguente. Ipotizziamo di riporre una grande fiducia nella conoscenza storica di Carlo; possiamo quindi dire che:

(1) Tutto quello che dice Carlo sull'Impero Romano è vero

In questo enunciato non possiamo eliminare la parola "vero" pena la perdita di forma corretta grammaticale. Naturalmente, questo implicherà la serie potenzialmente infinita di enunciati di forma:

(1a) Se Carlo dice p sull'Impero Romano, allora p
(1b) Se Carlo dice q sull'Impero Romano, allora q
...

Ma il punto è proprio questo; potremmo non sapere esattamente che cosa dirà Carlo a proposito dell'Impero Romano ma vogliamo, nondimeno, esprimere la nostra credenza che quello che dirà sarà comunque vero. Non abbiamo, cioè, a disposizione l'intera sequenza degli enunciati di forma "Se Carlo dice x allora x" perché potremmo non sapere i vari valori che assume x. La generalizzazione ci permette, quindi, di esprimerci anche su quegli enunciati che non conosciamo. Ancora Quine:

9 Per la discussione sulle congiunzioni infinite vedi Halbach 1999.
10 Quine 1970, 12.

> The truth predicate proves invaluable when we want generalize along a dimension that cannot be swept out by a general term[11].

La *raison d'etre* della verità, cioè il motivo per cui nelle lingue naturali è normalmente presente un predicato "vero" è proprio permettere la formulazione di generalizzazioni e cioè la riduzione di quantità infinite di enunciati. La questione che sorge a questo punto è se effettivamente il progetto deflazionista riesce a realizzare in maniera precisa questa riduzione. Come vedremo si tratta di analizzare due punti fondamentali: innanzitutto, in che senso intendere "una quantità infinita di enunciati"; in secondo luogo, chiedersi che cosa significa esprimere e cosa vuol dire che una teoria assiomatica deflazionisticamete accettabile deve poter esprimere generalizzazioni.

Quine, come abbiamo visto, parla di riduzione di quantità infinite di enunciati tramite l'impiego essenziale del predicato di verità. Quello che cerchiamo, sembra essere una funzione così formata:

(2) $f : \wp(X) \mapsto X$

dove X è una variabile per insiemi arbitrari di enunciati e $\wp(X)$ l'insieme potenza. Un esempio potrà chiarire il punto. Prendiamo l'insieme infinito di enunciati:

(1a) Se Carlo dice p sull'Impero Romano, allora p
(1b) Se Carlo dice q sull'Impero Romano, allora q
...

chiamiamo questo insieme Y. L'idea è quindi quella di rappresentare questo insieme infinito di enunciati con un enunciato in cui occorra il predicato di verità T, ottenendo qualcosa come $f(Y) = a_T$. Abbiamo parlato di una lista infinita di enunciati ma è la stessa cosa se prendiamo in considerazione un solo enunciato costituito da una congiunzione infinita di enunciati. Tuttavia, se l'insieme degli enunciati del nostro linguaggio è numerabile, le possibili liste infinite, ovvero le possibili congiunzioni infinite, avranno cardinalità $2^{\aleph_0}$. Ma, in base al teorema di Cantor, non ci può essere una funzione che associ a ciascuna congiunzione infinita un enunciato in cui occorre T.

11 Quine 1990, 81.

È necessario, allora, che gli insiemi di enunciati siano definibili nel nostro linguaggio, ovvero è necessario porre una restrizione circa l'arbitrarietà degli insiemi con cui abbiamo a che fare.

Diremo che una formula $\alpha(x)$ del nostro linguaggio definisce un insieme A se e solo se:

(3) $\forall x\big(\alpha(x) \leftrightarrow x \in A\big)$

o alternativamente

(4) $A = \{a \mid \mathcal{M} \vDash \alpha(\bar{a})\}$

La prima formulazione è sintattica: una formula definisce un insieme quando gli elementi dell'insieme soddisfano la formula. La seconda è chiaramente semantica; un attimo di riflessione mostra però che le due formulazioni, sotto alcune ipotesi generali, risultano del tutto equivalenti.

Assumiamo a questo punto che vi sia una formula $\alpha(x)$ che definisce A. In base alla tesi decitazionista, l'enunciato

(5) $\forall x\big(\alpha(x) \rightarrow Tx\big)$

esprime la congiunzione infinita di enunciati definita sugli elementi di A. Per esempio, se A contiene una sequenza infinita di enunciati come:

Se l'erba è verde allora l'erba è verde
Se la neve è bianca allora la neve è bianca
Se il mare è blu allora il mare è blu
...

l'enunciato (5) esprime, in maniera finitaria, la congiunzione infinita; (5) dice, infatti, che se x è un enunciato di forma $p \rightarrow p$ allora x è vero. Da un punto di vista intuitivo tutto ciò è estremamente ragionevole. Ma in che senso una proposizione universale esprime un insieme infinito di enunciati? È chiaro che non possiamo riferirci a un generico "voler dire" o "significare"; le strade possibili sono due: si può interpretare la relazione fra (5) e la congiunzione infinita o come una relazione essenzialmente *semantica* o come una relazione essenzialmente *sintattica*.

8.1. Caratterizzazione semantica

Da un punto di vista semantico non è difficile rendersi conto che (5) esprime in effetti la congiunzione di tutti gli elementi di A; ovvero che (5) vale in un modello se e solo se vale la rispettiva congiunzione infinita.

$\mathcal{M}$ sia il modello per il nostro linguaggio L senza predicato di verità. $\mathcal{E}$ sia l'estensione del predicato di verità definito (tradizionalmente) come $\mathcal{E} = \{\beta | \mathcal{M} \vDash \beta\}$. Inoltre, $\alpha(x)$ definisce semanticamente A se e solo se $A = \{a | \mathcal{M} \vDash \alpha(\overline{a})\}$; ovvero A risulta essere l'estensione in $\mathcal{M}$ di $\alpha(x)$.

Vogliamo quindi far vedere che $\forall x(\alpha(x) \rightarrow Tx)$ vale in un modello se e solo se, in quel modello, vale la rispettiva congiunzione infinita. Ovviamente dobbiamo scegliere un modello in cui sia presente l'estensione per il predicato di verità dal momento che la generalizzazione contiene, in maniera essenziale, per altro, il predicato di verità.

Assumiamo, quindi, che: $\langle \mathcal{M}, \mathcal{E} \rangle \vDash \forall x(\alpha(x) \rightarrow Tx)$ e $\alpha(x)$ definisce l'insieme A. Vogliamo dimostrare che $\forall \beta \in A, \mathcal{M} \vDash \beta$. Ora, se $\langle \mathcal{M}, \mathcal{E} \rangle \vDash \forall x(\alpha(x) \rightarrow Tx)$ e $\alpha(x)$ definisce l'insieme A, significa che l'estensione di $\alpha(x)$ in $\langle \mathcal{M}, \mathcal{E} \rangle$ è uguale ad A. Ma, per ogni x, $\alpha(x) \rightarrow Tx$ e quindi l'estensione di $\alpha(x)$ è inclusa nell'estensione di T. Cioè, $A \subseteq \mathcal{E}$. Pertanto per ogni $\beta \in A$ segue che $\beta \in \mathcal{E}$. Ma per definizione, se $\beta \in \mathcal{E}$ allora $\mathcal{M} \vDash \beta$.

Tuttavia, come fa giustamente notare Halbach[12], una simile argomentazione non è a disposizione del deflazionista. Come abbiamo già discusso in precedenza, il problema sta proprio nel riferimento semantico che si utilizza. Se si prende sul serio l'approccio assiomatico - e in questo caso anche la tesi deflazionista - non è poi possibile trascendere i limiti imposti dalla posizione assunta e adottare la semantica tarskiana. Le nozioni di modello, di estensione e di interpretazione, cruciali per l'argomento appena esposto, non sono disponibili a chi nega la validità stessa dell'approccio tarskiano. Per questa ragione, rimane solo l'altra via e cioè quella sintattica.

12 Halbach 1999, 8.

8.2. Caratterizzazione sintattica

Per l'approccio sintattico, l'intuizione secondo cui il predicato di verità è indispensabile per esprimere congiunzioni infinite di enunciati può essere tradotta come l'esigenza di *dimostrare* grazie al predicato di verità tali congiunzioni infinite. Ricordiamo, per chiarezza, che *PA* è la nostra teoria di base formulata in un linguaggio del primo ordine L_{PA}. Aggiungendo il predicato *T* a L_{PA} si ottiene il linguaggio L_{PA^T}. L'aggiunta dei bicondizionali di Tarski permette di ottenere la teoria *TB*, che abbiamo già analizzato. Pertanto $TB = PA \cup T(\ulcorner \alpha \urcorner) \leftrightarrow \alpha$, con $\alpha \in L_{PA}$.

Prendiamo allora in considerazione la congiunzione infinita di tutti gli enunciati di forma $\beta \to \beta$, appartenenti a L_{PA}. Indichiamo questo insieme infinito come:

$$\bigwedge \{\beta \to \beta | \beta \in L_{PA}\}$$

Sia $\alpha(x)$ la formula che rappresenta l'insieme di tutti gli enunciati $\beta \to \beta \in L_{PA}$. Bisogna, quindi, dimostrare che $TB \vdash \forall x(\alpha(x) \to Tx)$.

Se si riuscisse a ottenere questo risultato si potrebbe affermare a buon diritto che il predicato di verità permette di esprimere congiunzioni infinite dove, con 'esprimere, intendiamo una relazione sintattica estremamente forte e significativa, e cioè quella di derivabilità. Purtroppo le cose non stanno così. Vale infatti la seguente tesi.

Teorema 8. Se c'è un modello $\mathcal{M}$ di *PA* tale che l'estensione di $\alpha(x)$ in $\mathcal{M}$ è un insieme infinito, allora $TB \nvdash \forall x(\alpha(x) \to Tx)$. Per contrapposizione, poi, se $TB \vdash \forall x(\alpha(x) \to Tx)$, allora l'estensione di $\alpha(x)$ in $\mathcal{M}$ è un insieme finito.

Dimostrazione. La dimostrazione si basa sulla costruzione di un modello non standard $\mathbb{Q}$ per cui esiste un elemento $\overline{c}$ tale che $\mathbb{Q} \vDash \alpha(\overline{c})$ e $\mathbb{Q} \nvDash T(\overline{\ulcorner c \urcorner})$. Sia $\mathcal{M}$ il modello di *PA* con estensione infinita di $\alpha(x)$. Sia M il dominio di $\mathcal{M}$. Δ sia un insieme di enunciati così costituito:

$$\Delta = \begin{array}{l} \Gamma \cup \{a \in M | c \neq a\} \cup \{\alpha(c), \neg T(\overline{\ulcorner c \urcorner})\} \cup \\ \{T(\overline{\ulcorner \alpha \urcorner}) | \mathcal{M} \vDash \alpha\} \cup \{\neg T(\overline{\ulcorner \alpha \urcorner}) | \mathcal{M} \nvDash \alpha\} \end{array}$$

L'insieme Δ è costituito da cinque componenti; analizziamole singolarmente:

1. Γ è il diagramma elementare di $\mathcal{M}$ e contiene pertanto tutte le formule atomiche (e le negazioni) vere in $\mathcal{M}$.
2. $\{T(\overline{\ulcorner\alpha\urcorner}) \mid \mathcal{M} \vDash \alpha\}$ dice che l'estensione di T è l'insieme delle formule vere in $\mathcal{M}$.
3. $\{\neg T(\overline{\ulcorner\alpha\urcorner}) \mid \mathcal{M} \nvDash \alpha\}$ dice che il complemento dell'estensione di T è l'insieme delle formule false in $\mathcal{M}$
4. $\{a \in M \mid c \neq a\}$ dice che c è un elemento diverso da ogni elemento nel dominio di $\mathcal{M}$
5. $\{\alpha(c), \neg T(\ulcorner c \urcorner)\}$ dice che l'elemento c è nell'estensione di α ma non in quella di T.

Come è facile notare, i punti chiave sono il 4. e il 5. dove si stabilisce l'esistenza di un elemento non standard. Il procedimento è del tutto analogo a quello utilizzato per la costruzione di modelli non standard dell'aritmetica. Anche in quel caso, si postula l'esistenza di un numero *k* non standard, diverso cioè da 0, 1, 2, 3, ... [13]

Ora, Δ è un insieme infinito. È facile vedere che per ogni sottoinsieme finito di Δ esiste un modello (esattamente come nel caso dei numeri naturali). Ma, per il teorema di compattezza, Δ ha un modello $\mathbb{Q}$. Ed è immediato vedere che $\mathbb{Q} \vDash \alpha(c)$ e $\mathbb{Q} \nvDash T\left(\overline{\ulcorner c \urcorner}\right)$.

In base al teorema (8) possiamo dire che non sembra esserci un caratterizzazione precisa per interpretare, in maniera sufficientemente rigorosa, il motto deflazionista secondo cui la verità permette l'espressione di generalizzazioni infinite. L'approccio semantico, infatti, è ingiustificato

---

13 Per diagramma si intende la seguente costruzione modellistica (cfr. Chang Keisler 1973). Sia $\mathcal{M}$ un modello di $\mathcal{L}$. Espandiamo il linguaggio $\mathcal{L}$ a un nuovo linguaggio

$$\mathcal{L}_A = \mathcal{L} \cup \left\{c_a \mid a \in A\right\}$$

$\mathcal{L}_A$ è come $\mathcal{L}$ con una nuova costante c_a per ogni elemento a che appartiene ad A. Ovviamente, se $a \neq b$ avremo che $c_a \neq c_b$. Espandiamo a questo punto $\mathcal{M}$ al modello

$$\mathcal{M}_A = (\mathcal{M}, a)_{a \in A}$$

per $\mathcal{L}_A$. Il diagramma di $\mathcal{M}$, che indichiamo con $\Delta_{\mathcal{M}}$, è l'insieme di tutti gli enunciati atomici e delle negazioni degli enunciati atomici di $\mathcal{L}_A$ che valgono nel modello $\mathcal{M}_A$.

dal punto di vista filosofico perché mette in campo una serie di nozioni, in ultima analisi la caratterizzazione tarskiana di verità, che vengono esplicitamente rigettate da chi opta per un'assiomatizzazione decitazionale. D'altro canto, l'approccio sintattico è sicuramente disponibile in ottica deflazionista ma non riesce a rendere conto del fenomeno dei modelli non standard. Come si può vedere, il problema è lo stesso incontrato alla fine della sezione precedente a proposito della debolezza deduttiva delle teorie decitazionali. Anche in quel caso, come già preannunciato dallo stesso Tarski, ci troviamo nell'impossibilità di derivare una generalizzazione universale a partire da una serie - anche infinita - di enunciati. E, come abbiamo avuto modo di vedere, la radice del problema sta nell'impossibilità di fissare il modello standard nel quale non esistono recalcitranti elementi *c* che non compaiono negli enunciati di partenza e quindi invalidano la possibilità di ottenere la generalizzazione.

Ci si potrebbe chiedere, a questo punto, se il deflazionista non possa assumere proprio l'ottica del modello standard, eliminando, tramite un *fiat*, i modelli patologici. Ci sono (almeno) due problemi a tal riguardo. Innanzitutto, l'esclusione dei modelli non standard andrebbe in qualche modo giustificata e non semplicemente assunta; e in secondo luogo, non si capisce bene quale tipo di giustificazione possibile abbia a disposizione il deflazionista. In altre parole, come è possibile discriminare tra due o più (in effetti, infiniti) modelli di aritmetica senza fare appello, magari in maniera del tutto intuitiva, alla nozione di verità tarskianamente intesa?

> In other words, Tarski's model theoretic definition of truth would become a prerequisite of disquotationalism; but why should the Tarskian equivalences be called in, if a notion of truth is already available?[14]

8.3. Equivalenza rispetto alle conseguenze

Se le cose stanno così, quale possibilità per il deflazionista? Come giustificare cioè il fatto che una teoria deflazionisticamente accettabile, come la serie dei bicondizionali di Tarski, non sia in grado di rendere opportunamente la comunicazione di generalizzazioni. Una terza via è quella di considerare l'enunciato:

(1) Tutto quello che dice Carlo sull'Impero Romano è vero

---

14 Halbach 1999, 11.

e lo schema

(6) Se Carlo dice x sull'Impero Romano, allora x

come equivalenti rispetto alle loro conseguenze in cui non occorre il predicato di verità. In altri termini, (6) è uno schema di enunciati e corrisponde alla congiunzione infinita nella misura in cui noi sostituiamo a x gli enunciati appropriati; (1) al contrario è la generalizzazione in cui compare il predicato di verità. Ora, benché abbiamo visto che (6) non implica (1) tuttavia si può mostrare che (1) e (6) sono equivalenti rispetto alle loro conseguenze.

Teorema 9. Definiamo, a questo punto,
$TB \cup \vartheta =_{def.} TB \cup \forall x\big(\alpha(x) \to Tx\big)$

$TB' =_{def.} TB \cup \alpha\big(\overline{\ulcorner \beta \urcorner}\big) \to \beta$

dove $\beta \in L_{PA}$. Dobbiamo mostrare che $TB \cup \vartheta$ e TB' sono equivalenti rispetto alle loro conseguenze cioè che:

$\forall \alpha \in L_{PA}, TB' \vdash \alpha \Leftrightarrow TB \cup \vartheta \vdash \alpha$

Dimostrazione.

Ad $\Rightarrow$). Assumiamo che $TB' \vdash \alpha(\overline{\ulcorner \beta \urcorner}) \to \beta$. Si vuole dimostrare che $TB \cup \vartheta \vdash \alpha(\overline{\ulcorner \beta \urcorner}) \to \beta$. Ora, abbiamo $TB \cup \vartheta \vdash \forall x\big(\alpha(x) \to Tx\big)$, quindi $TB \cup \vartheta \vdash \alpha(\overline{\ulcorner \beta \urcorner}) \to T(\overline{\ulcorner \beta \urcorner})$. Ma $TB \cup \vartheta \vdash T(\overline{\ulcorner \beta \urcorner}) \to \beta$ per bi-condizionale di Tarski. E quindi, per logica, $TB \cup \vartheta \vdash \alpha(\overline{\ulcorner \beta \urcorner}) \to \beta$.

Ad $\Leftarrow$). Assumiamo che $TB \cup \vartheta \vdash \alpha$. Esiste quindi una dimostrazione in $TB \cup \vartheta$ di α che contiene solo un numero finito di T-enunciati:

$$T\big(\overline{\ulcorner \zeta_1 \urcorner}\big) \leftrightarrow \zeta_1$$

$$T\big(\overline{\ulcorner \zeta_2 \urcorner}\big) \leftrightarrow \zeta_2$$

$$\ldots$$

$$T\big(\overline{\ulcorner \zeta_n \urcorner}\big) \leftrightarrow \zeta_n$$

A questo punto è possibile definire in termini puramente aritmetici un predicato di verità parziale del tutto simile a quello ottenuto nella dimostrazione del teorema 4:

$$\tau(x) =_{def.} \left(x = \ulcorner \zeta_1 \urcorner \wedge \zeta_1\right) \vee \ldots \vee \left(x = \ulcorner \zeta_n \urcorner \wedge \zeta_n\right)$$

Basta poi sostituire nella dimostrazione di $\vartheta \to \alpha$ ogni occorrenza di T con un'occorrenza di τ e si giunge al risultato. In questo senso, si può dire, il sostenitore di una posizione decitazionale (deflazionista) della verità è in una posizione piuttosto confortevole. Da un lato egli deve ammettere come principi specifici della sua teoria della verità solamente i bicondizionali di Tarski; dall'altro la teoria *TB* risulta conservativa su *PA*. Anche il discorso sulle congiunzioni infinite pare accordarsi alle preferenze deflazioniste; ammesso un approccio del terzo tipo, si mostra come lo scopo principale del predicato di verità - quando cioè esso è indispensabile - è quello di esprimere generalizzazioni. Tuttavia le cose non sono così semplici. In particolare, dobbiamo ricordare, seguendo Tarski, l'estrema debolezza deduttiva di una teoria decitazionale; ma c'è di più. Le generalizzazioni, ovvero proprio il campo di applicazione specifico del predicato di verità, non sono sempre proposizioni *innocue*. Talvolta, anzi, veicolano una notevole quantità di informazioni, come nel caso in cui asseriamo che

(Corr) Tutti i teoremi di T sono veri

La forma logica di (Corr) è identica a quella di (1) ma enunciati come (Corr) hanno un ruolo fondamentale nella teoria della dimostrazione. Si tratta dei *principi di riflessione* che assicurano la correttezza della teoria in questione e tramite la correttezza, la sua consistenza. Il deflazionista deve rendere conto del fatto che tali generalizzazioni, in cui occorre in maniera essenziale il predicato di verità, risultano del tutto informative. E questo mal si accorda con l'idea della non sostanzialità della nozione di verità. È questa l'idea di fondo dell'argomento di conservatività che presenteremo e discuteremo nell'ultimo capitolo. Prima però ci concentreremo su teorie assiomatiche che non si basano sull'intuizione decitazionale ma accolgono l'idea di *composizionalità*, in ottica prettamente tarskiana.

VI
TEORIE COMPOSIZIONALI

1. *Dai bicondizionali alla teoria composizionale*

Una delle intuizioni fondamentali del lavoro di Tarski consiste nel riconoscere che la verità ha una natura composizionale; in maniera intuitiva ciò significa che la verità degli enunciati linguisticamente più complessi si costruisce a partire dalla verità degli enunciati più semplici che li compongono. La verità di una congiunzione equivale alla verità dei congiunti, la verità di una disgiunzione, alla verità dei disgiunti e così via. Come abbiamo visto nella prima parte del volume, la composizionalità è strettamente connessa con la procedura ricorsiva tramite la quale Tarski riesce ad assicurare una definizione di verità per un numero di enunciati potenzialmente infinito.

Anche per l'approccio assiomatico si può scegliere questa strada e costruire una teoria composizionale; si tratta della codifica formale delle clausole tarskiane. Ovviamente, in questo caso avremo - come al solito - il predicato primitivo (e indefinito) T mentre nell'originale è presente la nozione semantica di soddisfazione. Per il resto, la struttura è perfettamente simile.

In ciò che segue presenteremo la classica teoria composizionale (*CT*) per l'aritmetica di Peano che, come nel caso decitazionale, costituisce la teoria di base. Al fine di presentare questa teoria nella maniera più formale possibile, introduciamo ancora un predicato e una funzione aritmetizzati:

$ClTerm(x)$ x è il codice di un termine chiuso, cioè di un termine senza variabili libere.

$eq(x,y)=z$ z è il codice dell'identità tra i termini codificati da x e y. Detto altrimenti, $z = \ulcorner x = y \urcorner$.

CT possiede quattro assiomi specifici:

T1.
$\forall x \forall y \big(ClTerm(x) \wedge ClTerm(y) \wedge eq(x,y) = z \rightarrow Tz \leftrightarrow val(x) = val(y) \big)$

T2.
$\forall x \forall y \big(x = neg(y) \rightarrow Tx \leftrightarrow \neg Ty \big)$

T3.
$\forall x \forall y \forall z \big(conj(x,y) = z \rightarrow \big(Tz \leftrightarrow Tx \wedge Ty \big) \big)$

T4.
$\forall x \forall y \forall v \Big(y = all(v,x) \rightarrow \Big(Ty \leftrightarrow \forall z \big(ClTerm(z) \rightarrow Tso(x,v,z) \big) \Big) \Big)$

Formulazione alternativa di *CT*.

La presentazione precedente degli assiomi di *CT* è perfettamente rigorosa; tuttavia non risulta particolarmente leggibile. Talvolta è preferibile adottare una simbologia abbreviata di maggiore fruibilità:

T1. $\forall s \forall t \big(T \left(\ulcorner s = t \urcorner \right) \leftrightarrow val(s) = val(t) \big)$
T2. $\forall \alpha \big(T \left(\ulcorner \neg \alpha \urcorner \right) \leftrightarrow \neg T \left(\ulcorner \alpha \urcorner \right) \big)$
T3. $\forall \alpha \forall \beta \big(T \left(\ulcorner \alpha \wedge \beta \urcorner \right) \leftrightarrow T \left(\ulcorner \alpha \urcorner \right) \wedge T \left(\ulcorner \beta \urcorner \right) \big)$
T4. $\forall \alpha(x) \big(T \left(\ulcorner \forall x \alpha(x) \urcorner \right) \leftrightarrow \forall x T \left(\ulcorner \alpha(\dot{x}) \urcorner \right) \big)$

Si ricordi l'utilizzo delle variabili puntate; di fatto l'ultima clausola viene ad affermare che la verità di una proposizione quantificata è equivalente alla verità di $\alpha(\overline{n})$ per tutti i numerali.

CT è sensibilmente più potente rispetto alle teorie decitazionali *TB* e *UTB* che abbiamo visto nel capitolo precedente. In particolare, come si evince facilmente dalla presentazione formalizzata che abbiamo fornito, permette quelle generalizzazioni che sono, invece, precluse alle teorie decitazionali. Per esempio, da *CT* è immediatamente derivabile il principio di non contraddizione:

(NC) $\forall x \forall y \big(x = neg(y) \rightarrow \neg (Tx \wedge Ty) \big)$

Dimostrazione. Da T2, abbiamo: $\forall x \forall y(x = neg(y) \to Tx \leftrightarrow \neg Ty)$. Quindi, $\forall x \forall y(x = neg(y) \to (Tx \to \neg Ty))$. Ma per trasformazione dei connettivi, $\forall x \forall y(x = neg(y) \to \neg(Tx \wedge Ty))$.

Teorema 10. $CT \vdash TB$

CT è in grado di derivare tutte le istanze del T-schema e quindi abbiamo che $CT \vdash TB$. In realtà è sufficiente una sottoteoria di *CT* per ottenere il risultato desiderato; in effetti, il teorema 10 è ottenibile già in *CT* ristretta dal momento che è sufficiente, compiere un'induzione sulla complessità della formula che compare nel T-schema. Segue uno schizzo di dimostrazione:

Base.
$\alpha \equiv P^n x_1, \ldots, x_n$
La base segue direttamente da T1. Infatti, abbiamo che:
$CT \vdash T(\ulcorner P^n x_1, \ldots, x_n \urcorner) \leftrightarrow P^n(val(x_1), \ldots, val(x_n))$.

Passo.

Prendiamo il caso della negazione a titolo esemplificativo. Gli altri casi sono analoghi. Vogliamo dimostrare che $CT \vdash T(\ulcorner \neg\alpha \urcorner) \leftrightarrow \neg\alpha$. Per ipotesi induttiva, abbiamo che $CT \vdash T(\ulcorner \alpha \urcorner) \leftrightarrow \alpha$, con α atomica. Quindi, per logica, $CT \vdash \neg T(\ulcorner \alpha \urcorner) \leftrightarrow \neg\alpha$. Da T2, si ha che: $CT \vdash T(\ulcorner \neg\alpha \urcorner) \leftrightarrow \neg T(\ulcorner \alpha \urcorner)$. Concatenando, $CT \vdash T(\ulcorner \neg\alpha \urcorner) \leftrightarrow \neg\alpha$.

Una questione piuttosto interessante è se l'adozione di una teoria della verità come *CT* sia, di per sé, incompatibile con una posizione deflazionista. Sicuramente, *CT* viola la tesi II, poiché il contenuto della nozione di verità eccede quello dei bicondizionali. Tuttavia, *CT* permette, a differenza di *TB* e *UTB*, di ottenere generalizzazioni altamente informative, come sarà chiaro a breve. E, in questo senso, risponde pienamente alla tesi III. Ma quanto è forte *CT*? Sembra, infatti, che la plausibilità deflazionista di questa teoria dipenda anche dalla sua portata. Esiste una risposta precisa a questa domanda e sarà il primo risultato importante che dimostreremo in questo capitolo: la teoria assiomatica composizionale della verità è equivalente a un particolare sottosistema di aritmetica del secondo ordine (*ACA*). Il secondo risultato che ci interessa riguarda invece una sottoteoria di *CT*, e cioè $CT\restriction$ (*CT* ristretta), ed è una possibile risposta alla nostra prima questione. $CT\restriction$ ristretta è infatti deflazionisticamente accettabile, almeno secondo una certa interpretazione, perché risulta conservativa su *PA*.

2. *Definizione del predicato di verità nell'aritmetica del secondo ordine*

Il primo risultato che dimostreremo in questa sezione è la derivazione degli assiomi di *CT* nel sottosistema di aritmetica del secondo ordine *ACA*. Si tratta di un teorema interessante perché mette in luce l'effettiva potenza deduttiva di *CT*; inoltre è rilevante anche da un punto di vista filosofico perché, data l'equivalenza formale della teoria della verità assiomatica tarskiana e di un sistema di aritmetica del secondo ordine, si tratta di capire se e dove si possono riscontrare differenze tra i due.

Non sarà un compito piuttosto semplice; dovremo, infatti, introdurre rapidamente la nozione di sottosistema di aritmetica del secondo ordine, definire le proprietà di alcuni particolari insiemi e infine mostrare come l'esistenza di tali insiemi, dimostrabile in *ACA*, equivale alla dimostrazione dell'esistenza dell'estensione del predicato T tarskiano[1].

2.1. *Sottosistemi di aritmetica del secondo ordine*

Grazie alle ricerche pionieristiche di Harvey Friedman e poi all'enciclopedico lavoro di Stephen Simpson (1999), è ormai nota l'importanza nel dibattito sulle ricerche fondazionali dei sottosistemi di aritmetica del secondo ordine. Assumendo *PA* come teoria di partenza per le indagini in teoria della dimostrazione, si possono imboccare due sentieri diametralmente opposti. Si può infatti diminuire la potenza deduttiva di *PA*, tramite il decremento del principio di induzione, fino ad arrivare alla teoria *Q*, considerata uno dei limiti inferiori delle teorie aritmetiche non banali, oppure si può procedere in maniera diametralmente opposta e aumentare la potenza di *PA* fino a giungere alla sua versione nel linguaggio del secondo ordine PA_2. In ciò che segue ci interesserà un particolare sottosistema di PA_2, *ACA* (acronimo per *Arithmetical Comprehension Axiom*)

1 Nella letteratura la dimostrazione più nota è quella presentata da Takeuti (1987). Halbach (2011) ripropone la dimostrazione arricchendola con qualche dettaglio in più. Nel seguito ci siamo richiamati a questi due lavori aggiungendo alcuni passaggi lasciati impliciti nelle trattazioni precedenti.

nel quale otterremo la definizione del predicato di verità[2]. Per comprendere come effettivamente funziona un sottosistema di aritmetica del secondo ordine e quali sono le sue 'ragioni' concettuali non c'è miglior modo che presentarne uno.

Il linguaggio di *ACA*, $\ell(ACA)$, presenta due insiemi di variabili:

Variabili del primo ordine: $x_1, x_2, x_3, ...$
Variabili del secondo ordine: $X_1, X_2, X_3, ...$

Una costante predicativa relazionale: $\in$

Le solite costanti non logiche aritmetiche: 0, s, +, ·

$t \in X$ è una formula di $\ell(ACA)$ se t è un termine e X è una variabile del secondo ordine.

L'apparato logico di *ACA* deve comprendere un'estensione delle regole del calcolo dei predicati del primo ordine dal momento che, avendo variabili di ordine superiore, avremo anche quantificatori di ordine superiore. Le regole di introduzione ed eliminazione dei quantificatori di or-

---

2 È comunque utile chiarire, seppur molto brevemente, le intuizioni che sono alla base del primo programma di ricerca esposto. Nello studio delle teorie più deboli di *PA* si cerca, normalmente, di associare l'impegno teorico e formale di questi sistemi con intuizioni di stampo finitista. Il problema filosofico è, infatti, quello di caratterizzare al meglio la nostra concettualizzazione del finito (e, in certa parte, anche dell'infinito), segnatamente di numero finito. Di qui, l'attenzione all'interpretazione attualistica o potenzialistica dei quantificatori - rispettivamente esistenziale e universale - che richiama alcune delle considerazioni elaborate già da Hilbert ma più compiutamente dai suoi allievi Bernays e Gentzen. Il focus è ovviamente il principio di induzione perché i vari sistemi (PRA, I-Σ_1, I-Σ_2, e così via) si differenziano proprio per la complessità delle formule ammesse nell'induzione. Un caso limite è costituito da Q (aritmetica di Robinson) in cui manca il principio di induzione e costituisce il limite inferiore perché una teoria sia in grado di rappresentare la propria sintassi, essendo così la teoria più debole entro la quale sia possibile dimostrare il primo teorema di Gödel. A tal proposito confronta (Hajek Pudlak 1993), (Buss 1998), (Galvan 2011), (Tait 1981), (Parsons 1983).

dine superiore sono, però, del tutto simili alle controparti per le variabili del primo ordine. Per esempio:

$$\frac{\forall X \alpha(X)}{\alpha(Z)}$$

è la regola di eliminazione del quantificatore universale, con X e Z variabili di ordine superiore[3].

Gli assiomi specifici di *ACA* comprendono tutti gli assiomi di *PA* e il principio di comprensione; quest'ultimo riveste un'importanza cruciale e merita, per questo motivo, un po' di approfondimento. La forma generale dell'assioma di comprensione è la seguente:

(CA) $\quad \exists X \forall x \big(Xx \leftrightarrow \alpha(x) \big)$

Dove la variabile X non deve comparire libera in α pena il sorgere di contraddizioni. L'idea di fondo di (CA) è dichiarare l'esistenza di predicati la cui estensione è determinata da una certa condizione esprimibile nel linguaggio e simbolizzata dalla formula α. È una questione puramen-

3 In più punti (cfr., ad esempio, Murawski 2005, Smith 2007, Simpson 1999) si insiste che i sistemi di aritmetica del secondo ordine sono in realtà teorie al primo ordine multi-sortali. In che senso? Sembrerebbe che la presenza di variabili e quantificatori del secondo ordine con le relative regole qualifichi indiscutibilmente questi sistemi come teorie del secondo ordine. Questo perché normalmente, in una teoria assiomatica, si distingue tra assiomi logici (o regole) e assiomi specifici. E in questo caso, la base logica è di ordine superiore. Tuttavia, la demarcazione non è così netta - come non lo è del resto la demarcazione tra ciò che è logico e ciò che non lo è. Un caso emblematico è il principio di comprensione: è logico oppure teorico? Una risposta articolata è impossibile in questa sede ma normalmente si assume che il principio di comprensione abbia uno statuto aritmetico (anche se potrebbero essere svolte considerazioni molto plausibili in favore di uno suo *status* logico). In questo senso, si può proseguire dicendo che anche le regole di eliminazione e introduzione dei quantificatori non sono logiche in senso stretto. E quindi, la teoria rimane al primo ordine ma presenta due tipi di variabili messe in relazione dal predicato $\in$. La questione rimane però aperta, in mancanza di chiare ragioni per considerare i quantificatori del secondo ordine non logici (o logici) non ci sentiamo di affermare con sicurezza che si tratti di teorie del primo ordine multi-sortali. In fondo, si tratta di un altro aspetto della questione se la logica del secondo ordine sia davvero logica.

te notazionale (per ora) scrivere (CA) sostituendo la relazione di predicazione tra variabili del secondo ordine e variabili del primo ordine con la relazione di appartenenza:

(CA') $$\exists X \forall x \big(x \in X \leftrightarrow \alpha(x) \big)$$

Il punto fondamentale riguarda, invece, la complessità della formula α; questa infatti definisce, in un certo senso, l'ammissibilità (o meno) di determinati insiemi nel dominio inteso della teoria. Prendiamo, a scopo esemplificativo, una condizione estremamente semplice:

(1) $$\exists X \forall x \big(x \in X \leftrightarrow x > 0 \wedge x < 12 \big)$$

In questo caso, (1) dichiara l'esistenza di un insieme costituito dai numeri (naturali) maggiori di 0 e minori di 12. Ma le condizioni possono essere molto più complesse e, ovviamente, più interessanti. In particolare è possibile introdurre anche i quantificatori nella specificazione della condizione. Ci si può chiedere, in buona sostanza, se qualunque formula ben formata del linguaggio dell'aritmetica del secondo ordine possa entrare come condizione per l'esistenza di un insieme. Se rispondiamo affermativamente ammettiamo la massima potenza possibile per un sistema di aritmetica del secondo ordine, e cioè la teoria PA_2. Oltre alla generalità del principio di comprensione, un ruolo significativo lo gioca anche in questo caso, il principio di induzione:

(Ind) $$\forall X \Big(\big(X0 \wedge \forall x \big(Xx \rightarrow Xs(x) \big) \big) \rightarrow \forall x Xx \Big)$$

A differenza di *PA* espressa al primo ordine, in (Ind) non troviamo una variabile metateorica per formule ma l'indicazione (teorica) di una variabile del secondo ordine. Naturalmente, anche in questo caso, possiamo essere più o meno liberali nell'ammettere determinate proprietà nel nostro principio. Dalla interazione tra principio di induzione e assioma di comprensione è possibile stabilire la potenza deduttiva dei sistemi di aritmetica del secondo ordine. Abbiamo visto già il limite massimo, costituito da PA_2; quali allora i livelli intermedi e soprattutto quali le ragioni per fermarsi a un certo punto?

Rispondere a questa domanda in maniera esauriente implica affrontare uno dei grandi temi della filosofia della matematica e cioè la complessa questione della predicatività. Non possiamo soffermarci su questo punto ma proveremo a riassumerlo brevemente per motivare filosoficamente l'adozione di una teoria come *ACA*. Nel definire un determinato insieme di numeri si possono seguire due intuizioni filosofiche opposte: da un lato si può assumere un'impostazione realista (di stampo platonista) secondo cui gli insiemi sono già dati e la nostra attività definitoria è, in ultima analisi, una descrizione di essi. Secondo quest'ottica ha perfettamente senso ammettere insiemi arbitrari e infiniti perché si tratta di riconoscere strutture esistenti in maniera indipendente dall'attività del matematico e dal contesto teorico di riferimento. È questa la strada intrapresa da Gödel nella discussione della filosofia della matematica di Russell (1944) che infatti si esprime così:

> If, however, it is a question of objects that exist independently of our constructions, then there is nothing in the least absurd in the existence of totalities containing members, which can be described (i.e., uniquely characterized) only by reference to this totality. (Gödel 1944, 456)

Tuttavia chi non si sente a suo agio con l'impegnativa metafisica realista può concepire l'attività di definizione come una costruzione, per quanto idealizzata. In questo senso, definire equivale a porre in essere e non a descrivere qualcosa che è già dato. L'alternativa anti-realista presenta poi al suo interno una duplice opzione: si può mantenere l'assoluta libertà nella scelta degli insiemi, che saranno così completamente arbitrari oppure indicare una sorta di regola alla quale gli elementi degli insiemi devono obbedire. Nel primo caso, gli insiemi saranno il frutto di una serie di scelte, e quindi oltre che *arbitrari* saranno *finiti*, dal momento che l'idea stessa di un insieme infinito di scelte libere è filosoficamente sfuggente. Nel secondo caso, perdiamo la completa arbitrarietà - gli elementi dovranno infatti sottostare a una regola, cioè a una condizione di appartenenza - ma potremmo avere a che fare con insiemi *infiniti*, dal momento che siamo esentati dallo specificare ogni singolo elemento che appartiene all'insieme.

L'idea di fondo che accomuna però sia l'intuizione che privilegia l'arbitrarietà delle scelte sia quella che favorisce l'infinità degli assiomi è quella anti-realista e cioè che la definizione è un processo di costruzione. E come non è possibile costruire una casa, dando per scontata l'esistenza del tetto, quando questo non esiste ancora, così non è possibile definire degli insiemi dando per scontata l'esistenza della totalità degli insiemi. Ciò comporta formalmente la limitazione del principio di comprensione. Come è facile rendersi conto, infatti, il quantificatore esistenziale varia sul dominio di tutti gli insiemi che, secondo l'approccio predicativista, deve essere considerato *in fieri,* solo potenzialmente dato. Non è possibile, quindi, nello specificare le condizioni di esistenza di un insieme far riferimento a tutti gli insiemi perché 'tutti gli insiemi' non ci sono ma li stiamo appunto costruendo. Tradotto formalmente, questo requisito, equivale al non ammettere una quantificazione di ordine superiore nella formula del principio di comprensione. Infatti:

(CA-Imp) $\exists X \forall x \big(X \in x \leftrightarrow \forall Z \forall Y \forall y \exists x (Zy \rightarrow Yx \wedge \neg Zx)\big)$

In (CA-Imp) l'estensione dell'insieme X è definita mediante il riferimento agli insiemi Z e Y. Ma quale sarà il dominio di quantificazione inteso per questi quantificatori quando l'esistenza stessa dell'insieme X è data dal principio in questione?

Definizioni di questo tipo sono dette, per l'appunto, impredicative e sono considerate problematiche da coloro che hanno riserve circa la realtà degli oggetti matematici che compaiono nelle nostre teorie. La restrizione più naturale è allora quella di ammettere solo quantificazioni numeriche (cioè del primo ordine) nel principio di comprensione. Saranno definibili, ovvero costruibili, solo insiemi aritmetici, da cui Arithmetical Comprehension Axiom:

(CA-Pred) $\exists X \forall x \big(x \in X \leftrightarrow \alpha(x)\big)$, dove $\alpha(x)$ contiene solo quantificatori del primo ordine.

Un aspetto interessante di *ACA* è la possibilità di ammettere variabili del secondo ordine non quantificate all'interno del principio di comprensione, chiamate normalmente parametri:

(CA-Param) $\exists X \forall x \big(x \in X \leftrightarrow \alpha(x, Y)\big)$, dove Y è, per l'appunto, un parametro.

Vedremo che, nella definizione del predicato di verità in *ACA*, questo tratto sarà essenziale.

L'altra caratteristica essenziale di *ACA* è, come dicevamo, il principio di induzione. In questo caso, però, non vi è nessuna restrizione di sorta: la formula induttiva può possedere una complessità qualsiasi. Da un punto di vista filosofico, quindi, *ACA* sottolinea l'importanza ontologica del principio di comprensione ma non del principio di induzione. In effetti, mentre il primo è un'affermazione di esistenza il secondo ha la struttura di un condizionale; qualora la proprietà designata dal linguaggio fosse troppo complessa e senza un riscontro nel dominio inteso, la generalizzazione induttiva risulterebbe banalmente vera. Ma la potenza dell'induzione di *ACA* è necessaria per la dimostrazione del predicato di verità. Anzi, possiamo dire che è proprio lì la chiave di tutta l'argomentazione; torneremo su questo punto dopo aver esibito la dimostrazione.

2.2. Definizione del predicato di verità per PA in ACA

Teorema 10. In *ACA* è possibile definire il predicato *T* di *CT*.[4]

Dimostrazione. Il teorema[5] che ci approntiamo a dimostrare fa vedere come nel contesto di *ACA* sia possibile definire un predicato di verità per *PA*. Dal risultato di indefinibilità di Tarski segue che *ACA* è essenzialmente più potente di *PA* e infatti *ACA* riesce a dimostrare il principio di riflessione per *PA*. Dimostrare un predicato di verità significa derivare le condizioni T1-T4 di *CT* come teoremi di *ACA*. Prima di fare questo è necessario però ottenere una definizione esplicita di T in *ACA*. Vediamo dunque la struttura della dimostrazione:

1. Dimostrare che esiste un insieme che contiene esattamente i codici degli enunciati atomici veri, cioè degli enunciati di complessità logica minima.

2. Dimostrare che, dato l'insieme che contiene i codici degli enunciati veri di complessità n, è possibile definire l'insieme di tutti gli enunciati veri di complessità n+1.

3. Ma poiché (i) corrisponde alla base e (ii) al passo, si può concludere, per induzione, che per tutti i gradi di complessità, esiste un insieme di (codici di) enunciati veri.

---

4 Si può scrivere l'enunciato del teorema 10 come $ACA \vdash CT$; si tratta di una formulazione non del tutto rigorosa ma efficace.

5 Per la dimostrazione di questo teorema ci siamo riferiti alla trattazione di Halbach (2011, 111ss). Halbach cita anche la nota dimostrazione (peraltro solo abbozzata) di Takeuti (1987) e uno schizzo di dimostrazione - non pubblicata - di Feferman.

2.2.1. *Definizione di Tset*

Tset è un predicato relazionale di verità. $Tset(X,n)$ dice che X è l'insieme degli enunciati veri di complessità logica pari al massimo a n. La complessità logica di una formula è il numero dei simboli logici presenti in essa. Per simboli logici intendiamo, nel presente contesto, solo connettivi e quantificatori, senza considerare l'identità[6]. Per contare i simboli logici di una formula abbiamo bisogno di una funzione apposita.

2.2.1.1. *Definizione di lh*

La funzione *lh* è una funzione che applicata a una formula del linguaggio restituisce il numero dei simboli logici della formula. La definizione è induttiva:

Base. $\alpha \equiv P^n(x_1,...,x_n);\ \ lh(\alpha)=0$

Passo. $\alpha \equiv \neg\beta\ ,\ lh(\alpha)=lh(\beta)+1$
$\alpha \equiv \beta \wedge \gamma\ ,\ lh(\alpha)=lh(\beta)+lh(\gamma)+1$
$\alpha \equiv \forall x\beta\ ,\ lh(\alpha)=lh(\beta)+1$

A questo punto possiamo in effetti definire *Tset*:

$$Tset(X,n) \leftrightarrow$$
$$\forall x \begin{pmatrix} x = \ulcorner s = t \urcorner \rightarrow \big(x \in X \leftrightarrow val(s) = val(t)\big) \wedge \\ x = neg(y) \wedge Sent(x) \wedge lh(x) \leq n \rightarrow \big(x \in X \leftrightarrow y \notin X\big) \wedge \\ x = conj(y,z) \wedge Sent(x) \wedge lh(x) \leq n \rightarrow \big(x \in X \leftrightarrow \big(y \in X \wedge z \in X\big)\big) \wedge \\ x = all(v,y) \wedge Sent(x) \wedge lh(x) \leq n \rightarrow \big(x \in X \leftrightarrow \forall t(so(y,t,v)) \in X\big) \end{pmatrix}$$

La definizione di *Tset* presenta quattro clausole; si tratta di condizionali, per cui nel momento in cui x, cioè il codice della formula in questione, non corrisponde all'antecedente di una condizione, l'implicazione è banalmente vera. Vediamo quindi nel dettaglio le clausole:

---

6 Cioè diremo che un enunciato di forma $t = s$ ha complessità logica pari a 0. In realtà quand'anche si assumesse che l'enunciato avesse complessità logica 1 ciò non cambierebbe la situazione dal momento che ciò che è importante è che gli enunciati atomici abbiano la complessità logica minima.

1. Nel primo caso, se x è il codice di una identità, diremo che x è vero se e solo se il valore del termine s è uguale al valore del termine t[7].
2. Se x è la negazione di una formula, per esempio $\neg\alpha$, diremo che x è vero se e solo se α non è vera, cioè non appartiene a X.
3. Lo stesso discorso vale per la congiunzione; una congiunzione è vera se e solo se lo sono entrambi i congiunti.
4. Il caso della quantificazione universale presenta qualche tecnicismo in più. Se x è il codice di una formula $\forall v\alpha$ allora diremo che x è vero se e solo se, per ogni t, è vero $\alpha(t)$.

Lemma 1

Dimostriamo ora il seguente lemma:

$$Tset(X,n) \wedge Tset(Y,k) \wedge n \leq k \to \forall x\big(lh(x) \leq n \to (x \in X \leftrightarrow x \in Y)\big)$$

Il significato è abbastanza intuitivo. Dato un certo grado di complessità n, gli insiemi che soddisfano $Tset(X,n)$ sono identici. L'idea di fondo è che per ogni livello di complessità l'insieme degli enunciati veri è unico. La dimostrazione avviene per induzione sulla complessità della formula. L'assunzione generale è $Tset(X,n) \wedge Tset(Y,k) \wedge n \leq k$.

Base. È il caso in cui x è il codice di un'identità vera.

H:	$lh(\ulcorner s = t \urcorner) \leq n$
Dem:	$\ulcorner s = t \urcorner \in X \leftrightarrow \ulcorner s = t \urcorner \in Y$

$lh(\ulcorner s = t \urcorner) \leq n$	ipotesi
$\ulcorner s = t \urcorner \in X \leftrightarrow val(s) = val(t)$	da def. di $Tset(X,n)$
$\ulcorner s = t \urcorner \in Y \leftrightarrow val(s) = val(t)$	da def. di $Tset(Y,n)$
$\ulcorner s = t \urcorner \in X \leftrightarrow \ulcorner s = t \urcorner \in Y$	logica

Passo. Il passo ha la seguente struttura:

$$\big(lh(x) = j, j < n \to (x \in X \leftrightarrow x \in Y)\big) \to$$
$$\big(lh(x) = j+1, j \leq n \to (x \in X \leftrightarrow x \in Y)\big)$$

---

7 In ciò che segue ci riferiremo indifferentemente agli enunciati o ai codici che li denominano.

Ovvero, se a un certo livello di complessità j (minore di n) gli insiemi coincidono, allora coincideranno anche al livello j+1. Ma x è il codice di una formula e quindi può essere una negazione, una congiunzione o una quantificazione universale. Dimostriamo il caso della negazione e della congiunzione; la quantificazione universale è analoga.

Negazione.

H: $lh(\neg\alpha) = j+1, j+1 \leq n$
Dem: $\ulcorner\neg\alpha\urcorner \in X \leftrightarrow \ulcorner\neg\alpha\urcorner \in Y$

Ad $\rightarrow$)

$\ulcorner\neg\alpha\urcorner \in X \rightarrow \ulcorner\alpha\urcorner \notin X$	def. di *Tset*
$lh(\alpha) = lh(\neg\alpha) - 1$	da def. di *lh*
$lh(\alpha) = j+1-1$	da H
$lh(\alpha) = j$	aritmetica
$\ulcorner\alpha\urcorner \notin X \leftrightarrow \ulcorner\alpha\urcorner \notin Y$	def. di *Tset* e logica
$\ulcorner\neg\alpha\urcorner \in X \rightarrow \ulcorner\alpha\urcorner \notin Y$	logica
$\ulcorner\neg\alpha\urcorner \in X \rightarrow \ulcorner\neg\alpha\urcorner \in Y$	def. di *Tset*

Ad $\leftarrow$)

$\ulcorner\neg\alpha\urcorner \in Y \rightarrow \ulcorner\alpha\urcorner \notin Y$	def. di *Tset*
$lh(\alpha) = lh(\neg\alpha) - 1$	da def. di *lh*
$lh(\alpha) = j+1-1$	da H
$lh(\alpha) = j$	aritmetica
$\ulcorner\alpha\urcorner \notin X \leftrightarrow \ulcorner\alpha\urcorner \notin Y$	def. di *Tset* e logica
$\ulcorner\neg\alpha\urcorner \in Y \rightarrow \ulcorner\alpha\urcorner \notin X$	logica
$\ulcorner\neg\alpha\urcorner \in Y \rightarrow \ulcorner\neg\alpha\urcorner \in X$	def. di *Tset*

Congiunzione.

H: $lh(\alpha \wedge \beta) = j+1, j+1 \leq n$
Dem: $\ulcorner\alpha \wedge \beta\urcorner \in X \leftrightarrow \ulcorner\alpha \wedge \beta\urcorner \in Y$

Ad $\rightarrow$)

$\ulcorner\alpha \wedge \beta\urcorner \in X \leftrightarrow \ulcorner\alpha\urcorner \in X \wedge \ulcorner\beta\urcorner \in X$	def. di *Tset*
$\ulcorner\alpha \wedge \beta\urcorner \in X \leftrightarrow \ulcorner\alpha\urcorner \in X$	E$\wedge$
$\ulcorner\alpha \wedge \beta\urcorner \in X \leftrightarrow \ulcorner\beta\urcorner \in X$	E$\wedge$

$lh(\alpha) < lh(\alpha \wedge \beta)$	def. di *lh*
$\ulcorner\alpha\urcorner \in X \to \ulcorner\alpha\urcorner \in Y$	per ipotesi induttiva
$\ulcorner\beta\urcorner \in X \to \ulcorner\beta\urcorner \in Y$	per ipotesi induttiva
$\ulcorner\alpha\urcorner \in X \wedge \ulcorner\beta\urcorner \in X \to \ulcorner\alpha\urcorner \in Y \wedge \ulcorner\beta\urcorner \in Y$	logica
$\ulcorner\alpha \wedge \beta\urcorner \in X \to \ulcorner\alpha \wedge \beta\urcorner \in Y$	logica, def. di *Tset*

Ad $\leftarrow$)

$\ulcorner\alpha \wedge \beta\urcorner \in Y \leftrightarrow \ulcorner\alpha\urcorner \in Y \wedge \ulcorner\beta\urcorner \in Y$	def. di *Tset*
$\ulcorner\alpha \wedge \beta\urcorner \in Y \to \ulcorner\alpha\urcorner \in Y$	logica
$\ulcorner\alpha \wedge \beta\urcorner \in Y \to \ulcorner\beta\urcorner \in Y$	logica
$lh(\alpha) < lh(\alpha \wedge \beta)$	def. di *lh*
$\ulcorner\alpha\urcorner \in Y \to \ulcorner\alpha\urcorner \in X$	ipotesi induttiva
$\ulcorner\beta\urcorner \in Y \to \ulcorner\beta\urcorner \in X$	ipotesi induttiva
$\ulcorner\alpha \wedge \beta\urcorner \in Y \to \ulcorner\alpha\urcorner \in X$	logica
$\ulcorner\alpha \wedge \beta\urcorner \in Y \to \ulcorner\beta\urcorner \in X$	logica
$\ulcorner\alpha \wedge \beta\urcorner \in Y \to \ulcorner\alpha\urcorner \in X \wedge \ulcorner\beta\urcorner \in Y$	logica
$\ulcorner\alpha \wedge \beta\urcorner \in Y \to \ulcorner\alpha \wedge \beta\urcorner \in X$	def. di *Tset*

Quindi si ottiene,

$$Tset(X,n) \wedge Tset(Y,k) \wedge n \leq k \to$$
$$\forall x\big(lh(x) \leq n \to \big(x \in X \leftrightarrow x \in Y\big)\big)$$

Poiché non abbiamo compiuto ipotesi particolari circa la complessità della formula, il risultato è generalizzabile:

Lemma 2
$\forall n \forall x \forall X \forall Y \big(Tset(X,n) \wedge Tset(Y,n) \to \forall x\big(x \in X \leftrightarrow x \in Y\big)\big)$

A questo punto abbiamo tutto ciò che ci permette di derivare il teorema fondamentale e cioè che $ACA \vdash \forall n \exists X\big(Tset(X,n)\big)$; ovvero, per ogni grado di complessità della formula in questione, *ACA* dimostra l'esistenza dell'insieme di enunciati veri di quella complessità. La dimostrazione è per induzione.

Base. $ACA \vdash \exists X\big(Tset(X,0)\big)$

L'idea della dimostrazione è piuttosto semplice. Dal principio di comprensione di *ACA* riusciamo ad affermare l'esistenza di un insieme che contiene tutte le identità vere. Ma questo non è sufficiente; bisogna infatti dimostrare che le uniche formule atomiche, cioè le uniche formule con complessità pari a 0, sono le identità. Vediamo nel dettaglio la dimostrazione.

Assumiamo che $\forall x\left(x \in X \leftrightarrow \exists s \exists t\left(x = \ulcorner s = t \urcorner \wedge val(s) = val(t)\right)\right)$. Per logica abbiamo $x \in X \leftrightarrow \exists s \exists t\left(x = \ulcorner s = t \urcorner \wedge val(s) = val(t)\right)$; sciogliamo il bicondizionale: $x \in X \rightarrow \exists s \exists t\left(x = \ulcorner s = t \urcorner \wedge val(s) = val(t)\right)$ e $\exists s \exists t\left(x = \ulcorner s = t \urcorner \wedge val(s) = val(t)\right) \rightarrow x \in X$; assumiamo $x \in X$ e per modus ponens otteniamo $\exists s \exists t\left(x = \ulcorner s = t \urcorner \wedge val(s) = val(t)\right)$; per logica dei quantificatori abbiamo $\exists s \exists t\left(x = \ulcorner s = t \urcorner\right) \wedge \exists s \exists t\left(val(s) = val(t)\right)$. A questo punto assumiamo $Sent(x) \wedge lh(x) \leq 0 \leftrightarrow \exists s \exists t\left(x = \ulcorner s = t \urcorner\right)$; quindi $Sent(x) \wedge lh(x) \leq 0$, per modus ponens, e per introduzione della congiunzione $Sent(x) \wedge lh(x) \leq 0 \wedge \exists s \exists t\left(x = \ulcorner s = t \urcorner \wedge val(s) = val(t)\right)$. Quindi $x \in X \rightarrow Sent(x) \wedge lh(x) \leq 0 \wedge \exists s \exists t\left(x = \ulcorner s = t \urcorner \wedge val(s) = val(t)\right)$ per scaricamento dell'assunzione precedente; in base alla prima assunzione otteniamo $\exists s \exists t\left(x = \ulcorner s = t \urcorner \wedge val(s) = val(t)\right) \rightarrow x \in X$ e per logica $Sent(x) \wedge lh(x) \leq 0 \wedge \exists s \exists t\left(x = \ulcorner s = t \urcorner \wedge val(s) = val(t)\right) \rightarrow x \in X$. Quindi, $x \in X \leftrightarrow Sent(x) \wedge lh(x) \leq 0 \wedge \exists s \exists t\left(x = \ulcorner s = t \urcorner \wedge val(s) = val(t)\right)$ da cui si ha:

$$\exists X \forall x \begin{pmatrix} x \in X \leftrightarrow Sent(x) \wedge lh(x) \leq 0 \wedge \\ \exists s \exists t\left(x = \ulcorner s = t \urcorner \wedge val(s) = val(t)\right) \end{pmatrix}$$

e cioè $\exists X\left(Tset(X, 0)\right)$. Questo è il risultato che volevamo ottenere; tuttavia nella dimostrazione abbiamo fatto uso di due assunzioni:

A_1: $\forall x\left(x \in X \leftrightarrow \exists s \exists t\left(x = \ulcorner s = t \urcorner \wedge val(s) = val(t)\right)\right)$

A_2: $Sent(x) \wedge lh(x) \leq 0 \leftrightarrow \exists s \exists t\left(x = \ulcorner s = t \urcorner\right)$

Ora, A_1, è un'istanza del principio di comprensione di *ACA*. Mentre A_2 dipende dalla codifica aritmetica del linguaggio di *PA*.

Sia $Lib(x)$ una funzione che ha come argomenti formule e come valori il numero delle variabili libere in esse. Pertanto avremo che $Sent(x) \leftrightarrow Form(x) \wedge Lib(x) = 0$: gli enunciati sono formule chiuse del linguaggio. Assumiamo ora $Sent(x)$ e $lh(x) \leq 0$; poiché la complessità è pari a 0, abbiamo che $x = \ulcorner s = t \urcorner$. Viceversa assumendo proprio $x = \ulcorner s = t \urcorner$, abbiamo che $Sent(x)$, dal momento che, per ipotesi, i termini t e s sono chiusi e quindi $Lib(t = s) = 0$ e, per definizione, $lh(x) \leq 0$. Abbiamo così, $Sent(x) \wedge lh(x) \leq 0 \leftrightarrow \exists s \exists t\left(x = \ulcorner s = t \urcorner\right)$. Questo risultato sfrutta le funzioni di aritmetizzazione e quindi è sufficiente *PA* per ottenerlo. Ma poiché $PA \subseteq ACA$ segue che:

$$ACA \vdash Sent(x) \wedge lh(x) \leq 0 \leftrightarrow \exists s \exists t \left(x = \ulcorner s = t \urcorner\right)$$

Passo. L'ipotesi induttiva del passo è che $\exists X Tset(X, n)$. Chiamiamo questo insieme *Y*; *Y* è quindi l'insieme delle formule vere con complessità pari a n. Definiamo a questo punto un insieme *X* che contiene tutte le formule di *Y* e inoltre contiene altre formule secondo le clausole seguenti:

1. Se la formula α (di complessità n) non appartiene a *Y* (cioè non è vera), la negazione di α (ovvero $\neg\alpha$, di complessità $n+1$) appartiene a X.

2. Se le formule α e β (entrambe di complessità n) appartengono a *Y* allora apparterrà a *X* la congiunzione $\alpha \wedge \beta$, di complessità $n+1$.

3. Se, infine, per tutti i termini t, $\alpha(t) \in Y$ allora apparterrà a *X* la formula $\forall x \alpha$, di complessità $n+1$.

Definiamo pertanto l'insieme *X* in questo modo:

$$\forall x \left(\begin{array}{l} x \in X \leftrightarrow \left[Sent(x) \wedge lh(x) \leq n \wedge x \in Y\right] \vee \\ \left[\begin{array}{l} Sent(x) \wedge lh(x) = n+1 \wedge \\ \exists y \left(x = neg(y) \wedge y \notin Y\right) \vee \\ \exists y \exists z \left(x = conj(y,z) \wedge \left(y \in Y \wedge z \in Y\right)\right) \vee \\ \exists y \exists v \left(x = all(v,y) \wedge \forall t\, so(y,v,t) \in Y\right) \end{array} \right] \end{array} \right)$$

L'insieme *X* contiene tutti gli enunciati veri di complessità pari o minore di n che appartengono all'insieme *Y* e in più (e quindi è un'estensione) gli enunciati veri di complessità $n+1$.

Nota. Per renderci conto di questo vediamo il caso della congiunzione. Il passo dell'induzione afferma che se $\alpha \in Y$ e $\beta \in Y$, con α e β di complessità n, allora $\alpha \wedge \beta$ apparterrà a *X* se la complessità di $\alpha \wedge \beta$ è pari a $n+1$. Ma attenzione: non è detto che la congiunzione di due formule qualsiasi abbia una complessità pari a $n+1$. Anzi, questo è vero solo nel caso in cui almeno una delle due formule sia una formula atomica. Sia $lh(\alpha) = n$ e $lh(\beta) = 0$. Avremo che $lh(\alpha \wedge \beta) = lh(\alpha) + lh(\beta) + 1 = n + 1$. Che dire del caso in cui le formule hanno una complessità maggiore di 0? Ipotizziamo che $\alpha \in Y$ e $\beta \in Y$ e che α e β siano a loro volta due congiunzioni, $\alpha \equiv \alpha_1 \wedge \alpha_2$ e $\beta \equiv \beta_1 \wedge \beta_2$. Se $\alpha \in Y$, allora $\alpha_1 \in Y$ e $\alpha_2 \in Y$ e lo stesso valga per β_1 e β_2. Ora, ragioniamo sulle complessità: α è di complessità 1, mentre α_1 è di complessità 0. Per il passo dell'induzione, avremo che $\alpha \wedge \beta_1 \in Z$ dove Z è un insieme di complessità 2. Vediamo perché: $lh(\alpha \wedge \beta_1) = lh(\alpha) + lh(\beta_1) + 1 = 2$. Ma lo stesso ragionamento si può ripetere a proposito di *Z*, costruendo man mano *Tset* di complessità sempre più elevata.

L'insieme *X* è un'istanza del principio di comprensione di *ACA* e ciò conclude la dimostrazione. Si noti che la forma del principio è in questo caso: $ACA \vdash \exists X \forall x \big(x \in X \leftrightarrow \alpha(Y, x)\big)$ dove è essenziale la presenza del parametro *Y* nella formula di comprensione. Senza la possibilità di avere questi parametri, infatti, la dimostrazione non può essere portata a termine.

Un'ultima considerazione a proposito dei requisiti minimi per dimostrare $ACA \vdash \forall n \exists X \big(Tset(X, n)\big)$. Come abbiamo avuto modo di vedere, la dimostrazione si basa essenzialmente sull'induzione compiuta su una Σ^1_1- formula. Ciò è possibile perché l'assioma di induzione di *ACA* non presenta restrizioni particolari. Esiste però un sottosistema di aritmetica molto studiato in letteratura, ACA_0 dove il pedice 0 sta a indicare proprio la restrizione dell'assioma di induzione alle formule aritmetiche. In ACA_0 pertanto non è possibile compiere l'induzione necessaria alla dimostrazione del teorema.

È interessante vedere che, per ogni *n*, $ACA_0 \vdash \exists X \big(Tset(X, n)\big)$ ma non vale la generalizzazione universale *teorica*. Si tratta di un fenomeno analogo a quanto visto a proposito delle teorie decitazionali. Anche in quel caso, come si ricorderà, data una formula β, $TB \vdash T(\ulcorner \beta \urcorner) \vee T(\ulcorner \neg\beta \urcorner)$ mentre $TB \nvdash \forall \beta \big(T(\ulcorner \beta \urcorner) \vee T(\ulcorner \neg\beta \urcorner)\big)$.

2.3. Clausole tarskiane in ACA

L'ultima parte della nostra dimostrazione riguarda la possibilità di definire, entro *ACA*, le clausole tarskiane T1-T4. Innanzitutto, è possibile fornire la seguente definizione metateorica:

(2) $\quad Tx =_{def.} \exists Y \big(Tset(Y, lh(x)) \wedge x \in Y\big)$

Un enunciato (o meglio il codice di un enunciato) è vero se e solo se appartiene all'insieme degli enunciati 'veri' di complessità x. Le virgolette sono d'obbligo perché altrimenti si potrebbe pensare a una definizione circolare; in realtà la costruzione di *Tset*, come abbiamo visto non fa uso del predicato di verità (che non è nemmeno presente nel linguaggio di *ACA*). Rimane da discutere se e in quale misura la costruzione induttiva di *Tset* non faccia uso di una nozione di verità a livello base, cioè nel caso delle formule atomiche. Vediamo ora la derivazione delle clausole T1 e T2.

T1. $\forall s \forall t \left(T\left(\ulcorner s = t \urcorner \right) \leftrightarrow val(s) = val(t) \right)$

Dimostrare in *ACA* T1 equivale, in base alla definizione di *Tset*, a: $ACA \vdash \exists X \left(Tset(X, lh(s = t)) \wedge \ulcorner s = t \urcorner \in X \right) \leftrightarrow val(s) = val(t)$

Dimostrazione. La nostra ipotesi generale sia che $x = \ulcorner s = t \urcorner$ ovvero che la complessità della formula è minima. Dal lemma 2 sappiamo che ACA riesce a dimostrare l'esistenza di $Tset(X, 0)$ ovvero di $Tset(X, lh(s = t))$. Assumiamo quindi $x \in X$ e abbiamo immediatamente che $val(s) = val(t)$. Per l'altro verso, partiamo da $val(s) = val(t)$; ma in base a quanto detto prima, abbiamo che $Tset(X, lh(s = t))$ e quindi $x \in X$. Questo chiude la dimostrazione.

T2. $\forall x \left(Sent(x) \rightarrow \left(T(neg(x)) \leftrightarrow \neg Tx \right) \right)$

Dimostrazione. La dimostrazione è piuttosto impegnativa dal punto di vista tecnico anche se è intuitivamente chiara. Consta, nel dettaglio, di tre passaggi:

1. $\exists X \left(Tset(X, lh(neg(x))) \wedge neg(x) \in X \right) \leftrightarrow \exists X \left(Tset(X, lh(neg(x))) \wedge x \notin X \right)$

2. $\exists X \left(Tset(X, lh(neg(x))) \wedge x \notin X \right) \leftrightarrow \exists X \left(Tset(X, lh(x)) \wedge x \notin X \right)$

3. $\exists X \left(Tset(X, lh(x)) \wedge x \notin X \right) \leftrightarrow \neg \exists X \left(Tset(X, lh(x)) \wedge x \in X \right)$

Per logica, otteniamo immediatamente che:

$$\exists X \left(Tset(X, lh(neg(x))) \wedge neg(x) \in X \right) \leftrightarrow \neg \exists X \left(Tset(X, lh(x)) \wedge x \in X \right)$$

cioè la 'traduzione' in *ACA* di T2. Il problema è quindi quello di derivare le equivalenze (1) - (3). A titolo esemplificativo mostreremo solo il verso da sinistra a destra ma, come sarà chiaro a breve, ciò non comporta alcun problema dal momento che la struttura della dimostrazione è perfettamente speculare.

Ad (1). Partiamo dalla definizione di $Tset(X,n)$:

$$ACA, Tset\left(X, lh(neg(x))\right) \vdash$$
$$x = neg(y) \wedge Sent(x) \wedge lh(x) \leq lh(neg(x)) \rightarrow x \in X \leftrightarrow y \notin X$$

Per logica, otteniamo:

$$ACA, Tset\left(X, lh(neg(x))\right), x = neg(y), Sent(x), lh(x) \leq lh(neg(x)) \vdash$$
$$neg(y) \in X \leftrightarrow y \notin X$$

Assumiamo, a questo punto, $neg(y) \in X \vdash neg(y) \in X$ e scriviamo:

$$ACA, Tset(X, lh(neg(x))) \wedge neg(y) \in X \vdash$$
$$Tset(X, lh(neg(y))) \wedge y \notin X$$

Ma per sostituzione di y con x, abbiamo

$$ACA, Tset(X, lh(neg(x))) \wedge neg(x) \in X \vdash$$
$$Tset(X, lh(neg(x))) \wedge x \notin X$$

Infine, grazie all'introduzione dei quantificatori esistenziali predicativi (prima nel conseguente e poi nell'antecedente) otteniamo il risultato che ci interessa.

Ad (2). Partiamo dal lemma 1:

$$ACA \vdash Tset(X, lh(neg(x))) \wedge Tset(Y, lh(x)) \wedge lh(x) \leq lh(neg(x)) \rightarrow$$
$$\forall x\left(lh(x) \leq lh(neg(x)) \rightarrow (x \in X \leftrightarrow y \in X\right)$$

In base alla definizione di lh, otteniamo che $ACA \vdash lh(x) \leq lh(neg(x))$ e, quindi, $ACA, Tset(X, lh(neg(x))), Tset(Y, lh(x)) \vdash x \in X \leftrightarrow x \in Y$. Per $x \notin X \vdash x \notin X$, $ACA, Tset(X, lh(neg(x))) \wedge x \notin X, Tset(Y, lh(x)) \vdash x \notin Y$. Assumiamo poi $Tset(Y, lh(x)) \vdash Tset(Y, lh(x))$ e riscriviamo il tutto: $ACA, Tset(X, lh(neg(x))) \wedge x \notin X, Tset(Y, lh(x)) \vdash Tset(Y, lh(x)) \wedge x \notin Y$.

A questo punto possiamo chiudere esistenzialmente e abbiamo:

$$ACA, \exists X\left(Tset(X, lh(neg(x))) \wedge x \notin X\right), \exists X\left(Tset(X, lh(x))\right) \vdash$$
$$\exists X\left(Tset(Y, lh(x)) \wedge x \notin Y\right)$$

Poiché, *ACA* riesce a derivare l'esistenza di insiemi di enunciati veri di qualsiasi complessità (secondo il lemma (2)), otteniamo $ACA \vdash \exists X\big(Tset(X, lh(x))\big)$ e quindi:

$$ACA, \exists X\big(Tset(X, lh(neg(x))) \wedge x \notin X\big) \vdash \exists X\big(Tset(Y, lh(x)) \wedge x \notin Y\big)$$

come volevamo dimostrare.

Ad (3). L'ultima equivalenza si basa sull'utilizzo del lemma 2:

$ACA \vdash \forall n \forall x \forall X \forall Y\big(Tset(X, n) \wedge Tset(T, n) \rightarrow (x \in X \leftrightarrow x \in Y)\big)$, cioè $ACA, Tset(X, lh(x)), Tset(T, lh(x)) \vdash x \in X \leftrightarrow x \in Y$ per eliminazione dei quantificatori e logica. Assumiamo, quindi, $x \notin X \vdash x \notin X$ e, per logica, otteniamo $ACA, Tset(X, lh(x)) \wedge x \notin X, Tset(T, lh(x)) \vdash x \notin Y$. Per logica, $ACA, \exists X\big(Tset(X, lh(x)) \wedge x \notin X\big) \vdash \forall X\big(Tset(X, lh(x)) \rightarrow x \notin X\big)$ e, per trasformazione dei quantificatori, abbiamo quello che cercavamo: $ACA, \exists X\big(Tset(X, lh(x)) \wedge x \notin X\big) \vdash \neg\exists X\big(Tset(X, lh(x)) \wedge x \in X\big)$.

La struttura della dimostrazione è simile anche per le altre due clausole e cioè quella della congiunzione e quella della quantificazione universale. Nel primo caso, a titolo esemplificativo, le tre equivalenze avranno la forma seguente:

$$\text{(i)} \quad \begin{array}{l} \exists X\big(Tset(X, lh(conj(x, y))) \wedge conj(x, y) \in X\big) \leftrightarrow \\ \exists X\big(Tset(X, lh(conj(x, y))) \wedge (x \in X \wedge y \in X)\big) \end{array}$$

Anche in questo caso, da definizione di *Tset*, segue che $conj(x, y) \leftrightarrow x \in X \wedge y \in X$.

$$\text{(ii)} \quad \begin{array}{l} \exists X\big(Tset(X, lh(conj(x, y))) \wedge (x \in X \wedge y \in X)\big) \leftrightarrow \\ \exists X\big(Tset(X, lh(x)) \wedge (x \in X \wedge y \in X)\big) \end{array}$$

Poiché $lh(x) \leq lh(conj(x, y))$ segue che $Tset(X, lh(conj(x, y)))$ e $Tset(X, lh(x))$ coincidono per le formule di complessità minore o uguale a $lh(conj(x, y))$. Quindi, per il teorema dimostrato, $\exists X\big(Tset(X, lh(x))\big)$ e $x \in X \wedge y \in X$.

$$\text{(iii)}\quad \begin{array}{l}\exists X\big(Tset(X,lh(x))\wedge(x\in X\wedge y\in X)\big)\leftrightarrow\\ \exists X\big(Tset(X,lh(x))\wedge x\in X\big)\wedge\exists X\big(Tset(X,lh(x))\wedge y\in X\big)\end{array}$$

Da $\exists X\big(Tset(X,lh(x))\wedge(x\in X\wedge y\in X)\big)$ derivano, rispettivamente, $\exists X\big(Tset(X,lh(x))\wedge x\in X\big)$ e $\exists X\big(Tset(X,lh(x))\wedge y\in X\big)$. Per introduzione della congiunzione si ottiene il risultato voluto.

È interessante notare come nella derivazione delle clausole tarskiane in *ACA* si faccia un uso essenziale delle regole relative al segno logico indicato nella clausola. Così, per esempio, l'ultimo passaggio per T2 è una trasformazione di quantificatori, cioè una regola di negazione; in maniera analoga, in T3 si fa uso dell'introduzione della congiunzione. Ciò rispecchia in maniera perfetta, all'interno di *ACA*, l'impiego dei connettivi metateorici presenti nelle formulazioni usuali delle clausole semantiche tarskiane.

In conclusione, la possibilità di derivare la teoria tarskiana della verità (per *PA*, in questo caso) all'interno di un sottosistema di aritmetica del secondo ordine come *ACA* mette in luce due punti interessanti per la nostra discussione. In primo luogo il fatto che la teoria della verità non richiede un impegno ontologico proibitivo. L'assioma di comprensione è infatti limitato alle formule aritmetiche, che non presentano cioè quantificatori di ordine superiore. Ciò che invece è richiesto, in maniera essenziale, è il principio di induzione non ristretta: e infatti la formula su cui compiamo l'induzione descrive una proprietà di complessità pari a Σ^1_1. Proprio il principio di induzione sarà al centro dei paragrafi successivi, dove analizzeremo una versione di *CT* ristretta.

3. *CT e la consistenza dell'aritmetica*

Il secondo risultato tecnicamente significativo è che *CT* risulta essenzialmente non conservativa rispetto all'aritmetica di Peano. È infatti possibile dimostrare entro *CT* la consistenza di *PA*. In particolare, *CT* è in grado di derivare il principio di riflessione globale per *PA*,

$$\text{(RFN)}\qquad CT\vdash\forall x\big(\mathrm{Pr}_{PA}(\ulcorner\alpha(\dot{x})\urcorner)\rightarrow T(\ulcorner\alpha(\dot{x})\urcorner)\big)$$

Il significato di RFN è del tutto intuitivo: se una formula - o meglio la sua chiusura universale - è dimostrabile in *PA*, allora è vera. La portata, non solo logica ma anche filosofica dei principi di riflessione è un tema discusso a fondo nella letteratura contemporanea e verrà trattato nel prossimo capitolo. Delineiamo, ora, un abbozzo della dimostrazione di *RFN* in *CT*.

Teorema 11. $CT \vdash RFN$

Si tratta di una dimostrazione[8] piuttosto impegnativa: l'idea di fondo però è semplice. Cosa vuol dire, infatti, dimostrare la correttezza di *PA*? Significa far vedere che gli assiomi di *PA* sono veri e che tutto ciò che si riesce ad ottenere da essi tramite le regole del calcolo è vero anch'esso. La dimostrazione consta pertanto di tre parti:

1) mostrare che gli assiomi di *PA* sono veri
2) mostrare che lo schema di assiomi di induzione è vero
3) mostrare che la verità è preservata attraverso l'applicazione delle regole logiche.

Ad 1). Il primo punto è molto semplice da ottenere. *CT*, come si ricorderà, comprende *PA* e quindi, in base al T-schema, si ottiene la verità di ciascun assioma. Vediamo, come esempio, il primo assioma del successore: $PA \vdash \forall x\big(s(x) \neq 0\big)$. Ora, $CT \vdash \forall x\big(s(x) \neq 0\big)$ perché $PA \subset CT$. Ma, $CT \vdash \forall x\big(s(x) \neq 0\big) \leftrightarrow T\left(\ulcorner \forall x\big(s(x) \neq 0\big)\urcorner\right)$ per T-schema. Per logica, $CT \vdash T\left(\ulcorner \forall x\big(s(x) \neq 0\big)\urcorner\right)$. E così per gli altri assiomi di *PA*.

Ad 2). Per dimostrare che ogni istanza del principio di induzione è vera è necessario sfruttare il principio di induzione di *CT*. Abbiamo $CT \vdash \Big[T\left(\ulcorner \alpha(0)\urcorner\right) \wedge \forall x\Big(T\left(\ulcorner \alpha(\dot{x})\urcorner\right) \rightarrow T\left(\ulcorner \alpha(s(\dot{x}))\right)\Big)\Big] \rightarrow \forall x T\left(\ulcorner \alpha(\dot{x})\urcorner\right)$

Ora, in base alle clausole tarskiane, valgono le seguenti equivalenze:

$$CT \vdash \forall x T\left(\ulcorner \alpha(\dot{x})\urcorner\right) \leftrightarrow T\left(\ulcorner \forall x \alpha(x)\urcorner\right)$$
$$CT \vdash \Big(T\left(\ulcorner \alpha(\dot{x})\urcorner\right) \rightarrow T\left(\ulcorner \alpha(s(\dot{x}))\urcorner\right)\Big) \leftrightarrow T\left(\ulcorner \alpha(\dot{x}) \rightarrow \alpha(s(\dot{x}))\urcorner\right)$$

Otteniamo allora:

$$CT \vdash \Big[T\left(\ulcorner \alpha(0)\urcorner\right) \wedge \forall x\Big(T\left(\ulcorner \alpha(\dot{x}) \rightarrow \alpha(s(\dot{x}))\urcorner\right)\Big)\Big] \rightarrow T\left(\ulcorner \forall x \alpha(x)\urcorner\right)$$

---

8 Halbach (2011, 103ss) ne fornisce una presentazione esauriente; si rimanda anche a Takeuti (1987, 186) per una dimostrazione che impiega il calcolo dei sequenti.

Ma, sempre per la clausola della quantificazione universale:
$$CT \vdash \left[T\left(\ulcorner \alpha(0) \urcorner\right) \wedge T\left(\ulcorner \forall x\left(\alpha(x) \rightarrow \alpha(s(x))\right) \urcorner\right)\right] \rightarrow T\left(\ulcorner \forall x \alpha(x) \urcorner\right)$$

Per la clausola della congiunzione:
$$CT \vdash \left[T\left(\ulcorner \alpha(0) \wedge \forall x\left(\alpha(x) \rightarrow \alpha(s(x))\right) \urcorner\right)\right] \rightarrow T\left(\ulcorner \forall x \alpha(x) \urcorner\right)$$

Analogamente, per la clausola dell'implicazione:
$$CT \vdash T\left(\ulcorner \left[\alpha(0) \wedge \forall x\left(\alpha(x) \rightarrow \alpha(s(x))\right)\right] \rightarrow \forall x \alpha(x) \urcorner\right)$$

E questo conclude la dimostrazione.

Ad 3). Si tratta sicuramente del punto più complesso da dimostrare. Viene impiegata l'induzione sulla lunghezza delle dimostrazioni: da un punto di vista intuitivo ciò vuol dire che gli assiomi sono veri e se una formula, la cui dimostrazione ha una lunghezza pari a n è vera, sarà vera una formula la cui dimostrazione ha una lunghezza pari a n+1. Per questo motivo, introduciamo un nuovo predicato:

$PRV(x,y)$: y è il codice di una formula che è dimostrabile tramite una dimostrazione lunga al massimo x.

L'induzione si compone quindi di:
Base: $PRV(0,y) \rightarrow Ty$
Passo: $\forall x\left(\left(PRV(x,y) \rightarrow Ty\right) \rightarrow \left(PRV(s(x),y) \rightarrow Ty\right)\right)$

Base. La base è immediatamente dimostrata dal momento che nella prima parte della dimostrazione abbiamo mostrato come gli assiomi - e cioè le formule con dimostrazione di lunghezza 0 - siano veri.

Passo. Il passo si suddivide in tanti sottocasi quante sono le regole logiche primitive assunte nel calcolo. Ci limitiamo ad abbozzare la dimostrazione per la regola di introduzione della congiunzione. Gli altri casi sono analoghi.

Come ipotesi generali, assumiamo che $PRV\left(x,\ulcorner \alpha \urcorner\right) \rightarrow T\left(\ulcorner \alpha \urcorner\right)$ e che $PRV\left(x,\ulcorner \beta \urcorner\right) \rightarrow T\left(\ulcorner \beta \urcorner\right)$. Ora assumiamo che $PRV\left(x+1,\ulcorner \alpha \wedge \beta \urcorner\right)$ dal momento che, introducendo la congiunzione, c'è un segno logico in più e quindi la dimostrazione risulta di lunghezza $x+1$. Ma, poiché abbiamo

$T(\ulcorner\alpha\urcorner)$ e $T(\ulcorner\beta\urcorner)$, per clausola tarskiana abbiamo anche $T(\ulcorner\alpha\wedge\beta\urcorner)$ e quindi $PRV(x+1,\ulcorner\alpha\wedge\beta\urcorner)\rightarrow T(\ulcorner\alpha\wedge\beta\urcorner)$.

Teorema 12. $CT\vdash Cons(PA)$

Riuscendo a derivare il principio di riflessione per *PA*, *CT* dimostra immediatamente la consistenza di *PA*. Infatti, $CT\vdash \mathrm{Pr}_{PA}(\ulcorner\bot\urcorner)\rightarrow T(\ulcorner\bot\urcorner)$ da principio di riflessione. Ma, d'altro canto, $CT\vdash T(\ulcorner\bot\urcorner)\rightarrow\bot$ per T-schema. Logicamente, $CT\vdash\neg\bot$. E quindi, $CT\vdash\neg\mathrm{Pr}_{PA}(\ulcorner\bot\urcorner)$. Cioè, $CT\vdash Cons(PA)$. In base al secondo teorema di Gödel, abbiamo che *CT* è un'estensione non conservativa di *PA*.

4. *Teoria composizionale ristretta*

Infine, accenniamo a un terzo risultato significativo e cioè la conservatività di $CT\upharpoonright$ su *PA*. Come abbiamo avuto modo di vedere $CT\upharpoonright$ è costituita dalle clausole tarskiane classiche T1-T4 ma presenta l'assioma di induzione ristretto alle formule aritmetiche. Non è possibile cioè compiere l'induzione su proprietà che comprendono il predicato *T*. Come abbiamo già accennato in precedenza, questa caratteristica limita notevolmente la portata conoscitiva di questa teoria; il risultato che andiamo a dimostrare mostra in maniera precisa in che cosa consiste questo limite: $CT\upharpoonright$ è conservativa rispetto all'aritmetica, ovvero non consente di derivare enunciati appartenenti al linguaggio di *PA* che non sia possibile già derivare in *PA*. Segue immediatamente che $CT\upharpoonright$ non è in grado, a differenza di *CT*, di derivare la consistenza di PA. La dimostrazione è di natura semantica ed è stata fornita da Kotlarski, Krajewski, Lachlan (1981) che sfrutta un risultato precedente di Lachlan. Hartry Field (1999) fa invece riferimento a una dimostrazione di Charles Parsons (Parsons 1983); probabilmente però il ragionamento - non una vera e propria dimostrazione - di Parsons consiste nel riconoscere che $CT\upharpoonright$ è definibile in ACA_0 e che, notoriamente, ACA_0 è conservativa rispetto a *PA*.

5. *Conclusioni*

I risultati più importanti sono essenzialmente tre:

1. $ACA\vdash CT$

2. $CT \vdash Cons(PA)$
3. $CT \restriction \approx PA$

Il primo stabilisce che la teoria composizionale tarskiana può essere ottenuta in un particolare sottosistema di aritmetica del secondo ordine. Il secondo, conseguenza del primo, mostra che dalla teoria *CT* è possibile derivare la consistenza dell'aritmetica. Il terzo, infine, fa vedere come sia sufficiente restringere l'assioma di induzione di *CT* per ottenere una teoria conservativa rispetto a *PA*.

Si tratta di risultati tecnici, di logica matematica; tuttavia, la loro rilevanza filosofica è notevole. Sono al cuore infatti dell'argomento di conservatività, ovvero quella particolare critica (interna, secondo la terminologia introdotta nel primo capitolo della seconda parte del volume) alle posizioni deflazioniste circa il problema filosofico della verità.

VII
L'ARGOMENTO DI CONSERVATIVITÀ

1. *Formulazione generale*

Con questo capitolo chiudiamo la nostra analisi sull'approccio assiomatico al problema della verità. Come abbiamo detto in precedenza, l'interlocutore filosofico privilegiato è il *deflazionismo*, la posizione per la quale - in estrema sintesi - la verità non ha una natura propria. Benché non vi sia un'implicazione concettuale tra posizione deflazionista e approccio assiomatico, ciononostante alcuni influenti studiosi difensori della posizione deflazionista privilegiano la trattazione assiomatica del predicato di verità. I motivi per questo apparentamento li abbiamo esaminati nel primo capitolo; ora ci interessa tirare le fila del nostro discorso alla luce della notevole messe di risultati tecnici (e di relative glosse filosofiche) esposta nelle pagine precedenti.

L'argomento di conservatività costituisce proprio l'esito della nostra trattazione e mostra come a determinate condizioni, che ci riserviamo di esplicitare e commentare, la posizione deflazionista è difficilmente sostenibile. L'argomento di conservatività non presuppone che si assuma l'approccio assiomatico ma le tematiche ad esso collegate, per esempio, le interpretazioni del principio di induzione e dello schema di riflessione, suggeriscono una trattazione unitaria. Le prime formulazioni dell'argomento di conservatività sono di Ketland 1999 e Shapiro 1998 anche se si tratta della rielaborazione filosoficamente originale dei noti risultati di incompletezza e indefinibilità della verità scoperti da Gödel e Tarski negli anni '30.

L'argomento si basa su due premesse che il sostenitore della posizione deflazionista dovrebbe accettare ma che non possono essere soddisfatte contemporaneamente e che quindi pongono problemi. Vediamo quindi una formulazione molto discorsiva e in seguito un'analisi più dettagliata dell'argomento.

Struttura dell'argomento. Una teoria della verità (chiamiamola T) deflazionisticamente accettabile per una teoria di base (B) deve soddisfare alcune condizioni circa l'uso del predicato di verità. In particolare, deve permettere una serie di generalizzazioni informative. Per esempio, vogliamo poter dire che ogni teorema di B è vero. Ma B+T deve risultare a sua volta conservativa su B, cioè ogni enunciato appartenente al linguaggio di B che si ottiene da B+T deve essere derivabile dalla sola B. La conservatività assicura, cioè, che la nostra teoria della verità costituisca un'estensione conservativa della nostra teoria di base e che quindi la nozione di verità sia 'metafisicamente sottile', non abbia, cioè, proprietà specifiche. Tuttavia queste due condizioni, e cioè la possibilità di derivare la correttezza della teoria e la conservatività non possono essere vere entrambe: infatti, se B+T dimostra che tutti i teoremi di B sono veri, allora *ipso facto*, dimostra che B è consistente e per il secondo teorema di Gödel non è conservativa su B. Quindi, conclude l'argomento, se vogliamo che la nostra teoria della verità risulti informativa dobbiamo rinunciare alla conservatività, lasciando il deflazionismo in cattive acque.

Così Ketland riassume l'argomento:

> Stewart Shapiro and I have introduced an innovation in relation to understanding the notion of deflationism about truth. We have proposed that we define "substantial" (for a theory of truth) to mean "non-conservative" (over base theories). We proved that disquotation is conservative, and Tarski's inductive definition is sometimes non-conservative[1].

2. *Conservatività*

La prima premessa che vogliamo discutere tratta la questione della conservatività. Perché l'argomento funzioni è necessario che il deflazionista indichi nella conservatività della teoria della verità l'equivalente (formale) della non sostanzialità della nozione in gioco. Non si tratta di un'assunzione scontata e per questo motivo è necessario elaborarla filosoficamente. Vediamo come affronta la questione Shapiro:

> Suppose, for example, that [one holds] in a theory B in a language that cannot express truth. He adds a truth predicate to the language and ex-

1 Comunicazione personale (cfr. Vidal-Rosset)

> tends B to a theory B' using only axioms essential to truth. Assume that B' is not conservative over B. Then there is a sentence Φ in the original language (so that Φ does not contain the truth predicate) such that Φ is a consequence of B' but not a consequence of B. That is, it is logically possible for the axioms of B to be true and yet Φ false, but it is not logically possible for the axioms of B' to be true and Φ false. This undermines the central deflationist theme that truth is insubstantial. Before moving to B', ¬Φ was possible. The move from B to B' added semantic content sufficient to rule out the falsity of Φ[2].

Cioè, abbiamo il seguente caso:

$B \nvdash \phi$ e $B' \vdash \phi$. ϕ potrebbe essere, per esempio, $Cons(B)$. Per correttezza e completezza abbiamo la seguente situazione: $B \nVdash \phi$ e $B' \Vdash \phi$. Nel primo caso, è possibile che gli assiomi di B siano veri e ϕ falsa; nel secondo, no: se gli assiomi di B' sono veri, necessariamente ϕ deve essere vera. Che cosa è accaduto nel passaggio da B a B'? È stato conferito un aumento dell'informazione che ha permesso di escludere i modelli in cui è presente $\neg\phi$. Ma escludere modelli, ovvero delimitare la possibilità logica, implica proprio aggiungere qualcosa, e cioè contenuto. B' è in grado di dire qualcosa in più sul mondo, dal momento che riesce a eliminare la possibilità di $\neg\phi$ derivando così ϕ. Come sarebbe possibile tutto ciò se B' fosse una teoria ottenuta da B con l'aggiunta di principi deflazionisticamente accettabili?

Le alternative a disposizione del deflazionista sono, a questo punto, essenzialmente due: o cerca di superare l'impasse dichiarando che la non conservatività è plausibile anche in ottica deflazionista oppure può riformulare il concetto stesso di conservatività. E sarà quest'ultima opzione che andiamo a discutere brevemente.

La conservatività riguarda normalmente il nesso di derivabilità; tuttavia, possiamo formulare un'altra nozione che ha a che fare con il nesso semantico. B' è un'estensione semantica conservativa di B se e solo se:

$$\forall \phi \in \ell(B), B' \Vdash \phi \Rightarrow B \Vdash \phi$$

In questo senso, ϕ risulta vera già nei modelli che rendono veri gli assiomi di B. Naturalmente, qualcuno potrebbe obiettare che chi persegue un approccio puramente assiomatico in teoria della verità non può assu-

2 Shapiro 1998, 497.

mere la nozione di conseguenza logica che presuppone, tarskianamente, le nozioni di modello, interpretazione e struttura, tipiche della concezione semantica della verità. Tuttavia, le considerazioni in merito al problema della conservatività non sono prerogative del solo approccio assiomatico ma valgono anche per la concezione semantica della verità e in questo caso il ricorso alla conseguenza logica è legittimo.

Ora, il nesso di derivabilità e quello di conseguenza logica, per teorie espresse al primo ordine, coincidono, in virtù dei metateoremi generali di correttezza e completezza. Lo stesso non si può dire se ci si sposta a un linguaggio di ordine superiore. Cosa accade, infatti, se riproponessimo l'argomento di conservatività per una teoria espressa in un linguaggio del secondo ordine?

Sia PA^2 l'aritmetica di Peano del secondo ordine; sia CT^2 la teoria composizionale della verità per PA^2. Dal punto di vista sintattico le cose restano immutate. Infatti abbiamo che:

$$\exists\phi \in \ell(PA^2), PA^2 \nvdash \phi \,\&\, CT^2 \vdash \phi$$

da cui segue, per definizione, la non conservatività di CT^2 su PA^2. Infatti CT^2 è in grado di dimostrare, per esempio, la consistenza di PA^2. Se ci spostiamo però al livello semantico la situazione non è più speculare; PA^2, se interpretata con una semantica standard,[3] è una teoria categorica, ammette cioè solo modelli isomorfi. In altre parole, PA^2 possiede un solo modello, quello standard. Per cui, il nesso di conseguenza logica non è più coestensivo con quello di derivabilità. Di più. L'enunciato che descrive la consistenza di PA^2 è una conseguenza logica di PA^2 (anche se, per il teorema di Gödel non è derivabile e ciò comporta l'incompletezza del calcolo del secondo ordine). Ma, a maggior ragione, è una conseguenza logica di CT^2. Pertanto ci troviamo in questa situazione:

$$\forall\phi \in \ell(PA^2), CT^2 \Vdash \phi \Rightarrow PA^2 \Vdash \phi$$

La verità aritmetica è semanticamente conservativa nel senso che ogni modello di aritmetica (del secondo ordine) può essere esteso a un model-

---

3 Per "semantica standard" intendiamo le strutture piene (o full) ove il dominio dei quantificatori di ordine superiore comprende tutti i sottoinsiemi di individui. Per una discussione su queste tematiche si vedano, ad esempio, Shapiro 2005, Jané 2005 o anche i saggi contenuti in Shapiro 1996.

lo di teoria della verità aritmetica. In questo senso, passare a CT^2 non conferisce nessun incremento informativo dal punto di vista semantico dal momento che ciò che era vero nel modello di CT^2 è vero nel modello standard (l'unico) di PA^2. L'argomento prima proposto non funziona perché il passaggio a CT^2 non conferisce nuova informazione e quindi diventa deflazionisticamente accettabile.

Il deflazionista, per salvarsi, deve mettere in campo questa doppia strategia: *in primis*, riformulare il suo concetto di conservatività, ammettendone una variante di tipo semantico e, secondariamente, passare a teorie categoriche, come l'aritmetica del secondo ordine.

Ma si tratta di un'alternativa realmente praticabile per il deflazionista? Molto probabilmente no. E la ragione è piuttosto chiara. Il passaggio a teorie potenti come PA^2 comporta già un notevole incremento informativo; la categoricità di questa teoria consiste proprio nella possibilità di eliminare, in forza di una logica più potente ed espressiva, i modelli non standard. Ma se vale il discorso fatto in precedenza, l'eliminazione di alcune possibilità logiche coincide con l'incremento di contenuto informativo. Come è possibile dunque argomentare a favore della 'leggerezza' metafisica della nozione di verità quando si assume una teoria 'metafisicamente robusta' come PA^2 ?

Un'ultima considerazione. Il riferimento all'aritmetica del secondo ordine è una maniera tecnicamente elaborata per riuscire a isolare - *via* categoricità - il modello standard dei numeri naturali. Ed è abbastanza ovvio che, dal punto di vista di $\mathbb{N}$, la teoria della verità risulta conservativa su *PA*. Questo perché, $\mathbb{N}$ altro non è che la controparte semantica del concetto di verità che *CT* cattura assiomaticamente. Detto in altri termini, il deflazionista non può far riferimento al modello standard dell'aritmetica, sia direttamente, sia attraverso il *detour* di PA^2, perché il concetto stesso di modello (standard) non è deflazionisticamente accettabile, essendo più ricco della struttura sintattica cui viene associato. Il passaggio da una teoria dei numeri naturali alla teoria della verità per i numeri naturali non mostra incremento informativo solo quando la teoria di partenza contiene già, in maniera più o meno esplicita, il concetto di verità. Si capisce chiaramente perché questo tipo di argomenti sia, di principio, impraticabile per il deflazionista.

3. *Principi di riflessione*

Un altro tema su cui il dibattito è stato, ed è tuttora, piuttosto acceso riguarda la portata logico-filosofica dei principi di riflessione e del loro utilizzo nell'argomento di conservatività.

Come si ricorderà, una delle funzioni principali della nozione di verità che la rende, di fatto, ineliminabile e che differenzia pertanto la posizione ridondantista (*à la* Ramsey) dal fascio di posizioni deflazioniste, è quella di esprimere generalizzazioni. E, sempre in precedenza, ci siamo concentrati su un particolare tipo di proposizioni universali come la seguente:

(Corr) Tutti i teoremi della teoria T sono veri.

È evidente che (Corr) è un enunciato assolutamente importante per il nostro discorso. Ammettere (Corr) significa ammettere la nostra fiducia epistemica in T. Nel capitolo precedente, la dimostrazione di una versione formale di (Corr) in *CT* mette bene in luce la notevole potenza della teoria tarskiana della verità; è immediato concludere da (Corr) che T è consistente.

Neil Tennant, in un influente articolo[4] apparso su *Mind*, mette in discussione proprio questo punto dell'argomento di conservatività e cioè la necessità di assumere l'impegnativa e deflazionisticamente inaccettabile teoria tarskiana *CT* per ottenere un enunciato come (Corr). Alla proposta di Tennant è seguita la replica puntuale di Ketland che ha dato origine a una controreplica di Tennant. Nel dibattito si è inserito poi anche Cezary Cieśliński che ha proposto una modifica dell'argomento originale di Tennant oggetto di un'altra discussione con Ketland. In ciò che segue presenteremo le varie opzioni teoriche in gioco cercando di individuare i punti filosoficamente interessanti del dibattito.

Tennant non è un deflazionista. Lo dice esplicitamente all'inizio della sua risposta a Ketland:

> I am not a deflationist. I believe that truth and falsity are substantial. The truth of a proposition consists in its having a constructive proof, or truthmaker. The falsity of a proposition consist in its having a constructive disproof, or falsitymaker. Such proofs and disproofs will need to be

4 Tennant 2002.

> given modulo acceptable premisses. The choice of these premisses will depend on the discourse in question[5].

È abbastanza evidente l'orientamento costruttivista di Tennant; ma è ancora più significativo notare come Tennant apparenti il deflazionismo con una posizione anti-realista:

> Deflationism has its roots in Ramsey's contention that to assert that ϕ is true is to do no more than assert ϕ, unadorned. Truth is not a substantial property whose metaphysical essence could be laid bare. It has no essence; it is as variegated as the grammatical declaratives that would be its bearers[6].

E fino a questo punto, Tennant riprende perfettamente la tesi ontologica del deflazionismo, quella riguardante, cioè, la non sostanzialità della verità. Tuttavia, continua:

> There would therefore appear to be no gap, on the deflationist's view, between claims that are true and assertions that are warranted; or, generally, between *truth* on the one hand, and, on the other hand, grounds for assertion, or *proof*[7].

E questo è quantomeno strano. Tennant sembra asserire qualcosa come una implicazione così costituita:

(1) Verità non sostanziale $\Rightarrow$ Verità = Dimostrazione

Come rendere conto di (1)? A prima vista sembra poco convincente. Se la verità non ha una natura sua propria come può essere *identica* alla dimostrabilità? Ora, è chiaro che (1) varrebbe se l'antecedente dichiarasse che la verità non esiste. In tal caso, si verrebbe a dire che se la verità (inteso in senso intuitivo, corrispondentista) non esiste, allora quello che ci rimane è qualcosa come l'asseribilità garantita, cioè una forma di dimostrazione o giustificazione. Ma il deflazionismo non è nichilismo circa la verità, almeno da quanto si può inferire dalle tesi dei suoi sostenitori. La verità esiste ma è "sottile". Quindi non si capisce come possa coincidere con la dimostrazione la quale è invece, metafisicamente, *thick*. Seb-

5 Tennant 2005, 453.
6 Tennant 2002, 552.
7 Tennant 2002, 552.

bene anti-realiste, le teorie della verità costruttive, non sono affatto metafisicamente leggere, anzi. L'unico modo disponibile per interpretare plausibilmente l'argomento di Tennant sembra essere il seguente: se si riesce a mostrare che le richieste degli anti-deflazionisti (Ketland e Shapiro) possono essere soddisfatte tramite una concezione della verità di tipo costruttivo e quindi non tarskiano, allora il deflazionista non è costretto a cedere all'argomento di conservatività perché può sempre imbracciare la scappatoia aperta da Tennant. Rimane poi da far vedere in separata sede se e in quale misura la proposta di Tennant sia deflazionisticamente accettabile nonostante sia chiaramente non tarskiana. Vediamo quindi i dettagli dell'argomento di Tennant.

L'argomento di conservatività non è altro che una certa applicazione del teorema di Gödel: il fatto cioè che una qualsiasi teoria formale S, sufficientemente potente, non possa derivare l'enunciato che dichiara la sua correttezza, e cioè $\mathrm{Pr}_S(\ulcorner\alpha\urcorner) \rightarrow T(\ulcorner\alpha\urcorner)$ dal quale si ottiene immediatamente $Cons(S)$. Ma come è possibile rendersi conto della verità della proposizione gödeliana G? Tennant, citando Dummett, propone il seguente argomento semantico:

Argomento semantico. G è la chiusura universale di una formula primitiva ricorsiva (cioè una Π_1 formula). Ogni istanza numerica di G è dimostrabile all'interno del nostro sistema S. *Ma la dimostrabilità garantisce la verità*. Pertanto, ogni istanza di G è vera. Ma se ogni istanza di G è vera lo sarà anche la quantificazione universale. E quindi G è vera.

Questo ragionamento richiede una concezione robusta e sostanziale della verità e, nell'argomento di conservatività, viene impiegato a favore di una concezione anti-deflazionistica della verità. Quali sono i punti in cui emerge una concezione sostanzialistica della verità? Sono essenzialmente due: innanzitutto, si assume che la dimostrazione all'interno del sistema S garantisca la verità. Ma questo è proprio il contenuto del principio di riflessione e infatti l'argomento semantico non è riproducibile all'interno del sistema S. Inoltre, si afferma che se $\alpha(x)$ è vera per ogni numerale, allora sarà vera anche la chiusura universale, $\forall x\alpha(x)$. Ciò equivale esattamente alla clausola tarskiana della quantificazione che dice: $\forall x T(\ulcorner\alpha(\dot{x})\urcorner) \leftrightarrow T(\ulcorner\forall x\alpha(x)\urcorner)$.

Secondo Tennant è possibile riformulare l'argomento semantico senza dover per forza sposare la teoria della verità tarskiana, senza cioè aderire a quel «dogma semantico» che ci impegna nei confronti di una concezione sostanzialistica della verità.

Il punto fondamentale dell'argomento di Tennant è l'assunzione di un principio di riflessione particolare:

$$\text{(URPR)} \qquad \forall x\left(\mathrm{Pr}_{PA}\left(\ulcorner \alpha(\dot{x}) \urcorner\right) \rightarrow \alpha(x)\right),\ \alpha \in PR - \mathit{formule}$$

URPR (Uniform Reflection Primitive Recursive) è un principio di riflessione uniforme ristretto alle PR-formule. Ciò significa che se una formula α è dimostrabile in *PA* per tutti i numerali, allora α è vera. Non deve stupire l'assenza del predicato di verità. Come dice giustamente Ketland:

> Of course, $\mathrm{Pr}_S(\ulcorner \phi \urcorner) \rightarrow \phi$ is equivalent to $\mathrm{Pr}_S(\ulcorner \phi \urcorner) \rightarrow T(\ulcorner \phi \urcorner)$, if we have a disquotational truth predicate[8].

Ma il deflazionista *ha* a disposizione un predicato di verità decitazionale, come abbiamo più volte detto. Quindi, il fatto che nella formulazione di Tennant non compaia il simbolo T non è rilevante dato lo sfondo teorico deflazionista. Con a disposizione (URPR) è possibile riprodurre l'argomento semantico senza assumere la teoria tarskiana della verità. E quindi, conclude Tennant, la posizione deflazionista non è refutata dall'argomento di conservatività.

Il punto da discutere in merito alla posizione di Tennant è essenzialmente il seguente: il ricorso a un principio di riflessione è più o meno giustificato rispetto all'adozione della semantica tarskiana?

La replica di Ketland ruota di fatto intorno a questo tema. L'idea di fondo è che se qualcuno accetta una teoria, come *PA*, dovrebbe accettare anche le istanze del principio di riflessione $\mathrm{Pr}_S(\ulcorner \phi \urcorner) \rightarrow \phi$. Ma perché? Perché il principio di riflessione non è altro che la rappresentazione formalmente adeguata della nostra comprensione della nozione di verità. Ma se si accettano enunciati come $\mathrm{Pr}_S(\ulcorner \phi \urcorner) \rightarrow \phi$ e come G (la proposizione gödeliana) si è nella seguente situazione:

Obbligo epistemico. Se si accetta una teoria (non banale) S, è obbligatorio accettare anche altri enunciati appartenenti al linguaggio di S.

---

8 Ketland 2005, 453.

Seguendo l'analisi di Ketland (2010, 424), data una teoria della verità T* definita su una teoria T, possiamo prendere in considerazione due requisiti di adeguatezza: *adeguatezza deflazionistica* e *adeguatezza riflessiva*. L'adeguatezza deflazionistica coincide con il requisito di conservatività: una teoria della verità è adeguata deflazionisticamente se è conservativa rispetto alla teoria di base. L'adeguatezza riflessiva è qualcosa di più complesso. Per Tarski la definizione di verità deve essere materialmente adeguata e cioè in grado, per ogni $\alpha \in \ell$ di derivare il bicondizionale $T(\alpha) \leftrightarrow \alpha$. Ora, all'interno di una teoria aritmetica, per esempio, noi vogliamo che, per ogni teorema ψ, la teoria della verità sia in grado di ottenere $T(\psi)$. L'adeguatezza riflessiva è esattamente questo; con le parole di Ketland:

> Reflective adequacy is a generalization of material adequacy, which is the requirement on TL that from a single sentence φ we can prove 'φ is true' (and vice versa, of course); analogously, reflective adequacy is the requirement on TL that from a theory B we can prove 'all theorems of B are true' (Ketland 2010, 424).

L'argomento di conservatività mostra proprio come ogni teoria deflazionisticamente accettabile non sia riflessivamente adeguata e, viceversa, ogni teoria riflessivamente adeguata risulti inaccettabile dal punto di vista deflazionista perché non conservativa.

A questo punto si tratta di spiegare che cosa significa 'accettare una teoria'. Gli anti-deflazionisti, ragionano in questo modo: accettare una teoria significa, in ultima analisi, considerarla come vera. Ma dire che *PA*, per esempio, è vera vuol dire che tutti i suoi teoremi sono veri, cioè che se φ segue logicamente dagli assiomi di *PA* (e questi assiomi sono veri) allora φ è vera. Ciò significa, quindi, accettare una forma di principio di riflessione e, data l'inderivabilità di quest'ultimo all'interno di *PA* ma la sua derivabilità all'interno di *CT*, inferire la sostanzialità della nozione di verità. Naturalmente, è possibile dare un significato differente alla accettazione di una teoria. Tennant, per esempio, sostiene che c'è un'altra via. Il principio (URPR) è l'esito della nostra riflessione sulle procedure dimostrative condotte all'interno di *PA*. Il fatto che la rappresentazione formale dell'affidabilità della procedura di dimostrazione condotta in *PA* non sia dimostrabile all'interno di *PA* non tange - nell'opinione di Tennant - la non sostanzialità della nozione di verità. Al contrario secondo Ketland, il deflazionista (o Tennant) è in difetto, perché deve assumere, senza ulteriore giustificazione, l'adozione di un principio come UPRP.

La risposta di Tennant è chiara:

> No further justification is needed for the new commitment made by expressing one's earlier commitments. As soon as one appreciates the process of reflection, and how its outcome is expressed by the reflection principle, one already has an explanation of why someone who accept S should also accept all instances of the reflection principle[9].

Quindi, siamo giunti al nocciolo della questione: da un lato Tennant e la versione anti-realista del deflazionismo che dichiara come non ulteriormente giustificabile l'adozione del principio di riflessione; dall'altro, Ketland e gli anti-deflazionisti che reputano migliore e più fondato il ricorso alla teoria tarskiana della verità con la conseguente non conservatività. Chi ha ragione?

4. *Adeguatezza riflessiva*

Il problema più grande con la strategia di Tennant riguarda, essenzialmente, il suo essere *ad hoc*. In linea del tutto generale, la giustificazione per una determinata ipotesi (esplicativa) dovrebbe godere di una certa indipendenza rispetto al fenomeno e al contesto teorico per il quale viene introdotta. Altrimenti, si corre il rischio di assumere e postulare solo principi che *servono* ai nostri scopi; di qui l'accusa di essere *ad hoc*. Ebbene, anche nel caso dell'argomento di conservatività, l'ipotesi di assumere URPR sembra soffrire della medesima limitazione. Vediamo perché.

La giustificazione[10], sicuramente intuitiva e sfumata, della teoria tarskiana della verità *CT* dipende da due fattori distinti: una certa intuizione *corrispondentista* (che interviene specialmente nel primo assioma, cioè nella clausola atomica) e una certa intuizione *composizionale* (che permette di costruire enunciati specifici della teoria sempre più complessi). Ora, se ci fermassimo a questo punto, l'argomento di conservatività apparirebbe come superfluo dal momento che si presuppone già una posizione anti-deflazionista che invece si vuole dimostrare. E infatti seguiremo un'altra via. Ancora una volta dobbiamo partire da un punto condiviso sia da Ketland che da Tennant (testimoni rispettivamente di sostan-

---

9 Tennant 2005, 92.
10 Si tratta di una giustificazione filosofica.

zialismo e deflazionismo): il fatto che si possa tematizzare l'accettazione di una data teoria, *PA*, per esempio. L'esplicitazione di questa accettazione, cioè dell'assenso a ciò che *PA* dice, è codificata dalle istanze di un principio di riflessione. A questo punto, le strade si dividono: da un lato, Tennant sceglie di assumere il principio di riflessione come primitivo (si ricordi il suo riferimento alla non giustificabilità di URPR), mentre Ketland, dall'altro, decide di adottare *CT*. Ma in *CT* il principio di riflessione è *dimostrabile* e non semplicemente assunto. E quindi il problema, per Ketland, diventa quello di giustificare l'adozione di *CT*. *CT*, infatti, necessita a sua volta di una giustificazione; la differenza con URPR sta però nel fatto che *CT* non è unicamente formulata per derivare il principio di riflessione ma è l'assiomatizzazione di alcune intuizioni generali circa la verità. La possibilità di ottenere in *CT* il principio di riflessione è, se mai, una conseguenza (gradita) della sua stessa struttura e del suo significato.

Abbiamo già avuto modo di incontrare una questione analoga discutendo il requisito di adeguatezza materiale di una teoria della verità. Come è noto - e come abbiamo più volte ribadito - per Tarski una teoria S è materialmente adeguata se riesce a derivare tutte le istanze del T-schema:

$$\text{(T)} \quad T(\ulcorner\alpha\urcorner) \leftrightarrow \alpha^*$$

dove, ricordiamolo, $\ulcorner\alpha\urcorner$ è il nome dell'enunciato in questione e α^* la sua traduzione metalinguistica. Quindi, il criterio di Tarski può essere rappresentato come segue:

$$(1) \quad MAC(S) \Leftrightarrow S \vdash T(\ulcorner\alpha\urcorner) \leftrightarrow \alpha^*$$

dove *MAC* indica la proprietà di essere materialmente adeguata. Nel nostro caso, quindi, l'equivalenza dovrebbe essere posta nei termini seguenti:

$$(2) \quad RAC_{PA}(S) \Leftrightarrow S \vdash \mathrm{Pr}(\ulcorner\alpha\urcorner) \rightarrow T(\ulcorner\alpha\urcorner), \alpha \in \ell(PA)$$

dove RAC_{PA} indica la condizione di *adeguatezza riflessiva*, ovvero, come detto in precedenza, la tematizzazione dell'accettazione di una teoria. Si noti che, in questo caso, S è adeguatamente riflessiva rispetto alla teoria *PA* nel senso che S riesce a derivare il principio di riflessione per *PA*.

Per ottenere il requisito *MAC*, Tennant assume esattamente il contenuto di quel requisito:

(3) $PA \cup \mathrm{Pr}_{PA}(\ulcorner\alpha\urcorner) \to T(\ulcorner\alpha\urcorner) \vdash RAC_{PA}$

La teoria di Tennant è scelta *ad hoc* e la differenza con *CT* balza agli occhi. *CT* dimostra *genuinamente* il principio di riflessione, e questo perché dimostra molte altre cose. Un parallelo può essere utile. Come si ricorderà - sia dal primo capitolo, sia da altri punti nel volume - uno dei requisiti imposti da Tarski per una definizione della verità è che questa sia materialmente adeguata. Una teoria della verità, in altre parole, deve essere in grado di derivare i T-bicondizionali. Ora, se si assume, come *TB*, tutta la lista dei bicondizionali, chiaramente la teoria risulterà materialmente adeguata, ma in maniera banale: essa, infatti, coincide con il requisito stesso richiesto dalla condizione tarskiana. Si potrà poi giustificare l'adozione di *TB* in base a ragioni filosofiche, per esempio, sulla scorta di un'intuizione decitazionale in merito al predicato "essere vero" ma a nostro avviso non è una buona ragione per accettare *TB* invocare il requisito di adeguatezza materiale dal momento che, come detto prima, *TB* è in un certo senso ciò che il requisito richiede.

Quindi, a differenza della proposta avanzata da Tennant, *CT* invece dimostra tutte le istanze del T-schema in maniera rilevante. Per "rilevante" intendiamo qualcosa del genere: una teoria T dimostra in maniera rilevante un enunciato φ quando fa un uso non banale dei suoi principi. È chiaro che questa caratterizzazione di "rilevanza" dipende da quella di "banalità" e forse non è una via tanto più agevole. In ogni caso, sembra del tutto plausibile pensare che la teoria $PA \cup Cons_{PA}$ dimostri in maniera piuttosto banale l'enunciato $Cons_{PA}$ al contrario della teoria *ZF* (teoria degli insiemi di Zermelo-Fränkel) che dimostra anch'essa $Cons_{PA}$ ma non così scontatamente. In ogni caso, si può indebolire la nostra assunzione, dicendo che è possibile dare una connotazione ordinale alle dimostrazioni in gioco; in altri termini, possiamo affermare che delle due la prima è sicuramente più banale e quindi meno rilevante della seconda, senza prendere posizione circa il valore "assoluto" di rilevanza di questi risultati.

In conclusione, dal punto di vista prettamente formale, la mossa di Tennant è legittima: ciò che serve a caratterizzare l'accettazione di una teoria formale è il principio di riflessione e quindi è possibile assumerlo e non sposare, così, una teoria della verità impegnativa come quella tarskiana. Ma, in ottica filosofica, si tratta - lo ripetiamo - di una mossa *ad hoc*. Il che, ovviamente non significa che l'adozione di *CT* sia giustificata in maniera assoluta. La mossa speculare per Tennant sarebbe quella di esibire una teoria costruttiva della verità, nella quale, per esempio, la verità coincida con la dimostrabilità sotto determinate condizioni. Se le cose stessero così, allora sia *CT* che la teoria anti-realista sarebbero sullo stesso piano, abbisognando di una giustificazione. La prima, come dicevamo, si baserà su intuizioni di carattere realista, prima fra tutte la corrispondenza, la seconda farà leva sui concetti di evidenza, o costruibilità, o altri ancora. Ma, ed è questa la conclusione che davvero ci interessa, sia in un caso che nell'altro, la verità *è* qualcosa, ha, cioè, una natura specifica. E questo contro la tesi fondamentale del deflazionismo.

Nella discussione precedente un ruolo importante è stato giocato dai concetti che hanno a che fare con la dimensione epistemica dell'accettazione di una data teoria e con la dimensione semantico-informativa circa il contenuto concettuale del principio di riflessione. L'argomentazione di Tennant è stata ripresa e leggermente modificata da Cieśliński[11] che scompone la procedura di accettazione di una teoria nei seguenti tre passaggi. Sia Acc un predicato che si applica a oggetti come classi di formule chiuse (assiomi, teoremi) e classi di regole; si ha quindi che, dato un sistema formale S:

i. Acc(Ax(S)) & Acc(Reg(S))
ii. Acc(Tutti i teoremi di S)
iii. Acc($\mathrm{Pr}_S(\ulcorner\phi\urcorner) \to \phi$)

Chiariamo in breve la simbologia adottata. Acc(Ax(S)) significa che si accettano gli assiomi di una teoria S. Acc(Reg(S)) significa che si accettano le regole della logica in cui S è formulata ed eventualmente delle regole specifiche di S. Acc($\mathrm{Pr}_S(\ulcorner\phi\urcorner) \to \phi$) significa, infine, l'accettazione, per una formula ϕ qualsiasi, che se ϕ è dimostrabile allora è vera. Vedremo che sono problematici sia il passaggio da (i) a (ii) sia quello da (ii) a (iii)

In effetti perché valga il passaggio dall'accettazione degli assiomi e delle regole all'accettazione (ancorché potenziale) dei teoremi di S sono indispensabili due condizioni come:

(4) $\mathrm{Acc}(\alpha \vdash \beta) \wedge \mathrm{Acc}(\alpha) \Rightarrow \mathrm{Acc}(\beta)$

e

11 Cieslinski 2010.

(5) $\alpha \vdash \beta \Rightarrow \mathrm{Acc}(\alpha \vdash \beta)$

Il primo dichiara che se accettiamo l'inferenza da α a β, e accettiamo α dobbiamo accettare anche β. Il secondo, invece, afferma che se siamo in presenza di un'inferenza, questa deve essere accettata. Nel nostro caso, abbiamo che, $S \vdash \alpha \Rightarrow \mathrm{Acc}(S \vdash \alpha)$ da (5); Acc(S) per ipotesi generale e quindi, per (4), otteniamo $\mathrm{Acc}(\alpha)$.

Sotto l'assunzione che le nostre procedure di accettabilità soddisfino questi requisiti, possiamo affermare che accettare una teoria significa accettarne gli assiomi e le regole di deduzione e, *pertanto*, tutti i teoremi.

Più problematico è passare da (ii) a (iii). Infatti, l'enunciato $\mathrm{Pr}_S(\ulcorner \phi \urcorner) \rightarrow \phi$ non è, anche da un punto di vista intuitivo, un teorema del sistema S. Il suo significato è infatti meta-teorico: se qualcosa è un teorema di S allora è vero. E infatti è possibile ammettere (ii) ma non (iii). Vediamo come.

Il significato di (ii) può essere reso anche così:

(6) $\forall \alpha (S \vdash \alpha \Rightarrow Acc(\alpha))$

Se α è un teorema di S, allora deve essere accettato. Ma nessun teorema di S è $\mathrm{Pr}_S(\ulcorner \phi \urcorner) \rightarrow \phi$. Quindi, una mente ideale[12] potrebbe conoscere completamente tutta la serie dei teoremi di S senza per questo conoscere che $\mathrm{Pr}_S(\ulcorner \phi \urcorner) \rightarrow \phi$. E questo perché potrebbe non 'rendersi conto' che in effetti ogni teorema è vero anche se la conoscenza di ogni teorema implica, a un metalivello, che questo sia vero. Ma per compiere questa inferenza dobbiamo assumere una prospettiva metateorica che deve essere adeguatamente giustificata.

La strategia di Tennant è quella di postulare che il principio di riflessione, cioè $\mathrm{Pr}_S(\ulcorner \phi \urcorner) \rightarrow \phi$, *esprima* il contenuto di (ii). Ma quali ragioni si possono addurre in favore di ciò?

---

12 It is consistent to imagine a superhuman who is ready to accept all theorems of S, but is not ready to accept each instance of $\mathrm{Pr}_S(\ulcorner \phi \urcorner) \rightarrow \phi$. The total linguistic output of the superhuman might be precisely the theorems of PA, and these will not include all instances of $\mathrm{Pr}_S(\ulcorner \phi \urcorner) \rightarrow \phi$. (Ketland 2010, 429)

5. *Induzione matematica e contenuto di verità*

Field accetta molte delle assunzioni compiute da Shapiro; in particolare sono due i punti che meritano di essere ripresi: innanzitutto, egli ammette, con la maggior parte dei deflazionisti, che la nozione di verità aumenta il nostro potere espressivo e, conseguentemente, che lo scopo del predicato di verità è quello di permettere 'generalizzazioni feconde'.

Ora, la particolarità del argomento di Field sta nel fatto che egli accetta sia il fatto che la non conservatività di una teoria è prova certa della sostanzialità della nozione di verità, sia che la teoria della verità più adeguata debba essere di natura composizionale. Distanziandosi pertanto da Quine e dalla posizione decitazionalista, Field ricostruisce così l'argomento di Shapiro: per ottenere

(Corr) Tutti i teoremi di T sono veri

la teoria della verità deve possedere, oltre alle clausole tarskiane (T1 – T4) anche l'assioma di induzione non ristretto. Dobbiamo cioè essere in grado di compiere un'induzione su formule che, come abbiamo visto, contengono il predicato di verità *T*. Ed è proprio su questo punto che si impernia tutta l'argomentazione di Field. Secondo Field, la teoria *CT* non è conservativa su *PA* proprio *a causa* dell'assioma di induzione non ristretto.

Vediamo, dunque, nel dettaglio l'argomento. Per ottenere una teoria della verità della teoria di base che, assumiamo sia PA, è necessario aggiungere alcuni assiomi che risultino "*essential to truth*". Nel caso preso in esame, aggiungeremo all'aritmetica di Peano (PA) le clausole tarskiane T1-T4 e provvederemo a espandere il linguaggio della teoria risultante con il nuovo predicato *T*. Poiché, l'assioma di induzione non è ristretto segue che potranno occorrere formule in cui T occorre. Ma, ed è qui la chiave del ragionamento, la teoria della verità non conservativa *CT* è ottenuta aggiungendo a *PA* le clausole T1-T4 e l'assioma di induzione. Se a *PA* aggiungiamo le clausole tarskiane senza estendere il principio di induzione otteniamo la teoria $CT\upharpoonright$ che come abbiamo accennato nel capitolo precedente *è conservativa* su *PA*. Ora, continua Field, in base a quanto detto in precedenza, si deduce che è proprio l'assioma di induzione non ristretto che conferisce alla teoria di base $CT\upharpoonright$ l'incremento informativo da cui si deduce la sostanzialità della nozione di verità. Quindi tutta la questione si gioca, secondo Field, sull'assioma di induzione non ristretto. Ora, è noto che

> (A) The induction axioms are needed if we are to arithmetically derive important facts that involve the notion of truth[13].

E su questo punto non si può certo dissentire: gli assiomi di induzione cui ci si riferisce sono quelle particolari istanze del principio di induzione che contengono occorrenze del predicato di verità. Come abbiamo già avuto modo di vedere, l'impiego di queste istanze è essenziale per la derivazione di particolari generalizzazioni quali (Corr). Ma, secondo la ricostruzione che Field fa dell'argomento di conservatività, è necessario affermare qualcosa del genere

> (B) The truth of the induction axioms depends only on the nature of truth[14].

Infatti, se (B) fosse vera allora sarebbe del tutto pacifico ascrivere al predicato di verità la responsabilità della non conservatività dell'estensione. Tuttavia, quest'ultima affermazione appare chiaramente falsa. La verità degli assiomi di induzione dipende dalla struttura induttiva degli oggetti cui si applicano, e cioè dal fatto che

> Natural numbers are linearly ordered with each element having only finitely many predecessors[15].

Ma a questo punto Field ha buon gioco nell'affermare che proprio perché l'assioma di induzione non ristretto non è un principio che ha a che fare solamente con la verità, la non conservatività è un fenomeno aritmetico; d'altro canto, utilizzando solo principi 'puri' (quali le clausole tarskiane T1-T4) otteniamo un'estensione conservativa dimostrando, ancora una volta, la 'leggerezza' del predicato di verità.

Detto altrimenti, la non conservatività di *CT* su *PA* è un fatto e come tale ha delle ragioni: Field, correttamente, pone l'accento sull'impiego dell'induzione estesa al predicato *T* e si chiede: da che cosa dipende la verità di questo principio? Trattandosi di un assioma (o meglio, di uno schema di assiomi) la verità dell'induzione dipende dalla natura degli

13 Field 1999, 538.
14 Field 1999, 538.
15 Field 1999, 538.

oggetti cui si applica e quindi dalla natura (induttiva) dei numeri naturali. Ma allora - Field ribadisce - la validità dell'induzione è una questione aritmetica e non dipende dalla natura del predicato di verità. Infatti, e questo è assunto a riprova del suo ragionamento, gli assiomi "puri" che caratterizzano il predicato *T* e cioè le clausole T1-T4 sono conservative una volta aggiunte a una teoria di base *senza* estendere l'assioma di induzione l'aritmetica (è ciò che si fa, per l'appunto, nella teoria $CT\upharpoonright$).

L'argomento di Field può essere attaccato muovendo in base a due strategie differenti: la prima, più "rapida" fa leva su una sorta di 'simmetria' argomentativa in base alla quale la critica di Field può essere ribaltata e perdere così di efficacia; la seconda, invece, tocca una serie di tematiche più profonde e relativamente indipendenti dalla questione della conservatività e del deflazionismo. Vediamole dunque separatamente.

I critica

La classica controbiezione all'argomento di Field parte dalla seguente osservazione. Field, schematizzando un po', ragiona così: la teoria *CT* è composta da tre elementi (assiomi di *PA*, principio di induzione per *PA* e clausole tarskiane) che combinati comportano la non conservatività di *CT* su *PA*.

(7) *PA* + Induzione + Clausole tarskiane ⇒ non conservatività

La teoria $CT\upharpoonright$, invece, presenta sempre tre elementi ma, questa volta, l'induzione è ristretta, ovvero non comprende casi in cui nella formula induttiva occorra il predicato T. E, come abbiamo visto, si ottiene la conservatività di $CT\upharpoonright$ su PA.

(8) *PA* + Induzione$^{(-)}$ + Clausole tarskiane ⇒ conservatività

In questo caso il responsabile del passaggio da una teoria conservativa ($CT\upharpoonright$) a una non conservativa (*CT*) è l'induzione non ristretta. Ma in realtà, già il primo passo della critica di Field non è del tutto giustificato. Infatti, prendiamo in esame *TB*, cioè la teoria che codifica le intuizioni decitazionali. *TB* è costituita da:

(9) *PA* + Induzione + Bicondizionali ⇒ conservatività

Come è noto, infatti, *TB* risulta conservativa su *PA*. Ora, se al posto della serie dei bicondizionali, mettiamo le clausole tarskiane otteniamo:

(10) *PA* + Induzione + Clausole tarskiane $\Rightarrow$ non conservatività

cioè la teoria *CT*. Ma in questo caso, il responsabile della non conservatività, cioè, dell'aumento di informazione, non è certo l'assioma di induzione non ristretto, dal momento che questo è presente *già* in *TB*. Piuttosto, sembra essere la natura composizionale delle clausole tarskiane a comportare la non conservatività. Ma le clausole tarskiane, in quanto tali, sono principi che hanno a che fare esclusivamente con il predicato di verità e che quindi non riguardano la natura numerica degli oggetti cui si applicano. Contro Field, quindi, non è solo l'induzione a conferire incremento informativo: ci sono casi in cui, a parità di induzione non ristretta, il passaggio da teorie conservative a teorie non conservative è dato proprio dagli assiomi specifici della nozione di verità.

Ketland dedica a questo punto una serie di riflessioni piuttosto interessanti. Dice, infatti:

> In general, non-conservativeness occurs when we can prove 'all theorems of B are true'. We shall be able to do this if we can prove 'all axioms of B are true' and 'deductive inference preserves truth'[16].

Questo è proprio il risultato $CT \vdash Cons_{PA}$ presentato nel capitolo precedente. Ma, come Ketland sottolinea, l'aspetto interessante della dimostrazione sta nel fatto che bisogna far vedere che gli schemi di assiomi sono corretti, ovvero, nel nostro caso, che ogni istanza dello schema di induzione è vera.

> The subtlety that arises here concerns proving that all instances of some *axiom scheme* [...] are true. Given our truth theory [...] in the metalanguage, and an object language *scheme* [...] we aim to prove its soundness[17].

Ora, come nota Ketland, poco dopo, per dimostrare che lo schema di induzione è corretto noi assumiamo l'assioma di induzione della metateoria, ovvero la sua versione non ristretta:

---

16 Ketland 2010, 426.
17 Ketland 2010, 426.

> In general, we *assume* the scheme Φ to prove 'Φ is sound'[18].

In realtà, ciò non avviene solo per lo schema di induzione ma per tutte le clausole tarskiane. Cosa vuol dire, infatti, dimostrare che una determinata regola logica è corretta? Significa far vedere come, se le ipotesi sono corrette, lo sarà anche la conclusione. Il caso della congiunzione può chiarire la questione: per dimostrare che l'introduzione della congiunzione è una regola corretta, bisogna utilizzare il principio per cui, se $T(\alpha)$ e $T(\beta)$ allora $T(\alpha \wedge \beta)$.[19] Cioè esattamente la clausola tarskiana della congiunzione. Tutto questo a riprova del fatto che:

> [...] it is the compositionality of the principles governing truth which explains non-conservativeness, as the disquotational truth theory *does* remain conservative when induction and other schemes are extended[20].

II critica

Un'altra possibile contromossa all'argomento di Field consiste nel mettere in discussione l'assunzione implicita che l'autore compie; perché l'argomento abbia valore è necessario, infatti, poter distinguere, all'interno di una teoria formale assiomatica come *CT* (o *PA* stessa) il tipo di contenuto (logico, aritmetico, sintattico o veritativo) che un determinato teorema veicola. Si tratta di un tema che, come sarà evidente a breve, eccede la sua applicazione nella discussione circa le teorie formali della verità.[21]

Posta in termini un po' brutali la questione può essere riassunta come segue: sia *CT* la nostra teoria della verità per *PA*. Ora, abbiamo, rispettivamente:

(11) $CT \vdash 2+2=4$

---

18 Ketland 2010, 427.

19 Rimane aperta, naturalmente, l'interessante questione filosofica se la dimostrazione di correttezza attraverso l'utilizzo degli stessi principi (a livello della metateoria) sia tale da innescare un circolo vizioso o meno. Si tratta di un tema interessante che però non considereremo in questa sede.

20 Ketland 2010, 427.

21 Lo stesso Halbach accenna a questo fatto nelle riflessioni finali del suo lavoro. Cfr. Halbach 2011, 312-320.

e

(12) $CT \vdash T(\ulcorner 2+2=4 \urcorner) \leftrightarrow 2+2=4$

Il primo caso è un teorema di *PA*. Il secondo un'istanza del T-schema. Ebbene, perché l'argomento di Field abbia efficacia è necessario che il primo risultato sia di natura 'aritmetica' mentre il secondo di natura puramente '*truth-theoretical*'. Fino a che punto è possibile distinguere gli ambiti (quello aritmetico e quello veritativo) all'interno di una teoria della verità come *CT*?

La questione diventa ancora più complessa dal momento che praticamente tutte le teorie assiomatiche della verità fanno ricorso alla procedura dell'aritmetizzazione per avere a disposizione i nomi degli enunciati (cioè degli oggetti cui predicare la verità). Ma aritmetizzare vuol dire, in ultima analisi, tradurre in numeri strutture sintattiche. E, quindi, di cosa si sta parlando quando si afferma, per esempio, che *CT* dimostra la consistenza di *PA*? Di numeri? Di codici numerici che indicano strutture sintattiche? Di strutture sintattiche? Di proprietà di teorie?

L'intuizione dalla quale vogliamo partire (che non sappiamo se Field approverebbe) è che, normalmente, il contenuto di un enunciato di una teoria assiomatica dipende dal riferimento inteso della teoria e quindi dell'enunciato stesso. Detto in altri termini, il contenuto di un teorema di *PA* è determinato dal fatto che *PA* è una teoria che *parla* dei numeri naturali. Come *PA* sia in grado di parlare dei numeri, con quali limitazioni strutturali e che conseguenze ciò abbia per la filosofia della matematica sono problemi che meritano una discussione approfondita; resta il fatto che, anche a un livello minimale, possiamo concordare su questo punto e cioè che per stabilire il contenuto (o la natura) di un enunciato di una teoria dobbiamo andare a vedere il dominio oggettuale della teoria in esame.

Nel caso che ci interessa qui, però, le cose non sono così semplici e questo per due ordini di ragioni. Innanzitutto, per l'aritmetizzazione e poi per la presenza stessa di assiomi che hanno a che fare con la nozione di verità. Procediamo con ordine. Che cosa accade con l'aritmetizzazione? Una teoria numerica è in grado di rappresentare la propria sintassi. Abbiamo cioè a disposizione un codice che associa numeri naturali alle strutture sintattiche della teoria. Possiamo poi definire delle funzioni e dei predicati che esprimono proprietà sintattiche delle espressioni del linguaggi opportunamente aritmetizzate. Come avevamo visto nel capitolo introduttivo:

(13) $PA \vdash Sent\left(\ulcorner 2+2=4 \urcorner\right)$

dice che 2 + 2 = 4 è un enunciato di *PA*. Applichiamo il criterio intuitivo esposto in precedenza a (13): diremo che il contenuto di (13) è una certa proprietà di un oggetto sintattico: in questo caso infatti, *PA* non descrive la struttura dei numeri naturali ma, mediante il riferimento ai numeri naturali, dice qualcosa sulla sua struttura linguistica. Gli oggetti di *PA*, potremmo dire, non sono in questo caso i numeri ma alcuni elementi linguistici. Naturalmente possiamo dire ciò, solo perché noi ci collochiamo a un meta-livello: in altre parole, 'vediamo' *PA* e 'adoperiamo' l'aritmetizzazione per nominare gli oggetti della nostra teoria. A livello di *PA* abbiamo solo relazioni numeriche più o meno complesse che sono derivabili dagli assiomi di base: nel caso di (13), per esempio, si dichiara che un determinato numero *k* è uguale al codice di 2 + 2 = 4 e che questo numero appartiene all'insieme dei codici di enunciati ben formati della teoria. Questa relazione aritmetica è letta da noi come la predicazione di una proprietà sintattica ('essere un enunciato') a un oggetto sintattico (una sequenza di segni del linguaggio).

L'aritmetizzazione permette di "sdoppiare" il significato di un numerale: da un lato è, *at face value*, il nome di un numero cioè di un elemento di $\mathbb{N}$; dall'altro è il codice di una certa combinazione di segni del linguaggio di *PA*.

Che cosa avviene allora quando passiamo alla teoria formale della verità per *PA*, cioè alla teoria *CT*? Qui le cose si fanno ancora più complicate perché oltre agli assiomi aritmetici abbiamo anche le note clausole tarskiane T1-T4 e l'assioma di induzione che è al centro dell'argomento di Field. Per riprendere, però, i nostri esempi, possiamo avere qualcosa come:

(14) $CT \vdash T\left(\ulcorner 2+2=4 \urcorner\right)$

Anche in questo caso, *CT* afferma l'esistenza di una determinata relazione di appartenenza: il numero che codifica l'enunciato 2 +2 = 4 appartiene all'insieme $\mathcal{E}$ cioè all'estensione del predicato T. La differenza con (13) sta nel fatto che in questo caso abbiamo un predicato primitivo, la cui 'natura' non è aritmetica; ed è per questo motivo che parliamo di teorie della verità, intendendo così l'oggetto inteso degli assiomi tarskiani.

Ma il punto è esattamente questo. Qual è il dominio inteso di una teoria come *CT*? L'argomento di Field funzionerebbe, o acquisirebbe se non altro, una forte plausibilità, se e soltanto se fossimo in grado di distinguere sempre tra contenuto aritmetico e veritativo in un enunciato di *CT*. Infatti, se fosse così, l'assioma di induzione sarebbe un principio squisitamente aritmetico e quindi anche la non conservatività seguirebbe dall'incremento informativo degli assiomi matematici e non tarskiani.

Seguendo il ragionamento fino ad ora sviluppato, però, ci accorgiamo che un enunciato come (14) ripropone proprio quella molteplicità di livelli che avevamo già riscontrato in precedenza. Anzi, in questo caso, le cose si complicano maggiormente;

(14) $CT \vdash T\left(\ulcorner 2+2=4 \urcorner\right)$

afferma almeno *tre* cose: una relazione fra numeri naturali, una relazione sintattica e una relazione semantica. Da un lato, infatti, (14) dice che un determinato numero *k* appartiene a un insieme $\mathcal{E}$, cioè all'insieme costituito da tutti quei numeri che sono i codici degli enunciati veri di PA. Ma (14) afferma anche che un oggetto sintattico, e cioè un enunciato, è vero. E se questo vale per (14) che è, tutto sommato un esempio abbastanza semplice, varrà a maggior ragione per

(15) $CT \vdash Cons_{PA}$

Qui non abbiamo l'occorrenza del predicato T, anche se, come abbiamo visto è necessaria tutta la potenza di CT per ottenere (15). $Cons_{PA}$ è l'espressione formale della coerenza della teoria *PA*; ovvero la non esistenza di un numero che sia il codice della dimostrazione di una contraddizione in *PA*. Quindi, diremmo che il contenuto di $Cons_{PA}$ è essenzialmente 'sintattico' e non aritmetico e che il riferimento ai numeri naturali è solo *via codice* e non primario.

Tuttavia, quando abbiamo a che fare con il predicato di verità, siamo in presenza di un terzo livello di significato (o di contenuto). Dire infatti che è vero che 2 + 2 = 4 significa sia dire qualcosa su un enunciato aritmetico sia affermare qualcosa sulla realtà stessa dei numeri. Ed è questa, a nostro avviso, la lezione più profonda della teoria della verità tarskiana alla cui base sta il nocciolo di realismo che abbiamo analizzato nella prima parte del nostro volume. Ogni teoria della verità si muove su un doppio registro linguistico e semantico: da un lato, è una teoria di determinati oggetti (i truth-bearers) e dice quali di questi oggetti sono veri; dall'altro è una teoria degli oggetti (o delle entità) rappresentate dai primi. Semplificando al massimo, quando diciamo che

(16) È vero che la neve è bianca

stiamo sicuramente ascrivendo una certa proprietà a un'espressione del nostro linguaggio ma stiamo dicendo anche qualcosa sulla neve e sul colore bianco. Questo duplice livello diventa triplice nel caso delle teorie assiomatiche che hanno una base aritmetica e che sono in grado di rappresentare la loro sintassi. *CT*, come *PA*, in un certo senso parla *solo* di numeri ma questi numeri possono voler dire cose differenti. Noi notiamo la differenza proprio perché possiamo astrarre dal contesto teorico; detto altrimenti, noi non viviamo in *PA* (o in *CT*), ma poiché questa distinzione avviene a un livello che non è formalizzato nella teoria stessa, l'argomento di Field non sembra convincente. Insomma, non c'è un'etichetta che comunica il contenuto inteso degli enunciati di una teoria.

Anzi, come abbiamo visto, proprio la duplice 'natura' della nozione di verità che emerge dallo studio delle teorie assiomatiche coglie molto bene l'intuizione classica secondo cui la verità connette - in un certo modo - il linguaggio (o il pensiero) e il mondo. E di questa connessione ne è chiaro esempio il T-schema:

(17) $CT \vdash T(\ulcorner 2+2=4 \urcorner) \leftrightarrow 2+2=4$

A sinistra del bicondizionale abbiamo la predicazione sintattica, e cioè l'essere vero da parte di un enunciato. A destra, invece, la relazione fondamentale è di tipo aritmetico *puro*. Ebbene, proprio nella misura in cui è impossibile *teoricamente* (cioè a partire dalla teoria) discernere tra componenti veritative, numeriche e sintattiche emerge abbastanza chiaramente che la nozione di verità tarskiana, pur con tutte le cautele - sono molte - del caso, presenta un rimando a una qualche dimensione ontologica.

BIBLIOGRAFIA

AGAZZI, EVANDRO
- (1994), *On Formalism*, in G. FLØISTAD (ed.), *Philosophical Problems Today*, Kluwer Academic Publishers, Dordrecht, 1, pp. 74-137.

BAYS, TIMOTHY
- (2001), On Tarski on models, in “Journal of Symbolic Logic”, 66(4), pp. 1701-1726.
- (2009), Beth’s Theorem and Deflationism, in “Mind”, 118(472), pp. 1061-1073.

BENACERRAF, PAUL
- (1973), *Mathematical Truth*, in “Journal of Philosophy”, 70, pp. 661-679.

BUSS, SAMUEL
- (1998), *Handbook of proof theory*, Elsevier, Amsterdam.

CHANG, CHING; KEISLER, H. J.
- (1973), *Model Theory*, North-Holland, Amsterdam.

CIEŚLIŃSKI, CEZARY
- (2010), Truth, Conservativeness, and Provability, in “Mind”, 119(474), pp. 409-422.

COCCHIARELLA, NINO
- (1988), Predication versus membership in the distinction between logic as language and logic as calculus, in “Synthese”, 77, pp. 37-72.

ETCHEMENDY, JOHN
- (1990), The Concept of Logical Consequence, Harvard University Press, Cambridge (ma).

FEFERMAN, SOLOMON
- (2008), *Tarski's conceptual analysis of semantical notions*, in D. Patterson (ed.), *New Essays on Tarski and Philosophy*, Oxford University Press, Oxford.

FEFERMAN, SOLOMON; FEFERMAN, ANITA BURDMAN
- (2004), Alfred Tarski. Life and logic, Cambridge University Press, Cambridge.

FIELD, HARTRY
- (1972), Tarski's theory of truth, in "*The Journal of Philosophy"*, 69(13), pp. 347-375.
- (1999), Deflating the conservativeness argument, in "The Journal of Philosophy", 96(10), pp. 533-540.

FROST-ARNOLD, GREG
- (2004), Was Tarski's theory of truth motivated by physicalism?, in "History and Philosophy of Logic", 25(4), pp. 265-280.

GALVAN, SERGIO
- (1973), *Il concetto di verità di A. Tarski*, in "Verifiche", 2, pp. 3-66.
- (1983), *Teoria formale dei numeri naturali*, FrancoAngeli, Milano.
- (1992), *Introduzione ai teoremi di incompletezza*, FrancoAngeli, Milano.
- (2011), *Finitist Objects*, in C. KANZIAN (ed.), *The Ways Things Are. Studies in Ontology*, Ontos Verlag, Frankfurt, pp. 149- 166.

GÖDEL, KURT
- (1929), *Über die Vollständigkeit des Logikkalküls*, in S. FEFERMAN ET AL. (eds.), *Collected Works*, *Vol I: 1929-1936*, Clarendon Press, Oxford 1986, pp. 60-101.
- (1944), *Russell's Mathematical Logic*, in S. FEFERMAN ET AL. (eds.), *Collected Works*, *Vol II: 1936- 1974*, Clarendon Press, Oxford 1990, pp. 102-141.

GOMEZ-TORRENTE, MARIO
- (1996), Tarski on logical consequence, in "Notre Dame Journal of Formal Logic", 37(1), pp. 125-151.
- (1998), Logical truth and Tarskian logical truth, in "*Synthese"*,117(3), pp. 375-408.

GRATTAN-GUINNESS, IVOR
- (1979), In memoriam Kurt Gödel: his 1931 correspondence with Zermelo on his incompletability theorem, in "*Historia mathematica*", 6(3), pp. 294-304.

JANÉ, IGNASI
- (2005), *Higher-order logic reconsidered*, in S. SHAPIRO (ed.), *The Oxford handbook of philosophy of mathematics and logic*, Oxford University Press, Oxford, pp. 781-810.

HÁJEK, P., & PUDLÁK, P.
- (1993), *Metamathematics of first-order arithmetic*, Springer-Verlag, Berlin.

HALBACH, VOLKER
- (1994), A system of complete and consistent truth, in "*Notre Dame Journal of Formal Logic"*, 35(3), pp. 311-327.
- (1999), Conservative theories of classical truth, in "*Studia Logica"*, 62(3), pp. 353-370.
- (1999), Disquotationalism and infinite conjunctions, in "*Mind"*, 108(429), pp. 1-22.
- (2000), Truth and reduction, in "*Erkenntnis"*, 53(1), pp. 97-126.
- (2001), How innocent is deflationism?, in "*Synthese"*, 126(1), pp. 167-194.
- (2001), Disquotational truth and analyticity, in "*Journal of Symbolic Logic"*, 66(4), 1959-1973.
- (2011), *Axiomatic theories of truth*, Cambridge University Press, Cambridge.

HATCHER, WILLIAM
- (1968), *Foundations of mathematics*, Saunders, Philadelphia.

HILBERT, DAVID; ACKERMANN, WILHELM
- (1929), *Grundzüge der theoretishen Logik*, Springer, Berlin; eng. tr. R.E. LUCE, *Principles of Mathematical Logic*, Chelsea Publishing Company, New York 1950.

HINTIKKA, JAAKKO; HINTIKKA, MERRILL
- (1989), *Investigating Wittgenstein*, Blackwell, Oxford.

HODGES, WILFRID
- (1985), Truth in a structure, in "Proceedings of the Aristotelian Society", 86, pp. 135-151.

KETLAND, JEFFREY
- (1999), Deflationism and Tarski's paradise, in "Mind", 108(429), pp. 69-94.
- (2005), Deflationism and the Gödel phenomena: reply to Tennant, in "Mind", 114(453), pp. 75-88.
- (2010), Truth, Conservativeness, and Provability: a reply to Cielisnski, in "Mind", 119(474), pp. 423-436

KIRKHAM, RICHARD
- (1992), *Theories of Truth: A Critical Introduction*, MIT Press, Cambridge (MA).

KOLÁŘ, PETR
- (1999), Truth, correspondence, satisfaction, in J. Peregrin (ed), *Truth and its Nature (if any),* Kluwer Academic Press, Dordrecht, pp. 67-79.

KOTLARSKI, H., KRAJEWSKI, S., & LACHLAN, A. H.
- (1981), Construction of satisfaction classes for nonstandard models, in "Canadian Mathematical Bulletin", 24(3), pp. 283-293.

KUIPERS, THEO
- (2000), *From Instrumentalism to Constructive Realism: On Some Relations between Confirmation, Empirical Progress, and Truth Approximation*, Kluwer Academic Publisher, Dordrecht.

KUNNE, WOLFGANG
- (2003), *Conceptions of Truth*, Clarendon Press, Oxford.

KUSCH, MARTIN
- (1989), *Language as Calculus Vs. Language as Universal Medium: A Study in Husserl, Heidegger and Gadamer*, Kluwer Academic Publisher, Dordrecht.

MANCOSU, PAOLO
- (2008), Tarski, Neurath, and Kokoszynska on the semantic conception of truth, D. PATTERSON (ed.), *New Essays on Tarski and Philosophy*, Oxford University Press, Oxford, pp. 192-224.
- (2010), *Fixed-versus Variable-domain Interpretations of Tarski's Account of Logical Consequence*, in "Philosophy Compass", 5(9), pp. 745-759.

MANCOSU, PAOLO; ZACH, RICHARD; BADESA, CALIXTO
- (2009), *The development of Mathematical logic from Russell to Tarski: 1900-1935*, in L. HAAPARANTA (ed.), *The Development of Modern Logic*, Oxford University Press, Oxford, pp. 318-470.

MCDOWELL, JOHN
- (1978), Physicalism and primitive denotation: Field on Tarski, in "Erkenntnis", 13(1), pp. 131-152.

MORENO, FERNÁNDEZ
- (2001), *Tarskian truth and the correspondence theory*, in "Synthese", 126(1), pp. 123-148.

MURAWSKI, ROMAN
- (1998), Undefinability of truth. The problem of priority: Tarski *vs* Gödel, in "History and Philosophy of Logic", 19, pp. 153-160.
- (2005), *Recursive functions and metamathematics: problems of completeness and decidability, Gödel's theorems*, Kluwer Academic Press, Dordrecht.

NIINILUOTO, ILKKA
- (1994), *Defending Tarski against his critics*, in J. WOLENSKI (ed.), *Sixty Years of Tarski's Definition of Truth*, Philed, Cracow, pp. 48–68.
- (1999), Tarskian Truth as Correspondence-Replies to Some Objections, in J. PEREGRIN (ed), *Truth and its Nature (if any),* Kluwer Academic Press, Dordrecht, 91-104.

- (2004), *Tarski's definition and truth-makers*, in "Annals of Pure and Applied Logic", 126, pp. 57-76.

PARSONS, CHARLES
- (1983), *Mathematics in philosophy: Selected essays*, Cornell University Press, Ithaca New York.
PATTERSON, DOUGLAS
- (2003), What is a correspondence theory of truth?, in "Synthese",137(3), pp. 421-444.

POPPER, KARL
- (1963), *Conjectures and refutations: The growth of scientific knowledge*. Routledge & K. Paul, London.

QUINE, WILLARD VAN
- (1970), *Philosophy of logic*, Englewood Cliffs, Prentice-Hall.
- (1990), *Pursuit of Truth*, Harvard University Press, Cambridge (MA).

RAATIKAINEN, PANU
- (2008), Truth, meaning, and translation, in D. PATTERSON (ed.), *New Essays on Tarski and Philosophy*, Oxford University Press, Oxford, pp. 247-62.

ROGERS, ROBERT
- (1971), *Mathematical logic and formalized theories: a survey of basic concepts and results*, North-Holland, Amsterdam

ROJSZCZAK, ARTUR
- (2002), *Philosophical background and philosophical content of the semantic definition of truth*, in "Erkenntnis", 56(1), pp. 29-62.

SHAPIRO, STEWART
- (1998), Proof and truth: Through thick and thin, in "The journal of philosophy", 95(10), pp. 493-521.
- (1996) (ed.), *The limits of logic: higher-order logic and the Löwenheim-Skolem theorem*, Dartmouth, Cambridge.
- (2005), *Higher-order logic*, in S. SHAPIRO (ed.), *The Oxford handbook of philosophy of mathematics and logic*, Oxford University Press, Oxford, pp. 751-780.

SHER, GILA
- (1991), *The Bounds of Logic: a generalized viewpoint*, MIT Press, Cambridge (MA).
- (1998), *On the possibility of a substantive theory of truth*, in "Synthese", 117(1), pp. 133-172.
- (1999), What is Tarski's Theory of Truth?, in "Topoi", 18(2), pp. 149-166.
- (2004), In search of a substantive theory of truth, in "The Journal of philosophy", 101(1), pp. 5-36.

SHOENFIELD, JOSEPH
- (1967), *Mathematical Logic*, Addison-Wesley, Reading (MA)

SIMPSON, STEPHEN
- (1999), *Subsystems of second-order Arithmetic*, Spriger, Berlin.

SMITH, PETER
- (2007), *On some subsystems of second-order Arithmetic*, unpublished manuscript.

SMORYNSKI, C.
- (1977), *The incompleteness theorems*, in J. BARWISE (ed), *Handbook of mathematical logic*, North-Holland, Amsterdam, pp. 821-865.

SOAMES, SCOTT
- (1984), What is a Theory of Truth?, in "The Journal of Philosophy", 81(8), pp. 411-429.
- (1998), *Understanding Truth*, Oxford University Press, New York.

STEGMULLER, WOLFGANG
- (1968), *Das Wahrheitsproblem und die Idee der Semantik: eine Einführung in die Theorien von A. Tarski und R. Carnap*, Springer-Verlag, Wien *et al*.

TAIT, WILLIAM
- (1981), Finitism, in "The Journal of Philosophy", 78, pp. 524-546.

TAKEUTI, GAISI
- (1987), *Proof Theory*, North-Holland, Amsterdam.

TARSKI, ALFRED
- (1935), Der Wahrheitsbegriff in dem formalisierten Sprachen, in “Studia Philosophica”, 1, pp. 261-405 (eng. tr., *The concept of Truth in Formalized Languages*, in J. CORCORAN (ed.), *Logic, semantics, metamathematics: papers from 1923 to 1938,* Hackett Publishing Company Incorporated, Indianapolis 1983, pp. ??)
- (1944), *The semantics conception of truth and the foundation of semantics*, in “Philosophy and Phenomenological Research”, 4, pp. 341-376. (trad. it., *La concezione semantica della verità e la fondazione della semantica*, in L. LINSKY (ed.), *Semantica e filosofia del linguaggio*, Il Saggiatore, Milano 1969).
- (1969), Truth and proof, in “Scientific American”, June 1969, pp. 63-70, 75-77.

TENNANT, NEIL
- (2002), Deflationism and the Gödel phenomena, in “Mind”, 111(443), pp. 551-582.
- (2005), Deflationism and the Gödel Phenomena: Reply to Ketland, in “Mind”, 114(453), pp. 89-96.

VAN HEIJENOORT, JEAN
- (1967), Logic as Language and Logic as Calculus, in “*Synthese*”, 17, pp. 324–30.

WITTGENSTEIN, LUDWIG
- (1971), *Tractatus Logico-Philosophicus*, eng. tr. D.F. PEARS AND B.F. MCGUINNESS, Routledge and Kegan Paul, 1961.

WOLENSKI, JAN
- (1993), Alfred Tarski as a philosopher, in CONIGLIONE, F., POLI, R., & WOLEŃSKI, J. (Eds.), *Polish Scientific Philosophy: The Lvov-Warsaw School*, Rodopi, Amsterdam.
- (2002), From Intentionality To Formal Semantics (From Twardowski To Tarski), in “*Erkenntnis*”, 56(1), pp. 9-27.

WOLENSKI, JAN; SIMONS, PETER
- (1989), *De veritate: Austro-Polish contributions to the theory of truth from Brentano to Tarski*, in K. SZANIAWSKI (ed.), The Vienna Circle and

the Lvov–Warsaw School, Kluwer Academic Publishers, Dordrecht, pp. 391-443.

WRIGHT, CRISPIN
- (1994), *Truth and objectivity*, Harvard University Press, Cambridge.
- (1998), *Truth: a traditional debate reviewed*, in “Canadian journal of philosophy”, 24, pp. 31-74

FILOSOFIE ANALITICHE / METAFISICA

Collana diretta da *Massimiliano Carrara*

1. Andrea Bottani, Richard Davies (a cura di), *Ontologie regionali*
2. Richard Davies, *Gli oggetti della logica. Il folclore della filosofia analitica*
3. Pierdaniele Giaretta, Giuseppe Spolaore, *Esistenza e identità*
4. Vittorio Morato, *Modalità e mondi possibili*
5. Giuseppe Spolaore, *Logos in fabula. Un'indagine filosofica sui personaggi letterari*
6. Aldo Frigerio, *L'unità dell'enunciato. Indeterminismo semantico e determinazioni pragmatiche*
7. Massimiliano Carrara, Vittorio Morato (a cura di), *Verità, Annuario e Bollettino della Società Italiana di Filosofia Analitica (SIFA) 2012*
8. Barbara Giolito, *Epistemologia e Ragionamento*
9. Richard Davies (a cura di), *Analisi, Annuario e Bollettino della Società Italiana di Filosofia Analitica (SIFA) 2011*
10. M. Cristina Amoretti, Massimiliano Vignolo (a cura di), *Disaccordo, Annuario e Bollettino della Società Italiana di Filosofia Analitica (SIFA) 2012*
11. Giorgio Lando, *Forme, relazioni, oggetti. Saggio sulla metafisica del Tractatus Logico-Philosophicus*