



Università degli Studi di Torino  
Facoltà di Lettere e Filosofia  
Laurea Magistrale in Filosofia

**La competenza semantica negli esseri umani e nelle  
macchine. Teorie a confronto.**

Candidato: Degasperi Eleonora Anna

Relatore: Diego Marconi

Anno Accademico 2013 - 2014

## INDICE

1. INTRODUZIONE.....	pag. 2
2. LUDWIG WITTGENSTEIN.....	pag. 4
2.1. Il problema della denotazione,,.....	pag. 4
2.2. Significato e regole d'uso.....,,.....	pag. 6
2.3. Seguire la regola.....,,.....	pag. 8
2.4. Concordanza tra regole e usi.....	pag. 16
3. JERRY A. FODOR.....	pag. 21
3.1. Fodor e la Macchina di Turing.....	pag. 21
3.2. Il linguaggio del pensiero.....	pag. 23
3.3. Gli atteggiamenti proposizionali.....	pag. 30
4. JOHN SEARLE.....	pag. 35
4.1. L'esperimento della Stanza Cinese.....	pag. 36
4.2. Auto-obiezioni di Searle.....	pag. 40
4.3. Critiche e commenti al saggio di Searle.....	pag. 45
4.3.1. Searle sottovaluta l'importanza dell'applicazione delle regole.....	pag. 49
4.3.2. Critiche al concetto di intenzionalità.....	pag. 51
4.3.3. Critiche alla contrapposizione tra semantica e sintassi.....	pag. 54
4.4. Conclusione .....	pag. 57
5. CONCLUSIONE.....	pag. 63
5.1. Wittgenstein e Fodor.....	pag. 63
5.2. Turing.....	pag. 70
5.3. Wittgenstein e Searle.....	pag. 74
5.4. Riconsiderare il concetto di 'comprensione'.....	pag. 78

## 1. INTRODUZIONE

Può pensare una macchina? Può possedere stati cognitivi? Può comprendere il linguaggio così come lo comprendono gli esseri umani? Per rispondere a queste domande si sono prese in considerazione tre teorie, ciascuna delle quali ha trattato alcuni dei problemi che l'Intelligenza Artificiale ha affrontato (e sta affrontando) da quando è nata fino ad oggi: Wittgenstein e le problematiche legate al concetto di seguire la regola, la sua teoria del significato come uso e la sua teoria dei giochi linguistici, Fodor e la sua teoria del linguaggio del pensiero, per mezzo della quale l'autore sostiene una forte analogia tra il funzionamento della mente dell'uomo e quello della macchina, Searle e la critica all'idea delle macchine pensanti portata avanti a partire dal suo esperimento della stanza cinese e dalla sua teoria sull'intenzionalità.

Fodor e Searle si inseriscono in un panorama filosofico che è sicuramente più eterogeneo rispetto a quello in cui si inserisce Wittgenstein: le loro teorie si appoggiano, ma soprattutto si completano, con l'aiuto di altre discipline quali la psicologia, intelligenza artificiale, l'informatica, la neurologia, etc, discipline che sono molto più moderne e che sfruttano mezzi e concetti che ai tempi di Wittgenstein non erano ancora presenti nei dibattiti filosofici. Molto di ciò che viene detto da Fodor e da Searle sfrutta le competenze di rami scientifici e cognitivi che negli anni di Wittgenstein non si erano ancora sviluppati a sufficienza per poter diventare parte di un dibattito teorico quale quello che si è instaurato con la nascita dell'intelligenza artificiale. Fodor e Searle hanno teorie radicalmente diverse, il primo pensa alla mente dell'uomo come ad un elaboratore elettronico, il secondo pensa all'uomo come dotato essenzialmente di intenzionalità e quindi non paragonabile ad una macchina. Ma sono teorie che nascono su un terreno comune, cioè entrambe sfruttano quelle nuove conoscenze che scaturiscono da nuovi studi sulla mente umana e sui calcolatori elettronici e sulla possibilità o meno di porre una forte analogia tra queste due sfere.

Cosa c'entra, a questo punto, Wittgenstein, dal momento che egli non ha mai trattato problemi di questo genere? Wittgenstein, in effetti, non si è mai interessato alla possibilità o impossibilità delle macchine pensanti, ma nello sviluppo della sua teoria del linguaggio aveva colto ciò che c'è di essenziale nell'attività linguistica dell'uomo, cioè il fatto che gli uomini si comprendono tra di loro perché il linguaggio funziona in un certo modo e perché vi è una condivisione dei suoi meccanismi. Dal momento che l'analogia tra l'uomo e la macchina diventa problematica non tanto quando si parla a livello fisico ma quando si parla del problema a livello cognitivo, ciò che sembra interessante è quindi capire in che modo funziona il linguaggio per poter arrivare a rispondere se una macchina può o non può comprendere

ciò che l'uomo le sta trasmettendo. Il fatto che il linguaggio sia costituito da giochi linguistici governati da regole e che questi, se si seguono correttamente, possono far nascere una prassi linguistica condivisa dagli esseri umani, può ritenersi un punto di avvio per iniziare una riflessione rivolta a capire in che modo le macchine possano venir "addestrate" come gli esseri umani per arrivare a condividere la stessa competenza semantica.

I problemi che si affrontano per rispondere alle domande che ci si è posti vengono trattati principalmente sulla base di ciò che ci dicono le discipline scientifiche come la neurologia, l'informatica, la psicologia cognitiva, la biologia, ma penso che possa anche risultare utile affrontare queste discussioni sulla base di ciò che ci dice la teoria del linguaggio di Wittgenstein, cioè attraverso il problema del seguire la regola, la teoria dei giochi linguistici, l'aspetto dell'addestramento linguistico, e quindi affrontando questa questione non solo più in termini psicologistici, cognitivi e scientifici ma anche e soprattutto filosofici.

I primi tre capitoli si concentreranno su ciò c'è di significativo ed essenziale nella teoria di ciascun autore, mentre l'ultimo capitolo si concentrerà sul confronto tra le teorie di Fodor e di Searle con quella di Wittgenstein. Per mezzo di tale confronto si cercherà di porre una riflessione, da una parte, sulla attendibilità dell'analogia che Fodor sostiene con tanta forza e determinatezza tra l'uomo e la macchina e dall'altra, sulla poca chiarezza con la quale, così come Searle tanti altri autori, si utilizza il termine comprensione, concetto che nonostante non sia chiaro e distinto viene utilizzato come concetto fondamentale per trattare il problema delle macchine pensati.

## 2. LUDWIG WITTGENSTEIN

### 2.1. Il problema della denotazione

Le *Ricerche Filosofiche* si aprono con un passo delle Confessioni di S. Agostino, secondo cui il linguaggio è costituito da nomi che sono in un rapporto di designazione con gli oggetti. Secondo questa concezione denotativa, imparare un linguaggio significa imparare di che cosa ciascuna parola è un nome. Il linguaggio è costituito da nomi, da parole, a cui associamo significati; il significato è l'oggetto per cui la parola sta.

Quando [gli adulti] nominavano qualche oggetto, e, proferendo quella voce, facevano un gesto verso qualcosa, li osservavo, e ritenevo che la cosa di chiamasse con il nome che proferivano quando volevano indicarla. Che intendessero ciò era reso manifesto dai gesti del corpo, linguaggio naturale di ogni gente: dall'espressione del volto e dal cenno degli occhi, dalle movenze del corpo e dall'accento della voce, che indica le emozioni quando ricerchiamo, possediamo, rigettiamo o fuggiamo le cose. Così, udendo spesso le stesse parole ricorrere, al posto appropriato, in proposizioni differenti, mi rendevo conto, poco a poco, di quali cose esse fossero segni, e, avendo insegnato alla lingua a pronunziarle, esprimevo ormai con esse la mia volontà (RF, I, § 1)

Wittgenstein ritiene l'immagine agostiniana del linguaggio troppo semplicistica: è una "rappresentazione primitiva del modo e della maniera in cui funziona il linguaggio" (RF, I, § 2), nel senso che non tutto ciò che chiamiamo linguaggio è fatto così.

L'errore di Agostino è di trattare il linguaggio come puramente denotativo e le parole come semplici nomi, etichette, e che non ci sia nient'altro da aggiungere. Certamente, è perfettamente possibile immaginare un linguaggio fatto così, ma certamente le cose sono molto più complicate. Il problema in questione è se dare la denotazione sia dare il significato.

La parola 'designare' trova forse la sua applicazione più diretta nei casi in cui il segno è posto sull'oggetto che designa. Nel paragrafo 15 delle *Ricerche Filosofiche*, Wittgenstein prende in considerazione la metafora del riferimento "i nomi sono etichette".

La parola «designare» trova forse la sua applicazione più diretta nei casi in cui il segno è posto sull'oggetto che designa [...] Denominare una cosa è come attaccare a un oggetto un cartellino che reca il suo nome. (RF, I, §15)

Quando io attacco un cartellino ad un oggetto cosa faccio io da un punto di vista semantico? Non si sa, sto solo attaccando un cartellino. Perché il nome sul cartellino dovrebbe essere il nome dell'oggetto e non qualche cos'altro, per esempio le istruzioni per il suo utilizzo? Attaccare cartellini con il nome di un oggetto è una pratica semantica, ma c'è un problema: per attaccare un cartellino ad un oggetto si deve sapere già che il segno scritto su di esso è il nome di quell'oggetto. Supponiamo che sia in atto una certa pratica e che questa pratica funzioni così: ci sono degli strumenti sopra i quali ci sono incisi dei segni e poi c'è una tabella con riportati sopra questi stessi segni. A fa vedere a B un certo segno sulla tabella; B va a vedere se c'è qualche strumento su cui vi è riprodotto lo stesso segno e se c'è, lo porta ad A. Ma in cosa consiste, in questa pratica, il riferimento nome-oggetto? Qui vi è un intreccio tra nome e pratica: se un segno ha un certo ruolo, una certa funzione in una certa pratica è perché quel segno è un nome, cioè quello che chiamiamo nome proprio. Ma non vi è prima il nome e poi la pratica, cioè la pratica non si basa sul fatto che vi è un nome. Per Wittgenstein, il nome è un segno che ha un ruolo all'interno di una certa pratica, ovvero "essere un nome" è semplicemente avere un certo ruolo all'interno di una pratica come questa, o simili.

Wittgenstein vuole dimostrare che non si riesce a fondare la relazione di denominazione attraverso il gesto dell'ostensione. Non vuole negare la pratica dell'ostensione; il punto è se c'è nell'ostensione, in quanto tale, qualche cosa che, per ragioni di principio, non può che funzionare, cioè non può che instaurare, fissare, il riferimento che il parlante, che usa l'ostensione, ha in mente di instaurare. Vuole mostrare se l'ostensione è una pratica tale che non può essere fraintesa o se è possibile il fraintendimento del gesto ostensivo.

«Questo si chiama due» - e così dicendo si indicano due noci. Ma com'è possibile definire il due in questo modo? Colui al quale si dà la definizione non sa che cosa si voglia denominare con «due»; supporrà che tu denomini *questo* gruppo di noci! – può supporlo; ma forse non lo suppone. Al contrario, se voglio attribuire un nome a questo gruppo di noci, l'altro potrebbe anche scambiarlo per un numerale. E allo stesso modo colui al quale do una definizione ostensiva del nome di una persona potrebbe interpretarlo come il nome di un colore, come la designazione di una razza o addirittura come il nome di un punto cardinale. *Ciò vuol dire che la definizione ostensiva può in ogni caso essere interpretata in questo o in altri modi.* (RF, I, § 28)

Come viene detto esplicitamente in questo paragrafo, l'ostensione può essere fraintesa, anzi lo è spesso e volentieri. Si tratta quindi di capire che

cosa è che fissa il significato di un termine. In ogni caso, l'ostensione può al massimo fissare il riferimento, e per Wittgenstein il riferimento non è il significato, infatti, se muore il portatore di un nome, non per questo muore il significato del nome (RF, I § 40).

Di fronte ad una definizione ostensiva come faccio a capire a che cosa si riferisce? Ad esempio, mi viene detto, indicando con un dito un tavolo marrone: «Questo è marrone». Ma lo stesso gesto potrebbe accompagnare espressioni diverse come «Questo è un tavolo» o «Questo è di legno» o «Questo è rotondo» o «Questo è un oggetto». Per capire la definizione devo dunque sapere che chi si rivolge a me mi sta dando una definizione di colore e non, per esempio, di una forma o di un materiale. Per sapere a che cosa si riferisce la definizione ostensiva, devo sapere almeno di che cosa si sta parlando; ma allora devo già sapere molto per poter capire tale definizione e perciò la fissazione del riferimento mediante l'ostensione non può precedere ogni altro aspetto del linguaggio.

La definizione ostensiva spiega l'uso – il significato – della parola, quando sia già chiaro quale funzione la parola debba svolgere, in generale nel linguaggio [...] Per essere in grado di chiedere il nome di una cosa si deve già sapere (o saper fare) qualcosa. (RF, I, §30).

Secondo Wittgenstein il modo di pensare di S. Agostino, circa l'apprendimento del linguaggio, è errato perché presuppone che ci sia già una capacità di pensare nel bambino ancora prima di imparare a parlare, come se uno possedesse già un linguaggio ma non quello che gli si sta insegnando; questo modo di pensare comporta una descrizione errata dell'apprendimento del linguaggio umano: “[...] come se il bambino giungesse in una terra straniera e non possedesse la lingua del paese; vale a dire: come se possedesse una lingua ma non questa” (RF, I, § 32). Secondo Agostino l'origine del linguaggio consiste in un gesto che permette di pensare ad una cosa quando si sente una voce che la denomina e quindi pone l'essenza del linguaggio nell'atto del nominare. Secondo Wittgenstein, invece, dobbiamo sempre già presupporre il linguaggio quando vogliamo rendere conto del fatto che ciò che noi diciamo e ciò che noi facciamo abbia un significato. Non ci sono significati anteriori al linguaggio, ma solo significati creati dall'attività linguistica dei parlanti.

## **2.2. Significato e regole d'uso**

La discussione riguardante i problemi che sorgono dall'immagine agostiniana del linguaggio serve a Wittgenstein per introdurre la sua

concezione di significato inteso come *uso*. La concezione del denominare è connessa ad un processo occulto, appare come una strana connessione di una parola con un oggetto, una connessione che ha luogo quando si cerca di scoprire cos'è la relazione tra nome e nominato (la quale fissa un oggetto) e ripetendo più volte il suo nome. Si deve eliminare il primato della relazione di denominazione.

Se con la parola «significato» si designa l'oggetto che 'corrisponde' alla parola, allora la parola viene impiegata in modo contrario all'uso linguistico. Ciò vuol dire scambiare il significato di un nome con il portatore del nome (RF, I, §40).

Qualche volta, quando per esempio abbiamo a che fare con i nomi propri, per dare significato al nome proprio può essere utile indicarne il portatore. Ma questo è solo un modo. Il portatore non è il significato del nome proprio.<sup>1</sup> Il significato di un termine è l'uso che si fa di questo termine. Il linguaggio crea significati dal momento che è inserito in una prassi socialmente riconosciuta e condivisa, non li ha già preconfezionati e pronti per l'uso. Il gioco dell'ostensione può nascere solo all'interno di un contesto di interazioni sociali e concettuali. In realtà, solo gli adulti potrebbero apprendere significati in modo ostensivo. E questo perché sono già dentro una rete di significati, cioè dispongono già di una vasta costellazione di concetti inferenzialmente connessi tra loro. In questo senso l'immagine agostiniana non è una buona immagine dell'apprendimento del linguaggio: tutt'al più può

---

<sup>1</sup> La discussione sulla denominazione è legata a quella degli oggetti semplici di cui Wittgenstein ha parlato nel *Tractatus*. Nel *Tractatus* si chiede "perché devono esserci oggetti semplici?". Perché i nomi devono delineare qualcosa di semplice. Perché una proposizione abbia un senso determinato bisogna che sia analizzabile in segni che denotano qualcosa la cui esistenza non può essere messa in discussione, che non sia problematica. Perché? Perché altrimenti non potremmo mai sapere se l'intera proposizione ha un senso determinato perché ciò di cui la proposizione parla potrebbe non esistere e se ciò di cui la proposizione parla non esistesse allora la proposizione non avrebbe senso. (RF, I, §39). Nelle *Ricerche Filosofiche* però si ha un totale cambiamento di visione: domandare se qualcosa è composto non è sensato, cioè non è una domanda filosofica sensata. Quando diciamo che niente è intrinsecamente semplice o composto non stiamo facendo un'affermazione metafisica, cioè non stiamo caratterizzando una qualche proprietà del mondo. Si sta solo dicendo che parole come semplice o composto sono usate in modo diverso a seconda del tipo di complessità che abbiamo in mente e che uno stesso oggetto può essere considerato o semplice o composto in base a diversi punti di vista. Ha senso domandarsi se un essere umano è semplice o composto? No, non ha nessun senso. Dire che è semplice o composto significa assumere il punto di vista della metafisica. Non ha senso dire che quell'oggetto è intrinsecamente semplice. Esiste un problema scientifico della complessità e della semplicità, ma non un problema filosofico. Dal punto di vista filosofico si può dire che a seconda dei discorsi che ci interessa fare di volta in volta, alcune cose sono semplici, altre composte. (RF, I, § 47)



essere l'immagine dell'apprendimento di un linguaggio da parte di qualcuno che abbia già un linguaggio.

La parola 'uso' può essere intesa sia come uso comune, cioè l'impiego di un segno da parte di una comunità linguistica storica, sia come funzione, *regolata*, di un segno in un contesto. Wittgenstein non è interessato all'uso nel primo senso, ma all'uso inteso come funzione in un determinato contesto linguistico. Wittgenstein pensa all'uso regolato di una parola, cioè all'uso come espressione di regole che lo governano.<sup>2</sup>

Al concetto di uso è, infatti, intrinsecamente collegato quello di regola. Cos'è una regola? La regola può essere intesa come un'istruzione o una prescrizione dell'uso, cioè una norma che governa l'impiego delle nostre espressioni linguistiche. Una regola sta lì, come un indicatore stradale (RF, I, § 85). Il linguaggio è strettamente collegato ai concetti di gioco e di regola; Wittgenstein lo descrive come un insieme di giochi linguistici governati da regole. Come nel gioco, per poter giocare, si ha bisogno di regole, anche nel linguaggio, per poter comunicare, si devono conoscere e applicare determinate regole. Se non ci fossero regole nel gioco, ogni giocatore potrebbe fare qualsiasi mossa, con la conseguenza che il gioco perderebbe la sua particolarità di essere quel certo gioco; non si potrebbe neanche definire il gioco a cui si sta giocando se esso non fosse costituito da regole. Così come nel linguaggio, in assenza di regole, ognuno potrebbe parlare a suo modo senza farsi capire dagli altri, non si potrebbe neanche delimitare l'utilizzo di certe espressioni a certi contesti e non si potrebbero insegnare e apprendere proposizioni di alcun genere.

Wittgenstein cerca di chiarire la nozione di gioco linguistico, dicendo che non c'è un'essenza del gioco linguistico e quindi del linguaggio:

Invece di mostrare quello che è comune a tutto ciò che chiamo linguaggio, io dico che questi fenomeni non hanno affatto in comune qualcosa, in base al quale impieghiamo per tutti la stessa parola, - ma che sono imparentati l'uno con l'altro in modi differenti. E grazie a questa parentela, o a queste parentele, li chiamiamo tutti linguaggi". (RF, I, § 65).

I giochi linguistici non hanno un qualcosa che è comune a tutti, nel senso che non c'è un insieme di condizioni necessarie e sufficienti che caratterizzino i giochi. Vi sono «somiglianze di famiglia» (RF, I, § 67); noi applichiamo le parole, come per esempio "gioco", in base ad analogie e somiglianze.<sup>3</sup> Wittgenstein impiega quindi l'espressione somiglianze di

---

<sup>2</sup> D. Marconi, 1971, p.58

<sup>3</sup> La teoria delle somiglianze di famiglia ha una funzione *antiessenzialista*: serve a contenere la tendenza a porsi domande come "cos'è l'essenza del linguaggio?", cos'è l'essenza di una

famiglia per indicare la rete di somiglianze che si sovrappongono e s'incrociano tra gli usi di una stessa parola nei diversi giochi linguistici e ridefinisce il compito della filosofia nei termini di un'attività di descrizione di questo complesso intreccio di analogie e differenze. La filosofia si occupa infatti della grammatica del linguaggio, cioè descrive l'insieme di regole che usiamo per parlare, cioè le regole che governano l'uso del linguaggio. Che cos'è una parola? È un suono o segno che viene usato secondo certe regole. L'identità di una proposizione, cioè il suo significato, è determinata dal suo uso, ed è da ricercare nel suo uso quotidiano.

In altre parole, le regole del nostro linguaggio somigliano molto alle regole di un gioco e riguardano certe azioni all'interno di quella che Wittgenstein chiama "forma di vita". (RF, I, § 19). Il parlare un linguaggio fa parte di un'attività o di una forma di vita (RF, I, § 23), un'attività governata da regole. Dire che il significato è l'uso equivale a dire che esso coincide con la totalità delle funzioni di un segno in un contesto, che si potrebbe identificare con la totalità delle attività linguistiche degli uomini.<sup>4</sup>

Le parole del linguaggio, quindi, non sono tutte nomi che ricevono significato dal loro portatore, ma sono di tipi differenti e svolgono funzioni differenti. Wittgenstein le paragona agli strumenti che si trovano in una cassetta di utensili: un martello, una tenaglia, una sega, la colla, ecc.. «Quanto differenti sono le funzioni di questi oggetti, tanto differenti sono le funzioni delle parole» (RF, I, § 11). E tale differenza dipende da una differenza nelle regole che governano i loro impieghi. Come nel gioco degli scacchi l'identità di ciascun pezzo dipende dalle regole che ne determinano le mosse possibili, così in un gioco linguistico il significato di un termine dipende dalle regole che ne governano l'uso. Per comprendere il significato di un'espressione linguistica bisogna guardare anche alle sue circostanze d'uso e tenere conto delle azioni che si compiono nel proferirla, nonché delle reazioni e delle conseguenze che può essa produrre.<sup>5</sup>

---

proposizione?", e così via. Esempi come quello di 'gioco' fanno vedere come queste domande in realtà siano di fatto senza risposta. Fanno torto a quello che è di fatto l'uso del linguaggio. Noi non applichiamo la parola 'gioco' sulla base di condizioni necessarie e sufficienti che rendono qualcosa un gioco. La nozione di somiglianza di famiglia può essere presa come una teoria semantica che riguarda certi predicati del linguaggio naturale, ordinario, non di un linguaggio ideale. Assumere un atteggiamento essenzialista significa estrapolare parole come proposizione, linguaggio e regola, dal loro uso nel linguaggio comune, come se le si scarnificasse cercando di riportarle ad una qualche purezza estrema, dimenticandosi dell'uso che facciamo quotidianamente di questi termini. Fuori dal linguaggio comune queste parole non significano nulla. (Anche qui Wittgenstein sta criticando l'atteggiamento filosofico che era alla base del *Tractatus*).

<sup>4</sup> "Il linguaggio si riferisce ad un modo di vita" (OFM, VI, § 34)

<sup>5</sup> Pensa alle parole come strumenti caratterizzati dal loro uso e poi pensa all'uso di un martello, all'uso di uno scalpello, all'uso di una squadra, d'un barattolo della colla e della

Il significato di una parola quindi viene compreso se se ne comprendono le regole d'uso. L'intero apparato linguistico può funzionare se se ne comprendono le regole che lo governano. Un termine infatti può venir utilizzato solo se si conosce la regola di utilizzo, cioè se viene compreso il modo in cui lo si può impiegare.

### 2.3. Seguire la regola

Come imparo le regole, grazie alle quali comprendo quello che accade e posso farmi comprendere? Come imparo a riconoscere che quel determinato utilizzo è conforme alla regola? Wittgenstein parla di comprensione in termini di addestramento linguistico. Noi infatti veniamo addestrati ad usare un certo termine in un certo modo, secondo certe regole, all'interno di vari contesti. La comprensione del termine infatti è connessa con l'uso perché essa è la capacità di fare certe cose con i segni. Noi impariamo a giocare a scacchi imparando le regole del gioco. Molte attività governate da regole vengono apprese grazie all'addestramento e all'esercizio ripetuto. Che un'attività sia governata da regole implica una regolarità nel comportamento *nella sua esecuzione*, nel senso che, se seguo la regola, la mia condotta è vincolata a quella regola. Quando si segue una regola si fa sempre la stessa cosa, cioè, se si segue la regola allora si fa sempre la stessa cosa quando sorge l'occasione appropriata. Che un'azione di una persona è governata dalla regola, che una persona guida se stesso riferendosi alla regola, è manifesto nel modo in cui usa le regole, invoca la formulazione della regola, si riferisce alle regole spiegando ciò che fa, giustifica ciò che fa valutando, criticando e correggendo ciò che fa.

Ci si può ulteriormente interrogare su che cosa voglia dire comprendere un enunciato o anche soltanto una parola o una successione numerica. Una risposta sembra essere già pronta: comprendere una parola o una successione numerica vuol dire cogliere la regola che è sottesa dal suo uso. Ovvero: comprendere una serie significa cogliere, in una successione numerica, la regola che ci permette di continuarla. Ma il problema è proprio questo: *che cosa vuol dire 'cogliere una regola'?*

Vi è un maestro che cerca di insegnare ad un bambino la serie dei numeri. La prima mossa consisterà nello scrivere i numeri da 0 a 9 e poi nel guidare la mano del bambino nel copiarli. Poi il maestro mostrerà come la

---

colla. (Tutto quello che diciamo qui può comprendersi solo se si comprende che una grande varietà di giochi è giocata con gli enunciati del nostro linguaggio: dare ordini ed obbedire a ordini; fare domande, e rispondere a domande; descrivere un evento; narrare una storia; raccontare una barzelletta; descrivere un'esperienza immediata; far congetture intorno ad eventi nel mondo fisico; formulare ipotesi e teorie scientifiche; salutare qualcuno). (LB p.92)

successione che si ripete dalla parte delle unità si dia anche dalla parte delle decine. Infine il maestro dovrà farsi da parte e lasciare che sia il bambino a continuare la serie, ed il successo in questo compito varrà come una riprova del fatto che qualcosa è stato appreso e che lo scolaro ha capito. Ma che cosa è accaduto nella mente del bambino che esclama “Ora so andare avanti!”?

Una prima classe di risposte può essere fin da principio messa da parte: ogni tentativo di ricondurre il nostro comprendere ad un qualche evento fisiologico non è un tentativo soddisfacente. Quando comprendo la serie numerica accadrà certo qualcosa nella mia testa, ma il fatto di supporre una qualche modificazione a livello neuronale non spiega che cosa voglia dire comprendere una serie, tant’è che io debbo già sapere che cosa intendo con quell’espressione per cercare se – quando comprendo – accade proprio quello che credo debba accadere.

Se la via fisiologica non ci aiuta, anche l’indagine introspettiva non sembra darci una soluzione soddisfacente. L’indagine introspettiva ci riconduce, in generale, alla critica allo psicologismo (caratteristica della posizione wittgensteiniana): il significato di una parola o di una qualsiasi mossa di un gioco linguistico non può essere ricondotto ad un vissuto, perché ciò che accade nella nostra testa è un fatto e non una regola, e come tale non può legittimare il nostro comportamento. Può darsi che tutte le volte che comprendo come proseguire una qualche serie numerica io viva un’esperienza particolare, ma questo mio provare un determinato vissuto non è un argomento che mi consente di dire che ho davvero compreso la regola di progressione della serie. Quindi, per rendere conto del significato dell’espressione “ora comprendo (so continuare) la serie!” non si può additare un qualche vissuto. L’errore risiede nel pensare che la comprensione sia un evento psichico che il soggetto avverte immediatamente in se stesso.

Ma se la via fisiologica e l’indagine introspettiva non danno risposte accettabili al quesito “come si coglie la regola?”, si potrebbe, allora, tentare un’altra strada e sostenere che l’intendere un concetto o il comprendere una formula sia qualcosa che avvenga in un solo istante. Ma se posso comprendere in un attimo una parola o una serie numerica allora la comprensione di una regola deve fare tutt’uno con il fatto di afferrare un *oggetto ideale*, un *eidos* che, proprio come una formula, sappia idealmente abbracciare in una volta sola tutte le sue possibili applicazioni. Secondo questo modo di vedere le cose, le conseguenze di una regola sarebbero già esistenti in essa, in un senso, per così dire, ideale di “esistere”; «[...] La tua idea era che quell’intendere il comando avesse già eseguito, a suo modo, tutti quei passaggi: che la tua mente, quando intende, voli, per così dire, in avanti e compia tutti i passaggi prima che tu pervenga fisicamente a questo o a quel punto. Tu eri dunque disposto a usare espressioni come: “Per parlare propriamente i passaggi sono già compiuti; anche prima che io li conosca per

iscritto, oralmente o idealmente. E sembrava che fossero predeterminati, anticipati in un modo singolarissimo – come solo l'intendere può anticipare la realtà.» (RF, I, § 188).

In altre parole, se comprendo la serie di cui discorriamo, quei passi io debbo in qualche modo averli già compiuti, sia pure solo idealmente.

Il problema che si pone in questo caso è quello riguardante la legge di successione: si costruisce una successione secondo una regola, cioè: una volta che mi si è data la regola e il punto di partenza 0, è come se mi fosse già data tutta la successione. “Questa formula determina questa successione”: ovvero, vi è una regola per estrarre da questa formula l'intera successione, come se la formula stessa fosse la regola. In che senso però la formula è una regola? Se io comprendo la formula allora conosco la regola: la formula è una scrittura che ha un certo significato, se io conosco il significato allora comprendo la regola e sono in grado di sviluppare la successione. In altre parole, se io comprendo la regola non posso che sviluppare quella successione lì: è come se la successione fosse già data nel momento in cui è data la regola. «Il modo in cui la formula viene intesa determina quali passaggi si debbano compiere.» (RF, I, § 190)

Ma cosa vuol dire che i passi successivi sono determinati da una certa formula? Cosa vuol dire che questa conclusione è determinata da queste premesse? Analogamente: se comprendo una parola, allora so usare quella parola: tutti i suoi usi possibili sono dati nel significato della parola. Dicendo questo, si vorrebbe sostenere che la formula di una serie contenga idealmente in sé tutti gli infiniti passaggi di cui la serie consta. La formula e, con essa, il concetto divengono il contenitore finito di un'infinità di mosse reali, il luogo ideale e prima del tempo in cui è già da sempre racchiuso ciò che a noi, passo dopo passo, si manifesta.<sup>6</sup>

«È come se in un solo istante potessimo afferrare tutto quanto l'impiego della parola». – E anzi, diciamo che l'afferriamo. Vale a dire, qualche volta con queste parole descriviamo ciò che facciamo. Ma in ciò che accade non c'è nulla di sorprendente, nulla di strano. *Strano diventa quando siamo indotti a pensare che lo sviluppo futuro debba essere già in qualche modo presente nell'atto di afferrare, e tuttavia presente non lo è.* [corsivo mio] – Diciamo, infatti, che non c'è alcun dubbio che comprendiamo una certa parola, e d'altra parte il suo significato sta nel suo impiego. Non c'è alcun dubbio che io ora voglia giocare a scacchi; ma il giuoco degli scacchi è il giuoco che è in forza di tutte le sue regole (ecc). Dunque

---

<sup>6</sup> In questo modo, non si farebbe altro che riproporre l'immagine platonica delle idee: il pensiero della formula (e quindi del concetto) come un'entità ideale che racchiude in sé tutte le sue possibili applicazioni, anche se queste sono infinite.

prima di aver giocato non so che cosa voglio giocare? Oppure tutte le regole sono contenute nell'atto della mia intenzione? Ma è l'esperienza ad insegnarmi che a quest'atto dell'intenzione segue di solito questo modo di giocare? Non posso dunque sapere con certezza che cosa intendevo fare? E se ciò è privo di senso, - quale genere di connessione ultra-rigida sussiste tra l'atto e l'oggetto d'intenzione? – Dove viene effettuata la connessione tra il senso delle parole «Giochiamo a una partita i scacchi!» e tutte le regole del gioco? – Ebbene, *nell'elenco delle regole del gioco, nell'insegnamento degli scacchi, nella pratica quotidiana del gioco.* [Corsivo mio] (RF, I, § 197)

Quando qualcuno mi chiede se voglio giocare a scacchi, noi capiamo quello che costui mi sta chiedendo, cioè capiamo l'espressione 'giocare a scacchi'. Il gioco degli scacchi è costituito dalle regole degli scacchi. Ma se capiamo quello che ci viene detto vuol dire che dobbiamo avere, in qualche modo, presenti tutte le regole degli scacchi? Io posso non capire l'invito “giochiamo a scacchi” se per esempio, sono un bambino che non ne ha mai sentito parlare; ma poniamo che io lo capisca e che lo capisca perché so di che cosa si tratta, cioè so cos'è il gioco degli scacchi perché ne conosco le regole, ho imparato il gioco tanto tempo fa, ci gioco spesso, etc. Questo non vuol dire che nel momento stesso in cui sento le parole “Giochiamo a scacchi?” io sappia elencare tutte le regole, tutte le volte che ho giocato a scacchi, cosa ho letto nei manuali del gioco degli scacchi, etc. Noi non chiamiamo “saper giocare a scacchi” tutto questo, cioè l'aver presenti nella mente tutte queste cose. Io attribuisco una competenza a chi sa giocare a scacchi; ma l'attribuire una competenza è diverso dall'attribuire un insieme di conoscenze: le competenze presuppongono molte conoscenze, ma non la somma delle conoscenze pertinenti ad un certo particolare ambito. Un botanico competente, per esempio, non possiede tutte le conoscenze botaniche; un botanico competente in certe circostanze fa certe cose piuttosto che altre. Dove sta, quindi, la connessione tra il significato di “giocare a scacchi” e le regole del giocare a scacchi? Nella *competenza del parlante e dell'interlocutore*, cioè nel fatto che entrambi condividono una prassi comune, sanno fare certe cose e che sanno fare certe cose perché hanno avuto un certo addestramento.

Riassumendo, ogni gioco linguistico ci riconduce ad uno o più esempi paradigmatici di cui ci avvaliamo seguendo un criterio cui siamo stati addestrati, e ciò è quanto dire che le regole su cui si fondano i nostri giochi linguistici sono, per così dire, procedure operative, non entità ideali che possano esibire una qualche fondazione ultima.

Un altro tipo di risposta al problema del “cosa vuol dire cogliere la regola” nasce e si sviluppa dalle difficoltà appena emerse: se una regola non

anticipa le sue applicazioni, allora bisognerà ammettere che tra la regola e le sue applicazioni ci sia qualcos'altro. "Ma come può una regola insegnarmi che cosa devo fare a questo punto?" (RF, I, § 198). Viene ipotizzato che tra la regola e la sua applicazione intervenga un atto di interpretazione; ma se la regola deve essere interpretata allora, presa una regola, «[...] qualunque cosa io faccia, può sempre essere resa compatibile con la regola mediante una qualche interpretazione» (§ 198) ed inoltre, così come qualsiasi applicazione può risultare in accordo con la regola grazie a una certa interpretazione, così, grazie a un'altra interpretazione, quella stessa applicazione può risultare in disaccordo con la regola.

Il nostro paradosso era questo: una regola non può determinare alcun modo d'agire poiché qualsiasi modo di agire può essere messo d'accordo con la regola. La risposta è stata: se può essere messa d'accordo con la regola potrà anche essere messa in contraddizione con essa. (RF, I, §201)

Si possono richiamare alla memoria due esempi per comprendere meglio il perché del rifiuto del modello interpretativo.

Nell'osservazione 185 Wittgenstein porta l'esempio dell'allievo recalcitrante: una volta imparata la successione di "aggiungi n", dopo aver fatto esercizi e prove per stabilire la sua comprensione sul campo dei numeri fino a mille, scrive la successione "aggiungi 1" correttamente. Ma quando lo si invita a scrivere la progressione aritmetica di ragione 2 oltre il mille, lo scolaro sbaglia: scrive 1004,1008,1012...

Se ci si attiene al modello interpretativo, allora, nel caso dell'esempio dell'allievo recalcitrante, al quale si è insegnato la successione in ragione 2 ma che, una volta che arrivato a mille, comincia a scrivere una successione in ragione 4, c'è un modo di interpretare la regola che gli dà ragione, come se la regola fosse "aggiungi 2 fino a 1000 e poi comincia ad aggiungere 4". Ammettere questo, ovviamente sarebbe assurdo.

Analogamente, il modello interpretativo può venir confutato per mezzo dell'esempio dell'immagine del cubo, che Wittgenstein presenta nell'osservazione 139:

Ebbene, supponi che quando odi la parola «cubo» si presenti alla tua mente un'immagine, poniamo il disegno di un cubo. In che senso quest'immagine può convenire o non convenire all'impiego della parola «cubo»? – Forse tu dici: è molto semplice; - se questa immagine mi sta davanti alla mente e io indico, per esempio, un prisma triangolare dicendo che si tratta di un cubo, questo impiego non conviene all'immagine. – Ma è poi vero che non le conviene? Ho scelto l'esempio intenzionalmente, in modo che riuscisse assai

facile immaginare un *metodo di proiezione*, secondo il quale l'immagine conviene. L'immagine del cubo ci *suggeriva* bensì un certo impiego, ma io potevo anche impiegarla diversamente.

L'immagine determina il suo impiego? Secondo Wittgenstein, no. Di per sé, un'immagine non dice nulla, perché non dice quale sia il suo uso effettivo. Proporre un'immagine non significa avere detto qualcosa sulla relazione di proiezione che la rende tale, poiché in linea di principio sono possibili innumerevoli relazioni che ci consentono di connettere un'immagine ad una diversa interpretazione.<sup>7</sup>

Cerchiamo di comprendere meglio la tesi secondo la quale un'immagine di per sé non ha un senso definito poiché non è chiaro quale sia il criterio che ci guida nel suo impiego. C'è la parola 'cubo' e l'applicazione della parola. Cos'è che governa l'applicazione della parola 'cubo'? Qual è il suo significato? Secondo il modello interpretativo, quando vedo l'immagine del cubo o sento la parola 'cubo', io potrei disegnare prismi triangolari o parallelepipedi, sulla base di ciò che mi dice l'immagine, cioè sulla base del fatto che per me quelle cose lì vanno chiamate 'cubo'. Ci sarebbe quindi un'interpretazione della regola, un metodo di proiezione che in qualche modo giustifica la mia applicazione dell'immagine del cubo, sia che la applichi ai prismi triangolari, sia che la applichi ai parallelepipedi. La stessa applicazione della regola, cioè la stessa interpretazione, allora può essere considerata o coerente o incoerente: coerente secondo un certo metodo di proiezione, incoerente secondo un altro metodo di proiezione. Quindi il paradosso è presente nella misura in cui la medesima applicazione della regola può risultare sia corretta sia scorretta; come ci dice Wittgenstein stesso «qui non esistono, pertanto, né concordanza né contraddizione» (§ 201) e, dunque, non esiste propriamente, una regola.

Wittgenstein infatti rifiuta che la comprensione di un termine consista nell'aver in mente una certa immagine. Comprendere una parola vuol dire avere in mente un'immagine? No, perché un'immagine è solo una figura che può essere interpretata in vari modi; una figura non contiene la sua applicazione, non mi dice come dev'essere applicata. L'immagine non contiene la regola della sua applicazione. Io, infatti, posso avere un metodo di proiezione del tutto particolare che fa sì che io metta in rapporto la mia immagine del cubo con quella di un prisma triangolare. In questo senso si direbbe che il cubo, oltre ad essere un parallelepipedo, è anche un prisma triangolare, secondo una certa regola di proiezione. Un'immagine mentale non detta l'uso di un termine, dal momento che essa stessa è suscettibile di varie interpretazioni: non c'è un percorso logico dalle proprietà intrinseche

---

<sup>7</sup> Si può dire che un'immagine 'parla' soltanto all'interno di un qualche gioco linguistico.



dell'immagine al significato di una parola, a causa della possibilità di applicazioni devianti dell'immagine.

La critica di Wittgenstein si rivolge quindi contro l'idea che i significati siano identificabili con immagini mentali o rappresentazioni automaticamente associate alle parole e agli enunciati del linguaggio. Che cosa è che fa della mia immagine mentale di un cubo proprio l'immagine di un cubo? La risposta di Wittgenstein è che è solo *l'applicazione pubblicamente osservabile*, cioè il modo effettivo in cui usiamo tale immagine, a conferirle il valore di immagine di un cubo. Di per sé l'immagine mentale per quanto somigliante a un cubo può anche essere trattata come l'immagine di qualcos'altro, ad esempio, come l'immagine di un prisma triangolare.

Ciò che si può concludere, allora è che il paradigma del seguire una regola non è l'interpretazione:

Ogni interpretazione è sospesa nell'aria insieme con l'interpretato; quella non può servire da sostegno a questo. Le interpretazioni, da sole, non determinano il significato.” (RF, I, § 198).

In cosa consiste quindi il fraintendimento di cui parla Wittgenstein nel paragrafo 201? Si fraintende il rapporto tra regola e applicazione, cioè si pensa che il rapporto tra regola e applicazione sia un rapporto ermeneutico. Noi interpretiamo la regola, ma, così facendo, ci troviamo davanti a diverse interpretazioni possibili le quali sono tutte legittime. Si arriverebbe sempre e soltanto a nuove interpretazioni, ma così facendo, non si arriverebbe ad un'applicazione, ma solo a riformulazioni della regola, ovvero ad un'interpretazione costante e continua. Questo processo interpretativo continuo, in linea di principio si potrebbe anche fare, ma non è quello che accade di fatto. «[...] L'interpretazione ha un termine» (OFM, VI, § 38). Applicare una regola non è interpretare la regola. Per Wittgenstein, l'interpretazione è solo la traduzione di un segno in un altro segno. «Ogni volta che noi interpretiamo un simbolo in un modo o nell'altro, l'interpretazione è un ulteriore simbolo aggiunto a quello anteriore». (LB, p.48)

Riassumendo, nessuna interpretazione può determinare il modo dell'applicazione della regola. Ma questo non significa affatto che si possa affermare che qualunque cosa io faccia, può sempre essere resa compatibile con la regola, poiché le regole non sono (non vivono nelle) loro interpretazioni ma nel modo in cui noi ce ne avvaliamo. Il fondamento della regola non è neanche nel suo carattere ideale o nella sua capacità di anticipare idealmente il futuro delle sue applicazioni, ma risiede nel fatto che “sono stato addestrato a reagire in questo modo a questo segno, e ora reagisco così”.

«Dunque, qualunque cosa io faccia, può sempre essere resa compatibile con la regola?» Lasciami chiedere: Che cosa ha da spartire l'espressione della regola – diciamo, un segnale stradale – con le mie azioni? Che tipo di connessione sussiste tra le due cose? – Ebbene, forse questa: sono stato *addestrato* [corsivo mio] a reagire in un certo modo a questo segno, ed ora reagisco così.

Ma in questo modo hai solo indicato un nesso causale, hai soltanto spiegato come mai ora ci regoliamo secondo le indicazioni di un segnale stradale; non in cosa consista, propriamente, questo attenersi al segnale. (RF, I, § 198)

Poiché sono stato addestrato a reagire in un determinato modo al segno in questione, inteso come espressione di una regola, ora reagisco così e non diversamente. Certo, potrei reagire in contrasto con l'addestramento, ma in questo caso, appunto, diremmo che non starei seguendo la regola. Tuttavia questa risposta potrebbe risultare insoddisfacente. L'obiettore in questione potrebbe infatti contestare che con ciò non è stato ancora detto in che cosa consista effettivamente il seguire una regola, ma che, in realtà, si è solo indicato un "nesso causale"; cioè, è come se ci fosse stato un processo di condizionamento, che mi ha permesso di acquisire questa disposizione e di reagire istintivamente ad un certo segnale (come se ci fosse un certo stimolo che determina una certa azione, cioè, un processo causale). In questo senso non si è spiegato in cosa consiste propriamente questo "attenersi ad un segnale". Ma Wittgenstein replica dicendo che in realtà il perché delle mie azioni non è un perché causale: quando io vedo il semaforo rosso, mi attengo al segnale e mi fermo, ma questo perché non è un perché causale bensì la *ragione* del mio fermarmi: il semaforo rosso non è la causa ma la ragione del mio fermarmi alla visione del segnale. «[...] uno si regola secondo le indicazioni di un segnale stradale solo in quanto esiste un uso stabile, un'abitudine». (RF, I, §198) L'uso stabile o l'abitudine è la condizione necessaria dell'attenersi ad un segnale stradale: se non ci fosse in uso un certo criterio di codificazione a cui tutti noi siamo abituati, io non mi fermerei al rosso; il fatto che mi fermi richiede un'abitudine, una prassi consolidata.

Quindi, nel seguire una regola è decisivo l'addestramento, è decisivo il complesso di atteggiamenti e di risposte connesse a quella regola. L'addestramento crea un'abitudine. Wittgenstein associa infatti la nozione di comprensione con quella di capacità di fare qualcosa e con quella di sapere-come<sup>8</sup>: comprendere un segno è avere un'abilità, essere padroni di una tecnica. «La grammatica della parola «sapere» è, come si vede facilmente, strettamente imparentata alla grammatica delle parole «potere» ed «essere in

---

<sup>8</sup> C. McGinn, 1984, pp. 30-31

grado». Ma è anche strettamente imparentata a quella della parola «comprendere». ('Padroneggiare una tecnica').» (RF, I, §150) La comprensione è essenzialmente connessa con l'uso perché è precisamente la capacità di fare certe cose con i segni.

Ciò che chiamiamo "seguire una regola" è forse qualcosa che potrebbe essere fatto da uno solo, una sola volta nella sua vita? – E questa, naturalmente, è un'annotazione sulla grammatica dell'espressione «seguire una regola». Non è possibile che un solo uomo abbia seguito una regola una sola volta. Non è possibile che una comunicazione sia stata fatta una sola volta, una sola volta un ordine sia stato dato e compreso, e così via. – Seguire una regola, fare una comunicazione, dare un ordine, giocare una partita a scacchi, sono consuetudini (usi, istituzioni). Comprendere una proposizione vuol dire comprendere un linguaggio. E comprendere un linguaggio significa padroneggiare una tecnica (ivi, § 199).

Evidenziare il carattere prescrittivo dell'agire secondo una regola significa, per Wittgenstein, rammentare che non è possibile che una regola sia seguita una sola volta e da un solo uomo. Seguire una regola una sola volta, non vuol dire seguire una regola.

Wittgenstein dice che è nell'addestramento che si manifesta la dimensione della ripetitività della regola. L'abitudine si crea perché vi è una ripetitività che fa sì che l'agire secondo la regola diventi una sorta di "modello" di comportamento per le azioni future. La ripetizione gioca un ruolo essenziale nel concetto di regola, proprio perché è attraverso la ripetizione esplicita e manifesta che qualcosa può porsi come un paradigma. La regola richiede la ripetizione e la ripetizione si fonda sulla possibilità dell'accordo: il seguire una regola non è solo una consuetudine, ma è anche un'istituzione, e questo termine ci riconduce necessariamente sul terreno della prassi intersoggettiva. Seguire una regola non vuol dire soltanto ripetere un certo modo di comportarsi: significa anche aderire ad una consuetudine condivisa. Per questo seguire una regola è una prassi: seguire una regola significa agire in un certo modo. Prova del mio aver seguito la regola si dà solo nell'accordo, nell'approvazione del mio comportamento da parte degli altri.

Per questo 'seguire la regola' è una prassi. E credere di seguire la regola non è seguire la regola. E perciò non si può seguire una regola 'privatim': altrimenti credere di seguire la regola sarebbe la stessa cosa che seguire la regola (RF, I, § 202)

L'osservazione 202 aggiunge ancora un dettaglio: seguire la regola non può essere ridotto alle mie credenze; quello che io credo è irrilevante nel mio seguire la regola. Io posso credere certe cose, anche giuste, ma non seguire la regola, così come posso seguire la regola ma credendo cose sbagliate. Io non comprendo, per esempio, la regola "aggiungi 2" quando ho l'impressione di saper proseguire la successione numerica; "l'impressione di aver capito" non centra nulla con la comprensione, non è condizione né necessaria né sufficiente, ma corrisponde all'aver creduto di seguire la regola. Potrei credere di procedere correttamente nella serie fondata sull'incremento in ragione di 2, proponendo 5 come successore di 4, ma se non vi è un criterio per distinguere ciò che credo da ciò che è, non sarà in alcun modo possibile distinguere tra le mosse che concordano effettivamente con la regola e le mosse che invece vi contravvengono. Il punto non è che io comprendo la regola quando ho l'impressione di saper proseguire; io afferro la regola quando effettivamente proseguo tale successione.

Se uno segue correttamente la regola lo si vede da come si comporta. Seguire la regola è una prassi, o, in altre parole, afferrare la regola è un saper fare. Se si segue una regola si esemplifica una regolarità di condotta, la padronanza di una tecnica, una prassi appunto. Credere di seguire una regola non equivale a seguirla per lo stesso motivo per cui non si può seguire una regola "in privato". Possiamo affermare di qualcuno che sta seguendo una regola solo a patto che costui esibisca la sua abilità, solo cioè se il suo comportamento soddisfa determinati criteri di correttezza. Credere di seguire una regola, appunto, non è ancora seguirla. Ciononostante, se ci convinciamo che l'ultima parola al riguardo ce l'abbia l'intenzione soggettiva, intesa come qualcosa di essenzialmente privato, allora il credere di seguire una regola diventa sinonimo di seguirla, il che è assurdo.

Il seguire la regola ha luogo nella sfera dei comportamenti e non in meccanismi mentali o processi interiori delle persone. Le operazioni della comprensione consistono in ciò che uno fa coi segni. Quando una persona segue un cartello stradale, il suo comportamento è il suo seguire la regola, e non il suo seguire la regola ha luogo prima internamente e poi viene tradotto in un'azione. Il seguire la regola non è un processo interiore nascosto, è qualcosa che si osserva quando si osserva il comportamento nell'uso di un segno. Il comportamento non è un "sintomo" del seguire la regola, ma la prova della comprensione della regola. Cogliere una regola non è un'interpretazione (come si è già spiegato) nel senso di una traduzione di un segno in un altro, ma è ciò che noi esibiamo quando applichiamo la regola. E che cos'è che noi esibiamo? Una padronanza nella tecnica di applicare il segno, avendo appunto un'abitudine e una pratica acquisita nel tempo. Le regole possono essere colte solo se noi le seguiamo regolarmente e ripetutamente. L'esistenza del significato richiede infatti un uso ripetuto nel

tempo, e «[...] solo mediante una tecnica, infatti, possiamo concepire una regolarità» (OFM, VI, §2) in quanto «[...] il fenomeno del linguaggio riposa sulla regolarità, sulla concordanza nell'agire» (OFM, VI, § 39).

Cos'è quindi il seguire la regola? È una prassi, è l'aver una certa capacità che viene acquisita attraverso l'addestramento. «In che modo posso seguire la regola?» (RF, I, §217): questa non è una domanda riguardante le cause, cioè, non sto domandando attraverso quale processo di condizionamento arrivo a seguire la regola. «Se questa non è una domanda riguardante le cause, è una richiesta di giustificare il fatto che, seguendo la regola, agisco così.» (Ibidem) Ci si sta ancora domandando perché chiamiamo il fatto di fare quelle certe cose “seguire una regola”. In forza di che cosa quella determinata prassi costituisce quel determinato ‘seguire la regola’? Ci si sta dunque domandando qual è la connessione tra applicazione e seguire la regola. Wittgenstein usa l'immagine dello strato di roccia: «Quando ho esaurito le giustificazioni arrivo allo strato di roccia, e la mia vanga si piega. Allora sono disposta a dire: “Ecco, agisco proprio così”». (Ibidem)

Fino a qual punto si può descrivere la funzione del linguaggio? A chi non è padrone d'un linguaggio, insegnare a padroneggiarlo. A chi ne è padrone, posso ridestar la memoria del modo e della maniera in cui è stato addestrato o anche descriverglieli: per uno scopo particolare; e intanto posso farlo in quanto anch'io sono già padrone d'un linguaggio. Fino a qual punto si può descrivere la funzione della regola? Chi non ne padroneggia nessuna, questo posso addestrarlo. Ma come posso spiegare a me stesso l'essenza della regola?

Qui il difficile non sta tanto nello scavare fino al fondamento, ma riconoscer come tale il fondamento, che sta davanti a noi.

*La nostra malattia è il voler spiegare.* [corsivo mio ] (OFM, VI, § 31)

Con l'osservazione 206, Wittgenstein torna su una questione già sollevata nel paragrafo 198, ma che era stata momentaneamente lasciata in sospeso, ossia quella del rapporto tra il seguire una regola e l'addestramento:

Seguire una regola è analogo a: obbedire a un comando. Si viene addestrati a ubbidire al comando e si reagisce ad esso in maniera determinata.

Sussiste quindi una qualche analogia tra il seguire una regola ed obbedire ad un comando; in entrambi i casi, infatti, l'addestramento sembra

svolgere un ruolo cruciale. Il concetto di addestramento, quindi, è in qualche modo il punto d'accordo tra quello di seguire una regola e quello di obbedire ad un comando.

#### 2.4. Concordanza tra regole ed usi

Rimane però aperto un quesito: Wittgenstein si chiede, infatti, che ne sarebbe di quanto detto se le reazioni della gente all'addestramento e agli ordini non fossero abbastanza uniformi. Che dire se uno reagisce al comando e all'addestramento in un modo, e un altro in un altro modo? Se una persona, allo stesso addestramento, reagisse in un modo ed un'altra in un altro, il disaccordo diventerebbe la regola e non l'eccezione; in un simile caso sarebbero inconcepibili attività quali il seguire una regola. Lo stesso linguaggio, inteso come attività governata da regole, sarebbe impossibile. Le pratiche di seguire regole ed obbedire agli ordini, infatti, possono aver luogo solo nel contesto di una condivisione generale delle nostre azioni, la cui uniformità, come abbiamo appena visto, costituisce un presupposto fondamentale affinché l'addestramento umano possa raggiungere i propri scopi.

Il paragrafo 208 pone un'altra domanda:

Come faccio a spiegare a qualcuno il significato di “regolare”, “omogeneo”, “eguale”? Nel caso in cui la persona a cui intendo spiegare questi significati parli una lingua che già possiede tali concetti, come per esempio il francese, sarà sufficiente che faccia uso delle corrispondenti parole francesi. Che dire però del caso in cui costui non possieda ancora i concetti in questione? Gli si dovrà insegnare ad usare le parole mediante esempi ed esercizi, mediante cioè l'addestramento. “Il fondamento di ogni spiegazione è l'addestramento”. (Z, § 419)

La differenza messa in evidenza in questo passo è quella tra chi padroneggia una tecnica e chi invece non la padroneggia. Se la persona in questione possiede già il concetto che intendiamo spiegargli, se cioè possiede già una tecnica che include l'uso analogo di un termine differente, allora è sufficiente fornire una spiegazione che metta in relazione i due termini. Per esempio, ad un inglese potremmo spiegare il significato di “regola” dicendogli semplicemente “ ‘regola’ means ‘rule’ ”. L'intesa sarebbe immediata. Per quanto riguarda, invece, la persona che questo concetto ancora non lo possiede, le cose si fanno evidentemente più complicate, poiché costui necessiterebbe d'esser addestrato all'uso del termine. «Se però costui non possiede ancora questi *concetti*, gli insegnerò a usare le parole mediante

*esempi e con l'esercizio*". E nel far ciò non gli comunico nulla di meno di quanto sappia io stesso» (RF, I, § 208). Wittgenstein ci sta dicendo che l'esempio non costituisce solo il primo passo di una spiegazione del concetto, ma io stesso che fornisco gli esempi non so nulla di più di ciò che comunico a colui che li riceve. Si potrebbe aggiungere: non so nulla di più in termini "teorici", non ho, cioè, vere e proprie ragioni per usare il segno in questo modo piuttosto che in un altro (lo stesso si può dire per il seguire una regola); l'unica differenza tra l'addestratore e colui a cui l'addestramento viene impartito è che il primo possiede un "sapere pratico", se così lo vogliamo chiamare, che il secondo invece non possiede. Chi possiede un concetto, infatti, sa fare qualcosa che chi non lo possiede non sa fare, ossia usare un certo termine in un dato modo. Io che fornisco esempi non so nulla di più di chi li riceve, ma a differenza di costui li riconosco già come esempi dell'uso di una parola e uno degli scopi dell'addestramento è proprio quello di far sì che anche la persona in questione riesca a riconoscere negli esempi, con l'esercizio, ciò che vi riconosco io, ossia un certo uso regolare.

Il paragrafo 219 sottolinea un altro punto importante

Quando seguo la regola non scelgo. Seguo la regola *ciecamente* (RF, I, § 219)

Una volta compresa, la regola vincola il mio giudizio riguardo a ciò che concorda e ciò che non concorda con essa. «Una volta che ho afferrato una regola sono vincolato in quello che farò in seguito. Ma questo, naturalmente, significa soltanto che sono vincolato nei giudizi che pronuncerò su quel che è conforme alla regola e su quel che conforme non è.» (RFM, VI, § 27); se voglio seguire la regola devo fare quella data cosa che corrisponde ad essa, e non un'altra. In questo senso "seguo la regola *ciecamente*". Wittgenstein vuol mettere in evidenza il fatto che seguiamo la regola con completa familiarità, senza porsi problemi e senza essere chiamati a decidere alcunché; non abbiamo bisogno di riflettere sulle sue possibili interpretazioni, la seguiamo e basta, poiché, se la intendiamo, sappiamo esattamente quel che dobbiamo fare. Il concetto di "seguire una regola" è legato al concetto di abitudine, cioè al carattere abitudinario che è implicito nella prassi e che ci fa seguire la regola ciecamente.

La regola si segue meccanicamente [...] Meccanicamente: vale a dire: senza pensarci. Ma senza pensarci affatto? Senza rifletterci su. (OFM, VII, § 60)

C'è però un'ulteriore precisazione da fare e che ho già toccato precedentemente. Qualcuno, infatti, potrebbe essere tentato di interpretare

quanto detto come se si trattasse di cogliere, per esempio con un'immagine mentale, tutte le applicazioni della regola. Costui molto probabilmente sosterebbe che è per questo motivo che “quando seguo la regola non scelgo”; tuttavia non si tratta di ciò. Poniamo di avere un dato segno. Sono stato addestrato a comprendere il segno in questione come una regola ben specifica. Perché dovremmo assumere che la regola provochi in qualche modo le nostre azioni solo grazie ad un'immagine mentale? Se lo facessimo, infatti, saremmo poi costretti a spiegare in quale modo l'immagine stessa, a sua volta, possa tradursi in azione. Per mezzo forse di un'altra immagine mentale? Evidentemente, se optassimo per questa soluzione, verremmo risucchiati da un regresso all'infinito e l'azione non potrebbe mai compiersi:

L'assunzione di una prefigurazione del passaggio non ci porta avanti, poiché non colma lo iato tra la prefigurazione ed il passaggio reale. Se le mere parole della regola non potevano anticipare un passaggio futuro, allora non poteva fare ciò neppure un atto mentale che accompagnasse queste parole (LM, II, § 5)

A questo punto, non si può fare altro che assumere un collegamento diretto (messo in atto da un certo addestramento) tra regola ed azione, prescindendo quindi da qualsiasi *medium* (in questo caso un'immagine mentale). La relazione tra la regola e la sua applicazione è una connessione interna, non mediata: una regola non esiste senza le sue applicazioni e le sue applicazioni sono intrinsecamente applicazioni della regola.

Successivamente, Wittgenstein mette in evidenza la parentela sussistente tra la parola “regola” e la parola “concordanza”:

La parola «concordanza» e la parola «regola» sono tra loro imparentate; sono cugine. Se insegno ad uno l'uso di una delle due parole, egli impara contemporaneamente anche l'uso dell'altra. (RF, I; § 224)

La parola concordanza e la parola regola sono tra loro affini: sono cugine. Il fenomeno dell'agire e quell'agire secondo una regola sono tra loro connessi. (OFM, VI, § 41)

Ciò che viene rilevato in questi paragrafi è che la regola se viene appresa non può che comportare un comportamento ad essa conforme. Non può esserci una regola senza concordanza, perché quando si addestra qualcuno, per esempio ad un gioco, si insegnano contemporaneamente le sue regole ed è proprio dall'agire conformemente ad esse che si può capire se uno ha compreso o no quel gioco. Nell'addestrare qualcuno a seguire una regola, infatti, io lo incoraggio ad agire in accordo con il mio esempio, ad imitare,



cioè, il paradigma di comportamento che gli fornisco. Inoltre, imparare quali atti sono in accordo con una regola significa imparare anche quando due persone concordano nelle loro azioni. «Vero e falso e ciò che gli uomini dicono; e nel linguaggio gli uomini concordano. E questa non è una concordanza delle opinioni, ma della forma di vita» (RF, I, § 241). Secondo Wittgenstein, se una proposizione ha un determinato valore di verità dev'essere dotata di un certo significato, ma avere un significato condiviso richiede l'accordo di uso tra le persone. Ma da ciò non segue che questo accordo renda la sentenza vera, per così dire, di per sé: ciò che è importante sono le condizioni di espressione del significato non le condizioni di verità della proposizione.

L'ultima parte dell'osservazione 241 chiarifica la nozione di concordanza: questa concordanza non è nelle opinioni ma nella forma di vita. Qui Wittgenstein si sta richiamando ad una sua precedente affermazione secondo cui applicare un segno in un certo modo non è basato su ragioni o su propensioni naturali.<sup>9</sup> Per 'opinioni' lui intende qualcosa che risulta dal raziocinio, da qualcosa che è sostenuto da ragioni. Con 'forma di vita' lui intende una forma che fa parte della nostra natura.<sup>10</sup> Che noi seguiamo la regola ciecamente significa appunto questo: non c'è qualcosa per la quale noi possiamo o abbiamo bisogno di fornire delle giustificazioni.

Wittgenstein però aggiunge ancora un dettaglio al termine concordanza:

Della comprensione che si raggiunge tramite il linguaggio, non fa parte soltanto una concordanza nelle definizioni, ma anche (per quanto strana può sembrare) una concordanza nei giudizi. Ciò sembra togliere di mezzo la logica, ma non è così. – Una cosa è descrivere i metodi di misurazione, un'altra è ricavare ed enunciare i risultati della misurazione. Ma ciò che chiamiamo "misurare" è determinato anche da una certa costanza nei risultati delle misurazioni. (RF, I, § 242)

---

<sup>9</sup> Wittgenstein nega che il seguire la regola si basi in ultima analisi su delle ragioni: quando noi applichiamo una parola a una cosa non lo facciamo perché abbiamo una qualche ragione di pensare che questa è l'applicazione corretta. ««In qualche modo tu lo addestri a proseguire nell'esecuzione di un motivo ornamentale, - come può sapere in quale modo deve continuare da solo?». Ebbene, e io come lo so? – se questo vuol dire «Ho buone ragioni?» la risposta sarà: ben presto le ragioni mi verranno meno. E allora agirò senza ragioni.» (RF, I, § 211). «Deve proseguire così, *senza ragione*. Non perché la ragione non gli si possa ancora rendere comprensibile, ma perché – in questo sistema – non c'è nessuna ragione. (La catena delle ragioni ha un termine)» (Z § 301). La natura del seguire la regola è immediata, non riflessiva, automatica, ha carattere abituale.

<sup>10</sup> C. McGinn, 1984, pp.54-55

Diciamo che gli uomini, per comunicare tra loro, devono concordare tra loro sui significati delle parole. Ma il criterio di questa concordanza non è soltanto una concordanza relativamente alle definizioni, per esempio alle definizioni ostensive, ma anche concordanza nei giudizi. È essenziale per la comunicazione che si concordi in un gran numero di giudizi. (OFM, VI, 39)

Affinché vi possa essere comunicazione linguistica è ovviamente necessaria una concordanza nelle definizioni: ma le definizioni da sole non determinano il significato. Sia le definizioni verbali che quelle ostensive possono essere interpretate in vari modi, così ci deve essere qualche cos'altro che fissa il significato che le definizioni intendono fissare. Cosa vuol dire 'concordanza nelle definizioni'? Per capire cosa mi dice una persona che mi domanda: "Hai dato da mangiare al gatto?" io devo conoscere le definizioni di gatto, mangiare, etc, ma non devo essere solo d'accordo col mio interlocutore sulle definizioni di questi termini, ma anche su tutta una serie di giudizi, quali per esempio "questo è un gatto".

Questa concezione di significato però non abolisce la logica. «Ciò sembra togliere di mezzo la logica, ma non è così»: Wittgenstein dice che ciò sembra abolire la distinzione tra concettuale e fattuale (o grammaticale e empirico), ma non è così. La distinzione tra concettuale e fattuale non è una distinzione tra classi di enunciati, cioè tra tipi di proposizioni, ma una distinzione tra funzioni: lo stesso enunciato può essere usato o per asserire un fatto o come regola del linguaggio, cioè può avere o una funzione fattuale o una funzione grammaticale. Detto diversamente, una cosa è descrivere i metodi di misurazione, un'altra è enunciare i risultati ottenuti. La logica è come il metodo di misurazione, laddove il giudizio ne è il risultato. Il metodo e il risultato non sono indipendenti, perché noi potremmo considerare un'operazione come metodo di misurazione solo se ci è data una costanza nei risultati, ovvero ciò che noi chiamiamo misurare è determinato anche da una certa costanza nei risultati ottenuti. Il significato è, quindi, ciò che noi diamo per scontato quando giudichiamo, ma c'è significato solo perché le persone esibiscono una regolarità nei loro giudizi. La comunicazione dipende da un "background" di concordanza nei giudizi, così come la misurazione dipende da una costanza nei risultati.

In breve, comprendo una regola se so applicarla, ed essere in grado di applicare una regola significa anche riconoscere, nelle applicazioni altrui, esempi di tale regola. Se però ci trovassimo tutti sempre in disaccordo su ogni giudizio, verrebbe a mancare proprio quella concordanza nelle azioni che ci permette di vedere in esse l'applicazione di una regola. È impensabile che tutti possano applicare lo stesso metodo di misurazione in modo diverso. Infatti, in che cosa consisterebbe, in una simile situazione, l'attenersi al

metodo? «Potrebbe esserci aritmetica se chi calcola non concordasse?» (OFM, VI, 45)

Proviamo, per esempio, ad immaginare che su dieci persone ognuna, nel misurare la lunghezza di un tavolo, ottenga un risultato differente rispetto alle altre. Diremmo che almeno nove di loro si devono essere sbagliate, ma come facciamo a dirlo? Probabilmente, per esperienza con casi simili, sappiamo che una risposta corretta c'è, cioè che il risultato corretto di una misurazione è uno solo, poiché siamo certi che gran parte delle persone che abbiamo incontrato finora (tutte quelle effettivamente capaci di applicare il metodo di misurazione) avrebbe dato un'unica risposta, per esempio "un metro". Sappiamo infatti che la gente generalmente concorda nelle applicazioni delle regole di misurazione. Supponiamo, invece, che una tale concordanza non ci sia mai stata, ossia che ognuno ottenga un risultato di misurazione della lunghezza del tavolo differente dagli altri. Che significato avrebbero le parole "il tavolo misura  $x$  metri" in un contesto simile? Se, come ci dice Wittgenstein, "ciò che chiamiamo 'misurare' è determinato anche da una certa costanza nei risultati delle misurazioni", come possiamo parlare di "misurazione" nel caso in cui questa costanza venga a mancare? Si provi a generalizzare quanto detto ad ogni situazione in cui saremmo disposti a parlare dell'applicazione di una regola; se ciò fosse effettivamente possibile, allora lo stesso concetto di "applicazione di una regola" perderebbe il suo significato. In una tale situazione nessun caso potrebbe dirsi "seguire una regola" per il semplice fatto che non vi sarebbero regole da seguire. Infatti, non può esserci regola senza una certa costanza nella sua applicazione, come non può esserci concordanza nelle definizioni a prescindere da una vasta concordanza nei giudizi.

### 3. JERRY A. FODOR

#### 3.1. Fodor e la Macchina di Turing

Jerry A. Fodor rappresenta uno dei promotori di riferimento dell'ortodossia cognitivista, ma anche il promotore di ipotesi dissonanti rispetto a tale paradigma teorico. L'obiettivo della scienza cognitiva fu quello di scoprire le capacità di rappresentazione e di calcolo della mente e la loro rappresentazione strutturale e funzionale. Rappresentazione e calcolo sono infatti le due parole chiave del panorama cognitivista che, enfatizzando quello che avviene all'interno degli individui, segna una netta divisione con il comportamentismo, paradigma teorico che aveva rappresentato la teoria ufficiale fino alla metà degli anni Cinquanta e da cui Fodor prende nettamente le distanze. Secondo Fodor, la tesi comportamentista non riesce a dar conto della causazione tra stati mentali, la quale però sembra avere un ruolo primario nell'ambito del mentale (se si considerano i comportamenti come effetti di catene causali mentali). La critica al comportamentismo è rivolta sia al comportamentismo psicologico che a quello logico. Il primo è fondato sul presupposto della riducibilità del mentale al comportamento manifesto e l'analisi viene condotta sui comportamenti osservabili e verificabili intersoggettivamente. Il secondo invece poggia sull'idea della traducibilità delle espressioni mentali in espressioni riguardanti il comportamento.<sup>11</sup>

L'alternativa a queste due visioni Fodor la trova nel funzionalismo, il quale ha come fondamento il "principio della realizzabilità multipla del mentale", secondo cui sistemi fisici diversi, per esempio noi ed il computer, possono condividere lo stesso tipo di stato funzionale a prescindere dal sistema fisico che lo realizza. Il funzionalismo riconosce che i particolari mentali possono essere fisici ed è quindi compatibile con l'idea che la causazione mentale sia una forma di causazione fisica. Sostiene sia che le proprietà mentali siano definite in termini di relazioni, sia che le interazioni tra mente e corpo siano causali. Il comportamentismo logico condivide solo la prima affermazione, il fisicalista solo la seconda. Ne risulta che il funzionalismo sembra catturare le caratteristiche migliori delle due alternative materialistiche al dualismo.<sup>12</sup>

Questa cesura dai paradigmi teorici precedenti viene caratterizzata principalmente da due assunti:

---

<sup>11</sup> Fodor rifiuterà, oltre alla tesi comportamentista, anche la cosiddetta teoria dell'identità dei tipi in quanto essa ha come limite l'incapacità di spiegare come avvenga che sistemi fisici diversi tra loro possano trovarsi in uno stesso stato mentale. La teoria dell'identità dei tipi è una forma di riduzionismo che identifica tipi di stati mentali con tipi di stati cerebrali e nega che sostanze diverse da quelle neuroniche possano produrre stati mentali.

<sup>12</sup> F. Ferretti, 2001, p. XV

- La centralità accordata alle rappresentazioni mentali, in base all'assunzione che l'agire esterno dell'individuo sia mediato da stati e processi interni
- Il ruolo esplicativo della metafora del calcolatore, secondo cui i processi del pensiero sono analoghi a processi di calcolo.<sup>13</sup>

La tesi che sostiene Fodor è la cosiddetta *teoria computazionale-rappresentazionale della mente* (TCRM), secondo la quale i processi cognitivi sono computazioni su rappresentazioni mentali. Per Fodor, il sistema rappresentazionale della mente è un vero e proprio *linguaggio del pensiero* (LdP), nel quale i contenuti del pensiero sono espressi in strutture simboliche che, per quanto siano più astratte degli enunciati del linguaggio naturale, devono avere una sintassi e una semantica molto simile alla loro.

Fodor pensa all'uomo come ad una macchina biologica, cioè come un elaboratore di informazioni i cui meccanismi cognitivi operano nel cervello. Se si considerano gli stati mentali come strutture simboliche e i processi mentali come operazioni su queste strutture, allora l'analogia con il calcolatore risulta essere molto utile per capire la posizione della TCRM, poiché la mente individuale viene concepita come elaboratore di informazioni e il calcolatore come metafora per spiegare i processi mentali dell'uomo. I calcolatori possono, infatti, essere visti, oltre che come strumenti fisici che agiscono conformemente alle leggi della fisica, come un modello adeguato di dispositivi che eseguono operazioni su simboli e strutture simboliche e che agiscono nel rispetto delle leggi della logica, cioè quelle leggi che stanno a fondamento del pensiero. L'esemplificazione più appropriata dell'idea della condizione di formalità che sta alla base dei processi mentali come forma di calcolo, è la Macchina di Turing. «Se le menti sono dispositivi per manipolare simboli si può pensare alle menti nei termini del modello della macchina di Turing, poiché le macchine di Turing hanno la stessa generalità di qualsiasi dispositivo per la manipolazione di simboli».<sup>14</sup>

Brevemente, la macchina di Turing (MT) fu il risultato del tentativo di Alan Turing di produrre un modello formale dell'attività di un essere umano che esegue un calcolo. Essa è composta da un'unità di esecuzione, che esegue le operazioni specificate dallo stato in cui la macchina si trova e dal simbolo che sta leggendo; da un nastro infinito suddiviso in celle nelle quali può essere letto o scritto un simbolo; da una testina di lettura/scrittura, che ad ogni istante è posizionata su una casella del nastro. Oltre a scrivere simboli e cancellarli, l'unità di esecuzione può spostarsi sul nastro, fermarsi e porsi in un certo stato interno. Il comportamento della macchina è determinato da una

---

<sup>13</sup> F.Ferretti, 2001, p. VI

<sup>14</sup> J. Fodor, 1988, p.71

tabella di istruzioni che specifica, per ogni possibile stato<sup>15</sup> in cui si trova la testina e ogni possibile simbolo presente sulla casella corrente, qual è lo stato successivo e qual è l'azione da compiere. La capacità operativa di una MT, quindi, è determinata dalla tavola delle regole condizionali, cioè regole che definiscono il comportamento della macchina su un nastro di Input-Output (lettura e scrittura), le quali indicano cosa fare in corrispondenza della combinazione di stati e di simboli letti sul nastro: se la macchina si trova nello stato 0 e legge il simbolo A, allora passa allo stato 1 e scrive il simbolo B. Queste regole consentono alla macchina di produrre in output un certo risultato, sulla base dell'input opportunamente codificato.

Nonostante la sua semplicità, la MT è in grado di risolvere qualsiasi problema per il quale esista un algoritmo di soluzione; si tratta solo di scomporre il problema in passi così elementari da poter essere eseguiti tramite una delle operazioni alla portata della macchina. Ogni funzione computabile per mezzo di un algoritmo è, infatti, calcolabile da una MT, e tale capacità è resa ancora più evidente dal fatto che è possibile una MT universale, cioè una macchina capace di simulare il comportamento di qualsiasi MT e quindi di computare qualsiasi funzione calcolabile mediante un algoritmo.

Quindi, per tornare a quanto si diceva, se la MT può eseguire qualsiasi funzione calcolabile e se il pensiero è una forma di calcolo, allora il computer è una metafora adeguata dei processi mentali. Fodor è convinto che la mente umana funzioni come un computer e questa convinzione è strettamente legata alla necessità di postulare il linguaggio del pensiero, un linguaggio universale, innato. Egli ritiene che la mente umana sia una MT perché parte dal presupposto che nel cervello abbiano luogo dei processi fisici che, se solo ne avessimo una conoscenza adeguata, potremmo descrivere come trasformazioni di espressioni secondo regole determinate. Questo sistema di regole che governa i processi di trasformazione – una sorta di algoritmo della mente – è anch'esso parte del nostro patrimonio biologico, ma la sua complessità è enorme, quindi è decisamente improbabile che si possa giungere ad una sua caratterizzazione completa.

Quando si dice che la MT “legge” i simboli contenuti nelle caselle del nastro, ciò che si vuole dire è che essa ne percepisce la forma grafica, non che ne colga il significato. Stessa cosa per il cervello, in quanto lavora

---

<sup>15</sup> Non si deve pensare agli stati della macchina come a oggetti ad essa interni, ma come a forme delle diverse fasi della computazione. Tenere traccia dello stato è indispensabile per far dipendere l'azione della macchina non solo dal simbolo letto, ma anche dall'istante in cui viene letto, cioè dalla particolare fase della computazione in cui si trova. Una macchina priva di stati (cioè, una macchina che si ritrova sempre nel medesimo stato) è capace solo di rispondere in un certo modo a un certo simbolo e non potrebbe eseguire un procedimento, che è qualcosa che si dispiega nel tempo. Lo stato è una sorta di memoria cumulativa delle mosse compite fino a quell'istante.

analogamente ad un computer. Fodor sostiene che le espressioni del linguaggio del pensiero debbano sì avere un significato, ma che questo non svolga nessun ruolo nei processi di trasformazione che coinvolgono le espressioni. Tali processi sono di natura formale, cioè dipendono solo dalle loro proprietà sintattiche. Le computazioni sono sensibili solo alle proprietà formali delle espressioni o simboli del linguaggio del pensiero, non al loro significato. Questo è l'aspetto per cui l'analogia con la MT è particolarmente forte. Nella MT la testina di lettura che ispeziona una casella del nastro è ignara del significato che possiede quel simbolo: ciò che la determina è una differenza puramente fisica concernente la forma esteriore del simbolo. Analogamente i processi mentali sono sensibili solo alle proprietà fisiche dei simboli che manipolano: le espressioni del linguaggio del pensiero sono, per i processi cognitivi, pure iscrizioni materiali.

### 3.2. Il Linguaggio del Pensiero

Nel 1975 Fodor pubblicò *The Language of Thought*, uno dei suoi saggi più famosi, in cui sostiene che la mente umana dispone di un sistema di rappresentazioni che ha molte caratteristiche del linguaggio naturale e che lui chiama, appunto, linguaggio del pensiero. Si postula che il pensiero prenda posto in un linguaggio mentale e che questo linguaggio consista in un sistema di rappresentazioni che è fisicamente realizzato nel cervello e che ha una sintassi e una semantica combinatoria, tali che le operazioni sulle rappresentazioni sono causalmente sensibili alle proprietà sintattiche delle rappresentazioni.

In questo saggio, l'autore sviluppa una teoria computazionale e rappresentazionale della mente secondo la quale la mente è quella organizzazione funzionale del cervello che, indipendentemente dalla struttura fisica di questo, è capace di spiegare causalmente il comportamento dell'organismo: il cervello, simile ad un computer, elabora informazioni che vengono fornite dagli stimoli sensoriali e li converte in risposte motorie. È la mente che consente al cervello di svolgere tali operazioni. Le attività mentali sono per Fodor dei processi computazionali determinati da regole sintattiche e questi processi sono computazioni, calcoli, cioè una sequenza di operazioni su rappresentazioni mentali. La teoria della mente ipotizza che esistano realmente degli stati intenzionali scritti nel cervello, cioè dei simboli dotati di valore semantico combinabili secondo regole sintattiche. Questi stati intenzionali sono *efficaci causalmente* e *valutabili semanticamente* cioè, da un lato, le transizioni da uno stato intenzionale all'altro costituiscono una sequenza *causale* (un processo in cui ogni stato è causa del successivo) e dall'altro lato, dato un certo stato intenzionale, saranno le sue proprietà

semantiche (il suo contenuto) a determinare quali altri stati seguiranno da esso. Dunque il contenuto degli stati intenzionali causa altri stati intenzionali. Essi sono intesi come relazioni fra un organismo e rappresentazioni mentali e qualsiasi attività mentale risulta essere così una computazione che il cervello opera manipolando simboli, rappresentazioni che trova dentro di sé scritte in un linguaggio neurologico, cioè il linguaggio del pensiero (o *mentalese*).

Secondo l'autore, quindi, i processi cognitivi sono computazionali e ad essi viene attribuito un medium, cioè un sistema rappresentazionale nel quale tali processi vengono compiuti.<sup>16</sup> Ma il sistema rappresentazionale non può essere concepito come linguaggio naturale: ciò che può essere detto nel linguaggio naturale, pubblico, è il risultato di precedenti elaborazioni e computazioni che avvengono in un linguaggio interno (il Mentalese), un linguaggio che utilizza formule, simboli e regole rigide.

Rappresentazioni e computazioni sono i termini cardine della sua teoria computazionale-rappresentazionale. La TCRM, da una parte, spiega il modo di codificare l'informazione, che è indispensabile per dar conto dei processi mentali; essa è un'ipotesi empirica che spiega come funziona una mente-cervello, per la quale i processi cognitivi devono essere capaci di elaborare un contenuto e di produrre un comportamento ad esso conforme. Dall'altra, richiede sistemi rappresentazionali e che i processi di elaborazione abbiano accesso solo alle proprietà formali dei simboli.<sup>17</sup>

Si è già detto che Fodor presuppone che le menti siano dispositivi che manipolano strutture simboliche. Ciò che i simboli rappresentano è il loro contenuto ed il contenuto di un simbolo è determinato dal suo ruolo funzionale, cioè dalle relazioni causali che intercorrono tra quel simbolo e gli altri simboli all'interno del sistema rappresentazionale. Le relazioni causali possono infatti venir imitate da quelle sintattico-formali dei simboli ed è proprio in riferimento a queste ultime proprietà che è possibile dar conto del contenuto mentale nei termini di relazioni formali che i simboli e le loro parti intrattengono tra di loro.

Per quanto concerne il ruolo causale degli stati mentali, cioè tanto la possibilità che avvenga quanto il modo in cui avviene che il contenuto del nostro pensiero possa causare un certo comportamento, esso rimane una componente essenziale dell'immagine del linguaggio che Fodor sostiene.

Le rappresentazioni mentali, infatti, influenzano il comportamento in forza del loro contenuto – ciò rende plausibile che i processi mentali siano collegati a rappresentazioni mentali – ed esse hanno contenuti distinti solo se esse hanno anche forme distinte – ciò rende plausibile che i processi mentali

---

<sup>16</sup> Fodor, 1975, p. 27

<sup>17</sup> F. Ferretti, 2001, p. XVI



siano computazionali. In *The Language of Thought* (1975) Fodor sostenne che le computazioni sono esattamente processi in cui le rappresentazioni hanno conseguenze causali in virtù della loro forma.<sup>18</sup> Che la forma di un simbolo del linguaggio del pensiero (o Mentalese) sia soggetta a calcoli non significa però che il contenuto dello stesso simbolo non abbia rilievo: la forma è tale proprio in quanto veicola quel contenuto.<sup>19</sup> La TCM funziona indipendentemente dalla questione di come le rappresentazioni siano relate al mondo. Successivamente – a partire da *Psicosemantica* (1987) – cambiò idea, sostenendo, invece, che il contenuto dei simboli del linguaggio del pensiero deriva dalle loro relazioni causali con il mondo esterno e che, quindi, esso non coincide più con il loro ruolo funzionale. Riconosce che una concezione rappresentativa della mente non può più fare a meno di prendere in considerazione la questione del rapporto tra stati mentali e mondo esterno. Questo cambiamento di prospettiva non implica però l’abbandono della principale idea della TCRM, secondo cui il comportamento dell’agente è da trattare come il risultato di computazioni, i cui calcoli presuppongono un medium nel quale avvengono questi calcoli, cioè il sistema rappresentazionale. Gli stati mentali hanno natura rappresentazionale, e questo significa che essi veicolano informazioni sul mondo, ovvero ce lo presentano in un certo modo. Essi sono valutabili semanticamente, cioè sono suscettibili di essere veri o falsi: se io credo che Torino sia a Nord di Roma significa che io mi rappresento uno stato di cose nella forma di una certa relazione spaziale tra due “oggetti”. Tale credenza è vera perché c’è un fatto nel mondo che la rende vera.<sup>20</sup>

Essere in uno stato mentale equivale intrattenere una relazione con un simbolo del linguaggio del pensiero il quale, al pari di un’espressione del linguaggio naturale, ha sia proprietà causali che proprietà semantiche. Le proprietà causali sono determinate dalla sua forma, cioè dalle caratteristiche materiali del simbolo stesso. Le sue proprietà semantiche derivano dal fatto che il simbolo rappresenta qualcosa, cioè il fatto di essere vero o falso. Il contenuto semantico di uno stato mentale, per esempio la mia credenza che fuori piove, è reso vero o falso dalle proprietà semantiche dell’espressione in mentalese «fuori piove», cioè dalle sue condizioni di verità, il modo in cui il mondo dev’essere affinché l’espressione in questione sia vera.

Secondo Fodor, è necessario postulare un linguaggio del pensiero come preconditione di qualsiasi teoria psicologica. Non solo le azioni, ma

---

<sup>18</sup> F. Ferretti, 2001, p. XVII

<sup>19</sup> Sotto questo punto di vista, Fodor accetta le analisi della psicologia del senso comune riguardanti gli atteggiamenti proposizionali, cioè quegli enunciati mentali con i quali ci riferiamo a qualcosa che ha a che fare con il mondo esterno e con i quali ci rapportiamo nei modi più disparati.

<sup>20</sup> A. Paternoster, 2008, p.35

anche l'apprendimento e la percezione devono essere visti come basi dei processi computazionali. Non c'è computazione senza rappresentazione.<sup>21</sup>

Apprendere un linguaggio significa, infatti, apprendere cosa significano i predicati del linguaggio e questo implica che si possa comprendere un predicato solo se si conoscono le condizioni sotto le quali le proposizioni che lo contengono risultano essere vere. Apprendere un linguaggio vuol dire apprendere le regole di verità dei suoi predicati: *S apprende P se S apprende le regole di verità di P.*<sup>22</sup> In altri termini, non si può apprendere un linguaggio i cui termini esprimono proprietà semantiche che non sono espresse da termini di un qualche linguaggio che si è già in grado di utilizzare e non si può apprendere un linguaggio a meno che non si possieda un sistema capace di rappresentare i predicati e la loro estensione. Apprendere un linguaggio implica, dunque, apprendere le proprietà semantiche dei suoi predicati ed S apprende le proprietà semantiche di P solo se S apprende una qualche generalizzazione che determina l'estensione di P, cioè una serie di cose di cui P è vera. Si comprende un predicato solo se si conoscono le condizioni sotto le quali le proposizioni che li contengono sono vere. Apprendere le proprietà semantiche di P viene identificato con l'apprendere le regole di verità di P e non si può apprendere un linguaggio i cui termini esprimono proprietà semantiche non espresse dai termini di un linguaggio che si è già in grado di usare.

Apprendere un linguaggio implica apprendere cosa significano i predicati del linguaggio. Apprendere cosa significano i predicati di un linguaggio implica apprendere una determinazione dell'estensione di questi predicati. Apprendere una determinazione dell'estensione dei predicati implica apprendere che essi ricadono sotto certe regole (cioè, regole di verità). Ma non si può apprendere che P ricade sotto certe regole a meno che non si sia già appreso un linguaggio nel quale P e R possono essere rappresentate. Così, non si può apprendere un linguaggio a meno che non si abbia già un linguaggio. In particolare, non si può apprendere un linguaggio a meno che non si abbia già un sistema capace di rappresentare i predicati in quel linguaggio e la loro estensione. E quel sistema non può essere un linguaggio che è stato appreso. Ma i primi linguaggi sono appresi. Quindi, almeno alcune operazioni cognitive vengono effettuate in un altro linguaggio rispetto ai linguaggi naturali. (Fodor, 1975, p. 64)

Quanto detto risulta essere accettabile se ci si impegna nell'assunzione che l'organismo è capace di apprendere se esso ha precedentemente accesso

---

<sup>21</sup> Fodor, 1975, p.34

<sup>22</sup> Fodor, 1975, p. 59

ad un qualche sistema rappresentazionale nel quale le proprietà semantiche dei predicati sono già espresse. Si può apprendere L (= linguaggio naturale) se si conosce già un linguaggio abbastanza ricco per esprimere l'estensione di un qualsiasi predicato di L. Detto in altro modo, si può apprendere cosa sono le proprietà semantiche di un termine solo se già si conosce un linguaggio che contiene un termine che ha le stesse proprietà semantiche. Quindi, non si può apprendere L a meno che non si conosca già un linguaggio differente da L ma sufficientemente ricco per esprimere le proprietà semantiche dei predicati di L. Nulla può essere espresso in L se non può essere precedentemente espresso nel linguaggio del pensiero. Ciò che viene negato, quindi, è che si possa apprendere un linguaggio i cui predicati hanno estensioni non esprimibili dai predicati dal sistema rappresentazionale.

Il fatto che si presupponga un sistema rappresentazionale come base sulla quale si sviluppa il linguaggio naturale porta a considerare il linguaggio del pensiero come strutturato e costituito sia di proprietà sintattiche sia di proprietà semantiche simili a quelle del linguaggio naturale. Tali proprietà includono (a) avere parti costituenti, che si combinano tra di loro secondo le regole della sintassi; (b) avere parti atomiche che si riferiscono a cose e a proprietà del mondo; (c) la composizionalità, ovvero, le proprietà semantiche di una rappresentazione complessa dipendono dalle proprietà semantiche dei suoi costituenti atomici e dalle regole di composizione o, anche, nel senso che il linguaggio del pensiero include un numero infinito di espressioni semplici che, combinandosi, generano un'infinità potenziale di espressioni complesse.<sup>23</sup>

Ricapitolando,

- I modelli disponibili dei processi cognitivi si caratterizzano come computazionali e quindi presuppongono un sistema rappresentazionale nel quale vengono eseguite le computazioni;
- Il sistema rappresentazionale non può essere un linguaggio naturale;
- Le proprietà semantiche dei predicati di un linguaggio naturale devono essere esprimibili prima di tutto nel sistema rappresentazionale.

Se, dunque, il linguaggio del pensiero non dev'essere inteso come linguaggio appreso, allora questo dev'essere analizzato come linguaggio già conosciuto, cioè innato. Il linguaggio del pensiero è già compreso nel senso che è già disponibile all'uso come veicolo dei processi cognitivi.<sup>24</sup> Fodor nega che le strutture e i contenuti mentali possano dipendere da processi di apprendimento e che siano riconducibili all'esperienza. Le rappresentazioni che costituiscono il linguaggio del pensiero sono, ovviamente, anch'esse

---

<sup>23</sup> D. Marconi, 2008, p. 57

<sup>24</sup> Fodor, 1975, p.65

innate: ciascuno di noi nasce con un sistema completo di rappresentazioni, che corrispondono ai concetti espressi dalle parole di una lingua.

Una possibile critica che potrebbe essere mossa a Fodor è che il concetto di imparare o comprendere un linguaggio porterebbe ad un regresso all'infinito: se non si può pensare al linguaggio del pensiero come ad un linguaggio appreso, allora si potrebbe considerare un meta-linguaggio nel quale siano formulate le rappresentazioni dell'estensione di predicati del linguaggio. Ma se, secondo il principio di Fodor, anche questo meta-linguaggio non può essere appreso perché già precedentemente conosciuto, allora si pone il problema di un regresso infinito nella ricerca di un primo linguaggio nel quale vengono formulate le definizioni di verità.<sup>25</sup> La replica di Fodor riguarda la possibilità di stabilire una distinzione tra conoscere e apprendere: «Il mio punto di vista è che non si può apprendere un linguaggio a meno che non se ne conosca già uno. Non è quello secondo cui non si può apprendere un linguaggio a meno che tu non ne abbia già appreso uno. Quest'ultima affermazione porta ad un regresso infinito, ma la prima no» (Fodor, 1975, p.65). Ciò significa che il linguaggio del pensiero è già conosciuto e non lo si può aver appreso; esso è il medium delle computazioni che costituiscono i processi cognitivi e lo si possiede fin dall'inizio, cioè è innato.

La critica chomskiana alla tesi che la sintassi del linguaggio sia appresa dev'essere perciò estesa agli elementi essenziali del pensiero: ad essere innati sono i contenuti espressi dagli enunciati del linguaggio del pensiero. Lo sviluppo cognitivo del bambino, infatti, è fondamentalmente lo sviluppo di sistemi concettuali-rappresentazionali sempre più potenti e ciò che Fodor ritiene improbabile è che l'apprendimento dei concetti produca i meccanismi per le transizioni da stadio a stadio (come, invece, suppose Piaget). «Lo sviluppo cognitivo non può essere conseguenza dell'apprendimento di concetti. Le ragioni dovrebbero essere note: esse sono in fondo le stesse che portano alla conclusione per cui non si può apprendere un linguaggio i cui predicati esprimono estensioni che sono inesprimibili da un linguaggio appartenente ad un livello più basso» (Fodor, 1975, 89-90). Supponiamo che un bambino al primo stadio debba provare ad apprendere il concetto C. Ciò che deve fare è apprendere le condizioni sotto le quali una certa cosa è un'occorrenza di C. Presumibilmente, dovrà apprendere qualcosa della forma "(x) (x è C se e solo se x è F)", dove F è un qualche concetto che si applica ogni volta che si applica C. Per poter apprendere questo, il sistema concettuale deve contenere necessariamente F. Consideriamo un altro caso in cui C appartenga al secondo stadio. Se qualcosa è un concetto del secondo stadio, allora segue che esso non è coestensivo con i concetti del primo stadio,

---

<sup>25</sup> Fodor, 1975, p. 65

«altrimenti la differenza tra stadi non sarebbe una differenza del potere espressivo dei sistemi concettuali che li caratterizzano».<sup>26</sup> Ma se un bambino al primo stadio non può rappresentare l'estensione di C in termini di un qualche concetto del sistema di cui dispone, egli non può rappresentare tale estensione poiché, per definizione, «il suo sistema concettuale è proprio la totalità dei dispositivi rappresentazionali che egli può utilizzare per l'elaborazione cognitiva»<sup>27</sup>. Se non può rappresentare l'estensione di C, allora non può apprendere C. Così, o le condizioni per l'applicazione di un concetto del secondo stadio possono essere rappresentate come concetto del primo stadio (nel qual caso non c'è alcun senso ovvio per cui il sistema concettuale di secondo livello sia più potente del sistema concettuale di primo livello), o ci sono concetti del secondo stadio la cui estensione non può essere rappresentata nel primo stadio, ed in questo caso il bambino del primo stadio non può apprenderli.<sup>28</sup>

Questo vuol semplicemente dire che non si può apprendere un sistema concettuale più ricco del sistema concettuale da cui si parte. Può darsi che i concetti complessi siano composti di concetti semplici, ma ciò implica che nessuno può apprendere un concetto complesso senza possedere già i concetti semplici che lo costituiscono e le operazioni combinatorie che permettono la combinazione dei concetti semplici. Presumibilmente, il linguaggio del pensiero contiene un'espressione semplice per ogni concetto possibile (o quasi), il che implica che i concetti siano innati.

Come si è già detto, le rappresentazioni mentali sono oggetti interni che hanno molte proprietà delle espressioni del linguaggio verbale; in particolare, le rappresentazioni atomiche ('concetti') hanno molte proprietà delle parole di una lingua. Ma solo alcune proprietà sono realmente rilevanti, agli effetti della loro elaborazione da parte delle regole mentali. Le regole sono sensibili solo alla *forma* delle rappresentazioni, cioè alle loro proprietà sintattiche, mentre sono indifferenti alle loro proprietà semantiche. Questa è una conseguenza del fatto che le regole sono computazionali e i processi computazionali sono sensibili solo alla forma dei dati a cui si applicano.<sup>29</sup>

---

<sup>26</sup> Fodor, 1975, p. 90

<sup>27</sup> Ibidem

<sup>28</sup> Ibidem

<sup>29</sup> Si può rendere più semplice la comprensione di questo punto, riferendosi al modo di funzionare di una macchina di Turing. Quando la macchina aggiunge 1 ad un numero, il suo processo di computazione è del tutto indifferente che le sequenze di 1 scritte sul nastro stiano per numeri o per altre cose; noi le trattiamo come sequenze numerali, ma la macchina, di per sé, sa solo riconoscere ed elaborare dei segni sulla base della loro forma. Quando si dice che la macchina distingue 1 da \* non si vuol dire che essa sa che 1 sta per il numero 1 e che \* indica la casella vuota: si vuol solo dire che la macchina sa discriminare tra la forma 1 e la forma \*.

Le computazioni con cui la TCM identifica i processi mentali sono, dunque, sensibili solo alle proprietà formali delle espressioni o simboli del linguaggio del pensiero. Esse consistono tipicamente nella generazione di simboli, tale che la scelta del simbolo da generare ad un certo passo dipende esclusivamente dal simbolo generato al passo precedente e dalle regole sintattiche.

Ciò che è importante tener presente è che gli stati mentali ereditano dalle espressioni del linguaggio del pensiero con cui sono in relazione le loro proprietà sintattiche e semantiche. Il ruolo causale dello stato mentale è determinato dalle proprietà sintattiche dell'espressione con cui lo stato è in relazione. Ed è determinato dalla sintassi perché il fatto che uno stato mentale ne causi un altro dipende dalle regole sensibili esclusivamente alle proprietà formali delle iscrizioni, nello stesso modo in cui in una MT un certo stato è causato dallo stato precedente in virtù del fatto che nella MT è "cablata", cioè codificata internamente, una certa istruzione che dice alla macchina cosa fare in quella determinata situazione.

I processi cognitivi, dunque, non sono altro che manipolazioni di rappresentazioni in base a regole rigide. Le connessioni causali tra gli stati mentali sono realizzate grazie a regole sintattiche che governano la manipolazione delle rappresentazioni, cioè la sintassi del linguaggio del pensiero.

Ciò conduce alla conseguenza che le elaborazioni mentali, cioè i processi cognitivi, siano 'cieche al significato'.<sup>30</sup> Questo potrebbe risultare controintuitivo, dal momento che, quando noi pensiamo ai nostri processi cognitivi, prendiamo a modello quelli coscienti i quali non sono ciechi al significato. Ma quando Fodor parla di processi cognitivi non parla di quelli che sono coscienti, ma del funzionamento di certe elaborazioni inconscie. Inoltre, egli non nega che l'input e l'output delle elaborazioni mentali (cioè le rappresentazioni di partenza e di arrivo) abbiano significato; nega solo che le computazioni, cioè le elaborazioni mentali, si basino sui significati delle rappresentazioni su cui operano.

Ma in che modo uno stato mentale riesce ad essere causa di qualcosa, senza far intervenire il significato di cui è portatore ma essendo attivo solo attraverso la sua forma? Cioè, il fatto che le computazioni non siano "semanticamente controllate"<sup>31</sup> non comporta forse che i risultati siano casuali e arbitrari? Com'è possibile che i processi cognitivi funzionino e che diano risultati corretti? Fodor pensa che, nonostante che il sistema rappresentazionale proceda mediante l'elaborazione sintattica, ci sia comunque un collegamento tra sintassi e semantica del mentale e questo

---

<sup>30</sup> D. Marconi, 2008, p.59

<sup>31</sup> Ibidem

collegamento può essere chiarito sfruttando l'analogia con ciò che succede nel linguaggio formale. Dal punto di vista della TCRM il passaggio da una premessa ad un'altra viene compiuto sulla base di regole puramente formali e la capacità di fare un ragionamento non richiede la comprensione degli enunciati che lo compongono, ma la sola applicazione di queste regole. Tali regole sono del tutto indifferenti al contenuto delle premesse, ma sono sensibili solo alla loro struttura, ovvero a ciò che si può esprimere nelle schematizzazioni del tipo di  $p \& q$ ,  $\sim p$ , etc. È, infatti, la struttura proposizionale delle rappresentazioni mentali a garantire che queste ultime possano entrare in relazione.

Le regole hanno tuttavia la proprietà di conservare la verità, cioè, esse non portano mai da premesse vere a conclusioni false. Ad esempio, se è vero che a) «Tutti gli uomini sono mortali» e che b) «Socrate è un uomo», allora è vero anche che c) «Socrate è mortale». L'occorrenza «Socrate è mortale» è causata dalle occorrenze precedenti a) e b) in virtù di una regola sintattica cablata nella mente. Perciò, nonostante siano sensibili solo alle proprietà formali, le regole logiche elaborano risultati corretti, cioè dotati delle proprietà semantiche desiderate. Analogamente, si può pensare che le regole che governano le nostre elaborazioni mentali siano *buone regole*, capaci di elaborare risultati funzionali alla nostra sopravvivenza.

La correttezza formale dei processi cognitivi, dunque, è garanzia anche della loro capacità di conservare la verità, cioè di derivare conclusioni vere da premesse vere. Le regole logiche, semanticamente appropriate, valgono quale che sia il significato attribuito ai simboli che mettono in relazione.

Una rappresentazione mentale R1 causa un'altra rappresentazione R2 connessa alla precedente in modo semanticamente appropriato in virtù dell'applicazione di una regola sintattica del tipo 'Tutte le volte che c'è un'iscrizione R1, produci un'iscrizione R2' " [Paternoster, 2008, p. 38]

I sistemi rappresentazionali quindi consentono che i processi mentali siano definiti sì sulla sintassi ma rispettando comunque i vincoli della semantica, ovvero, le regole che governano le inferenze sono definite dalla sintassi ma rispettano i vincoli semantici: date certe premesse vere, la loro corretta applicazione porta solamente a conclusioni vere.

Il sistema rappresentazionale quindi possiede sia una sintassi che una semantica: una macchina (o un cervello) è in grado di manipolare simboli e di effettuare operazioni che corrispondono ad un calcolo logico semanticamente coerente ed i simboli, che vengono manipolati solo

formalmente, svolgono allo stesso tempo il ruolo di rappresentazione del mondo esterno, garantendo un collegamento tra mente e mondo.

### 3.3. Gli atteggiamenti proposizionali

Come si è già spiegato, secondo Fodor il pensiero e il ragionamento vengono svolti in un linguaggio mentale, cioè in un sistema rappresentazionale, simbolico, realizzato fisicamente. La mente è un sistema che manipola simboli e questi simboli funzionano come le parole di un linguaggio che si combinano per formare enunciati che esprimono un contenuto. Che cosa determina il contenuto delle espressioni del linguaggio del pensiero?

Ciascuno di noi, originariamente, possiede una varietà di pensieri detti “atteggiamenti proposizionali”. Gli atteggiamenti proposizionali sono del genere di “S crede che P”, “S desidera che P”, “S spera che P”, dove “S” si riferisce al soggetto dell’atteggiamento, “P” ad una qualsiasi proposizione e “che P” si riferisce alla proposizione che è l’oggetto dell’atteggiamento.

Il linguaggio del pensiero può essere formulato come un’ipotesi circa la natura degli atteggiamenti proposizionali e del modo in cui ci relazioniamo ad essi, laddove gli atteggiamenti proposizionali vengono interpretati come relazioni tra simboli e “credere che P” significa essere in una certa relazione con il simbolo che significa “che P”. Gli atteggiamenti proposizionali sono infatti relazioni tra organismi e formule interne e i loro oggetti sono da considerare come enunciati di un linguaggio non-naturale, cioè come formule di un sistema di rappresentazioni interne.

Secondo Fodor, infatti, gli stati mentali sono realizzati in una particolare forma di struttura simbolica e gli enunciati del linguaggio del pensiero sono necessari per spiegare la natura relazionale degli stati mentali. Avere un certo atteggiamento proposizionale significa essere in una certa relazione con una rappresentazione interna. E questa relazione è computazionale: «[...] gli stati mentali sono relazioni tra organismi e rappresentazioni interne, stati mentali causalmente intercorrelati si succedono gli uni agli altri secondo i principi computazionali che valgono formalmente per le rappresentazioni» (Fodor, 1975, p. 198). Gli stati mentali sono quindi conseguenze di processi mentali, cioè di processi di trasformazione delle rappresentazioni.

Fodor propone una concezione *triadica* degli atteggiamenti proposizionali: gli stati mentali sono una relazione R tra un organismo S e una proposizione P.

Gli atteggiamenti proposizionali devono essere quindi analizzati prima di tutto come relazioni, cioè come relazioni tra organismi e formule di



un linguaggio interno (enunciati interni o rappresentazioni interne). Un organismo sta in relazione con una proposizione se sta in una relazione causale/funzionale con una qualche occorrenza di una formula che esprime la proposizione.<sup>32</sup> Un enunciato come “John crede che piove” (S crede che P) esprime una relazione tra John (S) e qualcos’altro (P) e un’occorrenza di tale enunciato è vera se John (S) è in una relazione di credenza con quel qualche cosa (P).

Se, per esempio, si vogliono specificare le credenze e i desideri di John, il procedimento canonico consiste nell’usare farsi del tipo “John crede che P” / “John desidera che P”, dove gli enunciati subordinati ai verbi ‘credere’ / ‘desiderare’ esprimono appunto il contenuto delle credenze e dei desideri. L’idea di Fodor è che, affinché qualcuno possa credere o desiderare qualcosa, bisogna che da qualche parte nel suo cervello sia “scritto” un enunciato in mentalese (o nel linguaggio del pensiero), cioè una formula del sistema rappresentazionale. Gli stati di credenza e di desiderio comportano una realizzazione fisica nel cervello di enunciati del mentalese ed hanno anche un’efficacia causale, e questo dovrebbe spiegare come gli stati mentali possano avere un contenuto e come facciano ad interagire causalmente in virtù del loro contenuto.

La relazione tra un simbolo e la proprietà di ciò a cui si riferisce dev’essere una relazione causale: un simbolo mentale si riferisce a ciò che ha la proprietà P se è ciò che ha la proprietà P a causare nel cervello le iscrizioni del simbolo mentale. Supponiamo che nel nostro cervello appaia un’iscrizione di un certo simbolo  $x$  quando, e solo quando, vediamo un cane. Possiamo dire che sono i cani a causare in noi iscrizioni di  $x$  e che, quindi,  $x$  si riferisce ai cani.<sup>33</sup> È falso pensare che noi diciamo ‘cane’ tutte le volte che vediamo un cane. Ma, se vediamo un cane in circostanze normali, non possiamo fare a meno di pensare “Ecco un cane!”, il che significa che da qualche parte nel nostro cervello si materializza automaticamente un simbolo  $x$ .

Ciò che è importante da precisare è che il sistema di relazioni causali che fissa il riferimento di un simbolo del linguaggio del pensiero deve avere natura di “legge”. Possiamo affermare che il simbolo  $x$  si riferisce ai cani solo se è una legge che i cani causino  $x$ .

---

<sup>32</sup> F. Ferretti, 2001, pp. 11-35

<sup>33</sup> Questa affermazione si pone in contrasto con le tesi comportamentiste secondo cui la parola ‘cane’ si riferirebbe ai cani perché le nostre disposizioni al comportamento verbale sono tali che diciamo ‘cane’ in presenza dei cani e di nient’altro; questo perché sono i cani a causare il proferimento della parola ‘cane’. La differenza tra Fodor ed un comportamentista sta nel fatto che, mentre quest’ultimo considera il proferimento di parole del linguaggio naturale, Fodor considera invece la comparsa nel cervello di simboli in mentalese.

Secondo Fodor, per poter dire che gli stati mentali sono cause del comportamento è necessario che un singolo nesso causale sia l'esemplificazione di una legge generale. E poiché soltanto ciò che è fisico è dotato di poteri causali, allora gli stati mentali devono essere stati fisici (o devono essere realizzati da stati fisici).

Quindi:

- Gli stati mentali sono cause del comportamento
- Soltanto ciò che è fisico può avere poteri causali
- Gli stati mentali devono avere una realizzazione fisica

In altre parole, credenze e desideri devono essere proprietà della mente realizzate fisicamente nel cervello delle persone.

Sia  $S$  uno stato di credenza o di desiderio e sia  $\phi$  l'enunciato del mentalese che deve comparire nel cervello di un individuo affinché tale individuo possa essere nello stato  $S$ . In quanto enunciato del mentalese,  $\phi$  ha un contenuto, ed è questo il contenuto attribuibile allo stato  $S$ . D'altra parte, l'occorrenza materiale di  $\phi$  può essere coinvolta nei processi computazionali che hanno luogo nel cervello, può condizionarne lo sviluppo e, in definitiva, può influire, attraverso la mediazione di complicati meccanismi, sul comportamento osservabile: se questo capita, siamo autorizzati a dire che il comportamento ha tra le sue cause lo stato mentale di  $S$ . (Casalegno, 1997. p. 370)

Ciò che viene stabilito fin dal primo capitolo di *The Language of Thought* è che i comportamenti dell'agente sono connessi alle condizioni di credenza nelle quali esso si trova. Il processo che porta dalla credenza al comportamento prodotto viene schematizzato da Fodor in questo modo:

- L'agente si trova in una certa situazione  $S$ ;
- L'agente crede che una certa serie di opzioni comportamentali ( $b_1, b_2, \dots b_n$ ) siano disponibili in  $S$ , cioè, data  $S$ , tali opzioni comportamentali sono cose che lui crede di poter fare;
- Vengono previste le conseguenze dell'eseguire ciascuna delle opzioni, nel senso che l'agente calcola una serie di ipotesi alle quali verranno assegnate certe probabilità; tali ipotesi e probabilità dipenderanno da ciò che l'organismo crede e conosce in  $S$ ;
- La scelta del comportamenti sarà determinata come funzione di preferenze e probabilità assegnate alle conseguenze.<sup>34</sup>

In altre parole, secondo Fodor, per poter cogliere l'azione più appropriata a determinate circostanze l'organismo deve disporre di certi mezzi che gli consentano di descrivere la situazione in cui si trova ma anche

---

<sup>34</sup> J. Fodor, 1975, p. 28

a) le azioni possibili, b) le conseguenze di quelle azioni, c) un ordine preferenziale delle opzioni che egli vaglia quando deve scegliere l'azione da compiere. Affinché un comportamento venga prodotto, non è sufficiente dire che esso è determinato e guidato dalla realizzazione di un certo risultato; è inoltre richiesto che l'agente possa agire *nella mente* prima di agire effettivamente, cioè che l'agente percepisca il suo agire come capace di produrre il risultato che egli vuole; in altre parole, gli agenti possono rappresentare a loro stessi il comportamento grazie al sistema rappresentazionale che possiedono e alle operazioni computazionali che si svolgono su tale sistema: tali computazioni consistono in una successione logica di stati di credenza che consente ad un agente di arrivare a una certa conclusione partendo da certe premesse o di produrre un certo output comportamentale. Ne consegue che, per la TCRM, ogni volta che un agente si trova in un certo stato di credenza, egli intrattiene una relazione con un simbolo del linguaggio del pensiero.

[...] Sostenere di avere una credenza o un desiderio – o qualsiasi altra cosa – consiste nell'essere relati in un certo modo a una rappresentazione mentale. In accordo con la formulazione canonica di questo punto di vista: per ogni organismo O e per ogni proposizione P, c'è una relazione R e una rappresentazione mentale RM tale che: RM significa (esprime la proposizione che) P e O crede che P se e solo se per O vale R con RM (e, in modo simile, O desidera che P, se e solo se per O vale qualche diversa relazione R' con RM). E così via. [...] <sup>35</sup>

Fodor pensa che nella nostra testa esista una cosiddetta “scatola delle intenzioni”. Ogni intenzione non è altro che l'intenzione di rendere una qualche proposizione vera. Quando vogliamo rendere vero ‘che P’ ciò che facciamo è mettere nella scatola delle intenzioni un simbolo mentale che significa ‘che P’. Ciò che fa la scatola è computare, analizzare un comportamento che rende vero ‘che P’. Supponiamo che io abbia intenzione di alzare la mano sinistra (e voglia rendere vera la proposizione che io alzo la mano sinistra). Ciò che faccio è mettere nella scatola delle intenzioni un simbolo mentale che significa “io alzo la mano sinistra” e dopo varie computazioni ed elaborazioni, la mia mano si alza. <sup>36</sup>

Ciò che si ammette è che (1) ci sono stati mentali che sono associati ad oggetti intenzionali (per esempio, avere un simbolo che significa ‘io alzo la mano sinistra’) e (2) che questi stati mentali che sono associati ad oggetti intenzionali hanno anche ruoli causali. Ciò che però il linguaggio del pensiero

---

<sup>35</sup> F. Ferretti, p.122

<sup>36</sup> Fodor, 1990, p. 207-208

assume, per non ricadere nelle limitazioni del realismo intenzionale, è che questi stati mentali abbiamo sì un contenuto, ma anche una struttura sintattica. Nel realismo intenzionale ciò che si mette nella scatola delle credenze è un oggetto, mentre nella teoria del linguaggio del pensiero ciò che io inserisco nella scatola è una formula che contiene un'espressione che denota me e un'espressione che denota, per esempio, la mia mano sinistra. Le formule che si inseriscono nella scatola delle intenzioni contengono delle sotto-espressioni come costituenti, che significano, per esempio, 'io alzo la mia mano sinistra'.<sup>37</sup>

Un sistema rappresentazionale possiede, infatti, sia una sintassi combinatoria, la quale impiega due tipi di rappresentazioni, quelle atomiche e quelle composte, e in cui i costituenti delle rappresentazioni composte sono o atomici o composti; sia una semantica compositiva, cioè il contenuto semantico di una rappresentazione è funzione del contenuto semantico del suo costituente sintattico, della struttura complessiva della rappresentazione e dell'organizzazione dei costituenti nella struttura complessiva.

Le rappresentazioni strutturalmente complesse (molecolari) sono costruite sistematicamente a partire dai costituenti semplici (atomici) e il contenuto semantico delle rappresentazioni molecolari è una funzione del contenuto semantico dei loro costituenti atomici assieme alla loro struttura sintattico-formale. Le operazioni sulle rappresentazioni, che dominano il processo mentale, sono causalmente sensibili alla struttura sintattico-formale delle rappresentazioni definite da questa sintassi combinatoria. In altre parole, vi è una complessità degli stati mentali, cioè la struttura degli stati mentali è una struttura in costituenti, secondo cui i componenti atomici, che stanno alla base della costituzione di ogni enunciato, si combinano fra loro e formano nuovi enunciati complessi. Credenze e desideri sono, infatti, stati strutturati, hanno struttura combinatoria e avere una credenza o un desiderio implica avere una relazione con un oggetto strutturato semanticamente. Gli oggetti intenzionali e i ruoli causali degli stati dipendono da ciò che gli elementi contengono e da come vengono uniti insieme.

Questo porta a considerare un altro aspetto degli stati mentali, cioè la loro produttività, sistematicità e la struttura in costituenti. Sulla base di ciò che Chomsky aveva chiamato "creatività del linguaggio", Fodor chiama *produttività* la capacità di generare/comprendere un numero infinito di frasi a partire da un numero finito di componenti e di regole di combinazione. Tale capacità sta alla base della possibilità di generare/comprendere enunciati mai uditi in precedenza. Alla produttività si affianca la *sistematicità*: comprendere/produrre una frase è legato alla possibilità di comprenderne/produrne molte altre. Queste due capacità si fondano sulla

---

<sup>37</sup> Fodor, 1990, p. 208-209

struttura in costituenti del linguaggio, ovvero sulla realtà di componenti atomici che sono alla base della costituzione di ogni enunciato e che permettono, ricombinandosi tra loro, la formazione di nuovi enunciati.

Le proposizioni del linguaggio naturale però non hanno solo una sintassi compositiva ma hanno anche una semantica combinatoria e da questa dipendono la produttività e la sistematicità. Se la sistematicità delle capacità linguistiche dipende dal fatto che le proposizioni hanno una semantica combinatoria, il fatto che le proposizioni abbiano una semantica combinatoria dipende dal fatto che sono strutturate in costituenti. Dato che il linguaggio è l'espressione dei pensieri, anche i pensieri devono avere questa natura combinatoria: c'è una certa proprietà che le capacità linguistiche hanno in virtù del fatto che i linguaggi hanno una semantica combinatoria; anche il pensiero ha queste proprietà; quindi anche il pensiero deve avere questa capacità combinatoria.<sup>38</sup>

Ma le proprietà semantiche dei simboli possono essere connesse con le loro proprietà causali solo per mezzo della sintassi. Le rappresentazioni infatti sono formule del linguaggio del pensiero, enunciati mentali la cui sintassi dà luogo alla semantica combinatoria del contenuto che essi veicolano. Questo aspetto può venir spiegato più facilmente utilizzando la metafora del calcolatore. I calcolatori mostrano come connettere le proprietà semantiche dei simboli alle loro proprietà causali. Così se l'occorrenza di un atteggiamento implica l'occorrenza di un simbolo, allora è possibile avere un'idea di come funzioni un dispositivo in grado di connettere le proprietà semantiche a quelle causali dei pensieri. Si può pensare alla struttura sintattica come a una caratteristica astratta della forma di un simbolo. Poiché la sintassi si riduce alla forma e poiché la forma di un simbolo è una determinante del ruolo causale, è possibile immaginare occorrenze di simboli che interagiscono causalmente in virtù delle loro strutture sintattiche. «La sintassi di un simbolo può determinare le cause e gli effetti delle sue occorrenze in modo molto simile a quello in cui la geometria di una chiave determina quale serratura essa sarà in grado di aprire» [Fodor, 1985]. Si sa, dalla logica formale, che alcune delle relazioni semantiche tra simboli possono essere "imitate" dalle loro relazioni sintattiche. Così, la relazione semantica tra due simboli in cui la proposizione espressa da uno è implicata dalla proposizione espressa dall'altro può essere imitata dalle relazioni sintattiche in virtù delle quali un simbolo è derivabile dall'altro. Le operazioni della macchina consistono principalmente nella trasformazione di simboli; quando esegue queste operazioni, la macchina è sensibile solo alle proprietà sintattiche dei simboli e le operazioni che esegue sono limitate alla modificazione della forma dei simboli. La macchina tuttavia è concepita in modo tale che trasformerà un

---

<sup>38</sup> Fodor, 1990, p.226

simbolo in un altro se e solo se i simboli trasformati stanno in certe relazioni semantiche: le relazioni che in un argomento valido, per esempio, conducono dalle premesse alle conclusioni. I computer sono infatti quei tipi di dispositivi in cui il ruolo causale dell'occorrenza di un simbolo rispecchia il ruolo inferenziale della proposizione espressa da esso.

Se la mente è una sorta di computer, è possibile iniziare a vedere come sia possibile una teoria dei processi mentali che abbia successo: una teoria che spieghi come possono esserci relazioni di contenuto regolate e non arbitrarie tra pensieri in connessione tra loro. Evidentemente, tuttavia, affinché questa funzioni, devono esserci rappresentazioni mentali. Nella progettazione del computer il ruolo causale è messo in fase con il contenuto sfruttando il parallelismo tra sintassi di un simbolo e la sua semantica. Ma questa idea non renderebbe la teoria migliore se non ci fossero simboli mentali, entità mentali in possesso di proprietà sia semantiche che sintattiche. Devono esserci simboli mentali perché, in breve, solo i simboli hanno una sintassi e la nostra migliore teoria dei processi mentali ha bisogno di un'immagine della mente come di una macchina guidata dalla sintassi. [Fodor, 1985]<sup>39</sup>

---

<sup>39</sup> F. Ferretti, 2001, p.129

#### 4. JOHN SEARLE

Dopo aver presentato due immagini diverse del linguaggio, quella di Wittgenstein e quella di Fodor, ciascuna delle quali cerca di risolvere problemi legati alla comprensione del linguaggio naturale, e a come essa si articola, l'obiettivo che mi pongo ora è quello di prendere in considerazione il dibattito sull'Intelligenza Artificiale, concentrandomi principalmente sulla questione riguardante la possibilità o l'impossibilità delle macchine di comprendere un linguaggio naturale, come l'italiano o l'inglese.

Le questioni principali che animano l'impresa dell'IA e del dibattito intorno ad essa sono soprattutto incentrate sul quesito: "Una macchina può pensare, cioè può svolgere funzioni cognitive al pari di un cervello umano? I modelli computazionali che realizzano formalmente le attività cognitive tipicamente umane, quali l'apprendimento o il ragionamento, sono condizioni sufficienti per affermare che un computer opportunamente programmato comprende il linguaggio dell'uomo? Esistono, o sono teorizzabili, sistemi artificiali capaci di comprensione del linguaggio naturale?"

I sistemi artificiali attuali sono capaci di alcune prestazioni quali riassumere un testo, rispondere a domande riguardanti argomenti di cui il testo parla, tradurre il testo in un'altra lingua, etc. Un sistema che fa queste cose viene chiamato «di comprensione del linguaggio naturale» per il fatto che è in grado di fare una di queste cose.<sup>40</sup> I sistemi attuali, però, non svolgono le prestazioni in questione allo stesso modo di come le svolge un essere umano: sono imperfetti e sono limitati all'una o all'altra prestazione, ma non è solo a causa di questi limiti che spesso si afferma che i sistemi artificiali non comprendono il linguaggio naturale. Supponiamo di riuscire a costruire un sistema di comprensione molto perfezionato, la cui competenza inferenziale sia molto soddisfacente, cioè un sistema che riesce ad inferire in modo corretto una certa proposizione, o una serie di proposizioni, a partire da un'altra proposizione che gli si è precedentemente data. Ciò che molti continuerebbero a sostenere è che, neanche in questo caso, il sistema comprenderebbe il linguaggio che analizza.

A questo riguardo è fondamentale prendere in considerazione la posizione di John R. Searle, la quale viene esposta nel saggio "Minds, Brains and Programs" pubblicato nel 1980 dalla rivista scientifica americana *Behavioral and Brain Sciences*. Ciò che Searle pubblica in questa rivista è quello che viene chiamato l'esperimento mentale della "Stanza cinese", al quale seguono i commenti di filosofi sia a favore che a sfavore della tesi principale sostenuta in tale pubblicazione, cioè l'impossibilità che una macchina possa comprendere il linguaggio naturale sulla base della sola

---

<sup>40</sup> D. Marconi, *Competenza Lessicale*, Ed. Laterza, Roma-Bari, 1999, p.160

elaborazione sintattica di simboli; questo perché il sistema è un mero manipolatore di simboli che non sa e non può sapere in nessun modo che cosa essi significano.

Cercherò di mostrare in cosa consiste l'esperimento e quali sono i principali argomenti sostenuti da Searle; successivamente, mi concentrerò sulle obiezioni che gli sono state rivolte dai diversi filosofi, psicologi e informatici.

#### **4.1. L'esperimento della Stanza Cinese**

Nell'Abstract dell'articolo del 1980, Searle pone come assunti principali, evidenti e fondanti la sua argomentazione, due affermazioni:

(1) L'intenzionalità negli esseri umani (e animali) è un prodotto di proprietà causali inerenti il cervello. Io penso che questo sia un fatto empirico riguardante le relazioni causali effettive tra i processi mentali ed il cervello. Significa semplicemente che certi processi del cervello sono sufficienti per l'intenzionalità; (2) Istanziare un programma per il computer non è mai di per sé una condizione sufficiente d'intenzionalità. L'assunto principale di questo articolo è diretto a stabilire questa asserzione. Esso viene realizzato mostrandocome un agente umano potrebbe istanziare un programma e non avere tuttavia l'intenzionalità relativa. (Searle, 1980, p. 417)

Queste due affermazioni saranno le fondamenta su cui Searle incentrerà il suo esperimento e da cui dedurrà le sue conclusioni, ma anche i principali bersagli di tutte le obiezioni che seguiranno nei commenti.

L'articolo di Searle si apre con la distinzione tra due modi di intendere l'IA, cioè l'IA «debole» e l'IA «forte». Secondo l'IA debole, il computer è semplicemente uno strumento molto potente utile nello studio della mente, cioè uno strumento che, per esempio, offre la possibilità di formulare ed esaminare ipotesi in modo rigoroso e preciso. Secondo l'IA forte, invece, il computer non è solo uno strumento utile per studiare la mente, ma un computer appropriatamente programmato è realmente una mente, in quanto, se dotato dei giusti programmi possiede capacità e stati cognitivi al pari di un essere umano. Ciò che Searle mette subito in chiaro è che, secondo lui, non c'è nulla da obiettare all'IA debole; le sue obiezioni sono rivolte all'IA forte, cioè alle dichiarazioni secondo cui un computer appropriatamente programmato possiede stati cognitivi e i programmi possano realizzare le capacità umane, come la comprensione.

Searle prende in considerazione il lavoro di Roger Schank come esempio dei programmi di simulazione dei processi mentali umani. Lo scopo



del programma di Schank è quello di simulare la capacità umana di comprendere i racconti, cioè il fatto che, una volta compreso il racconto, si possa rispondere a certe domande anche se alcune informazioni non sono state esplicitamente formulate nel racconto. Affinché questo possa avvenire, le macchine di Schank hanno una “rappresentazione” del genere di informazioni che gli esseri umani hanno circa la storia che si sta raccontando, le quali rendono loro possibile rispondere a tali domande. Schank ritiene che il computer comprenda realmente la storia, che sia realmente capace di rispondere a domande riguardanti il suo contenuto e che ciò che fanno la macchina e il suo programma spieghi la capacità umana di comprendere una storia; al contrario, Searle pensa che queste convinzioni siano del tutto infondate.

A sostegno della sua posizione, Searle presenta il celebre esperimento mentale della stanza cinese. Supponiamo di essere chiusi in una stanza e che da una finestra mi vengano dati, inizialmente, dei fogli scritti in cinese. Io non conosco il cinese, non so assolutamente che cosa i simboli scritti sui fogli significano e non so neanche che si tratti di cinese. Successivamente, assieme ad altri fogli scritti in cinese, mi vengono dati anche dei fogli con scritti sopra delle regole scritte in inglese, utili per mettere in relazione il primo pacco di fogli con il secondo. Le regole, in quanto scritte in inglese, le capisco e perciò sono in grado di mettere in relazione i simboli del primo pacco con quelli del secondo, «[...] e "formale" qui significa soltanto che posso identificare i simboli interamente attraverso le loro forme». <sup>41</sup> Infine ci viene dato un terzo insieme di simboli cinesi, con le relative istruzioni in inglese per collegare questo terzo pacco di fogli con i primi due, istruendoci su come utilizzare certi simboli in risposta a quelli dati nel terzo gruppo. Quello che noi non sappiamo è che questa successione di simboli rappresenta una successione di eventi che caratterizza una certa situazione, ovvero l’atto di comprendere una storia e di rispondere a delle domande su di essa: «[...] le persone che mi danno tutti questi simboli chiamano il primo pacco di fogli “uno script”, chiamano il secondo “una storia” e il terzo “quesiti”. Inoltre, chiamano i simboli che rendo loro in risposta al terzo plico “risposte alle domande” e la serie di regole in inglese che mi hanno dato la chiamano “programma”». <sup>42</sup>

Si può presumere che, dopo un po’, io diventi così bravo nel manipolare questi simboli e nel seguire le istruzioni che i programmatori scrivono, che ad un osservatore esterno le mie risposte risulterebbero assolutamente indistinguibili da quelle di un madrelingua cinese. Se mi venisse dato un testo di una storia nella mia lingua madre e delle domande ad esso relative, per un osservatore esterno le mie risposte sarebbero equivalenti

---

<sup>41</sup> Searle, 1980, p. 418

<sup>42</sup> Searle, 1980, p. 417

a quelle date in cinese. Searle rileva però una differenza fondamentale a questo riguardo: «nel caso del cinese, io mi comporto semplicemente come un computer: eseguo operazioni calcolabili su elementi specificati formalmente. [...] Io sono semplicemente un'istanziamento di un programma del computer» (*Ibidem*). Le risposte sono prodotte eseguendo operazioni computazionali su elementi determinati formalmente; nel caso del cinese, noi mettiamo in atto semplicemente un programma, cioè il nostro comportamento è equivalente a quello di un computer. Secondo Searle, quindi nel caso del cinese l'uomo non comprende, mentre nell'altro caso sì e siccome nel caso del cinese il soggetto si comporta come un computer, ne deriva che un computer non comprende.

Secondo l'IA forte, un computer programmato capisce le storie e il programma spiega la comprensione dell'uomo. Ma alla luce di quanto detto, si può affermare che un computer col programma di Schank non comprende affatto nessuna storia, vi sono solo immissioni ed emissioni di dati che non si distinguono da quelle di un cinese. Il computer è come un uomo che agisce manipolando simboli secondo regole, ma questo non è una condizione sufficiente affinché avvenga comprensione. Il computer e il suo programma funzionano ma non c'è comprensione. Ciò che viene sostenuto dall'IA forte è che, quando capisco una storia, quel che faccio è semplicemente manipolare simboli formali e ciò viene sostenuto sulla base della supposizione che possiamo costruire un programma che ha le stesse immissioni ed emissioni di dati (input e output) dei nativi di madrelingua e assumendo che i parlanti possono essere descritti, a un certo livello, come istanze di un programma. In ogni caso l'esperimento mentale suggerisce che la sola esecuzione di un programma, cioè un'operazione computazionale su elementi definiti formalmente, non basta per produrre il processo intenzionale della comprensione.

Un programma di computer è semplicemente irrilevante per quanto riguarda la mia comprensione della storia. Nel caso della storia in cinese, ho tutto ciò che l'IA può mettere in me per mezzo di un programma, ma io non capisco nulla; nel caso della storia in inglese capisco tutto e non c'è alcuna ragione al mondo per supporre che il mio grado di comprensione ha qualcosa a che vedere con programmi per il computer, cioè con operazioni di calcolo su elementi specificati in modo puramente formale. Finché il programma è definito in termini di operazioni computazionali basati su elementi definiti solo formalmente, quello che l'esempio suggerisce è che questi, di per sé, non hanno alcuna connessione interessante con il comprendere in sé e per sé. Sono certamente condizioni non sufficienti e non c'è la minima ragione per supporre che siano condizioni necessarie o perfino che diano un minimo contributo

significativo al comprendere. Si noti che la forza di tale assunto non è semplicemente che macchine diverse possano avere lo stesso input o output mentre operano con principi formali diversi: non è assolutamente questo il punto. Piuttosto, qualunque principio puramente formale si mette nel computer, non sarà sufficiente per la comprensione, dato che nessun essere umano sarà in grado di seguire i principi formali senza capire nulla. (Searle, 1980, p. 418)

Secondo Searle, la comprensione può avvenire solo se, al livello della manipolazione sintattica degli elementi, si affianca il possesso di una semantica relativa agli stessi, una semantica che però rimane indipendente dall'area sintattica (come si vedrà in seguito). Noi comprendiamo i simboli non nel momento in cui siamo in grado di raggrupparli secondo certe regole, ma quando sappiamo cosa essi significano. Sulla base di questo assunto, la comprensione non può venir attribuita ad una macchina che opera secondo processi computazionali su elementi formali, cioè ad una macchina la cui operatività è definita come la messa in atto di un programma. Un uomo è in grado di capire una lingua ed avere intenzionalità perché è un organismo dotato di una certa struttura biologica, che produce causalmente percezione, azione, comprensione e apprendimento, etc. Solo ciò che è dotato di tali "poteri causali" può produrre intenzionalità;<sup>43</sup> un modello formale non sarà mai sufficiente di per sé per riprodurre l'intenzionalità, perché essa è un fenomeno biologico.<sup>44</sup>

Ciò che nega Searle non è che una macchina possa pensare: noi infatti siamo delle macchine pensanti. Se assumiamo che sia possibile produrre una macchina dotata di un sistema nervoso come il nostro, allora anche una macchina può pensare. Un computer digitale può pensare, se per "computer digitale" intendiamo una macchina che ha un livello di descrizione che può venir definito con l'istanziamento di un programma di computer (noi umani infatti siamo l'istanziamento di un certo numero di programmi e possiamo pensare)<sup>45</sup>. Ciò che nega Searle è che una macchina possa pensare, o comprendere, solo in virtù del fatto di essere un computer opportunamente programmato: l'esecuzione formale di un programma non è condizione sufficiente per la comprensione e la semplice manipolazione di simboli non riguarda nessun aspetto dell'intenzionalità e perciò non basta per spiegare la comprensione. L'esempio della stanza cinese «[mostra] che non appena mettiamo qualcosa nel sistema che realmente ha intenzionalità (un uomo), e poi lo programmiamo col programma formale, si può constatare che il

---

<sup>43</sup> Searle, 1980, p. 422

<sup>44</sup> I concetti di 'intenzionalità' e di 'potere causale' saranno concetti che rimarranno indefiniti per tutto il saggio e questa mancanza di specificità e di precisione nel trattarli sarà fonte di diverse critiche, come si vedrà nei paragrafi successivi

<sup>45</sup> Ibidem

programma formale non porta alcuna intenzionalità aggiuntiva»<sup>46</sup>. Ciò che viene mostrato, quindi, è che la semplice esecuzione di un programma non è condizione sufficiente per la comprensione, poiché il programma può essere perfettamente eseguito senza che vi sia nessuna comprensione.

Precisamente quella caratteristica dell'IA che sembrava così seducente – la distinzione tra programma e la realizzazione – risulta fatale alla tesi che la simulazione potrebbe essere un duplicato. La distinzione tra il programma e la sua realizzazione sembra essere parallela alla distinzione tra il livello delle operazioni mentali e il livello delle operazioni del cervello. E se potessimo descrivere il livello delle operazioni mentali come un programma formale, allora potremmo descrivere che cosa è essenziale alla mente senza fare né psicologia introspettiva né neurofisiologia del cervello (Searle, 1980, pp. 422,423).

Secondo Searle, l'errore che compie l'IA risiede nella confusione tra simulazione e duplicazione. Simulare un processo come quello della comprensione, per il quale sono sufficienti un input e un programma che lo trasforma nell'output corrispondente, è diverso dal mettere in atto il vero e proprio processo intenzionale della comprensione, cioè i suoi meccanismi – che non conosciamo – e i suoi effetti. La distinzione tra programma e realizzazione, secondo Searle, ha come effetto che lo stesso programma potrebbe avere ogni genere di «folli realizzazioni» che non hanno alcun tipo di intenzionalità. A prescindere da come viene realizzato il programma che comprende la storia in cinese, esso non comprenderà mai il cinese, perché qualsiasi altro materiale diverso dal cervello non può possedere intenzionalità: «[...] solo qualcosa che ha gli stessi poteri causali del cervello può avere intenzionalità»<sup>47</sup> ed il computer non può in alcun modo simulare le operazioni mentali, poiché per la simulazione ciò di cui si ha bisogno è semplicemente il giusto input o output e un programma che trasforma il precedente input nel seguente output. E questo è tutto ciò che il computer ha, qualunque cosa esso faccia.

L'errore dell'IA forte è, quindi, quello di pensare che «la mente sta al cervello come il programma sta all'hardware».<sup>48</sup> Questo assunto non è sostenibile per almeno tre motivi: 1) distinguere il programma dalla sua realizzazione nell'hardware vuol dire ammettere che per un certo programma è possibile qualsiasi realizzazione, anche attraverso ciò che è totalmente diverso dal cervello, ovvero attraverso ciò che non possiede intenzionalità ed è inadatto a simularla; 2) un programma è formale, mentre gli stati

---

<sup>46</sup> Ibidem

<sup>47</sup> Searle, 1980, p. 423

<sup>48</sup> Searle, 1980, p. 423

intenzionali sono definiti sulla base del loro contenuto; 3) gli stati e gli eventi mentali sono prodotti di operazioni del cervello, mentre un programma non è il prodotto del computer ma del programmatore.

Ci sono ancora due ragioni per cui l'IA ha creduto di poter riprodurre e spiegare i fenomeni mentali. La prima riguarda una confusione intorno alla nozione di trattamento dell'informazione: nel senso in cui gli uomini riflettono, leggono e rispondono a domande su una storia, etc, il computer programmato non fa trattamento di informazione, ma manipola simboli formali. Il programmatore è l'interprete dall'output ed usa simboli come sostitutivi degli oggetti del mondo esterno, ma questo non fa parte degli obiettivi del computer, il quale possiede una sintassi ma non una semantica. Il computer quindi non compie un'elaborazione di informazioni, ma si limita ad operare su simboli che non hanno significato.

O costruiamo la nozione di trattamento dell'informazione in modo tale che essa implichi l'intenzionalità come parte del processo o non lo facciamo. Nel primo caso, il computer programmato non tratta l'informazione, ma manipola solo simboli formali. Nel secondo caso, sebbene il computer esegua un trattamento d'informazione, lo fa solo nel senso in cui le macchine calcolatrici, le macchine da scrivere, lo stomaco, i termostati, i temporali e gli uragani trattano le informazioni; precisamente, essi hanno un livello di descrizione per cui possiamo accettarli come capaci di assumere informazione da una parte, trasformarla e produrre informazioni come output. In questo caso, spetta ad osservatori esterni interpretare l'input e l'output come informazioni nel senso ordinario del termine. E tra il computer e il cervello non si stabilisce alcuna similarità in termini di similarità nel trattamento dell'informazione. (Searle, 1980, p. 423)

La seconda ragione riguarda, invece, la presenza di residui di comportamentismo e operazionalismo, cioè l'idea che, poiché i computer opportunamente programmati possono avere modelli di input-output simili a quelli dell'uomo, questo ci permette di postulare stati mentali nel computer simili agli stati mentali dell'uomo. Tale operazionalismo è congiunto a residui di dualismo, secondo cui, laddove si ha a che fare con la mente, il cervello non c'entra; per l'IA forte ciò che importa sono i programmi i quali sono indipendenti dalla loro realizzazione nelle macchine. Questa concezione dualistica del rapporto tra mente e cervello non tiene presente che la mente umana è intrinsecamente connessa al cervello e che, quindi, non è possibile riprodurre i fenomeni mentali umani utilizzando programmi non intrinsecamente connessi all'hardware.

Searle conclude il saggio domandandosi ancora una volta: “Potrebbe una macchina pensare?”. La sua risposta è che una macchina che può pensare è precisamente quella che è dotata dei poteri causali del cervello. L'intenzionalità è un fenomeno biologico e quindi dipende causalmente dalla biochimica del cervello e la capacità causale del cervello di produrre intenzionalità non può consistere nell'istanziamento di un programma computazionale, dal momento che quest'ultimo può essere messo in atto in qualunque hardware senza che esso acquisti alcuno stato mentale. In qualsiasi modo il cervello produca intenzionalità, non lo fa solo esemplificando un programma poiché nessun programma è di per sé sufficiente per l'intenzionalità.

Naturalmente il cervello è un computer digitale. Poiché tutto è un computer digitale, anche il cervello lo è. Il fatto è che la capacità causale del cervello di produrre intenzionalità non può consistere nell'istanziamento di un programma di computer, poiché per qualunque programma si voglia, è possibile che qualcosa istanzi tale programma e tuttavia non abbia, per questo, alcuno stato mentale. Qualunque cosa faccia il cervello per produrre intenzionalità, questa non può consistere nell'istanziare un programma, poiché nessun programma, di per sé, è sufficiente per l'intenzionalità. (Searle, 1980, p. 424)

#### **4.2. Auto-obiezioni di Searle**

Searle, nel suo saggio, ipotizza e confuta alcune obiezioni che potrebbero venir mosse alla sua tesi, tentando di dimostrarne la validità argomentativa. Le obiezioni che gli vengono mosse sono suddivise come segue:

##### *1. La risposta del sistema*<sup>49</sup>

Un'obiezione che può venir mossa all'esperimento della stanza cinese consiste nel considerare l'individuo, chiuso nella stanza, il quale non capisce la storia, come parte di un intero sistema e sostenere che tale sistema effettivamente comprende la storia. Secondo questa obiezione, il comprendere non si attribuisce al singolo individuo ma all'intero sistema di cui esso fa parte. Come replica Searle? Secondo lui, anche supponendo che l'individuo possa interiorizzare tutti gli elementi del sistema, memorizzare regole e dati, svolgere mentalmente tutte le operazioni egli continuerebbe a

---

<sup>49</sup> Searle, 1980, p. 419

non comprendere nulla di cinese. E così anche il sistema, poiché in questa ipotesi non c'è nulla nel sistema che non sia anche nell'individuo. Secondo il punto di vista di chi sta obiettando, mentre l'uomo non capisce il cinese come un madrelingua, l'uomo come sistema di manipolazione di simboli comprende realmente il cinese. Per Searle, l'errore risiede nel porre come simili due "sottosistemi" nell'uomo: il sottosistema che comprende l'inglese e quello che comprende il cinese. Questi, in realtà, non hanno nulla in comune, poiché il primo ha capacità referenziale, cioè conosce ciò di cui parla la storia e può rispondere alle domande operando deduzioni, mentre il secondo è solo in grado di associare simboli privi di significato, cioè sa solo che squigglesquiggle è seguito da squogglesquoggle. L'esperimento mentale della stanza cinese ha come obiettivo il sostenere che tale manipolazione di simboli non è sufficiente di per sé per capire il cinese, cioè per comprendere un linguaggio naturale «perché l'uomo potrebbe scrivere squogglesquoggle dopo squigglesquiggle senza capire nulla di cinese». Postulare sottosistemi non aiuta poiché essi non sono in condizioni migliori di quanto lo sia un uomo; essi non hanno nulla di simile a ciò che ha l'uomo di madrelingua inglese. Ma per quali ragioni si può affermare che l'agente deve avere in sé un sottosistema che capisce la storia in cinese? Nell'esempio il soggetto ha lo stesso input e output dei nativi cinesi e un programma che li collega, ma il vero scopo dell'esempio era mostrare come questa condizione fosse insufficiente per capire le storie in cinese nello stesso senso in cui le si capisce in inglese, poiché una persona, e i sistemi che concorrono a costituire una persona, potrebbero sì avere la giusta combinazione di input, output e programma, ma senza capire comunque nulla. La motivazione che ci porta a supporre che il soggetto sia dotato di due sistemi è che esso possiede un programma e che sia in grado di superare il test di Turing, ma l'esempio mostra che i due sistemi possono sì superare tale prova, ma che solo uno dei due può comprendere. L'ipotetica obiezione che potrebbe venir rivolta a Searle condurrebbe a delle conseguenze assurde: se si concludesse che il soggetto può comprendere perché riceve input, emette output e utilizza un programma per combinarli, allora si potrebbe inferire che ogni sorta di sistema può essere considerato come cognitivo, così come lo stomaco, o i termostati. Ciò che deve fare l'IA, è distinguere i principi in base ai quali la mente opera da quelli in base ai quali operano i sistemi non mentali. Tale distinzione tra mentale e non mentale non può essere evidente solo agli occhi dell'osservatore, ma dev'essere intrinseca al sistema, se no il rischio sarebbe quello di trattare il mentale e il non mentale in modo arbitrario e secondo principi puramente soggettivi all'osservatore. Searle chiude questa parte confutando ancora una volta la tesi dell'IA forte secondo cui le macchine possiedono stati cognitivi dello stesso tipo dei nostri. Se l'IA forte sostiene questo genere di tesi, allora non può definirsi una scienza della mente poiché

nega il principio stesso su cui lo studio della mente si fonda, cioè il fatto che gli esseri umani hanno idee e credenze, mentre i termostati e le macchine in generale non li hanno

## 2. *La risposta del robot*<sup>50</sup>

Supponiamo di lavorare con un programma diverso da quello di Schank: mettiamo un computer all'interno di un robot, il quale non si limita a ricevere ed emettere simboli formali come input e output, ma fa funzionare il robot in modo che esso si comporti in modo molto simile, nel percepire, camminare, mangiare, etc, ad un essere umano. Questo robot è dotato di una telecamera che gli permette di “vedere”, di braccia e di gambe che gli permettono di “muoversi”, e il tutto sarebbe controllato dal computer. Tale robot avrebbe la capacità di comprendere e altri stati mentali.

Searle nota che ciò che in tale replica viene tacitamente concesso è che la cognizione non riguarda solo la manipolazione di simboli, ma anche «un insieme di relazioni causali inerenti al mondo esterno». Tuttavia, l'aggiunta di queste capacità percettive e motorie non aggiunge nulla alla comprensione, all'intenzionalità e al programma di Schank: lo stesso esperimento mentale si può applicare al caso del robot. Se al posto del computer interno al robot mettiamo l'uomo della stanza cinese, vediamo che il soggetto non capirebbe comunque nulla, nonostante i simboli cinesi e le regole in inglese vengono immessi per mezzo di una telecamera interna al robot ed elaborati per attuare movimenti. Tutto ciò che risulterebbe sarebbe solo una manipolazione di simboli formali: il soggetto riceverebbe informazioni dall'apparato “percettivo” del robot e distribuirebbe istruzioni al suo apparato motorio senza conoscere nessuno di questi fatti.

Non si produce alcuna intenzionalità istanziando un programma: tutto ciò che si è ottenuto è che il robot segua formalmente delle istruzioni formali.

## 3. *La risposta del simulatore del cervello*<sup>51</sup>

Supponiamo di mettere a punto un programma che non rappresenti l'informazione che abbiamo sul mondo ma simuli le sequenze dei collegamenti neuronali e la sinapsi del cervello di un cinese madrelingua nell'atto di comprendere una storia in cinese. Il programma simulerebbe la struttura formale del suo cervello quando comprende e quando risponde a domande sulla storia. Si può anche immaginare che la macchina lavori con più programmi operanti in parallelo, così come opera presumibilmente il

---

<sup>50</sup> Searle, 1980, p.420

<sup>51</sup> Searle, 1980, p. 420-21



cervello umano. Ora, si potrebbe affermare che la macchina capisce le storie e se negassimo questo si dovrebbe sostenere che anche il cinese non le capisce. Cosa distinguerebbe, a livello delle sinapsi, il programma del computer e il programma del cervello cinese?

Secondo Searle, l'IA forte sostiene che non è necessario conoscere come sia fatto un cervello umano per capire come opera la mente: ciò che interessa delle operazioni mentali sarebbero quei processi computazionali che operano su elementi formali, realizzabili dai processi cerebrali allo stesso modo in cui un programma può essere realizzato in diversi hardware. Secondo Searle, invece, il funzionamento del cervello non è ancora sufficiente per giustificare la comprensione. Immaginiamo un uomo che, invece di operare su simboli in cinese, operi su un sistema di tubi di acqua interconnessi, di modo che ciascuna giuntura dei tubi corrisponde ad una sinapsi del cervello: quando egli riceve input in cinese, sulla base delle regole del programma (scritte in inglese), apre e chiude le valvole che connettono i tubi ed escono le risposte in cinese dalla parte terminale della serie di tubi. Ora, questo sistema ha la capacità di comprendere? Certo che no, esso sta solo simulando la struttura formale del cervello, manipolando simboli come input e output. L'uomo non comprende, i tubi non comprendono e nemmeno l'unione dell'uomo e del sistema di tubi può possedere la capacità di comprendere. Il limite di questo simulatore è la riproduzione della sola struttura formale della sequenza delle aggregazioni neuronali e delle sinapsi, tralasciando ciò che è realmente essenziale al cervello, cioè «le sue proprietà causali, la sua capacità di produrre stati intenzionali».

#### 4. *La risposta della combinazione*<sup>52</sup>

Le tre precedenti risposte, prese insieme, possono risultare più convincenti di quanto lo possono essere se prese singolarmente. Immaginiamo un robot dotato di un computer che simula l'attività del cervello. Questo robot si comporta in modo indistinguibile da un essere umano e lo si deve pensare come un sistema unificato e non solo come un computer con input e output. In questo caso, non ci sarebbero problemi ad attribuire intenzionalità al sistema.

Searle è d'accordo che in casi di questo tipo si sarebbe tentati di attribuire intenzionalità al robot: per ascrivere stati intenzionali, in questo caso, basta l'apparenza del comportamento, gli altri elementi della combinazione sono inessenziali; se si costruisse un robot il cui comportamento è uguale a quello degli uomini, noi attribuiremo ad esso

---

<sup>52</sup> Searle, 1980, p.421

intenzionalità, fino a prova contraria. Ma questo non avvalorava la tesi dell'IA forte, secondo cui «istanziare un programma con i giusti input o output è una condizione sufficiente, anzi costitutiva, dell'intenzionalità». In questo esempio, però, l'attribuzione di intenzionalità al robot non ha a che fare con i programmi formali, ma si basa solo su condizioni comportamentali, cioè sul fatto che il robot si comporta come un uomo: noi supponiamo che, fino a prova contraria, il robot possieda stati intenzionali come i nostri che causano il comportamento e che esso possieda un meccanismo interno capace di produrre tali stati. Se però sapessimo che la macchina opera in questo modo, seguendo un programma meramente formale, non saremo di certo pronti ad attribuirgli stati intenzionali.

Ci sono casi in cui noi attribuiamo intenzionalità a certi animali, come le scimmie o animali domestici, come i cani. Le ragioni per cui lo facciamo risiedono sia in criteri comportamentali che in criteri fisiologici, cioè, non sappiamo spiegare il loro comportamento senza pensare che possiedano intenzionalità e siano fatti dello stesso nostro materiale causale. Noi ipotizziamo le stesse cose per il robot, ma non appena veniamo a conoscenza che il suo comportamento è il risultato di un programma formale e che le proprietà causali della sostanza fisica sono irrilevanti, siamo costretti ad abbandonare l'assunto dell'attribuzione dell'intenzionalità.

##### 5. *La risposta delle altre menti*<sup>53</sup>

Come sappiamo che le altre persone comprendono? Solo per mezzo del loro comportamento. Ora, il computer può anch'esso superare il test di Turing, quindi, se si vuole attribuire intenzionalità e capacità conoscitiva alle persone, lo si deve attribuire anche al computer.

Ma per Searle il problema in questione non verte sul come io so che le persone hanno stati cognitivi, ma su che cosa io attribuisco quando li accredito di stati cognitivi. Caratteristico dello stato cognitivo non può essere solo il procedimento computazionale e i suoi input e output, poiché essi esistono anche in assenza degli stati cognitivi.

##### 6. *La risposta delle molte dimore*<sup>54</sup>

L'argomento di Searle presuppone che l'IA si occupi solo dei computer digitali attuali. Ma, a prescindere da quali procedimenti causali si ritengono essenziali per l'attribuzione dell'intenzionalità, prima o poi si sarà in grado di costruire dispositivi capaci di quei procedimenti causali e questo

---

<sup>53</sup> Searle, 1980, p.421-22

<sup>54</sup> Searle, 1980, p. 422

si chiamerà Intelligenza Artificiale. Per Searle, però, questo non fa che banalizzare il progetto dell'IA forte la cui tesi originaria è che i procedimenti mentali sono processi computazionali che operano su elementi definiti formalmente. È a questa tesi che Searle rivolge le sue obiezioni. Se la tesi viene ridefinita in modo tale che risulta essere diversa da quella precedente, le sue obiezioni non sarebbero più valide in quanto non ci sarebbero più ipotesi da applicare ad essa.

### 4.3. Critiche e commenti al saggio di Searle

Il dibattito che segue la pubblicazione del saggio di Searle mette in luce i diversi problemi che studiosi di vari ambiti ritengono centrali in merito ai fenomeni dell'intelligenza e della comprensione, specialmente in riferimento al problema relativo alla possibilità che qualcosa costruito artificialmente possa avere capacità conoscitiva e stati intenzionali.

Gli studiosi che si dichiarano d'accordo con Searle sono relativamente pochi rispetto a coloro che si pongono in contrasto: essi si dichiarano favorevoli all'idea searlina secondo cui una macchina non potrà mai essere in grado di realizzare la comprensione, in quanto capace solo di eseguire formalmente programmi e mancante della componente essenziale dell'intenzionalità. La maggior parte dei commenti invece, tendono ad evidenziare le lacune, le inadeguatezze e le fallacie presenti nell'argomentazione di Searle. Alcuni autori lo accuseranno di aver scelto formule espressive ambigue e imprecise, volte a suscitare provocatoriamente nei lettori emozioni e consensi intuitivi, piuttosto che stimolarli in riflessioni e ragionamenti. Dennett, per esempio, parla di "*intuition pump*"<sup>55</sup>, generatore di intuizione, una strategia argomentativa che mette il lettore nelle condizioni di vedere le cose dal punto di vista di chi le scrive, impedendogli di notare le contraddizioni e le debolezze dell'impostazione teorica a favore di considerazioni intuitive e a scapito del rigore logico.

Le critiche che vengono mosse a Searle riguardano moltissimi argomenti, i quali però non verranno approfonditi tutti allo stesso modo: alcuni verranno solo menzionati senza ulteriori approfondimenti, altri verranno ulteriormente approfonditi e precisati in quanto utili all'elaborazione conclusiva del tema principale di questo scritto. Gli argomenti a cui seguirà un'elaborazione più dettagliata saranno quelli relativi all'importanza dell'applicazione delle regole, alla dicotomia che pone Searle tra semantica e sintassi e al concetto di intenzionalità.

---

<sup>55</sup> Dennett in "Open Peer Commentary", 1980, p. 429

I commenti che vengono fatti dai diversi autori della sezione “Open Peer Commentary” possono venir schematizzati come segue:

1. La critica di Searle non è pertinente a ciò che in realtà l’IA afferma – secondo Minsky<sup>56</sup>, Fodor<sup>57</sup>, Marshall<sup>58</sup> e Menzel<sup>59</sup>, Searle non avrebbe compreso appieno gli obiettivi e la metodologia dell’IA. Egli accusa l’IA di prestare poca attenzione al fatto che gli esseri umani hanno delle credenze, mentre i termostati, i telefoni e i calcolatori non ce le hanno, e che lo studio della mente inizi partendo da questo fatto;
2. Searle tratta un problema empirico come se fosse a priori – secondo Lycan<sup>60</sup>, Ringle<sup>61</sup>, Schank<sup>62</sup> e Haugeland<sup>63</sup>, Searle tratta ciò che è

---

<sup>56</sup> «Lo studio della mente non è lo studio delle credenze; è il tentativo di scoprire i concetti potenti che aiutano a spiegare perché le macchine e gli animali possono fare molte cose che altri non possono fare». Secondo Minsky, i concetti quali credenza e intenzionalità sono troppo vaghi per rientrare nella sfera della scienza della mente ed una teoria scientifica non può dare spazio a concetti imprecisi e approssimativi come questi. [Minsky, 1980, p. 439]

<sup>57</sup> Ciò che si propone l’IA è una riduzione del problema “Che cosa rende gli stati mentali intenzionali” al problema “Che cosa conferisce le proprietà semantiche ad un simbolo”. Gli stati mentali sarebbero da considerare intenzionale in quanto oggetti interpretati: «Se ci sono delle rappresentazioni mentali esse devono essere oggetti interpretati: è perché esse sono oggetti interpretati che quegli stati mentali sono intenzionali». [Fodor, 1980, p. 431]

<sup>58</sup> L’IA modella gli stati cognitivi in una sintassi formale, cerca per ogni distinzione semantica una corrispondente distinzione sintattica che la possa rappresentare a livello formale. L’obiettivo dell’IA non è costruire macchine simili all’uomo, ma riprodurre le strutture mentali da un punto di vista sintattico [Marshall, 1980, p. 436]

<sup>59</sup> «L’area dell’IA differisce da quella dell’intelligenza naturale in almeno tre aspetti. Primo, nell’IA si è limitati necessariamente all’uso di dati comportamentali formalizzati o “output” come base per fare inferenze circa il suo soggetto. [...] Secondo, nell’IA si deve assumere che, fino a prova contraria, il soggetto non ha più intenzionalità di una roccia, mentre nell’area dell’intelligenza naturale spesso ce la si cava con l’assunzione opposta, cioè che, fino a prova contraria, il soggetto può essere considerato senziente. Terzo, nell’IA l’analisi è ordinariamente limitata al problema della struttura dell’intelligenza, mentre una completa analisi dell’intelligenza naturale deve anche considerare i problemi riguardanti le funzioni, lo sviluppo e l’evoluzione». [Menzel, 1980, p. 438]

<sup>60</sup> Le argomentazioni di Searle non impediscono di pensare che in futuro sarà possibile attribuire stati mentali alle macchine, cioè a macchine che non replicano solamente l’organizzazione funzionale della macchina, ma che hanno rappresentazioni risultanti da processi causali che interagiscono con l’esterno. Questo sarà possibile, per Lycan, quando si sarà compreso in virtù di che cosa noi attribuiamo intenzionalità ai nostri stati cognitivi. Il fatto che i computer di oggi non possono avere stati intenzionali in virtù delle operazioni sintattiche che si eseguono su elementi caratterizzati formalmente, non è di per sé una condizione sufficiente per affermare che i computer non potranno mai avere stati intenzionali. [Lycan, 1980, p.435]

<sup>61</sup> Il problema, secondo Ringle, è che la disputa tra Searle e l’IA non ha nessun interesse filosofico poiché i problemi relativi alla possibilità di riprodurre il fenomeno dell’intenzionalità grazie all’implementazione di un programma, sono problemi che possono essere stabiliti soltanto empiricamente [Ringle, 1980, p. 444]

<sup>62</sup> «Possiamo sperare di ottenere programmi che comprendano? Possiamo creare la “vita”. Queste dopo tutto, sono questioni empiriche». [Schank, 1980, p. 447]

<sup>63</sup> «Non credo che ci sia una confusione concettuale nel supporre che i giusti poteri causali dell’intenzionalità originaria sono quelli che si catturerebbero specificando un programma (cioè, una macchina virtuale). [...] Mi sembra che la questione interessante sia molto più empirica: dato che i programmi sono il giusto modo per esprimere le strutture causali rilevanti, essi sono di fatto così? [...] Non mi curo del demone di Searle che lavora attraverso

prettamente empirico, cioè la possibilità della scienza di costruire artificialmente la comprensione, come se fosse decidibile con metodi a priori e che, quindi, l'impossibilità di replicare gli stati cognitivi umani è a priori impossibile

3. Searle ha sottovalutato l'importanza dell'applicazione delle regole – secondo Abelson, Bridgeman, Schank e Smythe, Searle trascura il ruolo dell'applicazione delle regole nel processo della comprensione, cioè non riconosce come aspetto fondamentale della comprensione l'applicazione delle regole<sup>64</sup>;
4. Searle ritiene erroneamente valida l'analogia tra l'uomo nella stanza e il computer – alcuni autori come Bridgeman<sup>65</sup> e Block contestano la validità dell'analogia che Searle ha istituito tra il cervello dell'uomo e il computer, sostenendo che il primo non differisce affatto dal secondo dal punto di vista delle funzionalità o capacità di manipolare le informazioni;
5. Searle ritiene erroneamente valida l'analogia dell'uomo nella stanza con l'uomo reale – alcuni autori ritengono che l'uomo nella stanza non possa venir assolutamente paragonato all'uomo reale perché a) ciò che fa l'essere nella stanza cinese verrebbe realmente fatto dall'essere umano in tempi estremamente più lunghi [Hofstadter, 1980, p. 434], b) sembrerebbe che in ogni uomo esista un singolo soggetto dalle capacità e proprietà misteriose che produce la comprensione e costituisce credenze, quando, al contrario, le caratteristiche essenziali del comprendere sorgono dalla totalità della mente e dalle sue iterazioni con il contesto [Minsky, 1980, p. 439], c) questo essere può essere visto come l'elemento sostitutivo del processore centrale di un computer, un essere "superveloce", una sorta di "demone" che si interpone tra i processi formali e la realizzazione dei processi di comprensione, il quale rende, però, gli stessi poteri causali, di cui si avvale Searle per dare ragione degli stati cognitivi umani, una condizione insufficiente per la comprensione [Haugeland, 1980, p. 432]

---

un programma per la perfetta simulazione di un parlante cinese – non perché non c'è tale demone ma perché non c'è tale programma. O piuttosto, se c'è un tale programma, e se non c'è, perché no, sono dal mio punto di vista, le questioni importanti». [Haugeland, 1980, p. 433]

<sup>64</sup> Questo punto verrà approfondito ed ulteriormente sviluppato nei paragrafi successivi.

<sup>65</sup> Bridgeman sostiene che un computer non è differente da un cervello umano per quanto riguarda il suo funzionamento o le sue capacità, ma per la sua incapacità di evolversi estensivamente, capacità che invece è costitutiva del cervello, non nega la capacità di apprendimento da parte del robot, così come non ne contesta le modalità. Ciò che sottolinea è che il cervello è geneticamente predisposto ad apprendere informazioni dal mondo esterno, grazie alle quali può svilupparsi e modificarsi, mentre il robot dev'essere costruito in un modo specifico e particolare, dotato di idee innate che gli permettano di capire e memorizzare i dati che arrivano dall'esterno. [Bridgeman, 1980, p. 427]

6. Le critiche al concetto di intenzionalità –questo termine rimane indefinito e indeterminato per tutto il saggio, nonostante venga posto da Searle come concetto cardine di tutta la sua argomentazione. Questo diventa motivo di critica da parte di diversi autori che accusano l'autore di utilizzare concetti - come quello di intenzionalità e poteri causali, che non sono affatto da prendere seriamente in ambito scientifico e nello studio della mente. Si vedrà nei paragrafi successivi quali saranno le accuse che i diversi autori muovono a Searle e quali saranno le loro tesi in merito a questi concetti.
7. Alcuni autori come Dennett<sup>66</sup>, Lycan<sup>67</sup>, Mendel, Sloman e Croucher, credono che l'errore di Searle consista nel porre l'analogia tra mente e computer al fine di spiegare la capacità di comprensione del sistema partendo dal livello della sua struttura fisica: chi comprende non è il cervello ma l'uomo, non è la sostanza fisica, ma l'uomo inteso come sistema generale, in relazione ad un contesto generale;
8. Critiche alla contrapposizione tra sintassi e semantica – la separazione tra queste due componenti essenziali della comprensione è il punto cruciale dell'esperimento mentale di Searle. Questa posizione non viene sostenuta da gran parte degli autori degli Open Peer Commentary, i quali, al contrario, non possono che dichiarare la necessaria correlazione tra sintassi e semantica come condizione per la comprensione. La mancata attenzione alla componente sintattica in tale processo cognitivo è un errore ritenuto molto importante e cruciale per la messa in questione della validità dell'esperimento. Nel paragrafo relativo a tale problematica, verranno approfondite critiche e tesi alternative;

---

<sup>66</sup>«Io comprendo l'inglese; non il mio cervello». La capacità di comprensione è propria dell'uomo, non della sua struttura celebrale. Ciò a cui si deve prestare attenzione non è il livello profondo della struttura fisica; la comprensione è un processo da spiegare a livello più generale, considerando tutte le operazioni e le interazioni con l'ambiente esterno all'agente. Nella discussione della replica del robot, Searle mostra di aver confuso questi due livelli, ma Dennett ne condivide i presupposti: la replica del sistema infatti sottolinea come sia necessario considerare i diversi elementi e lavorare ad un livello generale, senza limitarsi al supporto fisico: «Immaginiamo il nostro eroe nella stanza cinese "interiorizzare tutti questi elementi del sistema" così che egli "incorpori l'intero sistema". Il nostro eroe ora non è più una parte sub-personale non comprendente di un supersistema al quale la comprensione del cinese può essere propriamente attribuita, dal momento che non c'è nessuna parte del supersistema esterno alla sua pelle» [Dennett, 1980, p. 429]

<sup>67</sup> Lycan ritiene che se ci si concentra troppo sui dettagli relativi all'hardware si rimane ad un livello ancora troppo basso per capire come il sistema comprende. Nel suo saggio Searle si chiede dove si trovi la comprensione del sistema, ma tutto ciò che vede sono solo gli elementi fisici del sistema. Ciò che suggerisce Lycan è di guardare oltre i dettagli dell'hardware del sistema, perché questi sono troppo piccoli per vedere se il sistema sta comprendendo qualcosa. Se infatti diventassimo degli osservatori minuscoli che entrano nel cervello per scoprire cosa accade quando si comprende, ciò che noi noteremmo sarebbero solo neuroni che trasmettono meccanicamente scariche elettriche, ma ciò non ci illuminerebbe affatto sullo stato cognitivo della comprensione. [Lycan, 1980, p. 435]

9. Critiche alle sue auto-obiezioni – Molti commentatori (Fodor, Bridgeman<sup>68</sup>, Dennett<sup>69</sup>, Sloman e Croucher, Wilensky<sup>70</sup>, Walter<sup>71</sup>, Schank<sup>72</sup>) trovano nelle obiezioni all'esperimento mentale della stanza cinese, l'ispirazione per la loro critica alla posizione del filosofo, i quali, non solo sostengono tali obiezioni, ma le rivolgono direttamente contro la tesi searlina come prove della sua inconsistenza. Inoltre, le obiezioni che Searle muove a se stesso, le quali sono rivolte alla dinamica interna del suo ragionamento, non mettendo mai in questione la dicotomia che egli pone tra comprensione e manipolazione di simboli formali e questo gli permette di confutare ogni obiezione entro i binari del suo schema mentale. Ma è proprio mettendo in dubbio questa rigida divisione ed insistendo sui ruoli complementari delle

---

<sup>68</sup> Fodor e Bridgeman, per esempio, difendono la "risposta del robot": secondo Fodor, non è scontato che al robot non possa venir attribuita l'intenzionalità se esso si comporta in modo simile agli esseri umani. Dal momento che ci sono dei giusti tipi di collegamenti causali tra i simboli che il dispositivo manipola e le cose nel mondo è poco chiaro perché l'intuizione rifiuti l'attribuzione di atteggiamenti proposizionali ai simboli. Tutto ciò che dimostra l'argomentazione di Searle è che i tipi di collegamenti non sono quelli giusti, ma questo non è sufficiente per sostenere la tesi secondo cui l'insufficienza della competenza sintattica si traduce in una contrapposizione tra sintassi e semantica. Noi non sappiamo dire quali siano le giuste relazioni causali e non sappiamo rispondere alle domande su quale sia il giusto tipo di connessione tra una formula e il mondo esterno, ma data una certa rappresentazione mentale è naturale assumere che ciò che rende gli stati mentali intenzionali è che questi siano in relazione con oggetti mentali interpretati semanticamente. [Fodor, 1980, p. 431]

Secondo Bridgeman, un robot simile a quello ideato da Searle, capace di relazionarsi col mondo esterno, funzionerebbe esattamente come un cervello umano e la struttura fisica di questo sarebbe paragonabile al sistema elettrico del robot. Il robot ha solo bisogno di moltissime informazioni e può interiorizzare i significati grazie alle informazioni che riceve sul mondo esterno. Le proprietà del cervello sono genetiche ma nulla impedisce di replicarle nel programma del robot. [Bridgeman, 1980, p. 427]

<sup>69</sup> Dennett, ritiene valida la "risposta del sistema": secondo Dennett, questa è un'obiezione molto importante perché mette in luce come i processi cognitivi siano da studiare ad un livello più generale rispetto al mero livello chimico-biologico. [Dennett, 1980, p.429]

<sup>70</sup> Wilensky ritiene che questa la "risposta del sistema" sia in perfetta sintonia con ciò che afferma l'IA, cioè che la comprensione sia una proprietà del sistema che elabora le informazioni. Non è l'uomo preso singolarmente, ma il sistema intero che dev'essere considerato in grado di comprendere il cinese, viste le sue capacità di elaborazione dei simboli. Un sistema cognitivamente attivo deve quindi paragonarsi all'intero sistema "uomo + stanza" che mette in atto la comprensione. Searle pensa che il termine "comprensione" si riferisca alle proprietà fisiche del fenomeno, nello stesso modo del termine "fotosintesi". Ma, secondo Wilensky, la comprensione non coinvolge le proprietà dell'organismo biologico, ma coinvolge le proprietà logiche e comportamentali. [Wilensky, 1980, p. 450]

<sup>71</sup> Walter sostiene la "risposta della combinazione": se noi dessimo ragione a Searle quando dice che, se l'uomo nell'esperimento mentale non ha la comprensione e così neanche il sistema con cui opera la possiede, allora non la possiederebbe neanche l'unione dell'uomo con il sistema, dovremmo sostenere allo stesso modo che, per esempio, se la retina dell'occhio di un uomo e i suoi nervi ottici, presi separatamente, non vedono, allora neanche l'uomo in quanto tale non vedrebbe, il che è evidentemente falso.

<sup>72</sup> Schank sostiene la "risposta delle altre menti", sottolineando che, così come noi non possediamo nessuna evidenza che ci consenta di dire che gli uomini comprendono, allo stesso modo non abbiamo nessuna evidenza per attribuire la comprensione alle macchine. [Schank, 1980, p. 446]

facoltà semantiche e sintattiche che è possibile mettere in luce il problema di fondo dell'esperimento della stanza cinese.

#### **4.3.1. Searle sottovaluta l'importanza dell'applicazione delle regole**

Un'altra critica mossa al ragionamento di Searle è di trascurare il ruolo dell'applicazione delle regole nel processo della comprensione, cioè di non riconoscere come aspetto fondamentale della comprensione l'applicazione delle regole.

Robert P. Abelson<sup>73</sup> dice infatti che il processo di comprensione consiste nell'applicare regole manipolando simboli. Il fatto di scrivere regole per la trasformazione di simboli input in simboli output non è un aspetto banale dell'esperimento della stanza cinese. Più le regole sono precise, più l'output è convincente e questo è caratteristico della comprensione del contenuto di una storia. Quando un bambino impara a contare cosa fa se non applicare regole? Un esempio di comprensione potrebbe essere che i risultati di un'addizione si applicano indipendentemente dal contenuto, cioè:  $m+n=p$  significa che se si hanno  $m$  cose che si aggiungono ad  $n$  cose, si otterranno  $p$  cose. Ma questa non è altro che una regola. Quando vengono acquisite sempre più regole che riguardano un certo contenuto, e che connettono quel contenuto con altri domini, abbiamo l'impressione di un incremento della comprensione. Quando è che si passa da una "pura" manipolazione di regole a una "vera" comprensione?

Dello stesso avviso è anche Bridgeman, il quale, in riferimento all'esempio della manipolazione aritmetica, utilizzato da Searle per mostrare come gli umani comprendono qualcosa che invece le macchine non possono comprendere, parla di comprensione come mera "illusione cognitiva". Secondo Bridgeman né gli umani né le macchine comprendono i numeri nel senso in cui lo intende Searle.<sup>74</sup> La comprensione di un numero più grande di 5 è sempre un'illusione: gli umani possono avere a che fare con numeri più grandi perché usano "stratagemmi memorizzati" piuttosto che una reale comprensione.

«Se voglio addizionare 27 e 54 [...] io applico regole che ho memorizzato nella scuola elementare senza una reale conoscenza di ciò che significavano, e combino queste regole con fatti memorizzati relativi ad addizioni di numeri ad una cifra per arrivare alla risposta, senza comprendere i numeri stessi. Sebbene io abbia la sensazione

---

<sup>73</sup> Abelson, in "Open Peer Commentary", 1980, p. 424

<sup>74</sup> Bridgeman in "Open Peer Commentary", 1980, p. 427



di star eseguendo operazioni su numeri, guardando agli algoritmi che uso non c'è niente di numerico». (Bridgeman, 1980, p.427)

Anche Schank e Smythe sottolineano l'importanza dell'applicazione di regole formali nel processo di comprensione.

Schank, riferendosi all'esperimento mentale di Searle, ritiene che la comprensione consista nel trovare il sistema che sta dietro i simboli cinesi, cioè se questi sono scritti per il cervello o per il computer. Il linguista che scrive le regole comprende in un senso profondo come la lingua cinese funziona. E le regole che scrive incorporano quella comprensione. Le regole per la manipolazione dei simboli cinesi devono essere viste come un successo. Scrivere delle regole richiede una grande comprensione della natura del linguaggio: regole del genere risolverebbero molti problemi relativi alla filosofia, alla linguistica, alla psicologia e all'IA. Per Searle, chi scrive quelle regole non comprende, e non comprende né la configurazione dell'hardware del computer né quella del cervello. «Chi comprende allora? Di certo la persona che ha scritte le regole. E chi è? E' quello che viene chiamato un ricercatore di IA». <sup>75</sup>

Smythe <sup>76</sup> sottolinea che per Searle la realizzazione computazionale degli stati intenzionali è impossibile a priori e questa tesi si basa su un "gioco di simulazione", il quale dimostra che il seguire le regole, al fine di eseguire certe operazioni sintattiche su simboli senza significato, non è una condizione sufficiente per l'intenzionalità: data una serie di regole si può imitare un'attività intelligente ma senza che l'agente posseda stati intenzionali. Secondo Smythe, la difficoltà di questo argomento risiede nella mancata distinzione tra due modi fondamentalmente diversi di istanziare un programma di computer. Dare istruzioni ad un uomo rinchiuso in una stanza è del tutto diverso da programmare un computer, poiché nel primo caso si inserisce un programma in un sistema che è già in grado di interpretare e di seguire regole, mentre nel secondo si attua un programma come istruzione primitiva dell'hardware. In questo secondo caso le regole sono seguite non interpretandole, ma semplicemente eseguendo la procedura denotata dalla regola.

Un modo consiste nell'inserire il programma in un sistema che è già in grado di interpretare e seguire regole. Questo richiede che le procedure siano espresse in un "linguaggio" che il sistema può già "comprendere". Un secondo modo, consiste nell'istanziare il programma *direttamente* realizzando le sue regole come operazioni primitive dell'hardware. In questo caso, la regola è seguita non

---

<sup>75</sup> Schank in "Open Peer Commentary" 1980, 447

<sup>76</sup> Smythe in "Open Peer Commentary", 1980, p. 448-49

interpretandola, ma solo eseguendo qualsiasi procedura la regola denoti. La simulazione del gioco di Searle è parente del primo tipo di istanziazione, non del secondo. (Smythe, 1980, 448)

Il gioco dell'esperimento di Searle, in cui un uomo imita un computer, non distingue, dunque, due modi diversi di mettere in atto un programma: nel primo senso, il modo in cui agisce il sistema, in cui è stato inserito il programma, è già stato precedentemente programmato, come può essere l'uomo, il cui cervello è programmato geneticamente, mentre nel secondo senso, il programma si attua su un hardware vuoto (la macchina, in principio, è come una tabula rasa) e le istruzioni vengono eseguite semplicemente seguendo la procedura che è stata indicata.

La sfida che lancia Searle all'IA, quindi, è valida quando essa è avanzata in termini di capacità del programma, ma non quando è formulata in termini di sistemi computazionali. L'affermazione che il sistema computazionale non possiede stati intenzionali deve quindi procedere per differenti linee e l'approccio a tale problema richiama l'attenzione sulla relazione arbitraria tra il sistema computazionale e i referenti. Smythe conclude, dunque, che l'esperimento di Searle non mette in luce la distinzione tra i due modi fondamentali di mettere in atto le procedure computazionali e impedisce di notare la differenza tra dare istruzioni ad un computer e darle ad un essere umano, due cose estremamente differenti e contrarie l'una all'altra.

#### **4.3.2. Critiche al concetto di intenzionalità**

Il concetto di intenzionalità per Searle è peculiare della mente umana, ma è anche un concetto che rimane, per tutto il suo saggio, indefinito e molto misterioso, nonostante venga posto come elemento fondamentale della capacità di comprensione, e delle capacità cognitive in generale, dell'essere umano. Questa mancanza di definizione e chiarezza fa sì che l'intenzionalità non possa essere presa seriamente e come concetto realmente scientifico, per lo meno in riferimento al problema che si sta trattando; questo suscita diversi interventi critici, che vanno dall'accusa di misticismo<sup>77</sup>, all'accostamento di intenzionalità e anima<sup>78</sup>, a prese di posizione secondo cui la comprensione è mera illusione cognitiva<sup>79</sup>, o intesa come mezzo per ottenere il controllo del corpo da parte del cervello<sup>80</sup>. Le critiche relative a questo concetto sono quindi dirette ad evidenziare i problemi che sorgono dalla mancanza di

---

<sup>77</sup> Ringle in "Open Peer Commentary", 1980, p. 445

<sup>78</sup> Hofstadter in "Open Peer Commentary", 1980, p. 433

<sup>79</sup> Bridgeman in "Open Peer Commentary", 1980, p. 427

<sup>80</sup> Dennett in "Open Peer Commentary", 1980, p. 429

specificazione e dalle difficoltà di prendere seriamente in considerazione un concetto così vasto e aleatorio.

Abelson e Phylyshyn affermano che vi è una mancanza di definizione dell'intenzionalità e dei suoi poteri causali. Il primo dice, infatti, che per Searle l'ingrediente mancante del computer è il suo concetto di intenzionalità, ma nel suo scritto non giustifica questo assunto. Secondo Abelson sarebbe più ovvio che ciò di cui manca il manipolatore dei simboli cinesi sia la validità estensionale, cioè la conoscenza sensomotoria.<sup>81</sup>

Ma perché invece l'intenzionalità è così importante per Searle? Se noi recitiamo la sua litania, "speranze, paure e desideri", non centriamo il punto. Un computer o un umano non ha certamente bisogno di avere speranze o paure sul cliente al fine di comprendere la storia di una visita al ristorante. E l'uso inferenziale di questi concetti è interno alle capacità dei modelli di comprensione del computer. [...] Il ruolo cruciale dell'intenzionalità per la conoscenza è l'apprezzamento delle condizioni per la sua falsificazione. In che senso il computer realizza che quell'asserzione "John legge il menu" potrebbe o no essere vera, e che ci sono modi nel modo reale per scoprirlo? (Abelson, 1980, 425)

Phylyshyn è della stessa opinione di Abelson e, all'inizio del suo commento, prende in considerazione l'idea searliana secondo cui vi è una mancanza di intenzionalità dei computer dovuta al possesso di differenti poteri causali rispetto a quelli dell'essere umano. Solo i sistemi equivalenti agli esseri umani possono avere intenzionalità. La tesi di Searle è che l'intenzionalità è legata ad una specifica proprietà materiale, infatti, essa è letteralmente causata da questa proprietà. Da questo assunto si deduce quindi che ciò che non ha le proprietà materiali del cervello non può avere intenzionalità. Inoltre, Searle afferma che possedere un programma non è una condizione sufficiente per avere intenzionalità dal momento che si può implementare quel programma in una macchina di Turing realizzandolo con "un rotolo di carta igienica o un mucchio di piccole pietre". Tali macchine non avrebbero intenzionalità perché, in primo luogo, quegli oggetti non sarebbero la giusta materia che può avere intenzionalità.

«Ma qual è il giusto tipo di materia? Sono gruppi di cellule, singoli neuroni, protoplasma, molecole proteiche, atomi di carbonio e idrogeno, particelle elementari? [...] È ovvio in base all'argomento stesso di Searle che la natura della materia non può essere ciò che è

---

<sup>81</sup> «Ignorare che il simbolo che sta per "menu" si riferisce a quella cosa fuori nel mondo che si può tenere, piegare e guardare da vicino, significa mancare una comprensione reale di ciò che si intende per menu». [Abelson, 1980, p. 425]

rilevante, dal momento che si suppone che il parlante monolingue inglese che ha memorizzato le regole formali sia un esempio di sistema fatto della materia giusta eppure - si sostiene - ancora non ha l'intenzionalità pertinente.» (Phylyshyn, 1980, p. 442)

La conclusione di Phylyshyn, quindi, punta a sostenere che non possiamo affermare con nessuna precisione cosa è l'intenzionalità e che l'argomento contro l'intenzionalità dei computer è tipicamente un argomento a partire dall'ignoranza.<sup>82</sup> Secondo l'autore l'intenzionalità non può che essere la capacità di fare riferimenti diretti a oggetti e stati di cose del mondo, una visione che può essere vista come funzionalista, secondo cui vi è un ruolo funzionale che i simboli svolgono nel comportamento generale del sistema. Il comportamento generale del sistema infatti non preclude la specificazione delle sue interazioni causali con l'ambiente.<sup>83</sup>

Anche l'intervento di Ringle parte della poca chiarezza con cui Searle utilizza i suoi concetti fondamentali. Searle fonda il suo diniego nei confronti dell'adeguatezza del modello dell'IA sulla credenza che le proprietà fisiche del sistema neurale siano tali che esse non possano *in linea di principio* essere simulate da un sistema computazionale non protoplasmico. E questo viene chiamato da Ringle misticismo. Per quanto riguarda il concetto di intenzionalità, Ringle afferma che, nonostante Searle lo chiami in causa spesso e volentieri e lo intenda come fenomeno biologico, in realtà non ha mai dato una ragione del perché creda realmente che essa sia una proprietà inerente al sostrato fisico. Per ciò che concerne invece il concetto di poteri causali, Ringle prende in considerazione due modi in cui questi possono essere intesi: si può leggere questo concetto o come diretto collegamento del sistema fisico ai fenomeni fisici del mondo esterno, o come termine che si riferisce alle capacità del sistema neurale di produrre stati fenomenici. Secondo il primo modo, quando un uomo elabora un'immagine visiva, la ricchezza delle informazioni interne risulta da un'interazione fisica col mondo. Quando un computer elabora una scena, non ha bisogno di collegamenti reali tra il fenomeno della luce e le sue rappresentazioni interne: le sue rappresentazioni sono programmi stipulati e si possono inserire rappresentazioni senza che il computer posseda un apparato fisico. Ciò che si potrebbe quindi concludere è che da questo punto di vista, senza un collegamento causale, lo stato interno non può essere visto come stato cognitivo a causa della mancanza di un contenuto semantico indipendente dal programmatore. Ma l'IA potrebbe rimediare alla situazione introducendo

---

<sup>82</sup> Phylyshyn, in "Open Peer Commentary", 1980, p. 444

<sup>83</sup> Il funzionalismo a cui si riferisce Phylyshyn è quello secondo cui ciò che il simbolo rappresenta può essere visto esaminando come il simbolo entra in relazione con gli altri simboli e i con i trasduttori, anche se comunque ammette una certa incompletezza della teoria.

trasduttori e meccanismi appropriati nei sistemi. Ciò però continuerebbe ad venir criticato da Searle poiché continuerebbero a mancare i giusti poteri causali propri degli esseri umani. Ma se la causalità di cui parla Searle implica la diretta connessione tra i processi interni e gli stati sensomotori, allora sembrerebbe che stia parlando più di proprietà funzionali che di stati fisici.

Per quanto riguarda il secondo modo in cui si possono intendere i poteri causali, Searle argomenta che le cose come le automobili o le macchine da scrivere sono una categoria di cose che non possono provare sensazioni e che questo aspetto è cruciale per l'intenzionalità. Ma si potrebbe comunque sostenere che Searle sbaglia nel ritenere che l'intenzionalità richieda necessariamente provare sensazioni, e che le componenti funzionali delle sensazioni sono tutto ciò che è richiesto per il modello cognitivo. Anche se noi accettassimo la teoria dell'intenzionalità di Searle l'affermazione secondo cui l'intenzionalità richiede il provare sensazioni, sarebbe comunque insostenibile. Il fatto che il fenomeno mentale come il provare sensazioni sia stato confinato nell'organismo protoplasmico non dimostra in nessun modo che tale fenomeno non possa sorgere in un sistema non protoplasmico. «Si può solo concludere che la conoscenza della connessione necessaria tra l'intenzionalità è l'incarnazione protoplasmica sia ottenuta attraverso una qualche sorta di rivelazione mistica»<sup>84</sup>

Un altro autore che critica il concetto di intenzionalità di Searle è Dennett, il quale sostiene che, da quello che si può evincere leggendo il suo saggio, l'intenzionalità sembra essere una sostanza magica secreta dal cervello, nello stesso modo in cui il pancreas secerne l'insulina. Per Searle il cervello *produce* intenzionalità laddove i programmi del computer non lo fanno, anche se accade che siano designati per simulare il comportamento input-output del cervello. Ma quello che molti dei ricercatori dell'IA sosterebbero, secondo Dennett, è che ciò che viene prodotto è qualcosa come il controllo: il cervello è tale perché governa le relazioni giuste, appropriate e intelligenti tra input sensoriali e output comportamentali. Dal punto di vista di Searle le relazioni giuste tra input e output non sono il criterio di evidenza dell'intenzionalità, ma secondo Dennett la prova finale si può trovare nella presenza di proprietà causali interne alle operazioni del cervello, ed è questo che si deve mettere in luce. Quando Searle parla di proprietà causali si può pensare a quelle proprietà causali che sono cruciali per l'intenzionalità e che collegano le attività di un sistema alle cose nel mondo con le quali il sistema interagisce. Ma Searle insiste dicendo che queste non sono le proprietà causali rilevanti. Egli infatti concede la possibilità in linea di principio di duplicare le competenze input-output del cervello umano con un "programma formale, il quale guiderebbe un corpo attraverso il mondo così come farebbe il cervello

---

<sup>84</sup> Ringle in "Open Peer Commentary", 1980, p. 445

con il corpo di un uomo e ne acquisirebbe le proprietà causali. Ma tale cervello sostitutivo fallirebbe interamente nella riproduzione dell'intenzionalità. Tuttavia, Dennett si chiede in che modo noi potremmo venire a conoscenza di tali mancanze, se tutto ciò che sappiamo è che esso è un programma formale. Dal momento che Searle concede che le operazioni di qualsiasi cosa – come anche quelle del cervello umano – possono essere descritte in termini di esecuzione di un programma formale, la mera esistenza di un tale livello di descrizione del sistema non precluderebbe il suo avere intenzionalità. Sembra che solamente quando noi possiamo vedere che il sistema in questione è solo l'implementazione del programma formale possiamo concludere che esso non produce intenzionalità. «Ma nulla può essere solo l'implementazione di un programma formale; i computer trasudano calore e rumore nel corso delle loro operazioni – perché non anche l'intenzionalità?»<sup>85</sup>

In seguito a tali critiche, nella sezione relativa all' "Author'sResponse", Searle cercherà di chiarire meglio il suo concetto di intenzionalità (ma senza darne una definizione) e di rendere tale concetto più specifico, sostenendo di non poter definire gli stati intenzionali se non affermando che essi sono stati particolari causati e realizzati dalla struttura del cervello umano. Ciò che riuscirà a specificare meglio saranno semplicemente i modi in cui, secondo lui, si può prendere in considerazione l'intenzionalità, due modi che spesso sono stati confusi dai diversi autori dell' "Open Peer Commentary". Searle parla di intenzionalità intrinseca, grazie alla quale si hanno veri stati mentali, e di attribuzione di intenzionalità relativa all'osservatore, in cui l'attribuzione di stati intenzionali è un modo espressivo che i soggetti hanno per riferirsi a cose o entità presenti nella loro quotidianità, ma senza che queste abbiano realmente capacità intenzionali.

[...] Dobbiamo distinguere attentamente tra casi di ciò che chiamerò *intenzionalità intrinseca*, che sono casi di stati mentali effettivi, e ciò che chiamerò *attribuzioni di intenzionalità relative all'osservatore*, che sono modi che le persone hanno di parlare di entità che figurano nelle nostre attività, ma non hanno intenzionalità intrinseca. (Searle, 1980, p.451)

Secondo Searle, la vera e propria intenzionalità è quella intrinseca, cioè gli stati mentali possono essere caratterizzati solo da questa intenzionalità intrinseca (o *originaria*, come la chiama Haugeland,

---

<sup>85</sup> Dennett in "Open Peer Commentary", 1980, p. 430

distinguendola da quella derivata<sup>86</sup>), mentre le ascrizioni di intenzionalità alle cose sono sempre relative agli osservatori: è chiaro che la nostra attribuzione di intenzionalità a macchine, computer e calcolatori è relativa all'osservatore, e attribuire intenzionalità ai termostati, come fa McCarthy, significa confondere l'attribuzione di intenzionalità relativa all'osservatore come l'intenzionalità intrinseca. Questa confusione non deriva dal fatto che noi pensiamo che questi abbiano una vita mentale. Piuttosto commettiamo l'errore di designare ad essi la nostra intenzionalità «per servire certi nostri scopi [...] per eseguire il tipo di funzioni che noi eseguiamo sulla base della nostra intenzionalità».<sup>87</sup>

### 4.3.3. Critiche alla contrapposizione tra semantica e sintassi

Searle pone una netta contrapposizione tra la competenza semantica di un sistema e la sua capacità sintattica di manipolare simboli. Questa contrapposizione sembra mostrare che Searle ritenga la parte semantica e la parte sintattica due componenti separate, anche se costitutive della comprensione, cioè due processi paralleli, diversi e indipendenti l'uno dall'altro. Ciò che afferma Searle è che la componente semantica non è riducibile a operazioni formali di tipo sintattico: essa rimane una facoltà prettamente umana non riproducibile da macchine e programmi formali.

È bene notare che Searle non nega che la manipolazione dei simboli formali costituisca, in parte, il processo di comprensione: questa è costitutiva della comprensione, ma le carenze che sono proprie del processo di manipolazione vengono colmate dal contributo dato dalla semantica. Quel che nega Searle è che le manipolazioni sintattiche siano una condizione sufficiente della comprensione e l'esperimento della stanza cinese serve proprio a far emergere questa contrapposizione: se la manipolazione dei simboli fosse sufficiente affinché avvenga la comprensione, non ci sarebbe nessuna difficoltà a riconoscere che l'uomo nella stanza capisce il cinese.

L'argomentazione searliana serve quindi per arrivare alla conclusione che la comprensione non può essere manipolazione di simboli formali perché la competenza semantica, che sta alla base di tale processo cognitivo, è una capacità diversa da quella sintattica. La semplice elaborazione di simboli non

---

<sup>86</sup> Haugeland, in "Open Peer Commentary, 1980, p. 432.

Haugeland fa una differenza, molto simile a quella attuata da Searle, tra due tipi di intenzionalità: quella *originaria*, quella che una cosa ha per virtù propria, e quella *derivata*, cioè quella conferita da qualcos'altro. L'intenzionalità del pensiero cosciente e della percezione è originaria, mentre quella dei segni linguistici è derivata, cioè conferita dagli utilizzatori del linguaggio (le parole non hanno significato di per se stesse, ma grazie al fatto che noi glielo diamo).

<sup>87</sup> Searle in "Author's Response", 1980, p. 452

potrà mai conferire proprietà semantiche e perciò, dato che le nostre rappresentazioni hanno tali proprietà semantiche, nel nostro cervello avviene evidentemente qualcosa di diverso dall'elaborazione di simboli, anche se ancora non sappiamo cosa.

Tuttavia, non tutti sono così d'accordo nel sostenere questa indipendenza e separazione tra competenza semantica e capacità sintattica. Autori come Abelson, Dennett, Schank, Smythe, accusano l'autore di non considerare a sufficienza il ruolo fondamentale dell'operazione sintattica dell'applicazione di regole nei processi cognitivi.<sup>88</sup>

Per esempio, Abelson, sostenendo come la comprensione non sia altro che una connessione di contenuti tramite regole che si apprendono, afferma che nell'argomentazione di Searle la principale regola che viene imposta è che «se vedi il simbolo cinese che sta per “operazioni computazionali formali” allora emetterai il simbolo cinese che sta per “non comprendo affatto”». <sup>89</sup> Il programma ha bisogno di regole affinché possa avvenire una comprensione è questo non è sicuramente un problema da tralasciare. Le operazioni sintattiche e le regole che le governano non possono essere declassate a semplice banalità, perché regole mediocri producono output “mentalmente stupidi”<sup>90</sup> che dovranno essere riscritti e questa è una caratteristica molto importante della comprensione. Schank è del suo stesso avviso: le regole incarnano la comprensione<sup>91</sup> e dal momento che non sappiamo definire con precisione cosa essa sia, gli studiosi dell'IA possono trarre qualche vantaggio cercando di comprendere come funziona il linguaggio, le sue regole e le operazioni sintattiche. «Le persone scrivono le regole per Searle per usarle per emettere simboli cinesi nel tempo appropriato. [...] quali regole sono state scritte? Il linguista che scrive le regole “comprese” in un senso profondo come il linguaggio cinese funziona»<sup>92</sup>.

Un'altra questione è sollevata da Fodor, secondo il quale non è così scontato che la continua manipolazione formale di simboli non porti il sistema che la attua ad una forma di comprensione di essi, o che addirittura non costituisca già una forma di comprensione. Secondo Fodor, se si avessero le giuste relazioni causali tra i simboli formali e le cose nel mondo, questa potrebbe essere una condizione sufficiente per l'intenzionalità. In questo caso

---

<sup>88</sup> Searle non nega che molti processi cognitivi siano descrivibili come se seguissero le regole, ma sostiene che dire questo sarebbe semplicemente come dire che essi sono conformi a tali regole, cioè descrivibili mediante esse. In questo caso quindi, questi processi non sarebbero distinguibili dai processi meccanici, processi che a loro volta sono descrivibili tramite regole. Inoltre, per Searle la nozione di seguire una regola ha senso solo in riferimento ad agenti intenzionali coscienti, mentre sono si applica a processi di cui solo metaforicamente si può dire che obbediscano a certe direttive.

<sup>89</sup> Abelson in “Open Peer Commentary”, 1980, p. 424

<sup>90</sup> Ibidem

<sup>91</sup> Schank in “Open Peer Commentary”, 1980, p. 447

<sup>92</sup> Ibidem



le operazioni sintattiche avrebbero un ruolo fondamentale nella formazione della competenza semantica: la manipolazione dei simboli potrebbe essere determinante per l'attivazione dei processi di formazione e attribuzione di significati.

Altri autori come Block, Minsky, Marshall e Menzell mettono in dubbio questa rigida distinzione tra semantica e sintassi, sottolineando invece come esse abbiano ruoli complementari per la comprensione e queste critiche mettono in luce il problema principale che si presenta nell'esperimento mentale di Searle. Per esempio, secondo Block, anche se prendessimo per buona la risposta di Searle alla "replica del robot", secondo la quale nonostante il robot abbia atteggiamenti simili ai nostri non possiede comunque stati mentali, «la dottrina secondo cui la cognizione è una manipolazione di simboli formali rimane illesa».<sup>93</sup> Dalla sua argomentazione si evince che ci sono presumibilmente dei dati empirici che mostrano che pensare è una manipolazione di simboli formali e non è possibile rifiutare questa tesi per mezzo di intuizioni.

Ciò che Searle afferma è che i contenuti cognitivi e le costruzioni mentali della mente umana sono preclusi all'attività di qualsiasi macchina diversa da una copia perfetta del cervello umano, al di là di ogni processo tecnologico possibile: una macchina, anche la più tecnologicamente evoluta, non può essere nient'altro che un manipolatore di simboli dei quali ignora il significato.

[...] è nell'essenza di computazioni come il ragionamento o la comprensione del linguaggio di essere eseguibili solo da un supporto materiale che abbia le stesse proprietà materiali del cervello (i poteri causali del cervello). (Marconi, 2008, p.11)

L'autore considera le attività intenzionali degli stati cognitivi come imprescindibilmente legati all'"hardware" del cervello umano, cioè alla specifica macchina-uomo. Sembra, quindi, che egli consideri la comprensione non tanto come un'attività solamente mentale, ma come un'attività mentale-cerebrale umana. Per questo, per esempio, Hofstadter e Rorty <sup>94</sup> rigettano questa concezione come religiosa<sup>95</sup>, perché assume aprioristicamente che le macchine diverse dal cervello umano non possano avere una comprensione di un certo tipo.

---

<sup>93</sup> Block in "Open Peer Commentary", 1980, p. 426

<sup>94</sup> «Come sa questo Searle? Nello stesso modo, presumibilmente, di come il teologo ortodosso conosce le cose. Searle conosce cosa "mentale" e "cognitivo" significano, e così sa che non possono essere propriamente applicate in assenza del cervello – o, forse, in assenza di qualcosa che è molto simile al cervello rispetto ai "poteri causali"». [Rorty, 1980, pp.445-46]

<sup>95</sup> Hofstadter in "Open Peer Commentary", 1980, p. 433

Quello che viene evidenziato dai critici della posizione searliana è che una rigida contrapposizione tra semantica e sintassi, dove la semantica viene delegata a misteriose capacità intenzionali, non meglio specificate, è una posizione inadeguata allo studio della comprensione. Studiare i meccanismi sintattici che entrano in gioco nei processi cognitivi e nel processo di comprensione del linguaggio è fondamentale: entrambe gli aspetti sono estremamente importanti. Minsky, per esempio, facendo riferimento all'esperimento della stanza cinese, afferma che quella che sarebbe la sua esperienza consisterebbe, almeno inizialmente, nella comprensione della sintassi delle frasi cinesi, nella possibilità di una loro analisi a cui seguirebbero deduzioni. Ciò che rimarrebbe oscuro sarebbero i termini stessi, cioè come le parole "intendere" e "significare" si potrebbero intendere e significare.<sup>96</sup>

Ritenendo le procedure semantiche come le sole adatte a creare e ad attribuire i significati, separandole e contrapponendole a qualsiasi tipo di manipolazione sintattica, si rischierebbe di relegare la semantica dei processi cognitivi all'ambito dell'inconoscibile: si dovrebbe pensare alla semantica come alla messa in atto di una qualche capacità magica tipicamente umana e irriproducibile artificialmente.<sup>97</sup> Lo stesso Searle non riesce a fornire delle spiegazioni soddisfacenti riguardo a cosa siano e quale sia l'origine dell'intenzionalità e dei poteri causali, date anche le scarse conoscenze attuali sul funzionamento del cervello umano: esemplare è l'espressione che utilizza alla fine del suo saggio "qualunque cosa sia l'intenzionalità",<sup>98</sup> indicativa del fatto che Searle non sa cosa essa sia nonostante la ritenga, con assoluta certezza, una sostanza prodotta dal cervello; e nella parte relativa alle risposte che l'autore fornisce alle critiche e ai commenti, Searle risponde alla critica di Dennett ammettendo che non ha idea di come il cervello produca il fenomeno dell'intenzionalità, ma è certo che sia un fenomeno mentale e che le operazioni del cervello siano causalmente sufficienti per tale fenomeno è evidente.<sup>99</sup> Il fatto di non avere a disposizione una definizione chiara di questi concetti non gli sembra una ragione sufficiente per screditarne l'importanza: essi sono sufficientemente ricchi ed evidenti per spiegare la comprensione e muovere critiche alla posizione dell'IA forte. Tutto ciò che è essenziale sapere è che l'intenzionalità e i poteri causali sono peculiarità degli atti mentali degli umani e che la prima è un prodotto dovuto all'azione dei secondi.

---

<sup>96</sup> Minsky in "Open Peer Commentary", 1980, p. 439

<sup>97</sup> Nel saggio di Searle, infatti, la spiegazione della competenza semantica e delle capacità cognitive è lasciata a concetti oscuri come quello di intenzionalità, concetti che non sono indagabili dal punto di vista scientifico.

<sup>98</sup> «Qualunque cosa sia l'intenzionalità, è un fenomeno biologico, e quindi è verosimile che sia causalmente dipendente dalla biochimica specifica delle sue origini» [Searle, 1980, p. 424]

<sup>99</sup> Searle, 1980, p. 452

Ciò che però crea problemi, in mancanza di una definizione e una specificazione del concetto di intenzionalità, è che esso rimane un concetto discutibile, la cui stessa applicabilità risulta critica, indefinita e quindi modificabile all'occorrenza. Abelson dice infatti che Searle non è riuscito a rendere affatto convincente la sua tesi secondo cui fondamentale per la comprensione è l'intenzionalità.<sup>100</sup> L'imprecisione con cui viene trattata l'intenzionalità rischia di andare ad inficiare il concetto stesso poiché risulta estremamente versatile: non sapendo cosa essa sia, la si potrebbe applicare, per esempio, a qualsiasi capacità in grado di rendere efficienti i programmi dei computer che simulano la comprensione umana. Bridgeman parla dell'intenzionalità come "illusione cognitiva", Pylyshyn come di un concetto attribuibile arbitrariamente ad un tipo di materia o a un altro, rendendo così irrilevante la natura stessa della materia, Ringle parla di misticismo. Quello che viene sottolineato nei commenti è quindi l'impossibilità di trattare scientificamente un concetto che non ha nessuna rilevanza scientifica; questo invalida totalmente l'argomentazione searlina, rendendo il discorso religioso e aprioristico.

#### 4.4. Conclusione

Il dibattito presentato in questo capitolo ha cercato di mettere in luce i diversi punti di vista in merito alle questioni principali che animano l'impresa dell'IA. Le domande a cui si è tentato di dare risposta sono:

- Può una macchina, almeno in linea di principio, realizzare funzioni cognitive?
- I modelli computazionali che realizzano formalmente le attività cognitive, sono condizioni sufficienti?

Searle ha cercato di rispondere a queste domande tramite il suo esperimento mentale della stanza cinese. Secondo lui le risposte non possono che essere negative. Per quanto riguarda la prima questione, la macchina non può realizzare funzioni cognitive poiché manca dell'elemento fondante, cioè dell'intenzionalità. Per quanto riguarda la seconda domanda, i modelli computazionali non sono sufficienti perché sono solo una manipolazione di simboli, ovvero realizzano una sintassi ma non una semantica.

Si sono rilevate nel corso del capitolo alcune difficoltà che nascono da queste prese di posizione:

- 1) Il concetto di intenzionalità è un concetto vago e misterioso e non può essere preso sul serio in ambito scientifico,

---

<sup>100</sup> Abelson in "Open Peer Commentary, 1980, p. 425

- 2) La distinzione tra semantica e sintassi non è così netta, ma è possibile che siano due ambiti interdipendenti,
- 3) (Una terza questione che a parere mio è legata alla seconda) l'applicazione delle regole nel processo di comprensione del linguaggio, che Searle ha sottovalutato, è stata ritenuta molto importante da alcuni autori, i quali ritengono che regole e linguaggio siano due elementi strettamente connessi.

(1)

Secondo Searle a) l'intenzionalità è prodotta dai poteri causali del cervello; b) l'intenzionalità è possibile solo laddove il sistema ha le stesse proprietà fisiche e causali del cervello; c) le proprietà causali dipendono direttamente dal sostrato fisico.

C'è chi ha messo in discussione l'assunzione secondo cui il cervello è la sede dell'intenzionalità, chiedendosi perché proprio il cervello e non altro come il DNA o gruppi di cellule (Phylyshyn), chi denuncia la mancanza di una definizione precisa del concetto di intenzionalità (Abelson, Phylyshyn, Ringle), chi pensa che la connessione tra processi interni e stati sensomotori, in relazione alla causalità di cui parla Searle, sia riferibile più a proprietà funzionali che a stati fisici e che la connessione tra intenzionalità e sostanza protoplasmica sia un'assunzione mistica (Ringle), chi parla dell'intenzionalità come di una sostanza magica secreta dal cervello, il quale in realtà produrrebbe più un controllo delle relazioni input-output che questa sostanza misteriosa (Dennett).

In primo luogo, la nozione di intenzionalità sembra stare per qualsiasi processo che si voglia prendere come caratteristico della cognizione umana. Quindi assumere l'intenzionalità come premessa di tutta l'argomentazione non può che portare ad una conclusione vaga e imprecisa tanto quanto lo è la premessa di partenza.

In secondo luogo, che la sede dell'intenzionalità sia proprio il cervello è un'assunzione non dimostrabile – in buona parte, infatti, non si conosce come funziona il cervello. In questo modo, si fa derivare un concetto estremamente vago come l'intenzionalità da qualcosa di altrettanto misterioso come i processi cerebrali che realizzano la comprensione. Che esistano dei poteri causali del cervello è cosa ovvia, qualsiasi entità fisica li possiede, ma questi potrebbero semplicemente essere relegati alla sfera fisica e non a quella del mentale. Secondo Dennett, infatti, la capacità di comprensione è propria dell'uomo, non della sua struttura cerebrale: non si deve prestare attenzione tanto alla struttura fisica quanto all'individuo che opera ed interagisce con altri fattori, cioè col mondo esterno.

Non si discute che la sfera sensomotora sia determinata dal cervello, ma che lo siano le nostre idee, desideri, credenze, paure, etc... Possiamo decidere che la causa di tutto ciò sia o non sia il cervello, possiamo essere convinti dell'una o l'altra cosa, ma qualsiasi scelta si faccia è da ritenere aprioristica e religiosa (Hofstadter). Per quel che ne sappiamo, qualsiasi assunzione si faccia riguardo a tale problema sarebbe semplicemente da ritenere una scelta argomentativa strategica, utile per rafforzare la tesi da cui si è partiti, ma non scientificamente incontrovertibile.

(2)

L'intenzionalità è vista come propria degli umani anche perché essa è propria della sfera semantica del linguaggio. Secondo Searle le macchine non potranno mai arrivare ad avere stati cognitivi perché non sono dotate di intenzionalità e quindi non posseggono la sfera semantica, ma solo quella sintattica, cioè possono solo manipolare formalmente dei simboli. I simboli manipolati da un programma non hanno un contenuto ma solo una forma. Come l'uomo nella stanza cinese non sa che significato hanno quei simboli, così neanche il robot, anche se dotato di una telecamera che gli permetta di vedere, non può avere cognizione di ciò che sta vedendo e manipolando.

Ciò che viene fatto notare da diversi autori è che la distinzione che pone Searle tra semantica e sintassi non è così scontata, poiché la comprensione avviene grazie a entrambe. La sintassi e la semantica non si possono ritenere indipendenti l'una all'altra, non sono due processi paralleli, ma non interdipendenti. Relegare la semantica a misteriose capacità intenzionali significa assumere una posizione inadeguata rispetto al problema, anche per l'imprecisione con cui si utilizza il termine "intenzionalità".

Searle contrappone semantica e sintassi perché nota che il computer è in grado di manipolare simboli senza che ne abbia una vera e propria comprensione. Come però osserva Fodor, non è così scontato che la manipolazione di simboli non porti il sistema ad una comprensione di essi. Avere le giuste relazioni causali tra simbolo e mondo potrebbe essere una condizione sufficiente per l'intenzionalità ed in questo caso le competenze sintattiche avrebbero un ruolo fondamentale nella formazione della competenza semantica, cioè la manipolazione dei simboli potrebbe essere determinante per l'attivazione dei processi di formazione e attribuzione di significato. Secondo Fodor, il sistema rappresentazionale procede mediante elaborazione sintattica ed il collegamento tra sintassi e semantica è necessario affinché le computazioni non diano risultati casuali ed arbitrari. Le rappresentazioni mentali infatti sono vincolate dalla semantica nonostante siano definite dalla sintassi e ciò che conferisce le proprietà semantiche al

simbolo sono le sue proprietà sintattiche, cioè il fatto che la relazione tra simboli è determinata da regole che ne garantiscono le condizioni di verità.

Non viene messo in discussione il fatto che la sintassi presa da sola non sia sufficiente per determinare il processo di comprensione, ma rimane comunque una condizione necessaria affinché i significati, i contenuti dei simboli, vengano organizzati semanticamente in modo corretto. Come condizione necessaria non può venir separata da un'altra condizione altrettanto necessaria, cioè quella semantica, ciò che fornisce ai simboli un significato. La coppia semantica-sintassi diventa, in questo modo, una condizione necessaria e sufficiente per la comprensione; se si isolano non possono garantire la produzione e la comprensione del linguaggio. Ciò che è importante è studiare i meccanismi sintattici che entrano in gioco nell'attivazione del processo di comprensione per evidenziare la loro interdipendenza con i meccanismi semantici. Se si abbandona questo presupposto si rischia di relegare la semantica a processi misteriosi e sconosciuti, oltre a dover abbandonare la speranza di poter afferrare ciò che costituisce il processo di comprensione, quindi lo scopo dello studio della mente.

Come fa notare anche Marconi in *La competenza lessicale*, è che non è chiaro come la competenza linguistica potrebbe essere *altro* che la capacità di manipolare simboli, anche se le capacità di cui parla Searle nell'esperimento sono certamente insufficienti. Sottolineare l'insufficienza di quelle manipolazioni, però, non basta certo a screditare la connessione tra competenza linguistica e capacità di usare certi segni e suoni. Il problema non è se la conoscenza del significato possa essere ridotta alla manipolazioni, ma quali manipolazioni varrebbero come conoscenza del significato o comprensione del linguaggio.<sup>101</sup>

(3)

Questo secondo aspetto relativo alla sintattica e alla semantica è legato ad un altro aspetto che alcuni autori hanno rilevato come mancante nell'argomentazione di Searle: l'importanza delle applicazioni delle regole nel processo di comprensione.

Secondo Searle la comprensione non avviene quando sappiamo raggruppare i simboli secondo certe regole, ma quando sappiamo cosa essi significano, perché la comprensione non fa parte dell'aspetto sintattico ma di quello semantico.

Searle non nega che molti processi cognitivi siano descrivibili come se seguissero regole, ma sostiene che dire questo è semplicemente dire che

---

<sup>101</sup> Marconi, pp. 162-163

essi sono conformi a regole, cioè sono descrivibili mediante regole, e dunque non si distinguono da altri processi, che chiamiamo meccanici, i quali sono anch'essi descrivibili mediante regole.<sup>102</sup> La nozione di seguire una regola, per Searle, ha senso nel caso di agenti intenzionali coscienti. Si è già visto però che il termine intenzionalità è vago e impreciso, quindi la connessione che suggerisce Searle tra intenzionalità e seguire la regola costituisce un problema se non si ha una concezione precisa dell'intenzionalità.

Secondo Searle le operazioni computazionali su elementi formali non sono condizioni sufficienti per la comprensione. Autori come Abelson, Bridgeman, Smythe sottolineano invece come le regole di trasformazione di un input in un output siano un aspetto essenziale della comprensione: la comprensione è un processo che avviene grazie all'applicazione di regole: quando un bambino impara a contare cosa fa se non applicare regole? Quando si comprendono le regole relative a certi contenuti, queste rendono possibile il connettere quei contenuti con altri domini ed in questo modo si ottiene un incremento della comprensione (Abelson). Inoltre, queste regole vengono memorizzate e associate a fatti memorizzati relativi ai contesti in cui esse vengono applicate, cosicché il seguire la regola diventa parte integrante della comprensione e dei processi cognitivi (Bridgeman). Le capacità computazionali, seguire regole e istruzioni che mettono in relazione simboli, sono, quindi, fondamentali per la comprensione.

Per richiamare Wittgenstein, il seguire la regola è un aspetto essenziale poiché l'aspetto semantico di una lingua può essere realizzato solo laddove si conoscono le regole di utilizzo di un termine, cioè se viene compreso il modo in cui lo si può impiegare. La comprensione di un termine è connessa al suo uso perché questa è la capacità di fare certe cose coi segni; impariamo ad utilizzare certi segni perché impariamo le regole per utilizzarli ed un'azione è sempre governata da regole. La comprensione risiede nella competenza cioè nel saper utilizzare le regole nel modo corretto. In breve, il significato di una parola viene compreso quando se ne comprendono le regole d'uso.

Qui si contrappongono due concezioni. Secondo Searle, la comprensione è basata sull'attribuzione di significati ai simboli. Questa attribuzione non è identica a, né deriva da, la capacità di manipolare simboli secondo regole, *quali che siano queste regole*. Secondo Wittgenstein, comprendere un simbolo è saperlo usare, cioè essere in grado di seguire la regola per il suo uso.

La posizione di Wittgenstein viene definita da molti studiosi come comportamentismo; tale concezione viene nettamente rifiutata da Searle (e da molti altri – compreso Fodor), poiché secondo lui non si possono attribuire

---

<sup>102</sup> Marconi, 2008, p.54

stati cognitivi semplicemente sulla base dell'osservazione del comportamento dell'agente, cioè rispondere con certi output a determinati input non è una condizione sufficiente per dire che l'agente possiede stati intenzionali. C'è da chiedersi, però, sulla base di cosa si può attribuire comprensione: sulla base dei processi cerebrali? – viene ammesso dallo stesso Searle che non si conoscono ancora a sufficienza. Sulla base di concetti poco chiari come l'intenzionalità? Si è già fatto notare che una critica che parte da presupposti vaghi e imprecisi rischia di risultare inconsistente.

Separare la questione relativa all'applicazione delle regole dalla capacità di comprensione pone allora la domanda: come si apprende un linguaggio? Come si viene a conoscenza dei significati dei termini che si comprendono? I significati delle parole e dei termini che si comprendono vengono stabiliti ed utilizzati casualmente? O seguono regole che ne hanno permesso l'utilizzo e la condivisione?

Secondo Searle l'attribuzione di significato è indipendente dalle regole che si utilizzano per manipolare simboli, secondo Wittgenstein, il significato è l'uso inteso come funzione regolata di un segno in un contesto. Se le espressioni linguistiche non fossero determinate da regole, si porrebbe il problema relativo alla possibilità di utilizzarle in modo diverso a seconda delle occasioni, mettendo a rischio la possibilità stessa di comprendere la propria lingua (ammesso, non senza difficoltà, che in questo modo si possa costituire un linguaggio naturale). Se, al contrario, conoscere i significati di termini e parole accadesse grazie al seguire istruzioni e regole che consentono la connessione di certi significati a certi termini, si avrebbe come risultato una certa regolarità che permetterebbe la comprensione e l'utilizzo condiviso di una lingua. Fodor pensa che il sistema rappresentazionale poggi su un'elaborazione sintattica, cioè sull'applicazione di regole che sono sì sensibili solo alla forma, ma che hanno la proprietà di garantire la verità, cioè sono vincolate dalla semantica, poiché elaborano risultati corretti. Apprendere un linguaggio significa apprendere le regole di verità dei suoi predicati, cioè S apprende P se S apprende le regole di verità di P. In altre parole, apprendere un linguaggio vuol dire apprendere le proprietà semantiche dei suoi predicati. Pensare che la manipolazione dei simboli formali non sia una condizione sufficiente non è una posizione che risulta problematica, ma è da ritenere comunque una condizione necessaria che può essere resa anche sufficiente se affiancata alla semantica. Sarebbe possibile apprendere un linguaggio senza conoscere la semantica e la sintassi dei termini che lo costituiscono? In che modo potrebbe venir utilizzato e condiviso se il significato degli elementi che lo costituiscono non fossero organizzati secondo determinate regole? Come si può conoscere, comprendere e utilizzare una lingua se questa non è costituita da elementi che prima vengono elaborati formalmente ed organizzati per mezzo di regole? Si



dovrebbe pensare che il significato dei termini che costituiscono il linguaggio, nasca da una qualche processo magico e misterioso.

Secondo Schank, per comprendere il linguaggio è necessario comprendere prima di tutto le regole, cioè ciò che organizza i concetti, i termini e le idee relative agli stati di cose del mondo. Chi scrive le regole utili per mettere insieme i simboli cinesi ha una comprensione profonda del linguaggio cinese, non perché sa a cosa i simboli si riferiscono ma perché sa come i simboli interagiscono tra di loro. Il fatto che un input produca un certo output non è un caso; questo fatto è regolato da istruzioni, le quali non possono essere viste come elementi di secondaria importanza.

Se non ci fossero delle regole che permettono l'attribuzione di significati, si dovrebbe pensare che questi significati arrivano dal nulla, che siano in un certo senso già presenti e organizzati a priori nel nostro vocabolario linguistico. Wittgenstein non crede che sia così, come neanche Fodor e molti degli altri autori dell'Open Peer Commentary (Abelson, Dennett, Schank, Smythe, Fodor, Block, Minsky, Marshall e Menzell).

Nel libro *Filosofia e scienza cognitiva* Marconi prende come esempio la tesi del carattere astratto delle computazioni, la quale lascerebbe aperta la possibilità che l'impresa dell'IA sia possibile. La tesi della realizzabilità multipla dei processi cognitivi, dovuta originariamente a Putnam, sostiene che se è vero che un tipo di computazioni (i processi cognitivi sono computazioni) è identificabile non solo a livello dell'implementazione<sup>103</sup>, ma anche a livello dell'algoritmo<sup>104</sup>, allora è concepibile un punto di vista sui processi cognitivi che prescinde dalla particolare implementazione realizzata dal cervello umano. Se identifichiamo la funzione di comprensione di una lingua con un determinato algoritmo, niente vieta che quell'algoritmo sia realizzabile da vari hardware, oltre che dal cervello umano. Se l'algoritmo di comprensione venisse realizzato da un computer sarebbe appropriato dire che quel computer comprende l'italiano. La tesi del carattere astratto delle computazioni si limita, quindi, a svincolare concettualmente le computazioni dalle loro realizzazioni fisiche.

Marconi riprende, inoltre, l'esempio di SemyourPapert secondo il quale non c'è nessuna ragione di vincolare il concetto di volo alla particolare realizzazione che ne danno gli uccelli: il concetto di volo può essere identificato a livello più astratto, e questo si applica anche agli aeroplani e agli elicotteri, nonostante le differenze nel modo in cui viene realizzato.

---

<sup>103</sup> Le computazioni (per esempio la comprensione di una lingua, come l'italiano e l'inglese) possono essere realizzate ("implementate") da hardware diversi dal cervello umano e dai suoi scambi elettrochimici

<sup>104</sup> Un calcolo può essere semplicemente svolto, oppure può essere svolto esplicitando le regole che consentono di svolgerlo (fare e dire come si fa). Questo livello di elaborazione dei dati è definito "livello dell'algoritmo"

Analogamente, per la tesi del carattere astratto delle computazioni la comprensione del linguaggio, la visione, il ragionamento, non si identificano solo con ciò che fanno gli esseri umani quando comprendono, vedono, ragionano: comprensione, visione e ragionamento, sono identificabili a livello più astratto, perché godono di realizzabilità multipla, cioè possono essere realizzati, almeno in linea di principio, da macchie diverse dal cervello.

Le computazioni potrebbero, quindi, essere viste come una condizione sufficiente per la comprensione, concetto applicabile, almeno in linea di principio, anche alle macchine e non solo al cervello umano. E, sempre in linea di principio, potrebbe essere possibile che le macchine realizzino funzioni cognitive simili a quelle dell'uomo. Supporre il contrario, come è già stato fatto notare, significa assumere una posizione aprioristica e religiosa. Se si venisse a conoscenza di come funzionano i processi cognitivi, la comprensione ed il ragionamento, nulla vieterebbe di poter riprodurre tali processi in dispositivi fisicamente diversi dal cervello umano.

Per concludere, ciò che in questo scritto si è cercato di far emergere è che, affinché la realizzazione degli stati cognitivi in dispositivi artificiali possa risultare possibile, almeno in linea teorica, ci sono almeno tre aspetti dell'indagine che non possono venir trattati come inessenziali. Affinché lo studio su questi argomenti possa risultare scientificamente coerente non si deve:

- a) Assumere come premesse termini e concetti imprecisi e indefiniti, perché ciò rischia di inficiare la validità della conclusione a cui arriva la teoria;
- b) Separare e trattare come indipendenti l'uno dall'altro i due ambiti essenziali della lingua (semantica e sintassi);
- c) Tralasciare il ruolo fondamentale delle regole o istruzioni che regolano l'utilizzo e la condivisione dei significati di un linguaggio.

## 5. CONCLUSIONE

### 5.1. Wittgenstein e Fodor

Le teorie che forniscono Fodor e Wittgenstein sul linguaggio sono molto diverse e partono da presupposti diversi. Secondo Fodor, ciò che si manifesta nel linguaggio naturale si realizza prima di tutto fisicamente nel cervello, cioè le cause degli stati mentali devono avere prima di tutto una realizzazione fisica. Il linguaggio naturale sarebbe solo l'espressione di ciò che precedentemente (ed inconsciamente) si è formato grazie a computazioni che operano sulle rappresentazioni del linguaggio del pensiero, un linguaggio che è innato così come lo sono le regole sintattiche su cui operano tali rappresentazioni. Secondo Wittgenstein ciò che è essenziale al linguaggio non riguarda la sfera fisica, ma riguarda il modo in cui utilizziamo le parole a seconda dei giochi linguistici in cui i termini sono inseriti, cioè nulla di ciò che succede a livello concettuale può essere spiegato sulla base di processi fisiologici, ma soltanto sulla base di un apprendimento di regole che determinano l'uso di espressioni linguistiche che vengono insegnate, trasmesse da uomo a uomo.

Questi due presupposti portano a formulare concezioni radicalmente diverse e contrastanti per molti aspetti, ma ciò che qui interessa è capire in che modo possono essere utili per rispondere alle domande che anche Searle si era posto nel suo saggio sulla Stanza Cinese: Può una macchina avere stati cognitivi? Può una macchina comprendere un linguaggio umano?

Fodor ritiene che, almeno in linea di principio, la risposta a queste domande sia positiva. La teoria della mente di Fodor si muove, infatti, all'interno dell'orizzonte funzionalista, ossia di quella concezione che assimila la mente a un elaboratore elettronico: egli ritiene che non ci sia nessuna difficoltà nel pensare che la mente dell'uomo possa essere intesa come una macchina (e che, quindi, l'uomo sia una macchina pensante) e pone come elemento fondamentale della sua teoria questa analogia tra il funzionamento della macchina e il funzionamento della mente. Ritiene che ciò che succede a livello del LdP sia paragonabile a ciò che accade in una macchina di Turing: i processi computazionali si attengono alle condizioni di formalità, cioè utilizzano solo le proprietà sintattico-formali dei simboli; in altre parole, le computazioni non lavorano sui significati dei simboli, ma sulla loro forma così come le operazioni della macchina non avvengono perché essa riconosce il significato dei segni ma perché ne riconosce la forma. La macchina, quindi, è la metafora ideale per spiegare i processi della mente e del cervello: i calcolatori possono essere visti sia come meccanismi fisici che agiscono in conformità alle leggi della fisica, sia come dispositivi che, eseguendo operazioni su simboli e strutture simboliche, agiscono nel rispetto

delle leggi della logica. Dunque, se consideriamo gli stati mentali come strutture simboliche e i processi mentali come operazioni su tali strutture, il calcolatore diventa una metafora molto importante.

Due sono gli aspetti principali della teoria di Fodor: (1) il sistema di regole che opera sulle rappresentazioni del LdP è innato e (2) il sistema di rappresentazioni si realizza fisicamente nel cervello, base di ogni attività cognitiva. La sintassi, quindi, è innata, le regole che governano le rappresentazioni sono inviolabili, non modificabili e hanno valore di leggi di natura. Inoltre, in quanto gli stati mentali sono cause del comportamento e questi devono prima realizzarsi nel cervello, ne segue che causa del comportamento umano è la struttura fisica del cervello.

Secondo Fodor, il linguaggio non è appreso, non segue da un processo di apprendimento e di addestramento, ma noi possediamo già un linguaggio che ci permette di acquisire il linguaggio naturale. Il linguaggio naturale è possibile solo laddove esiste già in precedenza un linguaggio abbastanza ricco per esprimere l'estensione di un qualsiasi predicato del linguaggio naturale.

La semantica di Fodor è basata sulla sua teoria causale secondo cui i significati dei simboli si formano sulla base delle relazioni causali tra il mondo e i simboli del LdP: un simbolo del LdP ha il significato che assume grazie alle relazioni causali e quel significato non cambia; ciò che eventualmente cambia è il significato delle parole dei linguaggi esterni, nel senso che può cambiare il simbolo del LdP con cui una parola è associata. La relazione tra un simbolo del LdP e la proprietà a cui si riferisce dev'essere, quindi, causale: un simbolo mentale si riferisce a ciò che ha la proprietà P se è ciò che ha la proprietà P (e solo ciò che ha tale proprietà) a causare l'attivazione del simbolo mentale.

La semantica di Fodor si costituisce grazie ad una struttura sintattica le cui regole sono innate, cioè non sono il risultato di un processo di apprendimento e di addestramento e per questo motivo si può pensare che anche l'aspetto semantico dipenda da un processo innato: i significati sembrano essere pronti all'uso, già precostituiti, solo da rispolverare ed utilizzare, senza che intervenga un processo di apprendimento. Ciò che viene espresso nel linguaggio naturale non sono altro che predicati che sono già presenti in qualche modo in un linguaggio non naturale, innato e che si realizza prima di tutto fisicamente.

Wittgenstein non ha alcun tipo di affinità con il pensiero fodoriano: non pensa affatto che il sistema di regole che governa il linguaggio sia innato e che la spiegazione fisiologica sia utile per spiegare come pensiamo. Le regole che noi seguiamo nell'usare certi termini sono il risultato di un processo di addestramento che pone il soggetto nella condizione di saper usare una certa espressione linguistica nel modo corretto. La semantica e la sintassi si costituiscono, in un certo senso, insieme, non in modo immediato, ma nel

tempo. Le regole che si devono applicare per usare un dato termine sono insegnate e apprese, non sono già presenti nel soggetto; al contrario, per Fodor le regole esistono già nel soggetto ancor prima che funzioni il linguaggio naturale: sono le regole innate del LdP.

Wittgenstein spiega con la sua teoria dell'uso come si conoscono i significati, come si comprendono le parole, come si utilizzano e come si condividono tra i parlanti. Fodor, partendo da una base innatistica della sintassi, riesce a spiegare come avvengono i processi computazionali sulla rete di simboli che costituiscono il LdP, un linguaggio privato che rispecchia la semantica e la sintassi del linguaggio naturale. Fodor spiega che l'apprendimento consiste nell'associazione di suoni a simboli del LdP, cioè a concetti innati, ma oltre a questa base innatistica non da altre ragioni del perché si apprende in un certo modo: riduce ciò che a Wittgenstein appare come il risultato di un processo educativo che si estende nel tempo, ad una serie di elementi che sono già pronti all'uso, che trascendono l'importanza dei fattori esterni all'individuo che concorrono allo sviluppo del linguaggio naturale. Secondo Fodor, infatti, un linguaggio noi già lo possediamo, anche se non lo conosciamo, cioè il LdP. La semantica è come se venisse misteriosamente risvegliata grazie alla relazione causale che intercorre tra ciò che c'è nel mondo e ciò che c'è nella testa del soggetto, cioè simboli del LdP.

Al contrario di Fodor, Wittgenstein ritiene la spiegazione fisiologica una spiegazione incompleta e poco utile ai fini del problema del seguire la regola e del come noi comprendiamo ciò che comprendiamo. Secondo lui non è indispensabile sapere che a livello fisiologico c'è stata una qualche serie di eventi in cui consisterebbe il processo di comprensione del soggetto: questa è una spiegazione che non aggiunge nulla al problema. Ciò che è importante per rispondere al problema di come si comprende il linguaggio è fare attenzione a ciò che è pubblicamente osservabile, all'atteggiamento del soggetto in risposta all'addestramento che si sta compiendo per insegnargli le regole d'uso delle espressioni linguistiche applicabili nei diversi contesti linguistici.

Non credo che la struttura fisica del cervello sia del tutto inessenziale al problema della comprensione, come ritiene Wittgenstein, ma non credo neanche che sia una condizione sufficiente. Nonostante sia ovvio che succede qualcosa nel cervello di ciascuno quando ci si trova a comunicare con altri soggetti, noi non diciamo che qualcuno comprende perché sappiamo come funziona il suo cervello, ma perché il soggetto dimostra una certa competenza linguistica, cioè un saper fare. Le sue prestazioni sono il risultato di un addestramento rivolto al soggetto e sono la prova che le regole di utilizzo delle espressioni linguistiche sono state colte correttamente. Non diciamo che qualcuno comprende perché gli attribuiamo una struttura fisica che funziona in un certo modo, ma guardando a come egli usa i simboli. Ma grazie a che

cosa questi simboli vengono o non vengono usati correttamente? Grazie ad una struttura innata o grazie ad un processo di addestramento?

Ciò che quindi è da prendere in considerazione, in relazione al problema della macchina pensante, è se l'innatismo può ritenersi teoricamente interessante per spiegare come funziona una macchina e, di conseguenza, se si può parlare di innatismo anche nel caso delle macchine (dal momento che Fodor parte dal presupposto che la macchina può possedere stati cognitivi così come li posseggono gli uomini) o se la posizione di Wittgenstein apre più prospettive rispetto al problema se le macchine possano, almeno in linea di principio, possedere stati cognitivi, riprodurre il linguaggio ed interagire linguisticamente con gli esseri umani.

Come si è già spiegato, Fodor pensa che una macchina possa possedere stati cognitivi e questa sua idea viene spiegata grazie all'analogia che egli pone tra il funzionamento della macchina e il funzionamento della mente dell'uomo. Wittgenstein, invece, pensa che la macchina non possa pensare. Wittgenstein non si è particolarmente interessato al problema, ma da alcune osservazioni che fa sia nelle *Ricerche Filosofiche*, sia nella *Grammatica filosofica* si può evincere che, secondo lui, una macchina non può riprodurre il linguaggio e quindi il processo di comprensione che è proprio solo degli esseri umani.

«Ma una macchina potrebbe pensare?» - Potrebbe aver dolori? Importante qui è cosa s'intenda dicendo che «qualcosa ha dolori». [...] il fenomeno del dolore, che io descrivo quando dico, poniamo «ho mal di denti», non presuppone un corpo fisico. (Posso aver mal di denti senza avere denti). E qui la macchina non trova posto affatto. – È chiaro che la macchina può sostituire soltanto un corpo fisico. (GF, § 64)

Potrebbe una macchina pensare? – potrebbe provare dolore? Ebbene, dobbiamo dire che il corpo umano è una macchina così fatta? È certo, però, che gli manca molto poco per essere una macchina così fatta. (RF, I, § 359)

Ma una macchina non può certo pensare! – questa è una proposizione empirica? No. Solo dell'uomo, e di ciò che è ad esso simile, diciamo che pensa. Lo diciamo anche delle bambole e magari dei fantasmi. (RF, I, § 360)

Certamente, ai tempi di Wittgenstein lo studio della mente non era ancora rivolto al problema dell'IA, lo studio delle macchine non era avanzato come lo è oggi, ma viene comunque da chiedersi se l'autore avesse ragione o

no a sostenere questa posizione rispetto al problema delle macchine pensanti. Ciò che ci si può chiedere è se, quindi, secondo la sua teoria del linguaggio, del significato come uso, del seguire la regola, dei giochi linguistici, Wittgenstein avrebbe potuto sbagliarsi.

Come si è già detto nel primo capitolo, secondo Wittgenstein esiste uno stretto collegamento tra regola e gioco linguistico. Dalla fase di mezzo, quella fenomenologica, alla sua ultima fase, che inizia con le *Ricerche Filosofiche*, Wittgenstein capovolge la priorità tra regole e gioco linguistico, salvaguardando così il primato dei giochi linguistici: in questo modo non sono più le regole che dominano il linguaggio (nella fase fenomenologica il linguaggio viene inteso come calcolo, cioè un'attività governata da regole rigide, non violabili), ma sono i giochi linguistici, cioè la misura di tutte le cose, il concetto di modelli dell'uso linguistico tra loro differenti, sottoposti a regole. (Il linguaggio è un gioco, come è un gioco qualsiasi attività sottoposta a regole). Questo primato viene anche illustrato da Wittgenstein quando dice che le regole vengono seguite *ciecamente*. Dicendo che le regole vengono seguite ciecamente egli vuol dire che chiedersi quel che succede quando si segue una regola non aiuta a capire cosa significhi giocare un gioco linguistico. Al contrario, il seguire una regola può venir compreso solo in base ai giochi linguistici. Il gioco linguistico non è altro che il complesso di regole che si sta seguendo. Se i giochi linguistici ottengono questa priorità semantica ciò significa che non si imparano nuovi giochi linguistici imparando le regole, ma si imparano delle nuove regole imparando a padroneggiare i giochi linguistici di cui esse fanno parte. Il gioco linguistico non si impara, ma si addestra il soggetto a padroneggiarlo ed è l'addestramento che ha la funzione di stabilire il primato dei giochi linguistici.

La teoria dei giochi linguistici viene chiarita dallo stesso Wittgenstein attraverso l'analogia con il gioco degli scacchi. Le regole del gioco degli scacchi sono vincolanti per quel gioco, ma questo vincolo è forse una necessità immanente alla cosa stessa, come se esistesse una natura profonda del gioco degli scacchi e le regole del gioco circoscrivessero una essenza immutabile? Le cose non stanno così. Il gioco degli scacchi non ha affatto un'essenza. Il vincolo delle regole rimanda ad un accordo tacito tra i soggetti del gioco e non ad una pretesa oggettività del gioco stesso. (Quando comincia a funzionare l'analogia con il gioco, ogni tensione essenzialistica si sfalda). Allo stesso modo, non c'è un'essenza del linguaggio. Con il termine "linguaggio", piuttosto, designiamo una varietà di pratiche che, pur assomigliandosi, non hanno una natura comune. Il linguaggio è plurale. Poiché tutte le nostre pratiche sono intrecciate con l'uso del linguaggio, esse sono assimilabili a giochi linguistici. Non c'è un concetto universale e univoco di gioco linguistico, ma una rete di somiglianze di famiglia che collega un gioco all'altro. Se il linguaggio è, come ritiene Wittgenstein, l'insieme di tutti

i possibili giochi linguistici, il significato di una parola viene definito dalle regole che, all'interno del gioco, ne determinano l'uso, cioè è il modo in cui si utilizza quella certa espressione in un certo gioco che determina il significato di essa. Una parola o una proposizione assumono significati diversi in relazione ai diversi giochi, ai diversi contesti linguistici in cui vengono di volta in volta adoperate. Essendo i significati differenti da gioco a gioco, non si può più parlare del linguaggio come di ciò che è costituito dalle proposizioni elementari e dalle loro funzioni di verità (come faceva il *Tractatus*). Si dirà invece che un linguaggio è istituito da determinate regole: regole diverse istituiscono linguaggi diversi, conferendo senso, in un gioco, a espressioni che, in un gioco diverso, non ne hanno o ne hanno uno differente.

Alla luce di quanto detto, si può quindi iniziare a pensare che ciò che viene compiuto da una macchina quando elabora certe informazioni non sia altro che giocare un gioco linguistico, affine a quello degli esseri umani anche se diverso per alcuni aspetti e che, in quanto la macchina gioca quel gioco, per analogia, si possa parlare di comprensione da parte della macchina.

In che modo si potrebbe dire che una macchina comprende? In linea di principio, una macchina potrebbe interagire linguisticamente con gli esseri umani? Si è parlato prima di comprensione come competenza linguistica, cioè saper usare certi simboli, segni, o parole, in modo corretto. Se si pensa alla comunicazione tra esseri umani come a scambi di informazioni, dove gli scambi sono governati da regole e le informazioni sono intese come significati che vengono riconosciuti e condivisi, si può pensare alle interazioni tra macchina e uomo allo stesso modo: una macchina potrebbe interagire con un uomo perché verrebbe dotata di regole e quindi di significati che condivide con gli esseri umani. Così come gli uomini si comprendono perché condividono gli stessi significati e le stesse regole, utilizzano parole che comprendono e che applicano nei diversi giochi linguistici in modo competente, anche le macchine potrebbero comunicare con l'uomo se dotate degli stessi elementi che essi utilizzano.

Si deve pensare al gioco linguistico che si giocherebbe tra uomo e macchina come ad un gioco diverso da quello giocato tra gli uomini perché esso sarebbe caratterizzato da meno elementi linguistici, sarebbe più povero (almeno all'inizio), ma funzionerebbe comunque allo stesso modo, cioè sempre sulla base di regole che si applicano per utilizzare correttamente termini che sono stati trasmessi e appresi. Il processo di comprensione, nonostante il gioco linguistico sia diverso, sarebbe lo stesso: si direbbe che la macchina comprende se sa usare i segni in modo corretto, se dà risposte corrette e se riconosce ciò che viene detto da un essere umano allo stesso modo di come lo riconoscerebbe un altro essere umano. Si direbbe che la macchina comprende perché sarebbe linguisticamente competente.



Si può pensare che il gioco linguistico tra macchina e uomo sia un gioco linguistico primitivo: il linguaggio usato sarebbe più povero del linguaggio naturale che utilizzano gli uomini tra di loro, ma manterrebbe comunque degli elementi in comune. Se il significato di una parola si determina grazie alla grammatica del gioco in cui è inserito, si può anche pensare che gli stessi concetti di comprensione, di sapere, di comunicare, se applicati a un gioco linguistico diverso (cioè quello giocato tra macchina e uomo) possano assumere caratteri diversi, nonostante che mantengano delle somiglianze con altri giochi linguistici (cioè quelli giocati tra uomo e uomo). Il gioco linguistico che si instaurerebbe tra l'uomo e la macchina manterrebbe quindi delle somiglianze, quelle che Wittgenstein chiama somiglianze di famiglia, con il gioco linguistico giocato tra gli esseri umani, cioè l'applicazione di regole e l'utilizzo di espressioni che, grazie al modo in cui vengono usate, acquistano significati. È grazie a questa parentela, a queste somiglianze tra i giochi che si può pensare che la macchina possa interagire linguisticamente con gli esseri umani.

I giochi linguistici di cui parla Wittgenstein non hanno una essenza comune della quale tutti partecipano, tale da poter affermare che ciascun gioco ne sarebbe una particolare manifestazione.

Invece di mostrare quello che è comune a tutto ciò che chiamiamo linguaggio, io dico che questi fenomeni non hanno affatto in comune qualcosa, in base al quale impieghiamo per tutti la stessa parola, - ma che sono *imparentati* l'uno con l'altro in modi molto differenti (RF, I, § 65).

Non dire: «Deve esserci qualcosa di comune a tutti, altrimenti non si chiamerebbero 'giuochi'» - ma guarda se ci sia qualcosa di comune a tutti. - Infatti, se li osservi, non vedrai certamente qualche cosa che sia comune a tutti, ma vedrai somiglianze, parentele, e anzi ne vedrai tutta una serie (RF, I, § 66)

Tra questi giochi possibili Wittgenstein riconosce solo una "somiglianza di famiglia" fino ad affermare che, dopo tutto, «[...] i giochi formano una famiglia» (RF, I, § 67). Se tra i vari giochi esiste solo una "somiglianza di famiglia" non è possibile determinare con precisione i limiti della nozione di gioco. Si possono solo tracciare dei limiti arbitrari per stabilire ciò che può essere considerato un gioco o no. I limiti sono vaghi e difficilmente percepibili. «Infatti, - si chiede il filosofo -, in che modo si delimita il concetto di gioco? Che cosa è ancora un gioco e che cosa non lo è più? Puoi indicare i confini? No. Puoi tracciarne qualcuno, perché non ce ne sono di già tracciati» (RF, I, § 68).

Da ciò che Wittgenstein dice a proposito di cosa è il gioco non sembra difficile pensare che questo concetto di gioco può venir applicato tanto al linguaggio naturale degli uomini quanto al linguaggio-macchina e cioè, che il funzionamento della macchina può essere inteso e analizzato come gioco al pari di quelli che sono i giochi che caratterizzano il linguaggio degli uomini. Come gli esseri umani, la macchina verrebbe dotata di regole e significati necessari per comunicare ed elaborare ciò che le viene comunicato. In modo analogo agli esseri umani, si potrebbe sapere se essa comprende o no e se può interagire linguisticamente, se dimostrerà una certa competenza, cioè se le sue prestazioni saranno conformi a ciò che ci si aspetta, ovvero se si comporterà in modo corretto.

A questo punto, alla luce di quanto detto su Wittgenstein e Fodor e mantenendo l'analogia tra il funzionamento della macchina e il funzionamento della mente dell'uomo, mi chiedo se sarebbe corretto parlare di innatismo del sistema di regole che utilizza il computer per elaborare le informazioni che riceve o piuttosto di addestramento della macchina a fare queste cose. L'analogia per Fodor è fondamentale per dare ragione della tesi funzionalista secondo cui la mente dell'uomo è pensabile come elaboratore elettronico, ma quando sviluppa la sua teoria del LdP per dare ragione del funzionamento del linguaggio naturale, egli fonda la sua spiegazione sull'innatismo della sintassi. Credo però che questo discorso non possa essere sostenuto se ci si rivolge alla macchina e non più all'uomo. È difficile sostenere che il funzionamento della macchina derivi da un sistema di regole innato; semmai essa è debitrice di ciò che fa e le trasmette il programmatore (in molto molto simile a come ciò di cui viene a conoscenza il bambino è debitore di chi lo sta educando). Fodor cerca di spiegare che le macchine possono pensare, almeno in linea di principio, come gli esseri umani paragonandole a come funziona la mente dell'uomo: ma se per quest'ultimo, la teoria fodoriana si avvale della componente innatistica del sistema di regole che lavora sulle rappresentazioni del LdP, affiancata poi alla teoria causale, utile per dare ragione della semantica del linguaggio, per quanto riguarda il discorso della macchina pensante questo assunto innatista non può valere. Nelle macchine tanto la componente sintattica quanto quella semantica non possono considerarsi innate, semmai trasmesse dal programmatore, il cui ruolo sarebbe quello di addestrare la macchina a seguire certe regole nell'utilizzo di certi segni.

L'argomentazione di Fodor per mostrare l'innatezza del linguaggio risiede nell'affermare che le regole sintattiche non sono modificabili, appunto perché innate. Pertanto le relazioni causali che determinano i significati del LdP non sono modificabili. Ma questa impossibilità di variazione riguarda il LdP o il linguaggio naturale? Il LdP lavora su simboli, ma un simbolo è tale (1) se è simbolo di qualcosa, (2) se è un simbolo per qualcuno o qualcosa, che

pertanto deve già sapere che cosa sia un simbolo e come usarlo, e il sapere come usarlo non sempre coincide con il sapere cosa sia. Ma perché un simbolo sia un simbolo di qualcosa occorrerebbe che fosse inserito in un sistema in condizione di riconoscere un simbolo come tale, cioè un simbolo di qualche cosa. A questo punto parrebbe che il LdP, per poter eseguire quelle operazioni che gli attribuisce Fodor, dovrebbe essere inserito in un altro sistema che consenta al LdP di riconoscere qualcosa come un simbolo di qualcos'altro. Potrebbe instaurarsi un regressio ad infinitum, che potrebbe rendere arduo sostenere l'innatismo del linguaggio, così come delle sue regole. Se si presuppone invece che ciò di cui si avvale il linguaggio per funzionare fosse trasmesso e la cui applicazione venisse insegnata tramite un addestramento, allora l'innatismo non sarebbe più necessario per spiegare, per esempio, a cosa un simbolo si riferisce o a cosa una regola debba essere applicata.

Se si parte dal presupposto che le regole e la loro applicazione fanno parte, in modo innato, del sistema linguistico del soggetto, sia esso macchina che uomo, si può pensare all'innatismo come ad un punto di vista che pone al di fuori del sistema linguistico le regole stesse, cioè come se esse risiedessero in qualcosa che non è più linguaggio, ma che serve al linguaggio per funzionare. Se invece le regole venissero trattate come elementi acquisiti assieme ad un sistema linguistico che si sviluppa nel tempo e che si perfeziona assieme ad esse, si può pensare alla rete di regole che governa il linguaggio come inerente e facente parte del linguaggio stesso, cioè come ad elementi che non hanno più bisogno di qualcosa esterno al linguaggio che le definisca come tali, ma come ad elementi la cui definizione si dà nel momento stesso in cui si danno tutti gli altri elementi del sistema linguistico di cui esse fanno parte, cioè a qualcosa che è indiscernibile e che si crea e si perfeziona assieme al linguaggio stesso. Detto altrimenti, se le regole venissero pensate come innate allora queste dovrebbero far parte di altro rispetto al linguaggio, cioè sarebbero tali in virtù di altro rispetto al sistema in cui sono inserite, per esempio a qualcosa di fisico, come il cervello, mentre se le regole venissero analizzate come elementi acquisiti, per poi essere applicati assieme al sistema linguistico di cui fanno parte, queste potrebbero venir intese, assieme alla possibilità stessa del linguaggio, come elementi che hanno nel linguaggio ciò che le definisce e ciò che permette di riconoscerle e applicarle.

Se si vuole mantenere l'analogia tra il funzionamento della macchina e il funzionamento della mente dell'uomo si deve allora o assumere che l'innatismo vale anche per le macchine, cioè sia un presupposto utile per spiegare come esse possono interagire con l'uomo e sviluppare un loro linguaggio, o se non vale, questo presupposto non può essere ritenuto valido per spiegare il funzionamento del linguaggio dell'uomo. L'alternativa che ci si era proposti di valutare era quella per cui se l'assunto innatistico non fosse

stato teoricamente soddisfacente per dare una risposta al problema della macchina pensante, ciò che poteva essere teoricamente interessante da valutare era l'ipotesi secondo cui le macchine possono venir addestrate così come si addestra un essere umano, cioè prendere in considerazione un'alternativa che partisse dalle considerazioni di Wittgenstein a proposito dell'addestramento linguistico degli uomini. Secondo Wittgenstein tutto ciò che concerne il funzionamento del linguaggio è dovuto ad un processo di addestramento che rende i parlanti capaci e competenti ad usare i significati ed applicare le regole in modo corretto all'interno del gioco linguistico che si trovano a giocare. L'addestramento è il motore fondamentale della formazione del linguaggio naturale. Se tanto il bambino quanto la macchina, per poter iniziare a "funzionare", hanno bisogno di regole da applicare per poter utilizzare espressioni linguistiche e condividere significati con altri soggetti, allora si deve cercare di capire se può essere plausibile pensare che, così come nel bambino l'addestramento è fondamentale per far sì che egli diventi linguisticamente competente, allo stesso modo un certo tipo di addestramento possa essere altrettanto fondamentale anche per una macchina al fine di renderla capace di riprodurre la competenza linguistica dell'essere umano.

## **5.2. Turing**

A questo proposito è interessante prendere in considerazione la posizione che prese Alan Turing rispetto al problema delle macchine pensanti e al processo di addestramento di queste, nei due saggi "Macchine Intelligenti" (1948) e "Macchine calcolatrici e intelligenza" (1950). Turing prende in considerazione l'analogia con il cervello umano per spiegare come costruire e far funzionare una macchina (similmente a Fodor), ma non si ferma al mero livello fisiologico per sostenere questa analogia: lui pensa che per poter dire che una macchina è intelligente c'è bisogno di un periodo di addestramento (similmente a Wittgenstein), molto simile a quello degli esseri umani, che metta in grado la macchina di manifestare quelle caratteristiche e potenzialità che noi definiremmo intelligenti.

La sua ipotesi era che il segreto dell'enorme potenziale del cervello umano fosse legato alla nozione di universalità che lui aveva tenuto presente quando ideò la sua macchina universale. Secondo Turing, in qualche modo, il nostro cervello è realmente una macchina universale ed immaginava che questa potesse giocare a scacchi, imparare più o meno come impara un bambino e, in ultima analisi, si comportasse in una maniera che si sarebbe potuta chiamare intelligente.

Nel 1948 in “Macchine Intelligenti” Turing scrisse nel Sommario<sup>105</sup>:

Si prendono qui in considerazione e si discutono i possibili modi per manifestare alle macchine un comportamento intelligente. Come guida si usa l’analogia con il cervello umano. Si mette in evidenza che le potenzialità dell’intelligenza umana possono essere realizzate solo se viene fornita un’adeguata educazione. L’indagine si concentra soprattutto su un analogo processo di insegnamento applicato alle macchine. Si definisce la nozione di macchina non organizzata e si suggerisce l’idea che la corteccia celebrale del neonato sia di questa natura. Si presentano semplici esempi di tale macchine, e si esamina la loro educazione in termini di premi e punizioni.

In questo scritto Turing esplicita con convinzione la sua posizione circa le macchine intelligenti ed esprime un’ipotesi su come si sviluppa l’intelligenza e su come fare per riprodurla nei dispositivi meccanici. Si propone di indagare in modo generico e metaforico «se sia possibile per ciò che è meccanico manifestare un comportamento intelligente». <sup>106</sup>

Il progetto di una macchina intelligente si dovrebbe concentrare su pochi obiettivi, si dovrebbero, cioè, trovare i campi adatti del pensiero in cui la macchina possa esercitare i suoi poteri: i giochi, come per esempio, gli scacchi, l’apprendimento di linguaggi, la traduzione di linguaggi, la crittografia e la matematica. Ovviamente, sottolinea Turing, non ci si può aspettare da una macchina le stesse competenze di un adulto che ha ricevuto un addestramento per lungo tempo. L’uomo è costantemente in comunicazione con gli altri uomini e riceve sempre stimoli che costituiscono una forma di interferenza. «[...] nella misura in cui una persona è una macchina, lo è solo come macchina soggetta ad interferenze (l’interferenza è la norma, non l’eccezione)». <sup>107</sup> L’adulto è, infatti, nella condizione di tentare nuove combinazioni, apportare delle variazioni, applicarle in modi diversi, cioè si può dire che è autonomo e che si automodifica (la capacità di auto-organizzarsi era, infatti, considerata da Turing come una delle caratteristiche principali del pensiero). Se si cerca di costruire una macchina intelligente e se si è deciso di seguire il modello umano, allora si deve iniziare con una macchina che ha capacità minime di eseguire operazioni o di reagire ai comandi. In seguito, attraverso l’educazione e l’addestramento, si spera di modificare la macchina al fine di consentirle risposte precise a determinati comandi.

---

<sup>105</sup> Alan Turing, 1994, p. 88

<sup>106</sup> Ibidem

<sup>107</sup> Alan Turing, 1994, p. 105

Ciò che per Turing tiene in considerazione è l'importanza che acquisisce l'addestramento che può essere realizzato su una macchina, rispetto alla costituzione materiale della macchina stessa.

La differenza tra le lingue parlate ai due lati della manica non è dovuta ad una differenza nello sviluppo della “parte francese” e “parte inglese” del cervello, ma al fatto che i centri linguistici sono stati assoggettati a un diverso addestramento.<sup>108</sup>

Tutto questo suggerisce che la corteccia del neonato sia una macchina non organizzata, che può essere organizzata con un opportuno addestramento. L'organizzazione potrebbe risultare nella trasformazione della macchina in una macchina universale, o qualcosa del genere.<sup>109</sup>

Ciò che è fondamentale per Turing è trovare un metodo di programmazione di queste macchine affinché diventino abili per poter giocare a quello che lui presenta come “gioco dell'imitazione”<sup>110</sup>, quindi cercare di riflettere bene sulla mente umana e cercare di capire quale processo l'ha condotta allo stato in cui si trova. Si possono notare 3 componenti della mente umana con particolare sicurezza: a) lo stato iniziale, la nascita, b) l'educazione, c) le esperienze che ha vissuto. Ciò che sarebbe utile non è il tentativo di realizzare una macchina che simuli la mente di un adulto ma quella di un bambino, per poi, attraverso l'addestramento, ottenere una macchina con un cervello adulto. Il cervello infantile è una tabula rasa, una struttura dove c'è così poco meccanismo che può essere facilmente programmato; in quest'ottica la programmazione (o processo educativo) di una macchina risulta essere pressoché uguale a quello utilizzato per il bambino. Si dovrà quindi trovare un metodo di insegnamento apposito per raggiungere gli scopi. Un metodo sarebbe, per esempio, quello di associare premi e punizioni al processo di insegnamento. La macchina dev'essere costruita in modo tale che 1) non si ripetano gli avvenimenti che precedono il

---

<sup>108</sup> Alan Turing, 1994, p. 108

<sup>109</sup> Ibidem

<sup>110</sup> Questo esperimento venne in seguito chiamato “test di Turing”: esso consiste nell'ingannare un essere umano giocando al gioco dell'imitazione, inducendolo a credere di conversare con un altro essere umano e non, appunto, con una macchina. Il criterio di similitudine incorporato nel test, per valutare la somiglianza tra la macchina e l'essere umano, è la capacità di interagire linguisticamente, ovvero la capacità di usare un insieme di simboli (le lettere dell'alfabeto) e un insieme di regole (lessicali, grammaticali e logiche) per combinarli a formare parole e frasi. Tanto gli stati interni, tanto i segnali di ingresso e di uscita sono insiemi di concatenazioni di simboli; gli stati interni sono gli insiemi di proposizioni che esprimono le conoscenze possedute dall'essere umano, i segnali di ingresso sono le domande rivolte dall'interlocutore e i segnali di uscita sono le eventuali risposte date dall'essere umano in base alle sue conoscenze.

verificarsi di una punizione e 2) aumentino le probabilità di ripetizione degli avvenimenti ottenuti per mezzo del segnale di ricompensa.

Si intende che gli stimoli di dolore occorrono quando il comportamento della macchina è errato e gli stimoli di piacere quando è particolarmente buono. Con appropriati stimoli distribuiti secondo queste linee, giudiziosamente dispensati dall'«insegnante», si può sperare che il carattere tenda a convergere verso quello desiderato, cioè che il comportamento sbagliato tenda a diventare raro.<sup>111</sup>

Ovviamente l'uso di punizioni e ricompense sarà solo una parte del processo di insegnamento e sarà necessario trovare altri canali di comunicazione con i quali integrare il procedimento. Ciò che, comunque, sarà sempre importante da tenere in considerazione è lo stato della macchina in ogni momento del calcolo. Gran parte dei programmi che possiamo inserire nella macchina avranno come risultato il farle fare qualcosa che non possiamo capire, o prevedere, e che giudichiamo come comportamento casuale. Il comportamento intelligente consiste presumibilmente nello staccarsi dal comportamento prevedibile implicato nel calcolo in modo da non determinare un comportamento casuale o dei giri viziosi che si risolvono in ripetizioni inutili.

Turing pensa, quindi, che le macchine possano acquisire quell'intelligenza che è propria degli esseri umani attraverso un processo di addestramento che fa sì che la macchina risponda correttamente agli input esterni, limitando gli errori man mano che si compie questo processo. Secondo lui la macchina può imparare a comunicare con gli esseri umani se alla base c'è sì una struttura simile a quella del cervello, ma, non essendo questa di per sé sufficiente, ciò che conta per lo sviluppo di questo aspetto dell'intelligenza è l'insegnamento che può compiere il programmatore, definendo regole da applicare e segni da utilizzare per mezzo di queste regole.

Come si può vedere, l'idea che sta alla base dell'addestramento delle macchine di Turing non è molto lontana da quella di Wittgenstein. Ci sono delle regole che servono per organizzare i simboli che costituiscono il suo linguaggio, entrambi vengono trasmessi da un soggetto esterno, cioè il programmatore, il quale può educare, addestrare la macchina a comunicare con altri soggetti, in modo che ciò che viene comunicato venga compreso e ciò che viene elaborato venga fatto in modo che, con il continuo esercizio, la macchina arrivi a possedere una certa competenza linguistica, cioè risulti intelligente. Wittgenstein pensava che l'addestramento svolgesse un ruolo cruciale per comprendere il problema legato al seguire la regola tant'è che

---

<sup>111</sup> Alan Turing, 1994, p. 110

nelle *Ricerche* scrive: «Seguire una regola è analogo a: obbedire a un comando. Si viene addestrati a ubbidire al comando e si reagisce ad esso in maniera determinata» (§202). Allo stesso modo riteneva che se io comprendo un segno è perché sono stato addestrato a comprenderlo come quel certo segno. L'addestramento è ciò che rende possibile al segno l'acquisizione di un significato, nel senso che è il modo in cui si usa quel segno che gli conferisce significato ed il modo in cui si utilizza quel certo segno è il risultato di un processo educativo. Questo è un altro aspetto che differenzia la posizione di Wittgenstein da quella di Fodor, secondo il quale l'apprendimento non deriva da nessun tipo di processo educativo o esperienza: questa posizione certamente rispetta la premessa secondo cui il linguaggio naturale poggia su basi innatistiche della sintassi e della semantica, quindi mantiene una sua coerenza, ma credo che eliminare il concetto di addestramento dal problema che si sta trattando significhi di nuovo rimettere in gioco la possibilità che nella macchina esista un qualche cosa di innato che la faccia funzionare e che le permetta di eseguire correttamente le sue funzioni.

Dal momento che si è ritenuto che l'analisi fodoriana del linguaggio in riferimento all'analogia tra funzionamento della macchine e funzionamento della mente fosse poco convincente, vista l'improbabilità di poter parlare di innatismo del linguaggio-macchina, potremmo allora valutare la possibilità che l'analisi wittgensteiniana sia più interessante sia dal punto di vista della sua teoria dei giochi linguistici sia dal punto di vista del suo concetto di addestramento linguistico. Una macchina potrebbe venir addestrata a comunicare con gli esseri umani se ad essa venissero impartite le regole necessarie per poter utilizzare i significati che il programmatore gli fornisce insieme ai segni. L'essere umano ha bisogno di un altro essere umano per poter sviluppare una certa competenza linguistica, così come la macchina ha sicuramente bisogno del programmatore e di altri soggetti che gli insegnino ad interagire linguisticamente utilizzando i segni che le sono stati trasmessi. Il fatto che comprenda ciò che gli si trasmette può essere visto come gioco linguistico molto meno avanzato del nostro, così come la comprensione può essere intesa un processo molto meno raffinato rispetto al nostro. Ma un determinato processo di addestramento, affinché la macchina possa elaborare ciò gli viene trasmesso, è richiesto tanto quanto è richiesto in un essere umano che sta imparando a capire cosa gli si sta comunicando. Il fatto che un computer segua una regola per applicare un certo segno in base al contesto linguistico in cui si trova a lavorare, non deriva da un meccanismo misterioso, innato ma è il risultato di un processo educativo della macchina che il programmatore ha effettuato affinché essa possa dare risultati corretti.

Wittgenstein imputava a modelli di addestramento alternativi tra loro la possibilità della fondazione di apparati concettuali differenti. Allo stesso



modo, si può pensare che per mezzo dell'addestramento, l'apparato concettuale di una macchina sia differente dal nostro, cioè i termini quali "comprensione" e "saper fare", possano essere intesi differentemente da come li si intende nel linguaggio naturale dell'uomo, ma che mantengano sempre delle somiglianze, cioè delle affinità che ci permette di riconoscerli come processi simili, sebbene diversi.

Concludendo, si può affermare che la macchina può essere addestrata e può anche, in linea di principio, dimostrare una certa competenza linguistica che le permetterà di comunicare con gli esseri umani e comprendere ciò che essi le comunicano. La rete di significati con cui essa lavora e le regole per utilizzarli sarebbero indipendenti dalla struttura fisica del computer, nonostante questa sia necessaria affinché la macchina funzioni e questo lo pensava anche Turing, il quale, come si è già detto, ritiene processo il processo educativo più importante rispetto alla realizzazione fisica della macchina: la diversità tra le lingue per esempio è data non da una diversità fisica ma da una diversità dell'addestramento a cui è stato sottoposto il soggetto. Allo stesso modo, la competenza linguistica di un parlante è indipendente dalla struttura fisiologica del cervello, nonostante questa sia fondamentale affinché un essere umano posseda stati cognitivi. L'applicazione delle regole rispetterà le regole della logica, quindi alcune regole saranno non modificabili e inviolabili affinché il programma funzioni correttamente. Ma questo processo di applicazione e del seguire la regola non avrà una giustificazione innatistica poiché, se le regole verranno applicate in un modo o nell'altro, questo avverrà perché la macchina è stata educata, addestrata in un certo modo, cioè le saranno state trasmesse delle informazioni utili per funzionare correttamente. Le regole e le loro applicazioni non saranno, quindi, debitorie della struttura fisica della macchina, la quale, in modo innato, le possiede già precedentemente, ma del fatto che qualcuno le ha "insegnato" i metodi per accogliere certi input e a rispondere ad essi con certi output.

### **5.3. Wittgenstein e Searle**

Con l'esperimento mentale della Stanza Cinese, Searle arriva a sostenere che i computer non possono possedere stati cognitivi perché niente, che non abbia i poteri causali propri del cervello umano, può replicare il processo di comprensione tipico dell'uomo. L'uomo nella stanza non capisce il cinese: ciò che gli manca ha una spiegazione fondamentalmente biologica. Per Searle, infatti, solo le proprietà fisico-chimiche del cervello conferiscono dei poteri causali capaci di produrre i fenomeni intenzionali della nostra mente. Le macchine operano solo attraverso computazioni, cioè

manipolazioni di simboli formali che sono insufficienti per il processo di comprensione e le proprietà formali non hanno poteri causali.

Searle non nega che l'uomo sia una macchina che può pensare, ma in quanto macchina egli può pensare perché è dotato della struttura biologica propria del cervello. Una macchina, o manufatto, potrebbe pensare se e solo se in esso si replicassero le strutture neurali che caratterizzano il cervello, secondo il principio per cui se si ricostruiscono le cause si potranno riprodurre gli stessi effetti. Secondo Searle, alcuni aspetti della mente, come la coscienza e l'intenzionalità, sono semantici, ossia hanno a che fare con significati, non soltanto con simboli; una macchina opera, invece, sulla base di programmi sintattici e quindi, anche se corretta, la semplice manipolazione di simboli di cui non si conosce il significato (sintattica) non implica una loro comprensione (semantica); la sintassi non è di per sé sufficiente per la semantica. La prima non costituisce una condizione sufficiente per la comprensione ed è una sfera distaccata e indipendente dalla seconda. La semantica è caratterizzata da intenzionalità, cioè la caratteristica di un soggetto avente un cervello dotato di poteri causali. La semantica, la comprensione e il significato sono quindi tre elementi biologici, cioè non ci sarebbero se non ci fosse un cervello che ne permette l'esistenza. Gli eventi mentali sono letteralmente un prodotto del cervello e questo è per Searle un aspetto dal quale non si può prescindere se si parla di comprensione e se si analizza il problema delle macchine pensanti.

Ciò che propongo di fare è di prendere in considerazione due aspetti della posizione searliana e confrontarli con ciò che sostiene la teoria del linguaggio di Wittgenstein. I due aspetti che tratterò riguarderanno (1) l'idea secondo cui gli stati cognitivi sono causati dal cervello, il quale possiede i giusti poteri causali per l'intenzionalità e cioè per il processo di comprensione degli uomini; (2) l'idea secondo cui le regole e la loro applicazione, la sintassi, non sono un aspetto essenziale della comprensione, ma essenziale è solamente la semantica, che è indipendente dalla sintassi.

(1) Searle pensa che tutto ciò che fa parte della nostra vita mentale sia determinato da processi che avvengono nel cervello e che, quindi, tutti i fenomeni mentali sono semplicemente caratteristiche del cervello. Secondo Wittgenstein invece è profondamente sbagliato concepire la comprensione come un processo che si svolge nella mente, o che la mente svolge: secondo lui, comprendere per esempio una frase è essenzialmente essere capaci di certe prestazioni in certe circostanze e gli eventi mentali o cerebrali che sono alla base di questa capacità sono del tutto irrilevanti. Ciò che accade quando noi comprendiamo non è riducibile ad una spiegazione fisiologica, cioè non è riducibile a ciò che accade fisicamente nella nostra testa, come se dietro il processo di comprensione ci fosse un processo neurologico che ne permette lo sviluppo e ne pone le condizioni. La concezione searliana secondo cui gli

esseri umani sono in grado di comprendere perché sono dotati di una struttura biologica che causa percezione, azione, comprensione e apprendimento, per Wittgenstein, non sarebbe accettabile.

Che una certa sensazione sia prodotta da una modificazione neurologica non spiega *concettualmente* quella sensazione. Supporre che ci sia un qualche processo cerebrale nel momento in cui si comprende non spiega cosa significa comprendere quella certa cosa, perché io devo già sapere cosa intendo con una certa espressione per poter cercare se a livello fisiologico accade ciò che io credo debba accadere. Il fatto che il cervello funzioni in un certo modo quando noi agiamo, comprendiamo, apprendiamo non spiega il *perché* della comprensione e il *come* noi comprendiamo. Inoltre, la comprensione non è qualcosa che si può trovare oggettivamente in un substrato fisico. Cercando di spiegare la comprensione in termini fisici si incorre solo in una confusione concettuale, cioè una confusione tra concetti e fenomeni. Questo vuol dire che il tentativo di dare risposta a problemi di natura concettuale con metodi fisici e sperimentali non porta a soluzioni valide perché si sta parlando di due dimensioni differenti. La ricerca empirica può sì essere legittima, ma non ci spiega come e perché un certo termine si usa in quel modo e perché si accompagni a certi atteggiamenti. Tale ricerca indaga su fatti generali di natura, ma ciò che ci appare naturale, come osserva Wittgenstein, non è per questo anche necessario. La corrispondenza tra concetti e fatti generali di natura non dev'essere intesa, quindi, come ipotesi sulle cause possibili, cioè come se si volesse sostenere che se avessimo fatti naturali diversi avremmo anche concetti diversi. Una spiegazione fisica della comprensione e del linguaggio, quindi, non potrebbe essere ritenuta valida perché essa confonderebbe due livelli distinti, cioè quello concettuale e quello fattuale. Searle tratta i concetti di intenzionalità e di poteri causali come fatti naturali, mentre, in realtà, i fatti di natura non hanno rilievo concettuale. La distinzione che Wittgenstein pone tra fattuale e concettuale riguarda il fatto che le relazioni concettuali sono a priori, cioè sono proprietà del nostro modo di rappresentare il mondo e quindi non possono essere confermate o smentite dalla ricerca empirica (sarebbe come pensare di poter *scoprire* che gli scapoli sono sposati, o che non lo sono).

Ciò che, quindi, Wittgenstein vuole confutare è l'idea che possa esistere "qualcosa" dietro la nostra comprensione, qualcosa nella nostra testa, come un meccanismo fisico, che quando scatta fa sì che il soggetto comprenda. Al contrario, ciò a cui si deve guardare è il gioco linguistico entro cui si inseriscono i segni, affinché si possano cogliere i significati, cioè il loro utilizzo in relazione ad altri segni. Se noi comprendiamo qualcosa è perché abbiamo una certa competenza, un saper fare, padroneggiamo una tecnica, cioè siamo in grado di seguire regole e di riconoscere che altri le seguono.

La grammatica della parola «sapere» è, come si vede facilmente, strettamente imparentata alla grammatica delle parole «potere» ed «essere in grado». Ma è anche strettamente imparentata con quella della parola «comprendere» ('Padroneggiare' una tecnica). (RF, I, § 150)

Comprendere una proposizione significa comprendere un linguaggio. Comprendere un linguaggio significa essere padroni di una tecnica. (RF, I, § 199)

Lo scopo della filosofia di Wittgenstein non è ricercare su quali fatti di natura si fondino i nostri concetti, ma descrivere la grammatica del nostro linguaggio, cioè le regole che determinano l'uso dei segni, delle espressioni linguistiche. Il saper fare, il padroneggiare una tecnica è il risultato di un processo di addestramento ad un certo gioco linguistico, che ha messo in grado il soggetto di usare i termini in modo corretto applicando regole in modo corretto. Anche ammesso che riuscissimo a dare un senso all'idea per cui vi è un fenomeno interno del comprendere concomitante al nostro seguire una regola, perché mai il comprendere dovrebbe venir spiegato e giustificato per mezzo di fatti fisici interni? La comprensione – il posseder concetti – non può essere analizzata da una prospettiva “interna”; infatti «non analizziamo un fenomeno (per esempio il pensare), ma un concetto (per esempio quello del pensare), e dunque l'applicazione di una parola» (RF, I, § 383): “comprendere” significa “padroneggiare una tecnica”, e con ciò si intende che la comprensione non si riferisce alla classe di fenomeni naturali che accidentalmente si accompagnano a certe capacità. Non poter analizzare un fenomeno da una prospettiva “interna” significa anche che il comprendere non dipende concettualmente da qualche evento neuronale, cioè, quei processi interni potrebbero non esserci, ed esserci ugualmente comprensione.

Qual è il criterio per stabilire che quando dico mentalmente ABC faccio la stessa cosa che fa un altro, che dica silenziosamente ABC tra sé e sé? Può darsi che si scopra che nella mia laringe e nella sua ha luogo lo stesso processo. (E così pure quando pensiamo entrambi la stessa cosa, desideriamo la stessa cosa ecc). Ma quando ci sia stato indicato un processo che avviene nella laringe e nel cervello avremo forse imparato l'impiego delle parole: “dire questo e questo tra sé e sé”? Non è anche perfettamente possibile che alla mia rappresentazione del suono *a*, e alla sua, corrispondano processi fisiologici diversi? (RF, I § 376)

Per Wittgenstein, quindi, la posizione searliana non avrebbe rilevanza per quanto riguarda i problemi della mente. La sua posizione circa tali problemi viene chiaramente esposta in queste osservazioni:

Nessun'assunzione mi sembra più naturale di quella che all'associare e al pensare non è coordinato nessun processo nel cervello; e, precisamente, in modo che dai processi cerebrali sarebbe impossibile leggere i processi del pensiero. L'intendo così: Quando parlo o scrivo, allora, suppongo, non c'è un sistema di impulsi che partono dal mio cervello e sono coordinati al mio pensiero espresso a voce o per iscritto. Ma perché il sistema dovrebbe perseguire in direzione centrale? Perché questo coordinamento non dovrebbe, per così dire, scaturire dal caos? – il caso sarebbe simile a questo: – che certe specie di piante si moltiplicano per mezzo di semi, in modo che un seme produce sempre quella stessa pianta, da cui è stato prodotto; ma che nulla, nel seme, corrisponda alla pianta che diviene da esso, cosicché dalle proprietà o dalla struttura del seme è impossibile inferire a quelle delle piante che diviene da esso – che questo si può fare soltanto partendo dalla sua storia. In questo modo un organismo potrebbe divenire da qualcosa non completamente amorfo, per dir così senza cause; e non c'è nessuna ragione perché le cause non debbano stare effettivamente così con i nostri pensieri, e dunque con il nostro parlare, con il nostro scrivere, etc.. (Z § 608)

È quindi perfettamente possibile che certi fenomeni psicologici non si possano indagare fisiologicamente, perché ad essi corrisponde nulla di fisiologico. (Z, §609)

Ho visto quest'uomo molti anni fa: ora lo vedo, di nuovo, lo riconosco, mi ricordo del suo nome. È perché nel mio sistema nervoso dovrebbe esserci una causa di questo ricordo? Perché nel mio sistema nervoso dev'essere immagazzinata una cosa qualsiasi, qualunque cosa essa sia, in qualunque forma sia? Perché quell'uomo deve aver lasciato una traccia dietro di sé? Perché non devono esserci regolarità psicologiche a cui non corrisponda nessuna regolarità fisiologica? Se questo manda all'aria il nostro concetto di causalità, ebbene allora era ora che qualcosa lo mandasse all'aria. (Z, § 610)

Il pregiudizio in favore del parallelismo psico-fisico è un frutto di concezioni primitive dei nostri concetti. Infatti, se tra i fenomeni psicologici si ammette una causalità che non sia mediata fisiologicamente, si pensa con ciò di ammettere che fianco a fianco del corpo esista un'anima: una nebulosa entità spirituale. (Z, § 611)

(2) Come si è già esposto nel capitolo precedente, vari autori dell'Open Peer Commentary accusano Searle di prestare poca attenzione all'argomento dell'applicazione delle regole nel processo di comprensione. Questa

posizione è in netto contrasto anche con la posizione wittgensteiniana secondo cui la comprensione è una capacità di fare certe cose con i segni, la quale è essenzialmente connessa all'uso proprio perché è un saper fare. Essa è una capacità che si acquisisce con l'addestramento linguistico, nel senso che noi veniamo addestrati ad usare un certo termine in un certo modo, secondo certe regole, all'interno di giochi linguistici. Le regole, per Wittgenstein sono delle istruzioni, delle norme che governano l'impiego delle nostre espressioni linguistiche cosicché, se seguo la regola, la mia condotta è vincolata a quella regola.

L'aspetto del seguire la regola è essenziale affinché si possa sviluppare sia l'aspetto semantico che quello sintattico del linguaggio. Questi due aspetti sono coessenziali, cioè non si può prescindere né dall'uno né dall'altro: si costituiscono, in un certo senso, insieme. In altre parole, l'oggettività dei contenuti cognitivi è riconducibile all'oggettività delle regole del linguaggio. I primi non sono riconducibili al dominio fisico o mentale, ma coincidono con le espressioni linguistiche, considerate insieme alle regole per il loro uso. Le regole di cui parla Wittgenstein sono, infatti, regole che fanno sì che un significato si possa comprendere e che l'espressione linguistica ad esso corrispondente possa venir utilizzata nel modo corretto, cioè in modo competente. Se si è compresa la regola per utilizzare un certo termine, in un certo gioco linguistico, si è anche compreso che quel termine acquista un certo significato in relazione ad altri termini oppure che non lo acquista, o ne ha uno diverso, se in relazione ad altri. Quando si addestra qualcuno ad un certo gioco linguistico si insegnano contemporaneamente anche le regole che si applicano per utilizzare un certo termine ed è nell'agire conformemente a tali regole e cioè per il fatto che lo usa in modo corretto che noi diciamo che lo ha compreso. In questo caso, la competenza semantica è la conoscenza del significato, cioè, la competenza non è altro che il complesso di conoscenze e capacità che rendono possibile l'uso del linguaggio nel modo corretto.

Per Searle la mera manipolazione formale consiste nell'elaborazione di simboli secondo certe regole e questa non è affatto una condizione né necessaria né sufficiente per la comprensione. La macchina, infatti, non potrà mai replicare la competenza linguistica dell'essere umano perché è dotata solo di questa capacità di manipolazione formale e, per questo motivo, rimarrà sempre limitata e incapace di sviluppare una semantica adeguata. Per Wittgenstein, al contrario, questa sarebbe una condizione necessaria ma anche sufficiente perché grazie alle regole ed alla loro intrinseca possibilità di applicazione, si potrebbe riconoscere una certa competenza linguistica nel soggetto, cioè la capacità di fare certe cose con i segni. Affinché una macchina posseda la giusta competenza linguistica, i significati dei simboli verrebbero trasmessi così come vengono trasmessi agli esseri umani per mezzo di un processo di addestramento. Questi significati potranno essere compresi se

nella macchina verranno immesse anche le regole per poterli utilizzare. L'intrinseca relazione che sussiste tra regola e applicazione è da intendersi come condizione necessaria e sufficiente per la comprensione; un osservatore esterno potrà capire se una macchina comprende da come si comporta, cioè se gli output che vengono emessi in risposta a certi input sono corretti, ovvero se il suo comportamento, nella sua esecuzione, rispetta quelle regole che le sono state trasmesse. Come anche negli esseri umani, ci sono regole che proibiscono e regole che consentono l'uso di certi segni, o termini, in base al gioco linguistico che si sta giocando. Un certo segno ha acquisito un significato perché è stato più volte utilizzato all'interno di un certo gioco secondo certe regole. Anche nella macchina le cose potrebbero funzionare allo stesso modo: si inizia con giochi linguistici semplici, o primitivi, dotati di poche regole e pochi segni, ma man mano che si procede con l'addestramento, i giochi linguistici possono aumentare sia di numero che di complessità, e la macchina potrà riconoscerà da sé quando e se un certo segno potrà venir usato in quel certo contesto, quando una certa regola si potrà applicare o no: incrementando le regole e i giochi, il comportamento della macchina diventerebbe sempre più simile a quello di un essere umano ed in questo caso, la complessità del gioco linguistico che caratterizzerà la macchina sarà tale che un osservatore esterno potrebbe tranquillamente pensare che la macchina sia in grado di ragionare così come è in grado un uomo.

#### **5.4. Riconsiderare il concetto di 'comprensione'**

A questo punto però è necessario prendere in considerazione un problema che potrebbe aiutare a dare una risposta ai problemi che sono stati affrontati. L'esperimento mentale della stanza cinese è stato ideato da Searle per confermare l'idea di partenza che le macchine non possono, neanche in linea di principio, possedere stati cognitivi e che le computazioni non sono condizioni necessarie e sufficienti per la comprensione. Nel capitolo precedente si sono viste le varie obiezioni che sono state fatte contro questo esperimento e le tesi che avanzano i diversi autori. Ciò che si può notare è che ogni autore di cui si è parlato ha una sua diversa concezione di comprensione sulla base della quale modella le sue idee. La comprensione risulta spesso e volentieri un concetto di cui non si fornisce una definizione o se la si fornisce questa risulta essere poco chiara e poco precisa. Questo perché, a mio avviso, l'impiego di questo concetto è talmente esteso, cioè applicato a qualsiasi campo di studi, utilizzato in qualsiasi discorso, che ciò che risulta è una certa trascuratezza riguardo a ciò che si intende con tale concetto quando ci si imbatte in un campo come quello dell'intelligenza artificiale: non indicando

ciò che è essenziale di questo concetto, cioè gli elementi caratterizzanti questo processo, lo si lascia aperto a varie interpretazioni a seconda della teoria in cui lo si inserisce. Il problema che vorrei prendere in considerazione riguarda, dunque, questa indeterminatezza del concetto di comprensione che ciascuno utilizza per trarre le proprie conclusioni riguardo al problema delle macchine pensanti, cioè riguardo al progetto dell'Intelligenza Artificiale. Si deve cercare di dare una definizione di comprensione che sia chiara e non ambigua, cercando di individuare quegli elementi necessari e sufficienti affinché si possa dire che un soggetto comprende. Senza questa definizione si correrà ancora il rischio di dare risposte diverse ed elaborare teorie differenti a seconda di ciò che si ritiene essere la comprensione.

Tenendo presente questo aspetto, cercherò di riconsiderare il rapporto tra la comprensione e l'uomo e la comprensione e la macchina.

Lo sviluppo tecnologico e informatico è sicuramente molto avanzato e molte macchine riescono a fare cose che fino a 40-50 anni fa non ci saremmo neanche sognati di pensare, come, per esempio, interagire con esseri umani dando spiegazioni e rispondendo a domande in un modo che noi potremmo definire intelligente. Ci sono macchine che possono interagire con l'uomo per quanto concerne il loro dominio specifico – grazie a interfacce in linguaggio naturale – possono dare spiegazioni articolate di scelte operate e non operate. Le macchine di oggi, nonostante abbiano ancora enormi limiti rispetto all'uomo e alle macchine del domani, penso abbiano già eroso abbastanza i confini entro i quali Searle vorrebbe tenere protetto l'uomo e la sua esclusiva sulla comprensione. Per questo motivo credo sia necessario rivedere i confini di certe nozioni che sono invece state sempre intese come di esclusivo possesso dell'uomo. La semantica, il significato, la comprensione, possono essere riviste alla luce di tutte queste novità, evitando di mantenere una visione antropocentrica e permettendo una visione più generale di questi aspetti.

La questione si può affrontare dal punto di vista del test di Turing. Proviamo ad ipotizzare un sistema (computer più programma) che passa il test di Turing con le stesse probabilità di un uomo medio. In questo caso si avrebbe un sistema capace di dimostrare un comportamento linguistico non distinguibile da un essere umano. Cosa significa questo?

1. I racconti delle sue esperienze possono essere letti come racconti di un essere umano,
2. Le domande che esso pone sono tipiche di un essere umano,
3. L'elaborazione che esso opera sugli input provenienti da noi non sono distinguibili da quella di un essere umano, vale a dire, per esempio, che se noi gli poniamo una serie di problemi (che un uomo risolverebbe) anche il sistema li risolverebbe.



La domanda che ci si pone allora è: in che senso si può dire che un sistema di questo tipo non comprende, cioè non conosce i significati e non è dotato di semantica? Searle, come molti altri, dà molta importanza al fatto che un computer non sappia cosa sta succedendo dentro di lui quando elabora le sue mosse. Con ogni probabilità coloro che si occupano della programmazione di un computer sosterranno che il computer sa un sacco di cose; per esempio, nel caso di una partita a scacchi, affermerebbero che il computer *sa* che l'alfiere può andare in certe caselle e in altre no muovendosi in un certo modo e in altri no. Tutto dipende dal significato che diamo a "sapere" e "comprendere".

Penso che una risposta negativa alla domanda appena posta venga motivata soprattutto da una presa di posizione antropocentrica riguardo al significato di tali termini: si può dire che il sistema, cioè il computer più un programma, non conosca i significati così come li conosce un essere umano, cioè non è dotato di semantica umana, non comprende nello stesso modo e nello stesso senso di come comprende l'uomo. Ma, in un qualche senso, penso sia per lo meno interessante pensare che quel sistema conosca significati, abbia una semantica e cioè che comprenda quello che gli si sta trasmettendo e quello che sta emettendo. Il problema è proprio capire in quale senso.

Ciò che sappiamo è che si tratta di un sistema che non è un essere umano, quindi sappiamo che non comprende come fa l'uomo, ma il punto è che si comporta come si comporterebbe un uomo che comprende. Si può pensare che un punto di vista alternativo per rispondere positivamente alla domanda sia che il sistema conosca una rete di significati, possiede una semantica e che comprenda da un punto di vista comportamentista.

Searle tenta di proteggere quel *quid* irriducibile della mente umana associando termini come "coscienza" e "intenzionalità" a concetti come significato, semantica e comprensione, al fine di rendere specificamente umana tutta questa sfera concettuale. Se da una parte può apparire più problematico attribuire il pensiero e la coscienza anche alle macchine, togliendo ai relativi concetti la prerogativa di essere concetti specifici dell'uomo, dall'altro può apparire sensato, o per lo meno curioso, provare un nuovo tipo di approccio che non attribuisca più all'uomo l'esclusiva della comprensione e con essa quella del significato e della semantica, ma che attribuisca questi concetti anche a quei meccanismi che sono propri del funzionamento della macchina. In altre parole, se da una parte può essere più problematico mutuare l'aspetto semantico del concetto di "coscienza" perché con questo si è soliti indicare qualcosa che *nella nostra teoria* è unicamente pertinente all'uomo, d'altro canto non vi è nessun impedimento a estendere l'area semantica del termine "comprensione" a un nuovo dominio, che non è più ristretto alla sola sfera antropologica, ma che è quello dei sistemi artificiali

che trattano ciò che fino a ieri era sotto l'esclusiva sovranità dell'uomo: il linguaggio e i suoi simboli.

Wittgenstein aveva colto una evidente molteplicità di sensi del termine "comprendere" già nella sua applicazione a esseri umani: sono i modi di usare il termine "comprendere" che costituiscono il suo significato e che costituiscono il concetto di comprendere. I casi in cui si applica tale concetto hanno delle proprietà in comune ed è per questo che il concetto è applicabile a più casi che possono essere collegati tra loro. Qui si inserisce ancora una volta l'idea di somiglianze di famiglia. Si è già detto che le somiglianze di famiglia caratterizzano ciò che Wittgenstein denomina giochi linguistici ed è proprio grazie al concetto di somiglianze di famiglia che risulta legittimo ampliare la semantica di "comprendere".

Non posso caratterizzare queste somiglianze meglio che con l'espressione «somiglianze di famiglia»; infatti le varie somiglianze che sussistono tra i membri di una famiglia si sovrappongono e s'incrociano nello stesso modo: corporatura, tratti del volto, colore degli occhi, modo di camminare, temperamento, ecc.ecc. - E dirò: i 'giuochi' formano una famiglia. E allo stesso modo formano una famiglia, ad esempio, i vari tipi di numeri. Perché chiamiamo una certa cosa *numero*? Forse perché ha una - diretta - parentela con qualcosa che finora si è chiamato numero; e in questo modo, possiamo dire, acquisisce una parentela indiretta con altre cose che chiamiamo anche così. Ed estendiamo il nostro concetto di numero così come, nel tessere un filo, intrecciamo fibra con fibra. E la robustezza del filo non è data dal fatto che una fibra corre per tutta la sua lunghezza, ma dal sovrapporsi di molte fibre l'una all'altra (RF, I, § 67)

Si tratta in sostanza di prendere atto *wittgensteinianamente* che è emerso un nuovo uso del termine "comprensione" (perché è emerso un nuovo dominio per esso) e che tale uso è giustificato in quanto si basa su un criterio che coglie una profonda omogeneità rispetto ai due soggetti diversi (uomo e sistema artificiale) a cui è applicato.

Il termine "comprensione" si può, infatti, considerare in relazione alla teoria dei giochi linguistici. Un gioco linguistico è un linguaggio limitato, sono disponibili strutture sintattiche, certi enunciati e non altri e sono disponibili per certe funzioni, all'interno di determinate attività. Ogni gioco linguistico va considerato però come linguaggio completo. Secondo la teoria dei giochi linguistici potremmo allora dire che il computer comprende, cioè possiamo pensare che ciò che accade nel computer sia un gioco linguistico, differente da quello dell'uomo, ma che mantiene delle somiglianze. Tale

gioco linguistico chiama in causa i concetti di regola, parola, significato e addestramento, concetti che vengono facilmente individuati sia per quanto riguarda il funzionamento della macchina sia per quanto riguarda il funzionamento della mente dell'uomo. Si possono individuare così delle somiglianze di famiglia, cioè delle analogie nell'uso di termini, nell'applicazione di regole e nel processo di addestramento che viene compiuto su un uomo e su una macchina.

Se prendiamo come modello la teoria dei giochi linguistici di Wittgenstein, non è più necessario assumere quella visione antropocentrica per cui la comprensione è propria solo dell'essere umano, ma può essere estesa anche a sistemi diversi da esso, come quelli artificiali. Tutto dipende da come viene definita la comprensione. Se noi definissimo la comprensione come quel processo, consentito dalla struttura fisica del cervello, ma strutturato secondo regole che permettono l'utilizzo di segni, dal quale utilizzo si determinano i significati, i quali vengono trasmessi da un soggetto all'altro e la cui prova è data dalla correttezza delle risposte date in seguito a certi segnali (una correttezza delle azioni che è osservabile), allora si potrebbe dire che, almeno in linea di principio, i computer possono comprendere e che le computazioni, anche su simboli formali, possono venir ritenute una condizione necessaria e sufficiente per la comprensione, una condizione che si completerebbe con l'immissione dall'esterno di significati, per mezzo di un processo di addestramento, così come succede negli esseri umani.

Noi non possediamo significati precostituiti, ma li creiamo usando espressioni linguistiche secondo regole che ci vengono insegnate. Come un essere umano all'inizio non conosce significati e non possiede regole per comunicare, ma può diventare competente se viene addestrato ad applicare regole per utilizzare termini i cui significati derivano dal loro utilizzo all'interno di giochi linguistici, così le macchine, che all'inizio non possiedono né regole né significati, possono giocare il loro gioco linguistico solo se vengono loro fornite regole per elaborare simboli, possono cioè diventare linguisticamente competenti.

Concludendo, se vogliamo utilizzare la visione wittgensteiniana per trattare il problema delle macchine pensanti, la posizione di Wittgenstein sembra risultare teoricamente più interessante e apre a più possibilità teoriche. Nel senso di Wittgenstein, le macchine possono comprendere se il concetto di comprensione viene riferito a un gioco linguistico che, anche se differente per certi aspetti da quello degli esseri umani, mantiene comunque delle somiglianze. La macchina, infatti, può venir addestrata a giocare quel certo gioco, cioè ad usare certi segni secondo certe regole dimostrando così una certa competenza semantica. Essendo dotata di competenza semantica la macchina può interagire linguisticamente con l'uomo, il quale giocherà un gioco linguistico differente da quello giocato con altri uomini, ma comunque

giocato in modo competente. L'uso conforme alle regole che la macchina fa dei simboli sarà quindi la prova della sua competenza linguistica, cosicché, a certe condizioni, potremmo dire che la macchina comprende il linguaggio che usa nel senso che lo sa usare e lo usa in modo corretto: la macchina, cioè, mostra un saper fare, cioè un padroneggiare una tecnica.

## **BIBLIOGRAFIA**

- A. Paternoster, *Introduzione alla filosofia della mente*, Laterza, Roma-Bari 2008
- C. McGinn, *Wittgenstein On Meaning*, Basil Blackwell, Oxford 1984
- D. Marconi, *Filosofia e scienza cognitiva*, Laterza, Bari 2008
- D. Marconi, *Il mito del linguaggio scientifico*, U. Mursia & C., Milano, 1971
- D. Marconi, *La competenza lessicale*, Laterza, Bari 1999
- F. Ferretti, *Naturalizzare le menti: un percorso introduttivo al pensiero di Jerry A. Fodor*, Laterza, Bari 2001
- G. Lolli, (a cura di), *Alan M. Turing. Intelligenza meccanica*, Bollati Boringhieri, Torino, 2007
- J. A. Fodor, *Psychosemantics. The Problem of Meaning in the Philosophy of Mind*, The MIT Press, Cambridge 1987 [tr.it di G. Farabegoli, a cura di R. Luccio, Il Mulino, Bologna, 1990]
- J. A. Fodor, *The Language of Thought*, Harvard University Press, Hassocks 1975
- J. A. Fodor, *The Modularity of Mind. An Essay on Faculty Psychology*, The MIT Press, Cambridge 1983 [tr.it di R. Luccio, Il Mulino, Bologna 1988]
- J. Searle, "Minds, Brains and Programs", *The Behavioral and Brain Sciences*, 3, 1980, pp.417-457
- L. Wittgenstein, *Bemerkungenüber die Grundlagen der Mathematik*, Basil Blackwell, Oxford, 1956 [tr it. di M. Trinchero, Einaudi, Torino 1971]
- L. Wittgenstein, *PhilosophischeGrammatik*, Basil Blackwell, Oxford 1969 [tr.it di M. Trinchero, La Nuova Italia, Firenze, 1990]
- L. Wittgenstein, *PhilosophischeUntersuchungen*, Basil Blackwell, Oxford, 1953 [tr.it. di R. Piovesan (pp. 3-182) e di M. Trinchero (pp. 183-301), Einaudi, Torino, 1983]
- L. Wittgenstein, *The Blue and Brown Books*, Basil Blackwell, Oxford, 1958 [tr.it. di A. G. Conte, Einaudi 1983]
- L. Wittgenstein, *Zettel*, Basil Blackwell, Oxford, 1967 [tr.it di M. Trinchero, Einaudi, Torino 1986]
- P. Casalegno, *Filosofia del linguaggio: un'introduzione*, Carocci, Roma 1997 (pp. 365-385)

### **Testi e fonti non citati**

- "Introduzione all'intelligenza artificiale di Fabio Ciotti",  
[www.mediamente.rai.it/mediamentetv/learning/corsi/9912c2\\_1.asp](http://www.mediamente.rai.it/mediamentetv/learning/corsi/9912c2_1.asp)  
 M. Andronico, "Profilo di Ludwig Wittgenstein", [www.aphex.it](http://www.aphex.it), articolo del 6 giugno 2012  
 M. Barluzzi, "Wittgenstein e la natura umana", 2008, [www.tesi.cab.unipd.it](http://www.tesi.cab.unipd.it)  
 M. Davis, *Il Calcolatore Universale. Da Leibniz a Turing*, Adelphi, Milano 2003

M. Hintikka & J. Hintikka, *Investigation Wittgenstein*, Basil Blackwell, Oxford, 1986 [tr. It di M. Alasi, Il Mulino, Bologna 1990]  
M. Santambrogio (a cura di), *Introduzione alla filosofia analitica del linguaggio*, Laterza, Roma-Bari 1992 (pp. 89-133; 432-479)  
Paolo Spinicci “Lezioni sulle Ricerche Filosofiche di Ludwig Wittgenstein”, <http://www.lettere.unimi.it/dodeca/spini02/coperti02.htm>  
Rosaria Egidi (a cura di), *Wittgenstein e il Novecento. Tra filosofia e psicologia*, Donzelli editore, Roma 1996