

RAPPRESENTAZIONE DELLA CONOSCENZA

Giovanni Pezzulo

Fin dalle origini delle scienze cognitive e dell'Intelligenza Artificiale (IA) le rappresentazioni mentali sono state considerate un ingrediente centrale del pensiero. Secondo i pionieri dell'IA, come ad esempio McCarthy, Minsky e Newell, «l'intelligenza richiede conoscenza». L'attività cognitiva consiste proprio nella manipolazione di rappresentazioni interne del mondo *prima di* (ed a volte *al posto di*) agire direttamente sul mondo. A questo punto si pone una domanda: come sono formalizzate le rappresentazioni?

Secondo McCarthy (1959) un sistema intelligente deve essere capace di rappresentare sia i fatti che riguardano il mondo esterno, sia le proprie operazioni e le proprie strutture. Il formalismo deve dunque consentire di esprimere tutta la conoscenza e di eseguire calcoli su di essa. Vedremo come questa assunzione, molto focalizzata su rappresentazioni di tipo esplicito, sia stata variamente contestata nella letteratura successiva, ma tuttavia rimanga alla base di gran parte dei sistemi attuali di rappresentazione della conoscenza.

La modalità più caratteristica per gli esseri umani di trattare la conoscenza è il linguaggio naturale. Tuttavia, esso risulta avere una serie di problemi per il trattamento automatico attraverso calcolatori, il principale dei quali è quello della sua ambiguità. Ad esempio la proposizione: **la vecchia porta la sbarra** che può essere interpretata (almeno) in due sensi differenti. Vedremo che per risolvere il problema dell'ambiguità è spesso necessario riferirsi ai contesti, che sono tuttavia anch'essi difficili da esprimere formalmente. Esistono naturalmente molte differenze fra le rappresentazioni in linguaggio naturale e quelle usate tipicamente nella rappresentazione della conoscenza ad esempio in IA, che sono veri e propri linguaggi artificiali. Austin, Grice e Wittgenstein ad esempio mettono in evidenza come il linguaggio naturale sia concepito per comunicare e non per rappresentare. Le espressioni in linguaggio naturale dunque presentano caratteristiche di espressività che spesso mancano nei linguaggi artificiali, ma al tempo stesso hanno una serie di caratteristiche, come l'ambiguità e la forte dipendenza dal contesto, che renderebbero difficile ai linguaggi artificiali la rappresentazione adeguata e completa della conoscenza esprimibile attraverso il linguaggio naturale.

La conoscenza può essere rappresentata in strutture chiamate *basi di conoscenza* attraverso uno o più formalismi di rappresentazione. Nel corso del capitolo introdurremo il concetto di rappresentazione, mostrando quali siano le caratteristiche che una “buona” rappresentazione deve possedere, discutendo inoltre quali siano gli aspetti che possono motivare la scelta di un formalismo piuttosto che un altro.

I. COME SI RAPPRESENTA LA CONOSCENZA?

Una rappresentazione è un simbolo o un segno che “rappresenta” un oggetto o un evento. In IA tipicamente si assume che una rappresentazione sia un oggetto mentale, ovvero interno ad un agente, che si riferisce ad un fatto del mondo, ovvero esterno all’agente; questa posizione ha come corollario il fatto che ad una rappresentazione sia associato un certo formalismo. Ad esempio, dato il fatto del mondo: **il bicchiere è sul tavolo** la sua rappresentazione in linguaggio naturale avrebbe esattamente la stessa forma (**il bicchiere è sul tavolo**) mentre molti sistemi artificiali hanno bisogno di un formalismo più adatto al trattamento computazionale. Per questo motivo spesso gli agenti artificiali sono dotati di rappresentazioni interne ad esempio basate sulla logica come **sul(bicchiere, tavolo)** oppure **essere_sul_tavolo(bicchiere)**.

Ad ogni modo è molto rilevante distinguere fra *fatto rappresentato* e *rappresentazione*: il fatto rappresentato è un evento nel mondo (il fatto che il bicchiere sia sul tavolo), mentre possibili rappresentazioni di questo fatto sono le espressioni logiche o in linguaggio naturale precedentemente introdotte, come anche immagini o molti altri possibili formalismi.

La distinzione fondamentale fra *fatti* e *rappresentazioni* è introdotta ad esempio in IA da Newell (1982): mentre i primi sono l’oggetto che si intende rappresentare (ad esempio un evento nel mondo, ma anche una azione dell’agente), i secondi sono le rappresentazioni interne dei fatti, codificate in un certo formalismo, che un calcolatore ha e sulle quali può operare.



Figura 1 – Una rappresentazione è un simbolo o un segno che “codifica” un oggetto o un evento creando una corrispondenza fra strutture interne ed esterne all’agente cognitivo.

Tre sono le caratteristiche principali delle rappresentazioni intese come oggetti mentali. La prima di esse è chiamata da Brentano *intenzionalità*. Le rappresentazioni infatti “stanno per”, “vertono su” qualcos’altro. “Ogni fenomeno mentale è caratterizzato da [...] riferimento a un contenuto, direzione verso un oggetto (che non deve essere qui intesa come una cosa) o oggettività immanente. Ogni fenomeno mentale include in se stesso qualcosa come un oggetto [...]. Nessun fenomeno fisico mostra qualcosa di simile” (Brentano, 1985). Secondo Brentano, dunque, le rappresentazioni si riferiscono sempre a qualcos’altro; non è possibile che vi sia una rappresentazione mentale senza l’oggetto di tale rappresentazione. La seconda caratteristica fondamentale delle rap-

presentazioni può essere chiamata *distacco*: le rappresentazioni possono essere usate *in assenza* del loro riferimento. Ad esempio, posso dire che **il bicchiere è sul tavolo** anche se non lo vedo direttamente. La seconda caratteristica delle rappresentazioni, dunque, concerne il fatto che tipicamente gli agenti cognitivi agiscono su di esse al posto di agire sugli oggetti del mondo. Ad esempio, posso sapere che spostando il tavolo anche il bicchiere si sposterà, senza dover compiere materialmente tale operazione, ma solo ragionando e manipolando le mie rappresentazioni. Secondo Taylor (1971): «poter parlare di qualcosa significa essere potenzialmente consapevole di essa al di fuori di qualsiasi particolare interazione con essa; ovvero di essere potenzialmente consapevole di essa non soltanto nella sua rilevanza per una certa attività nella quale siamo in questo momento coinvolti, ma anche in una maniera distaccata (*disengaged*)». La terza caratteristica, centrale nelle scienze cognitive e nell'IA, è il potere causale delle rappresentazioni. Le rappresentazioni sono mantenute per guidare l'azione di un agente cognitivo. Mentre nell'approccio behaviorista (Watson, 1924) l'azione si riteneva guidata soltanto da semplici regole stimolo-risposta, nelle scienze cognitive si assume che fra lo stimolo e la risposta vi siano le rappresentazioni, che fungono da mediatori e da regolatori dell'azione, anche in mancanza di stimoli ambientali. Allo stesso modo nell'IA i *sistemi basati su conoscenze* agiscono in virtù di quello che sanno ed in particolare per realizzare i propri scopi, e si distinguono dunque dai sistemi puramente *reattivi* che invece agiscono soltanto sulla base di stimoli ambientali.

Secondo le scienze cognitive la possibilità di formarsi ed utilizzare rappresentazioni mentali è dunque un carattere distintivo degli agenti cognitivi; esse infatti rendono possibile parlare di *quello che non c'è* in molti diversi sensi: *quello che non percepisco ma so che c'è* (una palla non scompare se passa dietro un oggetto che la nasconde; gli oggetti non scompaiono se giro la testa dall'altra parte o se chiudo gli occhi); *quello che c'è ma non è percepibile attraverso i sensi* (questo è il caso di concetti astratti); *quello che percettivamente non c'è ma concettualmente c'è* (tutti comprendiamo la frase: **metti il vaso di fiori sul tavolo, dalla parte dove manca la sedia**; la sedia percettivamente non c'è, ma concettualmente c'è); *quello che non c'è ma vorrei che ci fosse* (questo è il caso dei veri e propri scopi i quali, per definizione, sono stati del mondo che non ci sono ancora).

Attraverso le rappresentazioni mentali è possibile dunque realizzare una serie di operazioni come categorizzare, immaginare, simulare, ragionare in astratto. Inoltre, come sottolinea Popper (1997), un agente cognitivo «lascia che siano le sue congetture a morire al posto suo»: questa capacità consiste nel poter formulare piani e valutare le conseguenze delle proprie azioni, ed in particolare quelle nocive, senza doverle necessariamente mettere in pratica. Allo stesso tempo, però, dato che le rappresentazioni hanno un carattere “distaccato” rispetto alla realtà, esse sono il presupposto per operazioni potenzialmente nocive come deliri ed illusioni.

1.1 Caratteristiche delle rappresentazioni

Ciascuna rappresentazione è, per forza di cose, un *modello semplificato del fenomeno rappresentato*, ovvero ne incorpora (attraverso la relazione di rappresentazione) alcuni aspetti a discapito di altri. Quali sono i criteri di scelta della rappresentazione? Ovviamente non esiste una risposta al problema generale, ma essa dipende dagli scopi di chi usa tale rap-

presentazione. Si tenga inoltre presente che, nel disegnare ad esempio una applicazione che consente di manipolare conoscenza, le scelte iniziali si ripercuotono successivamente su tutti gli sviluppi del modello. In questo senso, una buona rappresentazione deve avere fin dall'inizio una serie di caratteristiche:

- *Evidenziare oggetti e relazioni importanti.* La scelta del tipo di entità da rappresentare e delle relazioni fra le entità è ovviamente determinante. Tipicamente gran parte dei formalismi permette di riferirsi ad “oggetti” o a “fatti”; in alcuni di essi, come ad esempio le Reti Semantiche, è data particolare rilevanza alle relazioni fra le entità. Nella scelta delle entità e delle relazioni da modellare si deve ad ogni modo tenere ben presente quale sia lo scopo dell'applicazione, poiché vi sono molti modi di rappresentare la stessa conoscenza.
- *Fornire visioni complessive efficaci.* Uno dei punti chiave dei formalismi di rappresentazione della conoscenza è quello di permettere di rappresentare e recuperare in maniera semplice il tipo di conoscenza di cui si ha bisogno. Questa massima non vale soltanto per gli elementi, ma anche (e soprattutto) per il quadro di insieme, che deve essere funzionale allo scopo dell'applicazione.
- *Sopprimere dettagli irrilevanti.* Oltre a fornire visioni d'insieme i formalismi devono consentire una rappresentazione parsimoniosa dell'informazione. Tipicamente si assume che non tutta la conoscenza debba essere esplicitata e processata, anche se ovviamente stabilire la sua rilevanza è un compito molto complesso. Tutti i formalismi che si basano su “pacchetti” di conoscenze, come i frames e gli scripts, permettono di lasciare “sullo sfondo” una serie di conoscenze, che possono ad ogni modo essere recuperate successivamente, ad esempio con il meccanismo del “default”.
- *Consentire generalizzazioni.* Le rappresentazioni, per consentire la categorizzazione, non devono descrivere soltanto entità isolate (tokens) ma tipi (types) di entità simili, altrimenti ogni nuovo esemplare costituirebbe una categoria a sé. Per questo motivo poter generalizzare per includere entità “simili” è una caratteristica centrale dei sistemi di rappresentazione della conoscenza.
- *Essere comprensibile, completa, concisa.* Per poter essere manipolabile in maniera ottimale la conoscenza deve essere rappresentata in maniera comprensibile e parsimoniosa, pur garantendo la completezza. Come vedremo, i diversi formalismi forniscono approcci alternativi per realizzare queste condizioni. Alcuni, come gli approcci logici, sottolineano l'importanza di garantire formalmente la completezza a volte a discapito della parsimonia; altri, come gli approcci subsimbolici, utilizzano un formalismo compatto a discapito della comprensibilità.
- *Consentire una facile ed efficace manipolabilità della conoscenza.* Quando la conoscenza è organizzata in strutture come le basi di conoscenza, specialmente se esse sono molto grandi, divengono cruciali una serie di problemi: ad esempio come garantire un facile accesso all'informazione rilevante; oppure come strutturare e

memorizzare la conoscenza in maniera che essa possa essere facilmente ritrovata e modificata, evitando inoltre ridondanze ed incongruenze.

- *Essere utilizzabile anche se la conoscenza è incompleta.* Il formalismo deve permettere l'uso anche in assenza di alcune informazioni, che non sono disponibili al momento ma a volte possono essere integrate successivamente. Vi sono diversi modi di realizzare questa condizione, come ad esempio il meccanismo del default nei frames ed il processamento parallelo delle reti neurali, che permettono di completare l'input mancante.
- *Essere computabile.* Dato che la conoscenza è immagazzinata nelle basi di conoscenza per essere successivamente utilizzata mediante programmi al calcolatore, il formalismo deve consentire la computabilità, sia in termini assoluti (ovvero, non presentare funzioni non computabili) sia in termini di efficienza.
- *Avere potenza espressiva.* Il formalismo dovrebbe essere in grado di esprimere in modo efficace differenti domini di conoscenza aventi differenti caratteristiche, senza dover cambiare formalismo per ogni dominio.
- *Essere flessibile ed estensibile.* La conoscenza deve poter essere integrata in fasi successive; per questo motivo il formalismo deve consentire di aggiungere nuova conoscenza ed, in alcuni casi, di modificare quella preesistente (come vedremo questo problema si collega alla monotonicità della logica classica).

Ovviamente è molto difficile, se non impossibile, tener conto di tutti questi desiderata. Ad esempio, esiste un trade-off fra l'espressività del formalismo di rappresentazione e la complessità del meccanismo inferenziale. Il linguaggio naturale è un esempio di formalismo massimamente espressivo ma che proprio per questo motivo permette di generare espressioni vaghe ed imprecise e persino di omettere informazione. Tutte queste caratteristiche rendono estremamente difficile trattare la conoscenza espressa ed inferirne di nuova.

Ad ogni modo, a seconda dell'uso che si vuole fare delle rappresentazioni i desiderata avranno un peso maggiore o minore; l'uso dunque fornisce un criterio di scelta adeguato per il formalismo. Inoltre, qualunque sia il formalismo prescelto, vi sono una serie di problemi caratteristici della rappresentazione della conoscenza che devono essere conosciuti ed affrontati.

1.2 Problematiche della rappresentazione della conoscenza

Quando ci si appresta ad usare la conoscenza, ad esempio per descrivere una situazione, per scrivere una base di dati o per ragionare, si incontrano solitamente una serie di problemi che verranno qui introdotti per poi essere discussi nei dettagli durante il trattamento dei vari formalismi per la rappresentazione. Si tenga presente che tali problemi sono spesso collegati.

I primi tre problemi sono considerati classici del ragionamento su azioni in IA.

– Il *frame problem* (tradotto come *problema del contorno o del contesto*) (Dennett, 1978; McCarthy & Hayes, 1969) nella sua formulazione originaria riflette il fatto che nonostante sia possibile specificare tutti i fatti che cambiano in una situazione come conseguenza di una azione (ad esempio di un robot), è invece impossibile specificare tutti quelli che *non* cambiano, che sono peraltro verosimilmente molto più numerosi. Se il robot ha bisogno di conoscere lo stato di un oggetto che ha precedentemente esperito, deve sapere se esso da allora sia cambiato o meno; ad esempio, deve essere in grado di stabilire che se un oggetto si muove il suo colore non cambia. Il problema è che questa informazione rischia di non essere disponibile, poiché non è spesso possibile enumerare *tutte* le proprietà che non cambiano. Questo problema deriva soprattutto dalla forte enfasi sulla conoscenza esplicita che, come vedremo, è stato un assunto fondamentale dell'IA.

– Il *problema della ramificazione* consiste nello stabilire quali siano gli effetti indiretti di una certa azione. Ad esempio, se si acquista del pane una conseguenza diretta consiste nel fatto che successivamente si possiede del pane. Esistono tuttavia molte conseguenze indirette, come ad esempio il fatto che si possiede meno denaro; inferire tali conseguenze indirette può essere molto rilevante, specialmente dato il fatto che non tutte possono essere desiderate o desiderabili.

– Il *problema della qualificazione* consiste nello stabilire a quali condizioni un'azione abbia successo. Ad esempio, il successo dell'azione di acquistare il pane dipende da un grande numero di fattori, come ad esempio trovare il negozio aperto e possedere denaro, ma alcuni di essi possono essere difficilmente prevedibili o esplicitabili, come nel caso di ostacoli improvvisi come un terremoto.

Esistono altri problemi che sottostanno alla rappresentazione della conoscenza ed al modo in cui essa è utilizzata per ragionare.

– Il *problema della località o della rilevanza* è fortemente correlato al *frame problem* e consiste nella difficoltà di descrivere tutta e sola la conoscenza rilevante nel risolvere un problema. Teoricamente infatti per risolvere qualsiasi problema ciascuna conoscenza potrebbe essere utile, mentre ovviamente il formalismo deve consentire di processare tutta e sola la conoscenza rilevante, focalizzandosi dunque su di un sottoinsieme dell'informazione potenzialmente disponibile. Come vedremo, se un formalismo non include criteri di rilevanza esso porta ad una *esplosione combinatoriale*, in quanto il sistema cercherà di sfruttare troppa conoscenza (di fatto, tutta quella disponibile): questo è il caso di molti approcci logici che partono da un implausibile assunto di "onniscienza" che può ovviamente valere solo in mondi-giocattolo (*toy words*) molto idealizzati ma non per ragionare ed operare nel mondo reale. Allo stesso tempo, senza un criterio di rilevanza, l'acquisizione di nuova conoscenza procede in maniera cieca: se un sistema è privo di una certa conoscenza (per un certo suo scopo) deve sapere "cosa gli manca" in modo da poter cercare la nuova conoscenza necessaria. Un ulteriore aspetto di questo stesso problema riguarda non tanto la quantità di informazioni ma la quantità di risorse utilizzate: un sistema reale ha a disposizione una quantità limitata di risorse di calcolo e di tempo per

eseguire una operazione (come ad esempio derivare un teorema). Come vedremo spesso l'approccio logico si disinteressa di questo aspetto: per questo motivo alcuni problemi sono considerati potenzialmente risolvibili ma *intrattabili*, ovvero *non computabili* in un tempo umanamente accettabile. In psicologia un importante filone di ricerca sul problem solving umano si concentra proprio su questo aspetto, chiamato *bounded rationality* (razionalità limitata): gli esseri umani ragionano con risorse limitate e non illimitate.

– Il *problema della generalità*: qualunque definizione o assiomatizzazione si scelga per un concetto o per un principio, essa non coprirà tutti i casi possibili. Un esempio spesso citato è quello della relazione **sopra** (*above*): nessuna definizione di questa relazione sarà applicabile in tutte le circostanze. Come osserva McCarthy, spesso le conoscenze umane sono altamente contestuali e cercare di catturare tutti gli aspetti di un concetto con una singola definizione è una impresa vana. Vedremo come nell'affrontare questo problema emerga la necessità di ricorrere ad eccezioni o a trattare formalmente i contesti d'uso.

– Il *problema della vaghezza*, che Peirce (1897) formula in questo modo: «Una proposizione è vaga quando sono possibili stati di cose riguardo i quali è intrinsecamente incerto se, essendo stati essi contemplati dal parlante, siano considerati da lui come esclusi o ammessi dalla proposizione. Per intrinsecamente incerto non intendiamo incerto in conseguenza di una ignoranza dell'interprete, ma perché gli usi linguistici del parlante sono indeterminati». Vi sono diversi aspetti legati a questo problema. La vaghezza si riferisce alla difficoltà di stabilire dei confini netti per i concetti; Russell chiama questo aspetto "penombra". Ad esempio, nel categorizzare una serie di sedie che differiscono via via per un piccolo particolare è molto problematico comprendere dove e se ci sia un confine fra sedia e non-sedia. Zadeh (1975) ha sviluppato un "logica della vaghezza", la *logica fuzzy*, proprio per rispondere a questo tipo di problemi. Tuttavia ci sono una serie di aspetti ulteriori collegati alla vaghezza. Il primo è il problema della *generalità* (introdotto precedentemente): infatti, lo stesso simbolo si può applicare a molti oggetti nello stesso campo di riferimento. La teoria dei prototipi di Rosch (1975) spiega in parte questo fenomeno ricorrendo ad una gerarchizzazione delle categorie. Un ulteriore problema è quello dell'*ambiguità*: molti significati, appartenenti a differenti campi di riferimento, possono essere associati alla stessa forma fonetica. Questo è il caso della parola **pesc**a che assume significati molti differenti nei contesti di **frutta** e **sport**.

– Il *problema dell'incertezza*: nel mondo la conoscenza è incerta in molti sensi: il suo valore di verità può essere graduato (e non discreto) o probabile (e non certo); oppure può essere approssimata o incompleta, ad esempio perché alcune conoscenze possono non essere disponibili al momento o in linea di principio. Come vedremo sono stati proposti numerosi formalismi per far fronte alle diverse difficoltà; ad ogni modo, nelle situazioni del mondo reale esiste quasi sempre una situazione di incertezza e per questo motivo i sistemi di rappresentazione della conoscenza devono permettere di trattare conoscenza incompleta.

Infine esiste un problema che riguarda il rapporto fra le rappresentazioni mentali intese come simboli ed il mondo.

– Il problema dell'*ancoramento dei simboli* (*symbol grounding problem*): Harnad (1980) introduce il *grounding problem*, che è strettamente correlato a come i simboli debbano essere acquisiti per avere significato. Se infatti i concetti rimandano solo ad altri concetti, o i simboli ad altri simboli, come nel “sistema di simboli fisici” di Newell e Simon (1976) c'è il rischio di generare una circolarità. Per evitare la circolarità essi devono essere *ancorati* ad entità che non siano esse stesse dei simboli, ovvero direttamente al mondo. La proposta di Harnad di agganciare tutti i concetti al mondo però è sembrata per molti versi troppo forte, in quanto postula che tutti i significati debbano derivare da un processo di astrazione dalle esperienze sensoriali, una posizione molto vicina all'empirismo di Locke. Altri studiosi sottolineano l'esistenza di altri tipi di concetti; ad esempio all'interno di teorie scientifiche non tutti i concetti hanno significato perché riferiti a fatti del mondo, ma alcuni sono riferiti ad altri concetti, formando una struttura di legami concettuali via via sempre più complessa (Sloman e Chappel, 2005).

1.3 Criteri di valutazione dei sistemi di rappresentazione della conoscenza

Secondo McCarthy e Hayes (1969) i sistemi di rappresentazione della conoscenza possono essere valutati secondo due criteri fondamentali:

1. l'*adeguatezza epistemologica*, ovvero la capacità di rappresentare tutti gli aspetti del problema in esame;
2. l'*adeguatezza euristica*, ovvero l'efficienza nei tempi di processamento. Va notato che è presente un trade-off fra adeguatezza inferenziale (quello che è possibile inferire dato il formalismo) ed efficienza inferenziale (rapidità di inferenza).

Rich e Knight (1991) introducono tre altri criteri:

1. l'*adeguatezza inferenziale*, ovvero la capacità di ottenere nuove strutture a partire dalla manipolazione di quelle preesistenti;
2. l'*efficienza inferenziale*, ovvero la capacità di adattare le strategie di ricerca a seconda della nuova informazione che viene acquisita;
3. l'*efficienza di acquisizione*, ovvero la possibilità di inserire nuova informazione, sia per mezzo di operatori umani, sia per mezzo di tecniche di apprendimento automatico.

1.4 Tipi di conoscenza

Esistono vari tipi di conoscenze, che sono trattati mediante formalismi differenti. Le due distinzioni fondamentali qui presentate sono: *conoscenza dichiarativa e conoscenza procedurale; conoscenza esplicita e conoscenza implicita*.

1.4.1 *Conoscenza dichiarativa e conoscenza procedurale*

La distinzione fra conoscenza dichiarativa e procedurale, grosso modo, cattura la differenza fra “sapere” e “saper fare”. Già Russell e Ryle distinguevano fra *contenuto proposizionale* (come ad esempio **so che il mio nome è Giovanni**) e *competenza* (come ad esempio **so leggere e scrivere**). Si può definire come conoscenza dichiarativa una conoscenza espressa esplicitamente, di solito in un formato linguistico, che può essere direttamente utilizzata per fare “calcoli sulle rappresentazioni”. Ad esempio, **c’è una mela sul tavolo** oppure **Parigi è la capitale della Francia** sono tipici esempi di conoscenza dichiarativa. Non tutte le conoscenze sono di questo tipo. Ad esempio, la conoscenza su come piantare un chiodo o su come guidare una macchina coinvolgono soprattutto procedure che possiamo solo raramente e parzialmente riportare verbalmente. Si tratta in questo caso di *conoscenza procedurale*, che come vedremo è esprimibile attraverso regole procedurali del tipo **se il semaforo è rosso allora premi il freno**.

1.4.2 *Conoscenza esplicita e conoscenza implicita*

Un’ulteriore distinzione fondamentale è quella fra conoscenza *esplicita* ed *implicita*. Usualmente si definisce conoscenza esplicita quella che è esplicitamente espressa o usata, ad esempio verbalizzata in linguaggio naturale. Più problematica è la definizione di conoscenza implicita, poiché ci sono molte possibili forme di implicito. Vediamone alcuni esempi. (1) *Conoscenza inferibile* o deducibile dalle assunzioni o dalle regole che conosciamo, ma non (ancora) esplicitamente derivata (Dennett, 1987). Esempi di questo tipo sono il fatto che sappiamo che Parigi non è la capitale della Spagna anche se probabilmente non abbiamo mai esplicitato tale credenza. Questo aspetto coinvolge la nostra *conoscenza inferenziale*, visto che le rappresentazioni permettono di fare inferenze, ma di solito non facciamo tutte le inferenze che è possibile fare, ma solo quelle rilevanti. Un problema fortemente correlato è dunque quello del contesto (*frame problem*) che consiste nello stabilire cosa è rilevante. (2) *Conoscenza di sfondo*, che si assume in mancanza di ulteriori informazioni. Questa accezione è simile alla precedente, con l’ulteriore assunzione che normalmente “sappiamo” molte cose sui contesti senza doverle inferire, ma grazie al fatto che le rappresentazioni sono organizzate in “pacchetti” (e.g. frames o scripts) che contengono conoscenza per default. (3) *Conoscenza* (o procedura) *incorporata* nel funzionamento del sistema. Si tratta del “modo di funzionare” dei sistemi, che dunque non possono prescindere dal vincolare il tipo di conoscenza che viene prodotta. Ad esempio, se osserviamo un panorama con degli occhiali con lenti rosse, questo avrà una colorazione rossa a causa dei vincoli dati dal nostro modo di osservarlo. (4) *Conoscenza tacita*, che non sappiamo esplicitare: «sappiamo più di quanto sappiamo esprimere». Ad esempio Chomsky (1965) sostiene che «ovviamente ciascun parlante di una lingua ha padroneggiato ed internalizzato una grammatica generativa che esprime la sua conoscenza di tale linguaggio. Ciò non vuol dire che questi sia consapevole delle regole di tale linguaggio, né che possa diventarne consapevole». Si tratta dunque di una competenza internalizzata che non deve (o può) essere necessariamente esplicitata. (5) *Conoscenza procedurale*, ovvero “sapere come” piuttosto che “sapere che”, ad esempio espressa da regole di produzione (6) *Conoscenza* attiva ma attualmente *non sotto il*

controllo attentivo o cosciente. In questo caso un soggetto mostra sperimentalmente di avere una conoscenza attiva (ad esempio perché tale conoscenza innesca o inibisce altra conoscenza) ma non può esplicitarla. (7) *Conoscenza modulare*, ovvero incorporata in un modulo; se si assume una struttura modulare della mente, una determinata rappresentazione può essere esplicita (ed usata) anche soltanto nel contesto di tale modulo e non disponibile all'esterno (e non riferibile verbalmente).

La presenza di concetti è un carattere distintivo della cognizione; un concetto attiva rappresentazioni non soltanto di ciò che è percettivamente presente in una scena, o in un testo, ma anche di ciò che è “implicito”, in molti sensi: ad esempio ciò che è implicito, ciò che è assunto per *default*, ciò che ci si può aspettare, eccetera. A partire da queste considerazioni, si capiranno meglio alcuni formalismi presentati successivamente, come i *frames* e gli *scripts*, in cui la conoscenza è strutturata secondo “pacchetti” e che dunque permettono di tenere in conto la conoscenza di sfondo o implicita. Come vedremo i diversi formalismi per la rappresentazione della conoscenza permettono di modellare un certo tipo di rappresentazione implicita a discapito degli altri; come al solito, la scelta del formalismo dipende dunque dalle esigenze e dall'uso che se ne vuole fare¹.

1.5 Basi di conoscenza

È tipico organizzare la conoscenza all'interno di *basi di conoscenza* (*knowledge bases*, KB) e *basi di dati* (*data bases*, DB). Le basi di conoscenza racchiudono le rappresentazioni interne di un agente basato su conoscenze, ovvero l'insieme di “fatti del mondo” che esso conosce (e può comprendere, nello stile di McCarthy, una rappresentazione dichiarativa di ciò che l'agente sa fare). Per poter utilizzare la conoscenza l'agente deve dunque essere capace di interrogare la KB e ricevere una risposta, così come di aggiungere o modificare la conoscenza. Ciò può avvenire attraverso un insieme di operazioni come: **chiedi il valore di x, aggiungi la conoscenza y, cambia il valore di z in w**, ecc. La conoscenza consiste in una rappresentazione esplicita in un linguaggio simbolico; essa può contenere sia fatti su istanze (es.: **Furia è un cavallo**) che fatti su tipi (es.: **i cavalli sono animali quadrupedi**). Come vedremo, a seconda dei formalismi è possibile fare alcune ma non altre operazioni sulla conoscenza interna. Inoltre, ciascun formalismo permette un certo tipo di inferenze.

Ad ogni modo, è opportuno introdurre fin da subito una distinzione che si ritroverà in molti formalismi, quella fra ricerca “in avanti” (*forward*) ed “all'indietro” (*backward*).

¹ Gli approcci alle rappresentazioni si suddividono in due prospettive: la prospettiva esternista (o externalista) e quella internista (o internalista). La prospettiva esternista sostiene che le rappresentazioni si formino e siano utilizzate in particolare nel processo di interazione con il mondo; in questo modo l'enfasi è sul legame causale con il mondo esterno e su come le rappresentazioni servano a riferirsi ed a rimanere in contatto con esso. Allo stesso modo si enfatizzano gli aspetti intersoggettivi e sociali del significato delle rappresentazioni: ad esempio, per Saussure (1986) e per Wittgenstein (1953) il significato è funzione della comunità che lo usa ed il linguaggio è una pratica pubblica. La prospettiva internista invece fa risiedere le rappresentazioni nella mente degli individui, enfatizzando gli aspetti di autonomia dei processi cognitivi rispetto al mondo esterno. Sia Frege (1952) che Chomsky (1995) assumono che vi siano concetti tutti interni alla nostra mente, i quali possono essere idealisticamente o biologicamente innati; il linguaggio è dunque principalmente individuale.

Nel primo caso si parte dalla conoscenza che già si possiede e si deriva tutta la possibile nuova conoscenza applicando le regole in maniera iterativa. La ricerca in avanti si utilizza per conoscere tutto ciò che è derivabile dai fatti noti. Nel secondo caso si parte dal risultato che si vuole ottenere e, mediante l'applicazione ricorsiva delle regole, si cerca di verificare se è possibile ottenere tale risultato dalla conoscenza iniziale. La ricerca all'indietro si utilizza generalmente quando si conosce la meta e consente di attivare soltanto le regole dalle quali essa può essere dedotta, risultando dunque computazionalmente più rapida.

A differenza delle basi di conoscenze, le *basi di dati* contengono solo fatti positivi ed assumono una conoscenza completa del mondo (*closed world assumption*, CWA), una caratteristica che verrà investigata successivamente. Inoltre tipicamente le basi di dati forniscono un controllo sulla consistenza delle informazioni (che non possono essere contraddittorie) ma nessun meccanismo per inferire nuove conoscenze. Le basi di dati sono dunque più "passive", adatte ad immagazzinare informazioni ma non a manipolare la conoscenza di un agente.

2. CATEGORIE E CONCETTI

Come è possibile ricondurre gli oggetti di esperienza, ovvero ciò che vedo, agli oggetti di conoscenza, ovvero ciò che so? Gli esseri umani sono soliti costruire rappresentazioni mentali di fatti e situazioni in modo da raggruppare gli oggetti di esperienza in *categorie*. Gli esseri umani dunque possono riconoscere un nuovo oggetto d'esperienza riconducendolo appunto ad una categoria già nota, altrimenti bisognerebbe imparare a conoscere ciascun oggetto singolarmente. In psicologia, come sostiene Rosch, si ritiene che i correlati mentali delle categorie siano i *concetti*. In questo capitolo saranno introdotte le principali teorie psicologiche e filosofiche sui concetti, le quali hanno avuto un ruolo non secondario nello sviluppo dell'IA; ad ogni modo, i termini categoria e concetto saranno utilizzati come sinonimi.

Esiste un nesso inscindibile fra percetti e concetti: come sottolinea Kant (1781-87) i concetti permettono di categorizzare l'esperienza, mentre quest'ultima fornisce "materiale" ai concetti – che non sono dunque categorie astratte ma collegate al mondo (o, per usare un termine moderno, *grounded*). Infatti: «il percetto senza il concetto è cieco; il concetto senza il percetto è vuoto».

Gli esseri umani dispongono di un gran numero di categorie: *naturali* come quelle di predatore, preda o oggetti da arredamento; *artificiali* ed *astratte*, come quelle per gli artefatti, gli utensili e le operazioni matematiche, ma anche *sociali*, *convenzionali*, ecc. Mentre tradizionalmente nell'IA ci si è concentrati su categorie di tipo artificiale ed astratto, sulle quali è possibile operare attraverso le regole logiche di inferenza, recentemente si pone l'accento su *categorie naturali* legate dunque ad attività tipiche di un agente artificiale in un ambiente reale, come quello di navigare, riconoscere ed evitare ostacoli, eccetera. Questa impostazione segue un più generale approccio all'intelligenza, che sposta l'accento dagli aspetti astratti e non corporei a quelli legati alle attività situate (*situated*) ed immerse in un dominio naturale e sociale.

Alcune delle caratteristiche fondamentali dei concetti sono quella di rappresentare la conoscenza in maniera parsimoniosa e di permettere di fare inferenze. La prima caratte-

ristica è chiamata da Collins (1969) *economia cognitiva*: come vedremo, strutturando la conoscenza in concetti si può evitare di ricordare tutti i particolari degli oggetti e delle scene, mantenendo in memoria soltanto gli attributi più tipici di un concetto (che ne definiscono il prototipo) oppure soltanto alcuni esemplari tipici. Allo stesso tempo, i concetti permettono di inferire conoscenze che non sono immediatamente disponibili, ad esempio perché non attualmente percepibili. Ad esempio, so che una autovettura ha un motore anche senza poterlo vedere (con il cofano chiuso). Secondo Bruner *et al.* (1956) i concetti hanno anche un forte aspetto anticipatorio che poi si riflette nel nostro agire; ad esempio, se apro il cofano mi aspetto di vedere il motore; e spesso apro il cofano perché so che c'è il motore.

Una importante peculiarità dei concetti inoltre è quella di avere una *struttura interna* che può essere ad esempio utilizzata per descriverne le relazioni e le parti, per ricombinare i concetti e formarne di nuovi (Smith e Medin, 1981), o per usi metaforici. Ad esempio, il concetto di **automobile** ha le ruote come parti proprie, e può essere combinato con **arma** per formare il concetto di **carro armato** oppure utilizzato metaforicamente come in **macchina del tempo**.

2.1 Come sono strutturati i concetti? Le teorie filosofiche e psicologiche

Dagli anni '50 ad oggi nell'analisi filosofica e psicologica si sono susseguite una serie di teorie sui concetti, che qui ricapiteremo brevemente: 1) la *teoria classica* di Bruner *et al.* (1956); 2) la teoria delle *somiglianze di famiglia* di Wittgenstein (1953); 3) la *teoria dei prototipi* di Rosch (1975); 4) la teoria del *core più prototipo* di Armstrong *et al.* (1983); 5) la teoria dei *concetti come teorie* di Gopnik e Meltzoff (1997); 6) la *teoria degli esemplari* di Nosofsky (1988); 7) la *teoria dei concetti ad hoc e legati agli scopi* di Barsalou (1987).

Successivamente presenteremo la concezione dei concetti sviluppatasi all'interno della *semantica cognitiva* ed alcune ipotesi sulla nascita e lo sviluppo dei sistemi concettuali; infine forniremo una sintesi dei risultati e del loro impatto sulle teorie della rappresentazione.

2.1.1 La teoria classica di Bruner

La cosiddetta *teoria classica* si ispira al logicismo di Frege (1952) nel definire i concetti come un insieme di proprietà singolarmente necessarie e congiuntamente sufficienti. Ad esempio, il concetto di "scapolo" è definito come "maschio adulto non sposato"; sia l'essere maschio, sia adulto e sia non sposato sono attributi necessari, mentre il possedere congiuntamente le tre proprietà è sufficiente a definire uno scapolo. Sviluppata e testata empiricamente soprattutto per tener conto dei concetti "artificiali", questa teoria ha grossi problemi nel caratterizzare i concetti "natural", ovvero quelli più tipicamente utilizzati dagli umani e nel linguaggio. Il problema più evidente riguarda il trattamento delle eccezioni, che sono spesso presenti nei concetti di senso comune. Ad esempio, gli uccelli volano, ma il pinguino, che è un uccello, non vola. Il problema principale, però, è che questo approccio sembra non rendere conto degli usi quotidiani dei concetti. Come argomentano Murphy e Medin (1985), classificare

mediante regole porta a categorie artificiose: in linea di principio una categoria come **esseri con macchie, altezza superiore ad un metro e peso inferiore ai tre quintali** è perfettamente valida secondo questo approccio, mentre per il senso comune essa è del tutto implausibile.

2.1.2 *Le somiglianze di famiglia di Wittgenstein*

In opposizione alla teoria classica Wittgenstein (1953) sostiene che i concetti non siano definibili a partire da condizioni necessarie e sufficienti, ma siano invece accomunati da principi molto meno rigidi, le somiglianze di famiglia. Come i componenti di una famiglia si assomigliano tutti, pur senza condividere necessariamente gli stessi tratti, gli elementi di una categoria hanno alcuni tratti in comune, ma non altri, con altri elementi, ma non tutti, della categoria. Un esempio classico è quello del concetto di “gioco”: alcuni giochi sono competitivi, altri no; alcuni sono individuali, altri a squadre, ecc. Secondo questo approccio la definizione di un concetto diviene dunque molto più sfumata e fluida, senza confini netti.

In contrapposizione con la teoria classica oggettivista, per la quale ogni oggetto appartiene ad una sola categoria, Wittgenstein (1953) introduce inoltre il concetto di “vedere qualcosa come qualcos’altro”, sottolineando come a seconda dei contesti gli oggetti possano portare a categorizzazioni differenti. Ad esempio, un **tavolo** può essere categorizzato come **ostacolo** se si ha la necessità di attraversare una stanza, come **ripiano** se si devono posare degli oggetti. Come vedremo questo è un primo passo verso una concezione dei concetti come non assoluti ma relativi ai contesti ed agli usi².

2.1.3 *La teoria dei prototipi di Rosch*

Muovendosi da una concezione simile a quella di Wittgenstein, Rosch (1975) introduce nella teoria dei concetti un elemento nuovo rispetto alla teoria classica: quello di *tipicità* o *prototipicità*. In una serie di esperimenti emerge il fatto che alcuni esemplari tipici di una categoria siano classificati molto più rapidamente ed accuratamente di altri. Ad esempio, il pettirosso è considerato un esemplare con una tipicità maggiore della categoria **uccello**. Secondo questo approccio ciò dipende dal fatto che il pettirosso possiede un gran numero di attributi tipici degli uccelli (vola, ha una certa grandezza, ecc.). Secondo

² Per Wittgenstein (1953) il rapporto fra definizione di una parola e suo uso nel linguaggio è ribaltato: non è il significato di una parola a determinarne l’uso nel linguaggio (ad esempio: uso casa per riferirmi ad un certo oggetto perché tale è il suo significato), ma l’uso nel linguaggio a determinare il significato di una parola (**casa** si riferisce ad un certo oggetto perché il mio gruppo linguistico usa riferirsi ad esso in questi termini); peraltro l’uso non si può imporre al linguaggio, ma prende forma dalle consuetudini. Il linguaggio è una attività, una forma di vita, e molti giochi linguistici prendono forma al suo interno: descrivere un evento, raccontare una storia, ma anche dimostrare un teorema sono tutti esempi di pratiche linguistiche che hanno significato solo perché c’è una comunità che agisce secondo determinate regole. Per questo ha coniato il termine *giochi linguistici*; per riferirsi allo stretto rapporto fra pratiche sociali e linguistiche: il linguaggio si apprende solo all’interno di certe pratiche sociali ed i significati dipendono dall’uso che si fa delle parole all’interno di esse.

Rosch il prototipo è il principio organizzatore della categoria, i cui altri elementi sono giudicati a seconda della loro distanza da esso³.

Secondo la teoria del prototipo è possibile studiare le proprietà degli esemplari, come l'appartenenza ad una categoria o la vicinanza categoriale ad altri oggetti, usando leggi geometriche. I membri più tipici di una categoria sono infatti quelli che stanno al centro di queste strutture, mentre i membri sono via via più periferici se hanno uno o più attributi con valori lontani da quelli del prototipo. Inoltre le categorie *naturali*, ovvero quelle che includono oggetti naturali, hanno la proprietà geometrica di essere *convexe*: questo vuol dire che se due esemplari x ed y appartengono ad una categoria, e se un altro esemplare z è rappresentato fra di essi (in questo spazio a molte dimensioni), allora anche z appartiene alla stessa categoria.

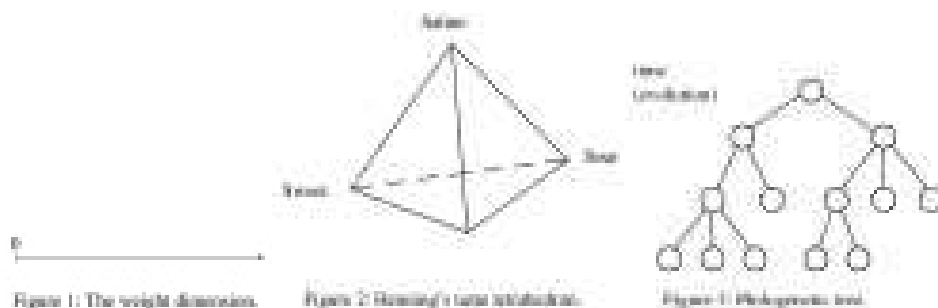


Figura 2 – Spazi Concettuali (Gardenfors, 2004).

Un assunto fondamentale della teoria dei prototipi è che, secondo Rosch (1975), i concetti sono organizzati gerarchicamente. Ad esempio, nel campo degli oggetti di ufficio, esiste un livello sopraordinato che include categorie come **mobilio**. A questo livello i membri hanno pochi attributi in comune: ad esempio, sedie e scrivanie sono mobili ma hanno pochi attributi in comune. Il livello più importante è quello dei concetti base, che comprende appunto le categorie di **sedia** e **scrivania**: fra le sedie ci sono molti attributi in comune, come fra le scrivanie. Ciò che è più rilevante, tuttavia, è che queste categorie sono molto omogenee anche per quanto riguarda gli usi (o meglio ciò che Gibson (1979) chiama *affordances*, veri e propri “inviti all’azione”). Poiché sono queste le categorie con le quali più spesso si interagisce, i concetti base sono quelli più facilmente classificabili. Vi sono inoltre le categorie subordinate, ad esempio **sedie con braccioli**, aventi molti attributi in comune non solo con oggetti della stessa categoria ma anche con quelli di categorie subordinate dello stesso concetto base.

³ La concezione “geometrica” degli spazi concettuali (*conceptual spaces*) nella quale possono esistere distanze fra concetti è formalizzata dal filosofo svedese Gardenfors (2000): le categorie sono rappresentate in spazi aventi un grande numero di dimensioni, una per ciascuno dei loro attributi (sia percettive, come forma e colore, sia funzionali, che riguardano i loro possibili usi), chiamati spazi concettuali. Le metriche per gli spazi concettuali non sono sempre le stesse; ad esempio (Gardenfors 2004) descrive il peso come un attributo monodimensionale, il sapore come un attributo a quattro dimensioni strutturate in un tetraedro (un sapore può trovarsi in un qualsiasi punto del tetraedro, ad esempio essere molto amaro e leggermente salino) e l’evoluzione come strutturata in una tassonomia, un albero che procede dal passato al futuro.

2.1.4 La teoria del “core più prototipo”

Rispetto alla teoria classica, quella del prototipo si adatta molto meglio ai concetti naturali; al contrario, per concetti di tipo artificiale o astratto risulta difficile individuare una struttura prototipica, anche perché spesso essi nascono a partire da vere e proprie definizioni. In questi casi dunque la teoria classica risulta più corretta. A partire da queste considerazioni Armstrong *et al.* (1983) propongono una teoria capace di riconciliare i due approcci, quella del *core più prototipo*.

Secondo questa teoria i concetti hanno un *core* (nucleo) (Miller e Johnson-Laird, 1976) ovvero un nucleo concettuale, una definizione che cattura gli aspetti “profondi” del concetto e permette di dare giudizi di appartenenza anche nei casi dubbi. Ad esempio, ci si può riferire al fatto di avere una certa storia evolutiva, che rende la balena un mammifero nonostante superficialmente sia più simile ad un pesce. I concetti inoltre hanno un *prototipo* che permette di esprimere giudizi di tipicità: ad esempio il pettirosso è giudicato un tipico uccello.

2.1.5 La teoria dei concetti come teorie di Gopnik e Meltzoff

Ispirandosi alla teoria del significato di Putnam (1975), Murphy e Medin (1985) sostengono che i concetti siano resi coerenti non da definizioni ma dalle “teorie ingenuè” che le persone hanno. Alla base dei sistemi concettuali vi sono dunque una serie di conoscenze di sfondo ereditate da una parte dalle abitudini linguistiche e dall'altra da una “divisione del lavoro” fra non esperti ed esperti: i primi, infatti, spesso ricorrono a concetti dei quali non capiscono bene l'estensione e, per ovviare a questa mancanza individuale, si fidano degli esperti. In primo piano c'è dunque la natura linguistica e sociale dei concetti. Gopnik e Meltzoff (1997) sostengono che i concetti siano in questo simili alle teorie: essi non sono soltanto collezioni di attributi, ma comprendono un sistema di relazioni, come ad esempio quella di causa-effetto, fra le loro parti. Un concetto, come una teoria, è più profondo se tali relazioni costituiscono un insieme coerente. Ad esempio, gli animali hanno le zampe e camminano, ed inoltre sono le zampe a consentirgli di camminare. Sono questi “principi ordinatori” a rendere possibile l'apprendimento dei concetti mostrandone i rapporti di somiglianza e le regolarità. Uno dei più potenti concetti ordinatori è quello di funzione: una bicicletta senza ruote non è un buon esemplare di bicicletta perché non può espletare la sua funzione che è quella di muoversi.

2.1.6 I modelli basati su esemplari

Secondo la teoria di Nosofsky (1988) i concetti non sono organizzati intorno a collezioni di attributi, come per i prototipi, ma su collezioni di esemplari concretamente incontrati; si tratta dunque di una teoria ancor molto vicina al modello delle rappresentazioni come *immagini mentali*. Al tempo stesso questa teoria descrive la costruzione dei concetti e la categorizzazione non come operazioni che costruiscono astrazioni (un prototipo può essere una astrazione che ha tutte le caratteristiche tipiche di un concetto), ma mediante la memorizzazione e la riattivazione delle tracce di memoria di esemplari tipici.

Un altro punto fondamentale che distingue la teoria degli esemplari da quella dei prototipi è che, poiché vengono riattivate tracce di memoria collegate ad esemplari concretamente esperiti, esse conservano (almeno in parte) informazioni sul contesto che invece andrebbero perse nel processo di ricodifica ed astrazione; a seconda del contesto, possono essere attivate differenti categorie. Le tracce di memoria inoltre sono specifiche della modalità (visiva, uditiva) utilizzata effettivamente per esperire gli esemplari.

2.1.7 Concetti “ad hoc” e legati agli scopi

Una teoria con un impianto differente è proposta da Barsalou (1983): essa non si focalizza sulla *struttura* dei concetti (una collezione di attributi, come nella teoria del prototipo, o di esemplari) ma mette in primo piano la loro *funzione*. Le teorie precedenti, infatti, non tengono conto di un aspetto fondamentale: a seconda dei contesti e delle necessità i giudizi di similarità fra oggetti possono differire in maniera significativa e gli oggetti possono essere classificati in maniera molto differente. A partire da queste considerazioni Barsalou (1983) propone una teoria molto differente, quella dei *concetti ad hoc* o *legati agli scopi*: «i concetti hanno origine in un processo estremamente flessibile che richiama dalla memoria a lungo termine un’informazione generica ed episodica per costruire concetti temporanei nella memoria operativa. Tale processo è vincolato dagli obiettivi, dal contesto e dall’esperienza recente». I concetti “ad hoc” non si organizzano intorno a principi astratti ma a scopi, e sono spesso costruiti al momento a seconda delle necessità. Ad esempio la categoria **cose utili da portare in campeggio** non deve necessariamente essere già pronta: lo scopo (andare in campeggio) viene attivato nella memoria di lavoro e funge da “attrattore” per le conoscenze nella memoria a lungo termine, determinando così una categoria che riceve forti influenze dal contesto e per questo può differire anche di molto a seconda delle circostanze. Esistono tuttavia alcune conoscenze più stabili: questo non accade perché vi sia una regola definitoria, ma piuttosto perché tali conoscenze sono state più volte associate alla categoria in contesti differenti e vi si associano dunque in maniera automatica. L’attivazione dei concetti dalla memoria è comunque basata su esemplari, in particolare sui *simboli percettivi* che Barsalou (1999) descrive in opposizione ai *simboli fisici* di Newell e Simon. Si tratta di una rievocazione delle stesse tracce sensomotorie coinvolte nella percezione originaria e dunque di una rappresentazione nella stessa modalità della percezione, senza il processo di “traduzione” tipicamente coinvolto nella costruzione dei simboli fisici. A differenza delle teorie basate su esemplari, nelle quali la categorizzazione è un processo “passivo” di confronto con gli esemplari in memoria, secondo questo approccio l’attivazione di un concetto è un processo attivo, guidato dalle aspettative. I simboli percettivi si organizzano in strutture complesse, chiamate *simulatori*, che sono capaci di generare simulazioni degli eventi attesi, riattivando le tracce di memoria corrispondenti, che possono avere anche più di una modalità (e.g. uditiva più visiva). Ad esempio, il concetto di **autovettura** è un simulatore capace di generare diversi tipi di rappresentazioni di una autovettura, a seconda delle azioni che è possibile fare su di essa: osservarla da fuori o da dentro, guidarla, ripararla, ecc. I concetti così attivati sono sempre “ad hoc” ed adatti alle circostanze,

pur restando tuttavia collegati allo stesso principio ordinatore, il simulatore, dunque ad un'unica categoria.

2.2 La rappresentazione della conoscenza alla luce dell'evoluzione della teoria dei concetti

Come si è modificato l'approccio alla rappresentazione della conoscenza alla luce dell'evoluzione delle teorie filosofiche e psicologiche sui concetti?

Come abbiamo precedentemente osservato McCarthy propone una visione astratta e non contestualizzata della conoscenza: essa è composta di fatti e di regole non adatte ad un singolo contesto ma valide in generale. È solo al momento dell'utilizzo che tali conoscenze sono contestualizzate. Tale approccio è fortemente collegato alle posizioni di Fodor rispetto ai simboli, che costituiscono un "linguaggio del pensiero" e sono arbitrari e per così dire "distillati" che non conservano alcuna caratteristica di tipo percettivo. Per questo motivo le rappresentazioni sottostanno ad un processo di "traduzione" da un formato modale (ovvero la modalità con la quale esse sono state esperite, ad esempio la visione o l'udito) ad un formato amodale, quello simbolico appunto, che ne permette una manipolazione con regole simili a quelle della sintassi del linguaggio naturale. Il processo di traduzione al tempo stesso distilla tutte le caratteristiche legate all'esperienza dell'acquisizione: dunque il contesto non viene conservato insieme al concetto. Allo stesso modo, l'impostazione fregeana poi ripresa dalla teoria "standard" dei concetti, che li categorizza a seconda di condizioni necessarie e sufficienti, ne mostra l'aspetto astratto e non legato all'esperienza.

Tutto questo impianto razionalista è recentemente messo fortemente in discussione sia da Kosslyn (1994), che mostra come le rappresentazioni mantengano una parte del loro aspetto percettuale (e per questo sono meglio definibili come *immagini mentali*), sia da numerose evidenze sperimentali che indicano come gli oggetti siano più facilmente riconoscibili nei loro contesti quotidiani che al di fuori di essi. Se dunque il precedente approccio di tipo razionalista aveva previsto per le rappresentazioni un processo di *traduzione* e *sublimazione* da percetto a concetto appunto, secondo questo approccio di tipo più empirista la traduzione non ha luogo affatto ed i concetti sono conservati insieme alle tipiche situazioni nei quali essi sono esperiti.

In contrapposizione all'approccio logico, in cui le rappresentazioni sono isolate, sono stati sviluppati formalismi come *frames* e *scripts* che cercano di associare varie rappresentazioni appartenenti allo stesso contesto, e di mostrare le loro relazioni ed i loro prototipi (Rosch, 1975). Tali formalismi permettono di stabilire quali siano le informazioni rilevanti e perciò disponibili, permettendo di affrontare (almeno in parte) il problema del contesto.

Dal punto di vista psicologico, la teoria degli esemplari di Nosofsky (1988) presenta i concetti come fortemente contestualizzati e relativi alle situazioni nelle quali essi sono stati realmente esperiti. Non c'è dunque un processo di traduzione dal percettivo al simbolico, e le rappresentazioni mantengono le loro caratteristiche percettuali.

Rimangono tuttavia due punti cruciali. Il primo riguarda il fatto che spesso tali teorie si focalizzano sulla conoscenza di tipo dichiarativo e non sull'uso degli oggetti. Le categorie sarebbero dunque raggruppate a seconda di *attributi* (come essere rosso, tondo,

ecc.) e non a seconda dei loro *usi*. Il secondo punto riguarda il fatto che effettivamente vi sono due casi di concetti al di fuori dei contesti di esperienza. Il primo caso riguarda gli oggetti che si possono utilizzare in molti contesti: ad esempio si può parlare di **sedile** in vari contesti: in cucina e dentro un'automobile ma il punto cruciale è che si tratta della stessa categoria, cosa che le teorie tutte basate sugli esemplari faticano a spiegare. Il secondo caso riguarda i concetti *astratti* (come verità e amicizia): qui è più complicato trovare "il" contesto di acquisizione, mentre l'analisi razionalista alla McCarthy spiega come gli esseri umani siano capaci di utilizzare tali concetti con disinvoltura.

Per entrambi questi problemi l'approccio di Barsalou (1999) permette di dare una risposta soddisfacente.

Egli mostra in una serie di esperimenti come vi siano categorie (chiamate "ad hoc") come quella di **oggetti utili per un campeggio** che raggruppano oggetti non tanto per i loro attributi quanto per il loro uso effettivo. Gli esseri umani sono particolarmente abili nel creare rapidamente tali categorie a seconda dei loro scopi: per questo motivo Barsalou (1987) presenta una teoria dei concetti *funzionalista* e non basata su attributi.

Allo stesso tempo egli affronta il problema dell'emergere di concetti che possono essere utilizzati in molti contesti o addirittura indifferentemente dal contesto. Egli mostra da una parte come la modalità per apprendere i concetti conservi la relazione con la situazione (come nella teoria degli "esemplari") ma poi alcuni oggetti che sono appresi in molti contesti possono in parte autonomizzarsi; si tratta dunque non di concetti non contestuali (come sostengono Fodor e McCarthy) ma piuttosto multi-contestuali.

La concettualizzazione è quindi un processo attivo che consiste nel riattivare endogenamente le rappresentazioni associate ad un episodio o un oggetto, dunque insieme al suo contesto ed in formato modale (nelle stesse modalità di acquisizione, ad esempio visiva o uditiva).

Tuttavia esiste una differenza fondamentale con la teoria degli esemplari: un concetto è un *frame multi-modale* di simboli percettivi che si organizza per formare un sistema "produttivo", dunque attivo e non passivo, che Barsalou chiama *simulatore*. I simulatori, a differenza degli esemplari, permettono di generare infiniti esemplari di oggetti legati appunto al concetto. A seconda del contesto corrente e degli usi gli esemplari generati saranno differenti, ad esempio se si tratta di riconoscere una macchina da davanti o da dietro o dall'interno – ma si tratta pur sempre dello stesso concetto di macchina.

Barsalou (1999) mostra anche come dall'esperienza possano emergere simulatori di concetti astratti come *verità*; questa è una vera e propria sfida per tutti gli approcci che si discostano dal formato simbolico, poiché mentre un approccio "simil-percettivo" o basato su immagini appare adeguato per concetti naturali esso appare molto più problematico per concetti astratti che non possono essere esperiti direttamente.

2.3 Rappresentazioni e linguaggio: orientamenti recenti

Alla luce delle recenti teorie filosofiche e psicologiche, così come delle necessità pratiche date ad esempio dall'enorme sviluppo della robotica e del World Wide Web si stanno recentemente sviluppando due nuovi approcci alla rappresentazione della conoscenza di tipo linguistico: un approccio *situato-sociale*, che si sta sviluppando principalmente nei settori dell'IA e della Vita Artificiale; ed un approccio *ontologico-semantico*, che sta

principalmente sviluppando strumenti tecnologici per gestire gli enormi quantitativi di conoscenza di cui la nostra società oggi si avvale, come ad esempio il *web semantico*. In alcuni casi questi due approcci seguono peraltro percorsi convergenti. Qui sarà trattato solo il primo dei due approcci. Il secondo verrà trattato nel capitolo 7.

2.3.1 L'approccio situato in IA

Recentemente all'interno dell'IA e della Vita Artificiale si stanno sviluppando una serie di teorie e di strumenti per il trattamento dell'informazione linguistica e per l'interazione uomo-computer (o uomo-robot) in linguaggio naturale.

Le assunzioni di base emergono dal paradigma *situato*, ovvero orientato all'azione in un mondo fisico, e *sociale*, che mette in evidenza l'importanza delle pratiche sociali per assegnare significato al linguaggio, sulla scia di Wittgenstein e Vygotsky. Questo approccio si allontana molto dall'impostazione logicista e sviluppa nuovi tipi di rappresentazioni, in particolare fortemente orientate all'azione e non come "specchio" del mondo. Un altro assunto fondamentale è l'importanza dell'apprendimento (delle rappresentazioni e del linguaggio) attraverso l'interazione con il mondo e con gli altri, che si oppone al paradigma *istruzionista* della prima IA secondo il quale la rappresentazioni di cui calcolatori e robot si servono sono costruite dal programmatore.

Il nuovo paradigma che si afferma (Clark, 1997) nega esplicitamente gran parte delle precedenti premesse teoriche e fornisce una versione molto differente degli assunti funzionalisti di partenza. Innanzitutto, le modalità del pensiero sono fortemente intrecciate con l'agente conoscente e la categorizzazione dipende in larga misura dalle strutture conoscitive: «il pensiero è dipendente dalla corporeità (*embodied*)». Viene inoltre riconosciuto un ruolo alla struttura del mondo che guida il formarsi delle rappresentazioni e delle categorizzazioni: il pensiero non è solo manipolazione di simboli astratti, ma si fonda su un precedente sostrato di significato residente nel mondo: «il pensiero ha struttura ecologica e non del tutto astratta».

Parallelamente all'interno della "nuova IA", a differenza da quanto avvenuto in precedenza, la principale area di indagine non sono i compiti astratti come il gioco degli scacchi, ma l'*azione situata*, ovvero compiuta in un dominio reale come il mondo fisico, ad esempio da un robot che ha sensori ed effettori. In questo senso, specialmente in contesti applicativi come quelli legati ai robot, la rappresentazione è vista come fortemente correlata all'azione ed all'uso. In particolare, è centrale il concetto di *grounding*: mentre tradizionalmente l'assegnazione del significato veniva considerata una operazione tutta interna al processamento dell'informazione (simbolica o linguistica) qui invece si riconosce che il significato può venire solo attraverso il rapporto con il mondo, dunque al di fuori della sfera dell'informazione.

Ad esempio, Roy (2005; ma si veda anche Roy, Hsiao, Mavridis, Gorniak, 2006) descrive una serie di esperimenti condotti con il robot Ripley, che consiste in un braccio mobile ed in una telecamera. Ripley agisce in un mondo composto da oggetti come tazze e mele, ma nel quale è presente anche uno sperimentatore che può dargli dei comandi verbali. Ripley mantiene un modello interno del suo ambiente attraverso una simulazione 3D, che gli consente di rappresentarsi anche gli oggetti fuori dal suo campo visivo (sebbene, se l'ambiente cambia, non necessariamente il modello interno

sarà aggiornato). Ripley inoltre sviluppa delle rappresentazioni sensomotorie degli oggetti, dei loro attributi, delle loro relazioni e delle azioni che può svolgere. Tali rappresentazioni sono grounded ed assumono significato attraverso l'interazione con il mondo e le aspettative. Ad esempio, gli aggettivi **pesante** o **leggero** sono compresi nei termini della resistenza attesa durante la manipolazione di un oggetto, mentre **rosso** e **verde** sono compresi nei termini delle aspettative sullo spazio (matematico) dei possibili colori così come misurati durante la visione di un oggetto. Verbi come "spostare" e "prendere" sono compresi nei termini delle possibili azioni e dei loro esiti attesi. Operando su tali rappresentazioni, ad esempio componendole, Ripley può costruire rappresentazioni complesse di oggetti e delle loro relazioni ed è capace di interpretare ed eseguire comandi come: **prendi la tazza rossa alla tua sinistra**. Per Ripley le asserzioni linguistiche acquistano significato in virtù del fatto che esso è capace di correlarle alle sue rappresentazioni sensomotorie, dunque ad una conoscenza *grounded*, basata sulle sue possibilità di azione sul mondo. Come si può notare questo approccio è molto vicino al concetto di *affordance* (invito all'azione) di Gibson (1979) che descrive il mondo nei termini delle possibilità che offre all'azione: ad esempio, una **tazza** è **afferrabile**, **spostabile**, **riempibile**, ecc. In maniera simile Narayanan (1997) descrive un'implementazione, quella degli *X-schemas*, che enfatizza le rappresentazioni sensomotorie e le organizza mediante conoscenza categoriale sfruttando gli "schemi" descritti nella semantica cognitiva (Lakoff, 1987).

Una differente linea di ricerca, più focalizzata sugli aspetti sociali, è quella di Steels (1996) che cerca di far "emergere" il linguaggio attraverso una serie di "giochi linguistici" fra robot che dunque costruiscono ed imparano una lingua interagendo non solo con oggetti del mondo ma anche fra di loro. L'enfasi dunque si sposta sulle pratiche sociali alla base non solo della semantica ma persino della grammatica delle lingue, fino a costruire nuovi "linguaggi naturali" per i parlanti-robot. Al tempo stesso, all'interno dell'IA più tradizionale si assiste ad una apertura alle problematiche sociali ed ecologiche attraverso lo sviluppo dei programmi di ricerca dell'IA Distribuita e dei Sistemi Multi Agenti, dove l'interesse si sposta verso lo studio del comportamento intelligente che deriva dalla cooperazione di un gruppo di agenti autonomi più semplici. Dunque diventano centrali tematiche di tipo sociale, come cooperazione (Huhns e Singh, 1998; Wooldridge e Jennings, 1995), fiducia, delega, autonomia (Castelfranchi, 1995), condivisione della conoscenza e piani condivisi (Grosz e Sidner, 1990).

Si deve inoltre segnalare la presenza di una corrente "anti-rappresentazionalista" che tende a minimizzare o eliminare del tutto il ruolo delle rappresentazioni interne. In opposizione all'IA classica, che enfatizzava il ruolo delle rappresentazioni e delle manipolazioni simboliche, Brooks (1991) propone il motto «intelligenza senza rappresentazione e senza ragionamento» e mostra come sia possibile costruire robots che esplorano un ambiente sconosciuto, evitano ostacoli ed imparano a «tornare a casa» anche senza fare a meno di rappresentazioni esplicite, ma soltanto con comportamenti di tipo reattivo. In maniera simile i proponenti della teoria dinamicista (Beer, 1995; Keijzer, 2001; Port e van Gelder, 1995) cercano di mostrare come molti compiti considerati cognitivi possano essere risolti da semplici sistemi di feedback in continua interazione con il mondo (coupling sensomotorio), senza il bisogno di rappresentazioni interne o di manipolazione di simboli. Dal versante psicologico O'Regan e Noe (2001) sotto-

lineano l'importanza dell'interazione con il mondo anche per la percezione: secondo tali autori, l'esperienza percettiva non consiste nel rappresentare internamente una realtà esterna, ma nell'interagire dinamicamente con la realtà esterna stessa. Allo stesso tempo Churchland (1986) propone un approccio ancora più radicale, «eliminativista» rispetto alle rappresentazioni. Tuttavia, nonostante questa serie di critiche serrate al concetto stesso di “rappresentazione” come utilizzato nelle scienze cognitive e nell'IA, esso continua ad avere un ruolo centrale e vi sono numerosi tentativi di riformularlo e specificarlo meglio (Clark, 1997).

3. FORMALISMI PER LA RAPPRESENTAZIONE DELLA CONOSCENZA

Vi sono molti modi di rappresentare la conoscenza. Prima di descrivere, nei prossimi paragrafi, i formalismi utilizzati in IA ed in linguistica per rappresentare la conoscenza, è opportuno introdurre alcune distinzioni di base.

Tipicamente quando si pensa ad una rappresentazione essa è concepita o come una immagine mentale di qualcosa di precedentemente esperito attraverso i sensi, oppure come una rappresentazione più astratta, potremmo dire in formato proposizionale o linguistico, del tipo **c'è una mela sul tavolo**. Negli anni '80 c'è stato un lungo dibattito su quale sia il formato più adatto per le rappresentazioni, in particolare fra i fautori delle *rappresentazioni proposizionali* (Pylyshyn, 1984) e quelli delle rappresentazioni come *immagini mentali* (Kosslyn, 1994). Il dibattito oggi è arricchito dalle evidenze di tipo neurobiologico e dall'uso in IA di molti diversi formalismi di rappresentazione della conoscenza: basate su logica, strutturate, procedurali, analogiche, modelli mentali, rappresentazioni neurali, etc. Ogni formalismo permette una serie di operazioni e presenta al tempo stesso una serie di vincoli. Nell'IA, fin dalle origini alla metà degli anni '50, l'intelligenza è stata concepita come la capacità di risolvere problemi, e si è data dunque grande importanza alle strategie di ragionamento, concepito essenzialmente come manipolazione sintattica di simboli, seguendo l'approccio del *sistema fisico simbolico* di Newell e Simon (1976); di qui il nome di *approccio simbolista o simbolico*.

Il formato simbolico dei dati permette di operare con grande semplicità e versatilità su strutture di dati complesse. All'interno del paradigma simbolico è opportuno rintracciare almeno due percorsi modellistici.

Storicamente il primo tipo di modello proposto ed implementato è focalizzato sulla *conoscenza procedurale*: si tratta di un *sistema di produzioni*, ovvero un set di regole condizione-azione: l'azione è attivata quando una condizione è soddisfatta. Si tratta del *General Problem Solver* (GPS) di Newell, Simon e Shaw (1958), capace di risolvere, mediante produzioni ed alcune euristiche, problemi formalizzati di tipo matematico; ma come suggerisce il nome, nelle intenzioni degli autori il GPS fornisce un modello di tutta l'attività di *problem solving*; se l'attività razionale umana può essere ricostruita come analisi e soluzione di problemi, ad un certo livello di astrazione si tratta della meccanica di base della razionalità. Dunque ciò che è prioritario è il processo, l'algoritmo, in particolare la minimizzazione della distanza fra stato origine e stato obiettivo (goal). Su queste basi la cognizione è una attività formale ed astratta di processamento di simboli, non importa cosa essi rappresentino e quale sia il loro contenuto seman-

tico. Un discendente diretto di questa impostazione modellistica, e della conseguente priorità data alla ricostruzione razionale dei processi di pensiero ed al problem solving, è State, Operator, and Result (SOAR) di Newell e Laird, che integra meccanismi per la presa di decisioni, l'applicazione ad una gerarchia di goal e subgoal ed il learning di produzioni, secondo quanto descritto nella *Unified Theory of Cognition* di Newell (1990). Fra gli eredi di questa tradizione vi sono inoltre i sistemi esperti.

Il secondo percorso modellistico è più focalizzato sulla conoscenza di tipo *dichiarativo*. I primi formalismi utilizzati in questo senso sono quelli di tipo logico, con molte varianti. Come naturale espansione dei modelli più tradizionali, ed in risposta ad una serie di problemi difficili di rappresentazione della conoscenza (ragionamento di senso comune; *frame problem*) si sviluppano modelli e formalismi che si discostano dalla logica classica per vari aspetti: in modo da tener conto di conoscenza di *default*, situazionale, dinamica, pragmatica, eccetera. Inoltre la semantica interna dei sistemi e l'agire sul mondo si riavvicinano in qualche modo, attraverso un processo di integrazione fra conoscenza dichiarativa e procedurale. In un certo senso si tratta di spostare il carico computazionale e persino cognitivo sul formato dei dati, alleggerendo la computazione grazie ad una più sofisticata rappresentazione della conoscenza. Allo stesso tempo, ciò sposta l'attenzione sulle strutture semantiche della cognizione, dunque sul contenuto dei simboli e sulla loro modalità di aggregazione precedente alla computazione stessa. Fra i formati di dati più rilevanti, i *frames* di Minsky (1975), gli *schemata* di Rumelhart ed Ortony (1977), gli *scripts* di Schank ed Abelson (1977), gli *schemi* di Drescher (1991) per la rappresentazione di oggetti, eventi, e sequenze di azioni. Nella prospettiva dell'attenzione alla verosimiglianza, i principali referenti nel versante psicologico sono i lavori sugli schemi di Piaget (1936) e Bartlett (1932). Sebbene tali formati siano differenti, essi hanno in comune un meccanismo di rappresentazione per *default*; la possibilità di rappresentare classi astratte e loro relazioni e non solo istanze grazie all'uso di variabili; la possibilità di incassarsi e di costruire rappresentazioni a diversi livelli di astrazione.

Di seguito introdurremo una serie di approcci alla rappresentazione della conoscenza, mettendone in evidenza le peculiarità.

3.1 Gli approcci di tipo logico

Molti formalismi di rappresentazione della conoscenza si basano sull'uso della logica; in IA (ad esempio nel lavoro di uno dei pionieri, McCarthy) l'approccio logico è stato il primo ed il più rilevante a causa delle interessanti proprietà della logica come quello di consentire di determinare interamente la sintassi e la semantica e di permettere, mediante un apparato deduttivo generale, di generare un numero illimitato di possibili espressioni a partire da un numero limitato di postulati di partenza.

Nell'approccio logico rappresentare la conoscenza in un dominio significa tradurla in un insieme di fatti che descrivono il mondo e di regole che permettono di manipolare tali fatti. Il ragionamento consiste dunque in tale manipolazione; secondo Frege (1952), infatti, le leggi del pensiero sono le leggi della logica. Come vedremo, la plausibilità psicologica dell'approccio logico è stata fortemente contestata; tuttavia esso è frequentemente utilizzato in IA a causa della sua eleganza e praticità.

3.1.1 Calcolo proposizionale

L'approccio logico, la cui forma più semplice è il *calcolo proposizionale*, rende possibile rappresentare eventi del mondo semplici come **contiene(bicchiere, acqua)** oppure complessi come **contiene(bicchiere, acqua) AND sopra(bicchiere, tavolo)**. Le espressioni complesse si costruiscono a partire da quelle semplici utilizzando una serie di connettivi logici come **AND** (congiunzione), **OR** (disgiunzione), **NOT** (negazione), \Rightarrow (implicazione)⁴.

Il valore di verità di una espressione complessa dipende dal valore di verità dei suoi componenti in funzione dei connettivi usati; ad esempio, se sono veri sia **contiene(bicchiere, acqua)** che **sopra(bicchiere, tavolo)** allora sarà vera anche l'espressione **contiene(bicchiere, acqua) AND sopra(bicchiere, tavolo)**. Il modo canonico di attribuire una semantica alle proposizioni composte è quello delle così dette *tavole di verità* che permettono di stabilire quale insieme di interpretazioni soddisfi una certa formula. Ad esempio, data la formula **a OR b** essa sarà vera se **a** o **b** sono veri, falsa altrimenti. La formula **a AND b** sarà invece vera se e solo se entrambi **a** e **b** sono veri, falsa altrimenti.

3.1.2 Calcolo dei predicati

Il *calcolo dei predicati* o *logica del primo ordine*, che è una estensione del calcolo proposizionale, è spesso inteso come una formalizzazione del linguaggio naturale che ne evita le ambiguità ma al tempo stesso ne riduce il potere espressivo. Esso permette di descrivere la conoscenza come composta di oggetti, proprietà e relazioni fra oggetti. Successivamente è possibile manipolare tale conoscenza, ad esempio compiendo inferenze di tipo deduttivo, in un linguaggio espressivo e con una semantica ben definita.

Nel calcolo dei predicati un predicato, come ad esempio: **sopra(bicchiere, tavolo)** è una formula atomica che descrive un fatto elementare del mondo. I predicati possono essere ad uno o più posti (il termine usato è *arietà*); ad esempio il predicato **mortale(x)** ha arietà uno, mentre **sopra(x, y)** ha arietà due.

A partire dai singoli predicati si possono costruire poi espressioni via via più complesse utilizzando gli stessi connettivi del calcolo proposizionale. Esistono anche due quantificatori, espressi con i simboli \forall ed \exists , chiamati *quantificatore universale* e *quantificatore esistenziale*. Il primo sta a significare “per ogni”, come nell'espressione **per ogni oggetto x è vera la proprietà p**; il secondo sta a significare *esiste* come nell'espressione: **esiste almeno un oggetto x per il quale è vera la proprietà p**. I due quantificatori possono essere espressi l'uno nei termini dell'altro in quanto $(\exists x) P(x)$ equivale a **NOT(($\forall x$) (NOT P(x)))** e $(\forall x) P(x)$ equivale a: **NOT(($\exists x$) (NOT P(x)))**. Attraverso i quantificatori è possibile esprimere conoscenze come **tutte le città vicino al mare sono calde** in questo modo: $\forall x(\text{città}(x) \text{ AND vicina_al}(\text{mare}, x) \Rightarrow \text{calda}(x))$.

Un enunciato atomico è vero o falso a seconda che la relazione in esso espressa fra il predicato e i suoi argomenti sia vera o falsa. Ad esempio, **sopra(bicchiere, tavolo)** è vero se c'è un bicchiere sul tavolo. Il valore di verità di espressioni complesse dipende solo dal valore di verità delle loro parti, secondo le regole correlate ai connettivi ed ai quantificatori. Esiste una corrispondenza biunivoca fra predicati e quella di *funzioni*.

⁴ Si noti che esiste una equivalenza formale fra $a \Rightarrow b$ e **(NOT a) OR b**.

Le funzioni stabiliscono le condizioni che rendono vera la relazione fra i propri termini; per esempio, se **sopra(bicchiere, tavolo)** è vero allora esiste una funzione **sopra(x,y)** tale che **sopra(bicchiere,tavolo)** vale **vero**.

Il principale meccanismo inferenziale del calcolo dei predicati è la *deduzione*: a partire da un insieme di conoscenze iniziali, essa permette di esplicitare tutta la conoscenza implicitamente contenuta. Ad esempio, dalle conoscenze **Socrate è un uomo** e **tutti gli uomini sono mortali** si può dedurre la conoscenza **Socrate è mortale**. Purtroppo questo meccanismo inferenziale non è decidibile ma semidecidibile: ciò vuol dire che non esiste un algoritmo che sia sempre in grado di stabilire se una asserzione nel calcolo dei predicati sia vera o falsa in quanto, nel caso in cui essa sia falsa, il tempo di calcolo della risposta potrebbe essere indefinitamente lungo.

Un ulteriore meccanismo inferenziale è la *risoluzione*, che è possibile se si utilizza un sistema a clausole come il PROLOG, che sarà presentato successivamente.

3.1.3 Limiti del calcolo dei predicati

Il problema principale del ragionamento attraverso i metodi di deduzione logica consiste nella *esplosione combinatoria*. I sistemi logici sono infatti basati sul presupposto dell'onniscienza: tutta l'informazione (compresa quella derivabile) è data per conosciuta ed è disponibile per essere usata. I sistemi logici inoltre non forniscono un criterio di rilevanza dell'informazione: in questo senso non forniscono una indicazione di quale sia l'informazione rilevante da utilizzare e nemmeno di quali fra le inferenze possibili siano quelle rilevanti. Dal punto di vista di un sistema computazionale ciò si traduce in una mole di informazioni enorme che rende pressoché impossibile da computare un gran numero di problemi interessanti: se tutte le conseguenze delle conoscenze possedute devono essere derivate, senza un criterio di rilevanza, il tempo di calcolo cresce in maniera esponenziale.

Un altro attributo molto rilevante della logica classica, che la rende al tempo stesso potente in certi ambiti e limitata in altri, è la sua *monotonicità*: se vengono aggiunte nuove conoscenze queste non possono contraddire (o invalidare) la conoscenza preesistente. Dunque, non è possibile modificare la conoscenza ma solo integrarla. Un problema di questa assunzione è che, come mette in evidenza Minsky, tipicamente nel senso comune si fanno inferenze sulla base di conoscenza incompleta, e dunque le conclusioni sono spesso successivamente riviste sulla base di nuova conoscenza. La revisione della conoscenza è una pratica comune (e piuttosto complessa) nel ragionamento umano. A partire da questo problema è stata sviluppata una nuova famiglia di logiche, quelle *non monotone*.

Un'altra caratteristica molto importante dell'approccio logico, messa in evidenza come un suo vantaggio da McCarthy ma successivamente fortemente contestata, è quella della *non contestualità*. Le rappresentazioni (fatti e regole) sono mantenute dal sistema in formato dichiarativo indipendentemente dall'uso che se ne farà. È soltanto al momento di utilizzarle che esse saranno adattate alla situazione corrente. Questa caratteristica consente una certa parsimonia e generalità nel mantenere le rappresentazioni, poiché non si devono mantenere separatamente le conoscenze di casi differenti. Tuttavia essa comporta una serie di problemi, primo fra tutti il problema di dover adattare rapidamente le rappresentazioni ai differenti contesti, visto che a seconda di essi la concettualizzazione può variare anche radicalmente. Recentemente sono state introdotte una serie di logiche adatte a trattare i contesti (Giunchiglia, 1993).

Un ulteriore problema della rappresentazione di tipo logico è che se anche si dispone di conoscenza completa, esistono spesso delle *eccezioni*. Ad esempio, tipicamente gli uccelli volano, ma non se sono dei pinguini o se hanno un'ala spezzata. Questo problema riflette la crisi in psicologia del modello "classico" dei concetti, secondo il quale essi possono essere definiti tramite condizioni necessarie e sufficienti. A partire da questo problema sono state sviluppate logiche alternative come la logica default e quella non monotona.

Per evitare le limitazioni del calcolo dei predicati rimanendo all'interno del paradigma simbolicista ci sono tre strade. La prima strada consiste nell'uso di linguaggi nei quali sia possibile limitare il tipo di inferenze, che hanno dunque proprietà computazionali migliori, come la *programmazione logica* e le *logiche descrittive*. La seconda strada consiste nell'uso di particolari logiche chiamate logiche non classiche, come quella non monotona e default, che permettono tipi di inferenza differenti, oppure la logica fuzzy ed il ragionamento probabilistico che permettono di trattare l'incertezza. Tutte queste logiche nascono per rappresentare forme di ragionamento non deduttivo; tali forme sono considerate più "deboli" nel senso che le inferenze portano a conclusioni che sono o probabili, o rivedibili, o valide solo in un contesto e non in assoluto. La terza strada consiste nell'uso di formalismi di rappresentazione "schematici", come ad esempio le *reti semantiche* ed i *frames*, nei quali la conoscenza è strutturata in maniera da facilitare particolari forme di inferenza, come ad esempio il *default* e l'ereditarietà. Nei prossimi paragrafi presenteremo brevemente tutte queste possibilità.

3.1.4 La programmazione logica

La *programmazione logica* di cui l'esempio principale è il linguaggio di programmazione PROLOG, sviluppato negli anni '70 da Colmerauer e Kowalski sfrutta la possibilità di esprimere le asserzioni logiche in forma di programma, contrapponendosi al linguaggio funzionale LISP di McCarthy che era utilizzato da quasi tutti gli sviluppatori nel campo dell'IA. Eseguire un programma in PROLOG infatti equivale a dimostrare un teorema sulla base di assiomi di partenza, espressi in maniera dichiarativa.

In PROLOG la conoscenza è espressa mediante il formalismo delle *clausole di Horn*. Ad esempio, ricordando che una clausola di Horn ha sempre una sola testa, è possibile tradurre la rappresentazione di tipo logico:

$$\forall(x) (\text{mortale}(x) \ \& \ \text{razionale}(x) \Rightarrow \text{uomo}(x))$$

uomo(giovanni)

in una rappresentazione equivalente in PROLOG:

**uomo(X):- mortale(X),
razionale(X).
uomo(giovanni).**

ed inferire in entrambi i casi:

mortale(giovanni).

La maggiore peculiarità del PROLOG è quella di consentire processi di inferenza decidibili in un tempo finito, tuttavia sottoposti al problema dell'esplosione combinatoria. Questo risultato è ottenuto diminuendo il potere espressivo del linguaggio: infatti, nelle clausole Horn non è possibile esprimere qualunque formula: in una clausola si possono esprimere o un termine positivo (e.g. **mortale(x)**) oppure l'implicazione di un termine positivo (e.g. **b ⇒ a**) ma non **NON b ⇒ NON a** e nemmeno **NON b ⇒ a**.

In PROLOG determinare il valore di verità di un termine è una procedura computazionale, una vera e propria ricerca fra le clausole Horn che prende il nome di *risoluzione*. La ricerca è effettuata "all'indietro", ovvero dallo stato obiettivo allo stato iniziale; è tuttavia possibile anche ragionare "in avanti", ovvero a partire dai dati.

3.1.5 Logica non monotona e logica default

A partire dalle limitazioni della logica classica sono stati sviluppati una serie di altri formalismi; ad esempio, per affrontare i problemi della monotonicità sono state sviluppate sia la *logica non monotona* che la *logica default*, che consentono di introdurre conoscenze che successivamente, in presenza di nuova informazione, possono essere eliminate.

La *logica non monotona* (McDermott e Doyle, 1980) permette di trattare i casi di *default* introducendo un operatore di consistenza (M) che consente di aggiungere nuova conoscenza a patto che essa sia compatibile con quella preesistente. Ad esempio:

$$(\forall x) \text{UCCELLO}(x) \text{ AND } M \text{ VOLA}(x) \Rightarrow \text{VOLA}(x)$$

può essere interpretato come: se c'è un esemplare di uccello, per il quale sia ragionevole assumere che voli (ovvero: il fatto che vola non contraddice le conoscenze preesistenti), allora si può assumere che voli.

In questo modo è possibile risolvere il problema delle eccezioni, esprimendo senza contraddizioni:

$$(\forall x) \text{PINGUINO}(x) \Rightarrow \text{NON VOLA}(x)$$

A questo punto, si può derivare sia il fatto che un passero voli:

$$\begin{aligned} (\forall x) \text{PASSERO}(x) &\Rightarrow \text{UCCELLO}(x) \\ \text{PASSERO}(\text{cip}) &\Rightarrow \text{VOLA}(\text{cip}) \end{aligned}$$

sia il fatto che un pinguino non voli:

$$\begin{aligned} (\forall x) \text{PINGUINO}(x) &\Rightarrow \text{UCCELLO}(x) \\ \text{PINGUINO}(\text{ciop}) &\Rightarrow \text{NON VOLA}(\text{ciop}) \end{aligned}$$

In caso due conoscenze siano compatibili con quelle preesistenti ma fra di loro contraddittorie, entrambe sono scartate.

Nella *logica default* (Reiter, 1980) esiste una regola di inferenza che permette operazioni simili a quelle attraverso l'operatore M. Mediante tale regola è possibile concludere C se, potendo dimostrare A ed assumere consistentemente B, ne deriva C. Dunque è possibile dedurre un teorema soltanto fino a quando non viene asserito esplicitamente qualcosa di contrario. Un problema di questo approccio è che spesso tutta la conoscenza sulle eccezioni deve essere già stabilita a priori; per questo motivo il formalismo risulta essere poco flessibile.

Un ulteriore modo per trattare il problema della non monotonicità è assumere l'*ipotesi del mondo chiuso* (*closed world assumption*). Essa, utilizzata anche nel PROLOG, consiste nell'attribuire il valore "falso" a tutto ciò che non è esplicitamente asserito o dedotto nel programma. È chiaro dunque che l'assunzione di un nuovo fatto rende vero un enunciato che prima era falso. L'ipotesi del mondo chiuso permette dunque di trarre inferenze rivedibili (*defeasible*) in presenza di nuove informazioni

3.1.6 Logiche per il trattamento dell'incertezza

Mentre nella logica classica le proposizioni sono vere o false, vi sono vari tipi di logiche che permettono di trattare conoscenza incerta (in vari sensi, come discusso precedentemente).

Le *logiche multivalenti* ammettono più di due valori di verità. Nella logica classica si assumono solo due valori di verità, il vero ed il falso; in questo modo, implicitamente si assume di poter sempre conoscere quale sia tale valore. Nelle logiche multivalenti vi sono almeno tre valori di verità (vero, falso, non so): questo formalismo permette di trattare meglio situazioni di incertezza o ambiguità.

La *Logica Fuzzy* (Zadeh, 1965) è stata principalmente sviluppata per trattare in rigorosi termini formali i problemi della vaghezza e dell'incertezza del linguaggio umano e dei fenomeni naturali. Essa nasce in risposta alla semplificazione della logica classica per la quale un enunciato è vero o falso, senza sfumature né gradi. Al contrario, nella logica *fuzzy* i valori di verità degli enunciati hanno gradi e non sono binari. Ad esempio un uomo può essere **alto** ad un certo grado: più precisamente può appartenere all'insieme degli alti ad un certo grado. Ad esempio un uomo alto 1,80 metri vi appartiene al grado 0,8 mentre un uomo alto 2 metri vi appartiene al grado 1 (se si assume che il dominio dei valori di verità sia l'intervallo fra zero ed uno in R).

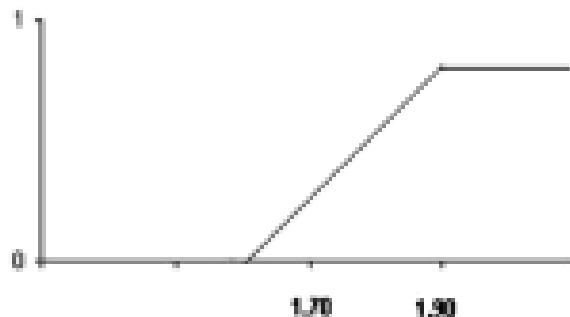


Figura 3 – Fuzzy membership per l'altezza.

3.1.7 Logiche modali

Mentre la logica classica permette di trattare la verità e la falsità di enunciati, le logiche modali introducono nuovi operatori che permettono di descrivere i “modi” in cui un enunciato può essere creduto vero o falso. Esistono molte logiche modali, ma i due operatori più tipici sono quelli che permettono di trattare i concetti di necessità (NECESSARIO) e di possibilità (POSSIBILE). Per trattare questi concetti si introduce a livello semantico la nozione di *insieme dei mondi possibili*. Formalmente l'insieme dei mondi possibili è un insieme W tale che in ciascun elemento di W vi siano le stesse variabili proposizionali (ad esempio a , b e c); tuttavia, in ciascun mondo esse possono avere valori di verità differenti. A seconda di tali valori ovviamente in ciascun mondo espressioni complesse come a OR b assumeranno valori differenti. Fra i mondi esiste anche una relazione di accessibilità, che determina se due mondi siano o meno collegati. Tale regola serve anche per determinare l'interpretazione di NECESSARIO e POSSIBILE: una proposizione è necessaria in *un mondo possibile* W se e solo se essa è vera in tutti i mondi possibili *accessibili* da W ; una proposizione è possibile in W se e solo se essa è vera almeno in un mondo accessibile da W . A livello predicativo, inoltre, si assume che nei mondi possibili vi siano gli stessi individui, ma non con gli stessi predicati.

Le logiche modali permettono inoltre di ridefinire il concetto di implicazione in maniera da soddisfare meglio le nostre intuizioni. Un paradosso della logica classica è infatti che, date due proposizioni qualsiasi a e b , è una tautologia che $a \Rightarrow b$ OR $b \Rightarrow a$. Non esistono dunque in logica classica proposizioni indipendenti l'una dall'altra. Al contrario nelle logiche modali l'implicazione $a \Rightarrow b$ viene riformulata come NOT POSSIBILE (a AND NOT b) (non è possibile che a sia vero e b sia falso) ma non vale la precedente tautologia: tipicamente le proposizioni in differenti mondi possibili sono fra loro indipendenti.

Data l'estrema versatilità della logica modale e la possibilità di costruire nuovi operatori, sono state sviluppate molte logiche modali per trattare problemi relativi ai contesti (intuitivamente, ogni mondo possibile equivale ad un contesto), oppure per trattare formalmente credenze, desideri, obblighi, intenzioni ecc., introducendo operatori come CREDE (x,p). Vi sono infatti numerosi formalismi che introducono operatori differenti, riflettendo diverse intuizioni filosofiche su tali concetti. Ad esempio, il concetto di “obbligo” dovrà essere formalizzato molto diversamente a seconda che si vogliano trattare casi come **un triangolo deve avere tre lati e non si deve attraversare un incrocio con il rosso**. Si tenga presente che queste logiche modali non sono vero-funzionali: ad esempio, **Maria crede che Giovanni sia inglese** può essere vero indipendentemente dal fatto che Giovanni sia davvero inglese.

Esistono inoltre altri tipi di logiche modali adatte a trattare situazioni differenti: ad esempio le *logiche temporali* che permettono di rappresentare il fatto che alcuni enunciati possono essere veri nel presente e falsi nel passato o viceversa; oppure le *logiche condizionali* che permettono di rappresentare il fatto che alcuni enunciati sarebbero stati veri solo se si fossero realizzate condizioni che invece non si sono realizzate, ecc.

3.2 Conoscenza procedurale e regole di produzione

Un approccio differente alla rappresentazione della conoscenza ed alla sua manipolazione consiste nell'utilizzare rappresentazioni procedurali. Al contrario della conoscenza di tipo

dichiarativo, centrata sugli oggetti, la conoscenza procedurale è focalizzata sulle azioni da compiere: ad esempio, oltre a saper riconoscere e categorizzare un bicchiere sappiamo anche come utilizzarlo. Questo secondo tipo di conoscenza, la conoscenza procedurale, può essere formalizzata attraverso regole di produzione, che hanno la forma:

SE vuoi accendere la luce ALLORA premi l'interruttore

Questo stile di computazione fu introdotto nell'IA da Newell e Simon (1965) per l'analisi del gioco degli scacchi e successivamente come modello del problem solving degli umani (Newell e Simon, 1972). Le due più note architetture cognitive, SOAR (Rosenbloom, Laird e Newell, 1992) ed ACT-R (Anderson e Lebiere, 1998), utilizzano entrambe sistemi di produzione. A differenza di SOAR, nella quale tutto è rappresentato attraverso regole di produzione, in ACT-R si utilizzano anche conoscenze dichiarative in formato proposizionale (organizzate in maniera simile alle reti semantiche).

Come si vede dall'esempio precedente, le regole di produzione hanno due parti, la sinistra e la destra, chiamate *antecedente* e *conseguente*, o *premessa* e *azione*. Il meccanismo inferenziale (chiamato *interprete*) fa in modo che quando la premessa è soddisfatta l'azione sia eseguita. L'azione può essere o una vera e propria azione sul mondo (ad esempio realizzata da un robot) oppure l'aggiunta o l'eliminazione di una conoscenza dalla base di conoscenza. Ad esempio, se il sistema contiene la regola: **ha_tre_lati(x) AND poligono_chiuso(x) \Rightarrow triangolo(x)** e due fatti che soddisfano la premessa **ha_tre_lati(figura1)** e **poligono_chiuso(figura1)** allora la regola è applicabile e come conseguenza una nuova conoscenza **triangolo(figura1)** sarà aggiunta alla base di conoscenza.

In un sistema di regole di produzione l'interprete può agire *in avanti* (*forward*) o *all'indietro* (*backward*), come descritto precedentemente. A differenza di quanto avviene con il meccanismo inferenziale deduttivo descritto precedentemente le regole di produzione vengono sempre attivate anche se generano conoscenza inconsistente, poiché lo scopo è di generare tutta la conoscenza per la quale sono disponibili produzioni adatte. Se si vuole mantenere la consistenza ci sarà dunque bisogno di un meccanismo di filtraggio successivo. Spesso molte produzioni sono applicabili contemporaneamente; per questo motivo possono essere aggiunte delle regole che risolvono i conflitti, ovvero stabiliscono quali produzioni sono applicate per prime.

Un punto a favore dei sistemi basati su regole di produzione è la facilità nella scrittura di nuove regole, che possono essere aggiunte in maniera modulare, ovvero senza interferire con le altre conoscenze. Tuttavia la forte modularità e la possibilità di inserire indipendentemente regole e conoscenze rende poi difficile comprendere nei dettagli cosa accadrà come conseguenza dell'aggiunta di una nuova regola, poiché essa può portare a derivare molta altra conoscenza. Infatti, i sistemi di produzione sono utilizzati in compiti nei quali le conoscenze sono fra loro indipendenti, come ad esempio nei dimostratori automatici di teoremi (come il GPS) nei quali gli assiomi e le regole di inferenza sono indipendenti.

3.3 Conoscenze come sistemi di relazioni

Come è stato presto riconosciuto nell'IA, esiste una fondamentale differenza fra esseri umani e sistemi artificiali. Mentre questi ultimi permettono di svolgere operazioni molto

complesse in domini altamente astratti e specializzati come la matematica, essi sono molto meno versatili ed hanno grandi difficoltà nell'affrontare i cosiddetti “problemi di senso comune” come ad esempio pianificare una gita e muoversi in un ambiente dinamico, nei quali invece gli esseri umani sembrano operare senza sforzo apparente.

Esistono molte spiegazioni di questo fenomeno. La prima riguarda la *versatilità*: mentre la conoscenza dei calcolatori è dominio-specifica, spesso gli esseri umani sono capaci di trasferire le conoscenze da un dominio all'altro e sfruttarle in domini differenti da quelli nei quali esse sono state apprese. Un altro aspetto fondamentale riguarda invece la quantità di informazione disponibile: gli esseri umani hanno una enorme quantità di conoscenza, compresa la conoscenza su quali metodi siano efficaci in quali situazioni, quali portino al fallimento, quali siano le eccezioni, eccetera.

Di fronte al *problema del senso comune* si è ipotizzato che la chiave fosse dotare i sistemi di un enorme numero di conoscenze ed al tempo stesso di strategie per ragionare su di esse. Non deve sorprendere dunque che uno dei grandi campi di indagine dell'IA sia stato quello della rappresentazione della conoscenza, visto che il formalismo con cui essa è espressa pone poi dei vincoli su come essa possa essere utilizzata e visto che diversi tipi di rappresentazioni consentiranno meccanismi di risoluzione dei problemi più o meno potenti.

Sono stati dunque proposti moltissimi formalismi per la rappresentazione della conoscenza, come le *reti semantiche* (Quillian, 1969), i *frames* (Minsky, 1975), gli *scripts* (Schank ed Abelson, 1977). Al tempo stesso si è cercato di codificare esplicitamente una gran mole di conoscenza di senso comune da utilizzare nei calcolatori; l'esempio più rilevante di questo approccio è il progetto *CYC* (Lenat, 1990) che dal 1984 sta costruendo un enorme database concettuale che comprende moltissima conoscenza di senso comune. *CYC* comprende migliaia di concetti e di asserzioni interconnesse, codificati in un formato omogeneo allo scopo di definirli parzialmente secondo una ontologia del senso comune. Tale conoscenza dovrebbe consentire al calcolatore di comprendere i domini di senso comune e di ragionare al loro interno per mezzo di un motore di ragionamento.

Collins e Quillian (1969) propongono che i significati non siano rappresentati in formato proposizionale ma attraverso strutture tassonomiche, le *reti semantiche* (*semantic networks*). Il significato di un concetto dunque dipende principalmente dalla struttura delle sue connessioni con altri significati. Le reti semantiche comprendono nodi, che stanno per oggetti, e legami fra nodi, che codificano le relazioni fra oggetti; alcune fra le più usate sono: **esemplare di** (KIND_OF), **è un** (IS_A), **è parte di** (PART-OF), **appartiene a** (BELONGS-TO). In generale, reti semantiche possono contenere relazioni di qualsiasi tipo.

Ad esempio, una rete semantica può contenere una serie di conoscenze come: **furia KIND_OF cavallo, cavallo IS_A mammifero, mammifero IS_A animale, coda PART_OF cavallo, furia BELONGS_TO zorro**. Una rete semantica può anche essere utilizzata per compiere inferenze mediante il meccanismo dell'*ereditarietà*. Ad esempio, gli attributi possono essere ereditati attraverso i legami **KIND_OF ed IS-A**. Utilizzando il meccanismo di ereditarietà dalla rete semantica precedente si ottiene: **coda PART-OF furia**.

Ad ogni modo, l'ereditarietà nelle reti semantiche deve essere circoscritta definendo quali siano le relazioni che la permettono, altrimenti qualunque attributo (anche quelli non rilevanti) potrebbe essere ereditato da nodi a livello più basso.

Una evidenza empirica per l'utilizzo in memoria di strutture simili alle reti semantiche è portata da Collins e Quillian (1969), che mostrano sperimentalmente come i tempi che i

soggetti impiegano per decidere ad esempio se un canarino sia un uccello o se un canarino sia un animale sono proporzionali alla distanza dei due nodi nella rete semantica, suffragando l'ipotesi che a guidarli sia una "ricerca" all'interno di tale struttura tassonomica. Un'ulteriore evidenza riguarda i tempi di accesso in memoria agli attributi degli oggetti: i soggetti impiegavano più tempo a rispondere che un cavallo respiri (caratteristica tipica di tutti gli animali) che non che abbia una coda (caratteristica tipica soltanto di alcuni mammiferi).

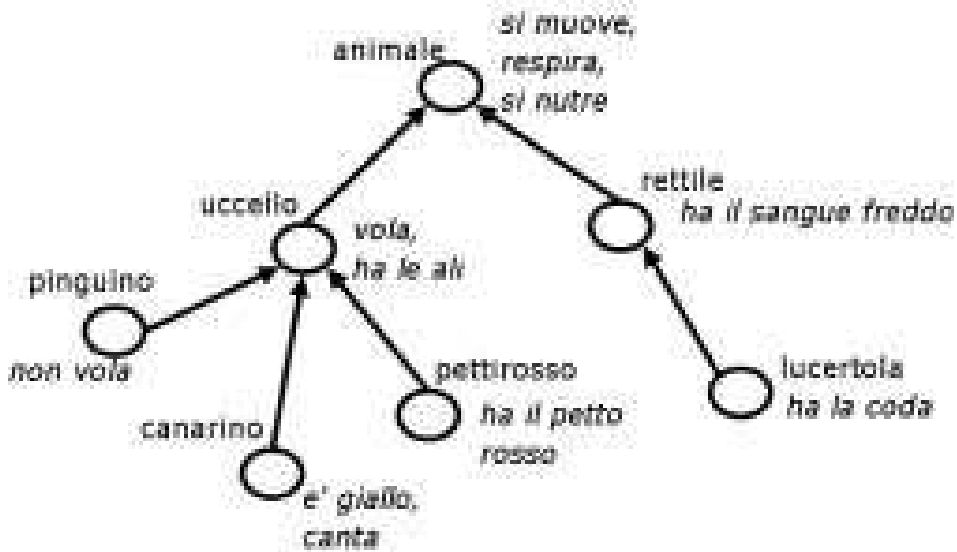


Figura 4 – Reti semantiche.

3.4 Strutture schematiche

Come abbiamo visto l'approccio logico formalizza le conoscenze mediante fatti e regole, ma senza contestualizzarle. Un approccio differente consiste nel raggruppare l'informazione in strutture di conoscenza più ricche, che possiamo chiamare *schemi* in riferimento ai lavori di Kant (1781-87), Bartlett (1932) e Piaget (1952), che sono specifiche per i vari contesti e dunque forniscono un modo per affrontare il *problema del contorno* (*frame problem*) (Dennett, 1978; McCarthy e Hayes, 1969).

Gli schemi rappresentano situazioni prototipiche e sono dunque collezioni di conoscenze contestualizzate e strutturate. L'importanza dei contesti può risultare da una considerazione di ordine pratico: per risolvere i problemi che Minsky chiama «di senso comune» gli umani non compiono elaborati calcoli logici, ma attivano «pacchetti» di informazioni stereotipiche associate alle situazioni ed ai contesti, evitando dunque anche il problema dell'esplosione combinatoria. Ad esempio, per muoversi in una stanza o per ordinare un pranzo al ristorante possono essere attivati pacchetti di informazioni utilizzate in casi simili in passato, che mettono in evidenza come nelle stanze vi siano porte e finestre e come nei ristoranti generalmente ci si siede prima di ordinare, ecc.

Questo approccio si ispira a numerose teorie sull'organizzazione della memoria, basata su prototipi (Rosch, 1975) o esemplari (Nosofsky, 1988). Bartlett (1932) aveva

provato sperimentalmente che è più facile riconoscere un oggetto nel suo contesto ordinario che fuori contesto.

In questo senso le operazioni di **categorizzare**, **riconoscere** ed **agire in una data situazione** sono tutti aspetti dello stesso meccanismo. Un agente cognitivo è infatti capace di attivare un pacchetto di informazioni adatte alla situazione, mediante la quale categorizza la situazione stessa ed ottiene immediatamente, senza la necessità di complessi calcoli logici, tutte le informazioni disponibili per riconoscerne i componenti e per agire appropriatamente. L'operazione di categorizzazione e riconoscimento consiste semplicemente nel confrontare uno o più schemi (fra quelli disponibili) con la situazione corrente per vedere se c'è compatibilità. Lo schema più appropriato è dunque prima selezionato e poi adattato alla situazione corrente. Si noti che in questo senso nel riconoscimento e nella categorizzazione c'è una compresenza di elementi *bottom-up*, che dipendono da cosa c'è nel mondo, e *top-down*, che dipendono da quali siano gli schemi disponibili. Se nel contesto della situazione si include inoltre l'intenzione o l'attività corrente dell'agente cognitivo, questo meccanismo diventa guidato dagli scopi. Ad esempio, nel categorizzare un tavolo al centro di una stanza possono essere attivati differenti schemi a seconda delle necessità dell'agente: se questi deve attraversare la stanza lo schema appropriato sarà quello di **ostacolo**, mentre se deve scrivere una lettera quello di **supporto**.

Un altro aspetto fondamentale è quello della predizione: agli schemi sono associate aspettative che verranno via via controllate. La predizione ha diversi ruoli: ad esempio quello di selezionare il comportamento più appropriato a seconda dei suoi risultati attesi, ma anche quello di cambiare categoria se le aspettative sono disattese.

Gli schemi non soltanto coinvolgono una struttura di relazioni fra oggetti (come le reti semantiche), poiché differenti schemi possono essere collegati, ma anche un insieme di proprietà di tali oggetti, in maniera tale che gli oggetti siano dunque forniti di una *struttura interna*. Ad esempio, l'oggetto **casa** ha una serie di attributi come quello di **possedere porte e finestre**.

Inoltre gli schemi forniscono una indicazione di quali siano gli elementi contestualmente rilevanti in una data situazione e di quali conoscenze siano *implicite*, ovvero non esplicitate ma tuttavia presenti sullo sfondo. Un esempio di Minsky è quello della comprensione del racconto: **Maria fu invitata alla festa di Pietro. Si chiese se lui avrebbe gradito un aquilone**. In questo caso, attivando un pacchetto di informazioni, posso sapere non soltanto che Maria e Pietro sono bambini, ma anche che l'aquilone è un regalo, nonostante queste informazioni non siano presenti nel racconto stesso.

Un meccanismo correlato è quello del *default*: ci sono conoscenze che possono essere assunte in assenza di ulteriori informazioni, ma possono successivamente essere riviste. Ad esempio, poiché tutti gli uccelli volano posso assumere per *default* che anche i pinguini volino, salvo poi ottenere la nuova informazione che i pinguini non volano, che rimpiazza il *default*. Questa operazione sarebbe impossibile in logica classica, che è monotona. Come Minsky osserva, è spesso molto più parsimonioso strutturare le conoscenze attraverso valori di *default* e conservare a parte una lista di eccezioni, piuttosto che costruire ogni volta una nuova categoria. Gli esseri umani sono dunque provvisti non soltanto di conoscenze positive, ma anche di questa "expertise negativa" che copre le eccezioni, i casi limite, ecc.

3.4.1 I frames di Minsky

I frames, descritti da Minsky (1975), sono forse l'esempio più noto di formalismo di tipo schematico. Essi consentono di organizzare la conoscenza sugli oggetti mediante una collezione di attributi, chiamati caselle o *slots*, aventi determinati valori. Alcuni esempi di coppie attributo-valore sono evidenziati nella seguente figura.

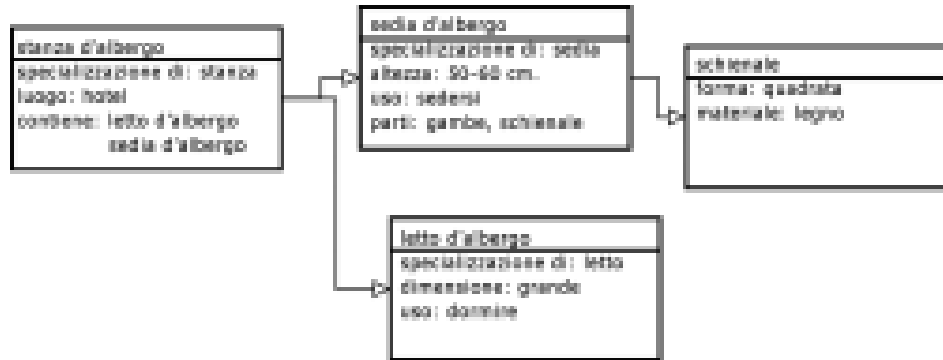


Figura 5 – Esempio di *frame*.

I valori possono essere anche dei *default*, permettendo di ragionare anche in mancanza di tutta l'informazione. Inoltre il valore di un attributo può anche essere un altro *frame*: l'informazione può dunque essere organizzata gerarchicamente ed in una rete.

Una ulteriore proposta di Minsky è quella dei *sistemi di frames* che permettono di rappresentare oggetti da differenti punti di vista. Un esempio è nella sfera della visione: una stanza può essere rappresentata da diversi punti di vista a seconda dell'orientamento di chi guarda. Una caratteristica molto importante dei sistemi di *frames* è che possono essere collegati mediante regole che definiscono il passaggio dall'uno all'altro. Ad esempio, entrando in una stanza si vede una certa scena, mentre girandosi all'indietro si vede la porta. Tale conoscenza può essere espressa mediante due frames e le regole che definiscono le conseguenze di ciascuna azione: mentre il primo *frame* è attivo, girarsi all'indietro determina l'attivazione del secondo *frame*. Questo meccanismo permette di affrontare il problema del contorno, ovvero capire cosa cambia in una scena come conseguenza delle proprie azioni, fornendo al tempo stesso una serie di aspettative (sotto forma di *frames* che si attivano).

L'approccio basato su frames, che comprende una rete di somiglianze e non vere e proprie regole definitorie per i concetti, è correlato con le idee di Wittgenstein sull'importanza delle *somiglianze di famiglia*.

3.4.2 Gli scripts di Schank ed Abelson

Un altro formalismo schematico è quello degli *scripts* di Schank e Abelson (1977). Essi descrivono situazioni tipiche, come quella di **mangiare in un ristorante**, suddividendole in una sequenza di scene. Gli script comprendono sia conoscenze sulle azioni (o meglio sulla sequenza di azioni), come quelle di sedersi, ordinare e mangiare, sia sugli

oggetti e gli agenti che assumono un ruolo, ad esempio quello di cameriere. Gli script possono essere utilizzati nella comprensione di storie, come ad esempio: **Maria entrò al ristorante ed ordinò un piatto di spaghetti. Aspettò a lungo ed alla fine, spazientita, se ne andò.** In maniera simile a quanto descritto per i *frames*, attraverso gli *script* è possibile comprendere una serie di elementi non esplicitati, come il fatto che Maria si spazientisce perché il piatto di spaghetti non arriva o come il fatto che Maria se ne sia andata senza mangiare.

Anche mediante questo formalismo è possibile accedere alla conoscenza rilevante senza complessi calcoli logici, ma semplicemente attivando lo *script* più appropriato.

Schank (1973) propone anche un formalismo per la *dipendenza concettuale* nel quale esiste un numero limitato di azioni e ruoli primitivi che possono essere combinati per interpretare qualsiasi verbo. Esistono differenti tipi di azioni (ACT) come gli *emotional ACT* (EACT), ad esempio amare o invidiare; i *communicative ACT* (CACT), come leggere ed proferire; i *physical ACT* (PACT), come spostare o costruire. Vi sono anche astrazioni come il trasferimento, che possono prendere differenti forme: trasferimento fisico (PTRANS), trasferimento di possesso (ATRANS) e trasferimento di informazione (MTRANS)

Una evoluzione del concetto di script è quella di MOP o «pacchetto di organizzazione della memoria» (Schank, 1982). In questo formalismo le scene non fanno più parte propria degli script; al contrario, molti script possono condividere una stessa scena, riferendosi ad essa. Ad esempio, la scena **sedersi al tavolo** può essere la stessa sia per lo script di **mangiare in un ristorante** sia per lo script di **partecipare ad una riunione**.

3.5 Alberi di decisione

Un tipo differente di rappresentazione simbolica è costituito dai *cosiddetti alberi di decisione*. Un albero di decisione permette di derivare una decisione (ad esempio binaria) quando si conosca di un elemento un insieme di proprietà, ovvero una serie di coppie attributo-valore.

Un semplice albero di decisione può essere quello per decidere quale linguaggio di programmazione utilizzare per scrivere una applicazione:

Problema: utilizzare C++ o Perl?

L'applicazione è complessa?

SI → C++

NO → L'applicazione deve essere estesa in futuro?

SI → C++

NO → Il programmatore è esperto?

SI → C++

NO → Perl

Gli alberi di decisione hanno lo stesso potere espressivo dei linguaggi predicativi; i due formalismi sono intertraducibili. Ad esempio, l'albero di decisione descritto precedentemente può essere anche espresso come una serie di regole, come ad esempio: **(NOT complessa (applicazione) AND NOT estensibile (applicazione) AND NOT esper-**

to (**programmatore**) \Rightarrow **utilizzare(perl)**, ecc. Più formalmente l'albero di decisione opera esplicitando una tabella che include tutte le coppie attributo-valore⁵.

COMPLESSO	ESTENDIBILE	ESPERTO	C++	PERL
true	?	?	true	false
false	true	?	true	false
false	false	true	true	false
false	false	false	false	true

La rappresentazione attraverso alberi di decisione permette inoltre di sfruttare efficaci algoritmi di apprendimento che consentono di estendere la base di dati, caratteristica molto importante per grandi sistemi basati su conoscenze per i quali non è plausibile che sia il programmatore a fornire tutta l'informazione. Attraverso appropriati algoritmi di apprendimento⁶ l'albero di decisione può inoltre trovare un sottoinsieme degli attributi tali da permettere di classificare con il minor numero di operazioni possibili.

3.6 Rappresentazioni distribuite connessioniste

Tutti i sistemi simbolici di rappresentazione della conoscenza si basano su di una serie di assunzioni che sono state invece contestate dai fautori dei sistemi connessionistici. Si tratta principalmente del fatto che le rappresentazioni sono discretizzate e localizzate (ad esempio in una proposizione o in un nodo di una rete semantica). Questa assunzione rende i sistemi simbolici molto rigidi; ad esempio nuove conoscenze possono essere aggiunte, o conoscenze passate modificarsi e scomparire (ad esempio a causa di una dimenticanza o di una malattia) ma solo in una modalità "tutto o niente", mentre al contrario vi sono evidenze psicologiche e neurobiologiche che le funzioni cognitive tendano piuttosto a migliorare e degradarsi piuttosto che apparire e scomparire di colpo come succede in un sistema simbolico.

Inoltre i sistemi simbolici sono poco resistenti agli errori ed all'incertezza proprio in virtù della loro rigidità. Al contrario il cervello è plastico e questa caratteristica permette sia l'apprendimento che la tolleranza di errori o mancanza di informazione (tutte funzioni che i sistemi connessionistici infatti mostrano).

Se una delle metafore che ha dato inizio al funzionalismo è "la mente sta al cervello come, in un computer, il software sta all'hardware", e se il principio di indipendenza delle funzioni (intelligenti) dal supporto fisico e dal meccanismo che implementa la funzione è uno dei capisaldi del cognitivismo, nei sistemi connessionistici invece il meccanismo e persino in alcuni casi il supporto fisico contano. Ancora, nella corrente *situata* sviluppatasi in particolare nel connessionismo anche il fatto di avere un corpo ed essere immersi in un contesto di interazione con un mondo fisico sono fattori determinanti.

Come abbiamo inoltre già evidenziato i sistemi simbolici, agendo tipicamente attraverso conoscenza esplicita, rendono molto difficile trattare il problema del con-

⁵ Il simbolo "?" indica che il valore è indifferente.

⁶ Si veda a tal riguardo il contributo di Basili e Moschitti contenuto nel presente volume.

torno (*frame problem*), poiché nonostante l'uso di particolari formalismi come *frames* e *scripts* rimane tuttavia difficile esplicitare tutta e sola l'informazione disponibile.

Un approccio radicalmente differente alla rappresentazione della conoscenza è quello connessionista o subsimbolico (Rumelhart e McClelland, 1986), ispirato alle evidenze neurobiologiche sulla struttura del cervello.

La principale novità di questo approccio è che la conoscenza non è *localizzata* in un elemento specifico (come in una proposizione o in un nodo di una rete semantica) ma *distribuita*, ed in particolare legata alle dinamiche di attivazione dei nodi delle reti neurali (che sono il prototipo di sistema connessionista). Lo stesso nodo di una rete neurale partecipa, attivandosi, a diverse rappresentazioni, anche simultaneamente. Le rappresentazioni complesse, che comprendono ad esempio diverse parti, non codificano esplicitamente le relazioni fra di esse (come ad esempio avviene nelle etichette degli archi nelle reti semantiche), ma al contrario le rappresentano in maniera largamente implicita nella struttura della rete.

Una *rete neurale* (*neural network*) costituisce un modello semplificato della struttura dei neuroni e delle sinapsi. Strutturalmente una rete neurale è simile ad una rete semantica in quanto comprende nodi e legami fra di essi; tuttavia, a differenza di quanto avviene nelle reti semantiche, nelle reti neurali nodi e legami non rappresentano concetti o entità, ma hanno come unico attributo un valore di attivazione il quale, attraverso funzioni matematiche (di differenti tipi) si propaga fra i nodi. Un esempio tipico di rete neurale è quella *feed-forward* costituita di tre strati: uno strato di input, uno nascosto (*hidden*) ed uno di output. I nodi dei tre strati hanno un valore numerico che ne rappresenta l'attivazione; essi possono trasmettere attivazione o inibire altri nodi (diminuendone l'attivazione) attraverso connessioni fra i nodi stessi, che hanno un peso (che rappresenta la forza del legame stesso). La quantità di attivazione o inibizione trasmessa dipende dai valori numerici dei nodi e dei pesi della rete; per questo motivo i pesi delle connessioni codificano il contenuto rappresentazionale della rete, determinando la risposta in termini di output ad un dato input. Il processamento nella rete neurale inizia quando i neuroni dello strato di input ricevono un segnale numerico; a seconda della funzione matematica che essi incorporano essi assumono dunque una attivazione, che è propagata al secondo e poi al terzo strato attraverso le connessioni. Come risultato, i neuroni dello strato di output assumeranno dunque un valore numerico, che è la "risposta" della rete all'input.

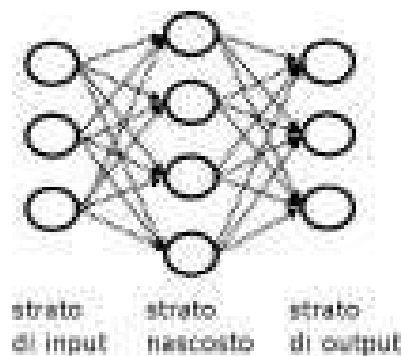


Figura 6 – Esempio di rete neurale.

Vi sono una serie di algoritmi di apprendimento per le reti neurali (Rumelhart e McClelland, 1986), anch'essi ispirati più o meno plausibilmente ad evidenze neurobiologiche, come la regola di Hebb e la *backpropagation*: attraverso una serie di esempi di "risposte corrette" la rete impara a produrre un output appropriato a partire da certi input, mostrando tipicamente una buona capacità di generalizzare rispetto a casi simili.

Un esempio può chiarire il funzionamento delle reti neurali. Immaginiamo una rete neurale che controlla i movimenti di un veicolo; in questo caso lo strato di input codifica stimoli sensoriali (i neuroni potrebbero essere recettori sensibili ai colori), mentre quella di output comandi motori (i neuroni potrebbero controllare le ruote del veicolo). Se lo strato di input riceve una stimolazione sensoriale che codifica la presenza di un "predatore" (ad esempio un certo pattern di colori), dopo un adeguato apprendimento lo strato di output potrebbe generare una risposta motoria, ad esempio muoversi velocemente in avanti.

Una differenza molto rilevante fra sistemi simbolici e subsimbolici riguarda la manipolazione delle rappresentazioni. Nei sistemi connessionisti, infatti, non c'è una vera e propria distinzione fra le rappresentazioni ed i processi di manipolazione, o fra fatti e regole, in quanto tutto è codificato all'interno della stessa rete neurale. Nelle reti neurali, poiché l'informazione è distribuita, non si manipolano simboli, ma strutture matematiche (come i pesi dei nodi e degli archi) che stanno al di sotto dei simboli: di qui il nome di *sistemi subsimbolici*. I simboli emergono soltanto come *patterns* di attivazione della rete stessa. Per questo motivo i connessionisti (Smolensky, 1988) sostengono che i fenomeni cognitivi emergano manipolando tale livello subcognitivo, la "microstruttura della cognizione", piuttosto che da operazioni su rappresentazioni localizzate.

Un punto critico dei sistemi connessionistici è tuttavia la grande difficoltà (e spesso impossibilità) di analizzare le rappresentazioni all'interno di una rete neurale, che sono spesso simili a "scatole nere" in quanto è possibile osservarne soltanto il risultato senza poter dare una semantica chiara alle operazioni intermedie. In maniera simile Fodor e Pylyshyn (1988) affermano che i sistemi connessionisti non siano capaci di composizionalità e sistematicità, come i sistemi simbolici, in quanto è impossibile estrarre "porzioni" di rappresentazione ed operare su di esse, così come più in generale manipolare la loro struttura, in quanto essa non ha semantica composizionale (ad esempio, sommando due reti neurali non si ottiene banalmente un'altra rete neurale il cui significato è la somma delle due). Molti connessionisti, come ad esempio Smolensky (1988), contestano invece queste affermazioni, mostrando come anche le rappresentazioni di tipo implicito possano avere una semantica combinatoria.

Vi sono dei pro e dei contro nell'utilizzo delle reti neurali come sistema di rappresentazione della conoscenza. I principali vantaggi consistono nella loro estrema adattività e nei potenti strumenti di apprendimento, che permettono di modellare un dominio a partire da poca o nessuna conoscenza codificata. Il principale svantaggio consiste nella mancanza di una chiara e riconosciuta metodologia per decidere quale rete utilizzare, come disegnare e controllare i numerosi parametri; dal punto di vista pratico a causa di questo problema può diventare estremamente difficile ottenere risultati soddisfacenti. Un ulteriore svantaggio è che, come osservato precedentemente, è spesso alquanto difficile analizzare i processi all'interno delle reti neurali; seppure si raggiunga il risultato desiderato, a volte è impossibile stabilire perché.

BIBLIOGRAFIA

- Anderson, J.R., Lebiere, C. (1998), *The atomic components of thought*, Mahwah, NJ, Erlbaum.
- Armstrong, S.L., Gleitman, L.R., Gleitman, H. (1983), "What some concepts might not be", *Cognition*, 13, pp. 263-308.
- Barsalou, L.W. (1987), *Concepts and conceptual development*, Cambridge, MA, Cambridge Univ. Press.
- Barsalou, L.W. (1999), "Perceptual symbol systems", *Behavioral and Brain Sciences*, 22, pp. 577-600.
- Bartlett, F.C. (1932), *Remembering*, Cambridge, MA, Cambridge University Press.
- Beer, R. (1995), "A dynamical systems perspective on agent-environment interaction", *Artificial Intelligence*, 72, pp. 173-215.
- Brentano, F. (1985), *Psychology from an Empirical Standpoint*, translation by A.C. Rancurello, D.B. Terrell, L.L. McAlister, London, Routledge.
- Brooks, R.A. (1991), "Intelligence without representation", *Artificial Intelligence*, 47, pp. 139-159.
- Bruner, J.S., Goodnow, J.J., Austin, G.A. (1956), *A study of thinking*, New York, Wiley.
- Castelfranchi, C. (1995), "Guarantees for autonomy in cognitive agent architecture", in M. Wooldridge, N.R. Jennings (eds.), *Intelligent Agents: Theories, Architectures, and Languages*, n. 890 in *LNAI*, Berlin, New York, Springer-Verlag, pp. 56-70.
- Chomsky, N. (1995), "Language and nature", *Mind*, 104, pp. 1-61.
- Churchland, P.S. (1986), *Neurophilosophy: toward a unified science of the mind-brain*, Cambridge, MA, MIT Press.
- Clark, A. (1997), *Being There. Putting Brain, Body, and World Together*, Cambridge, MA, MIT Press.
- Collins, A.M., Quillian, M.R. (1969), "Retrieval time from semantic Memory", *Journal of verbal learning and verbal behavior*, 8, pp. 240-248.
- Dennett, D. (1978), *Brainstorms*, Cambridge, MA, MIT Press.
- Dennett, D. (1987), *The intentional stance*, Cambridge, MA, MIT Press.
- Doyle, J., McDermott, D. (1980), "Nonmonotonic logic", *Artificial Intelligence*, 13, pp. 41-72.
- Drescher, G.L. (1991), *Made-Up Minds: A Constructivist Approach to Artificial Intelligence*, Cambridge, MA, MIT Press.
- Fodor, J., Pylyshyn, Z. (1988), "Connectionism and cognitive architecture: a critical analysis", *Cognition*, 28, pp. 3-71.
- Frege, G. (1952), "On sense and reference", in P. Geach, M. Black (eds.), *Translations from the Philosophical Writings of Gottlob Frege*, Oxford, Basil Blackwell.
- Gardenfors, P. (2000), *Conceptual Spaces: The Geometry of Thought*, Cambridge, MA, MIT Press.
- Gardenfors, P. (2004), "Conceptual spaces as a framework for knowledge representation", *Mind and Matter*, 2(2), pp. 9-27.
- Gibson, J. (1979), *The ecological approach to visual perception*, Boston, Houghton Mifflin.
- Giunchiglia, F. (1993), "Contextual reasoning", *Epistemologia*, 16, pp. 345-364.

- Gopnik, A., Meltzoff, A. (1997), *Words, Thoughts and Theorie*, Cambridge MA, MIT Press.
- Grosz, B.J., Sidner, C. (1990), *Intentions in Communication*, Cambridge, MA, MIT Press.
- Harnad, S. (1990), "The symbol grounding problem", *Physica D*, 42, pp. 335-346.
- Huhns, M., Singh, M. (1998), "Agents and multi-agent systems: Themes, approaches, and challenges", in M. Huhns, M. Singh (eds.), *Readings in Agents*, San Francisco, Morgan Kaufmann Publishers, pp. 1-23.
- Kant, I. (1781-87), *Kritik der reinen Vernunft* (trad. it., *Critica della Ragione Pura*, a cura di G. Colli, Milano, Adelphi, 1995).
- Keijzer, F. (2001), *Representation and behavior*, Cambridge, MA, MIT Press.
- Kosslyn, S. (1994), *Image and Brain: The Resolution of the Imagery Debate*, Cambridge, MA, MIT Press.
- Lakoff, G. (1987), *Women, fire, and dangerous things: What categories reveal about the mind*, Chicago, University of Chicago Press.
- Lenat, D., Guha, R.V. (1990), *Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project*, Reading, MA, Addison-Wesley.
- McCarthy, J. (1959), "Programs with common sense", in *Proceedings of the Teddington Conference on the Mechanization of Thought Processes*, London, Her Majesty's Stationary Office, pp. 75-91.
- McCarthy, J., Hayes, P.J. (1969), "Philosophical problems from the standpoint of artificial intelligence", in B. Meltzer, D. Michie (eds.), *Machine Intelligence*, vol. 4., Edinburgh, Edinburgh University Press, pp. 463-502.
- Miller, G.A., Johnson-Laird, P.L. (1976), *Language and perception*, Cambridge, Eng., Cambridge University Press.
- Minsky, M. (1975), "A framework for representing knowledge", in P.H. Winston (ed.), *The Psychology of computer vision*, New York, McGraw-Hill, pp. 211-280.
- Minsky, M. (1988), *The Society of Mind*, New York, Simon & Schuster.
- Minsky, M. (2007), *The emotion machine*, New York, Simon & Schuster.
- Murphy, G., Medin, D. (1985), "The role of theories in conceptual coherence", *Psychological Review*, 92(3), pp. 289-316.
- Newell, A. (1982), "Knowledge level", *Artificial Intelligence*, 1(18), pp. 87-127.
- Newell, A. (1990), *Unified Theories of Cognition*, Cambridge, MA, Harvard University Press.
- Newell, A., Simon, H.A. (1972), *Human problem solving*, Englewood Cliffs, NJ, Prentice-Hall.
- Newell, A., Simon, H.A. (1976), "Computer science as empirical enquiry: Symbols and search", in *Communications of ACM*, 19(3), pp. 113-126.
- Newell, A., Simon, H.A., Shaw, J.C. (1958), "Elements of a theory of human problem solving", *Psychological Review*, 65, pp. 151-166.
- Nosofsky, R.M. (1988), "Exemplar-based accounts of relations between classification, recognition, and typicality", *Journal of Experimental Psychology: Learning, Memory and Cognition*, 14, pp. 700-708.
- O'Regan, J., Noe, A. (2001), "A sensorimotor account of vision and visual Consciousness", *Behavioral and Brain Sciences*, 24, 5, pp. 883-917.

- Piaget, J. (1952), *The Origins of Intelligence in Children*, New York, International Universities Press (originally published 1936).
- Popper, K. (1997), *Myth of the Framework*, New York, Routledge.
- Port, R.F., van Gelder, T. (eds.) (1995), *Mind as motion: Explorations in the dynamics of cognition*, Cambridge, MA, MIT Press.
- Putnam, H. (1975), *Mind, language and reality*, Cambridge, Eng., Cambridge University Press.
- Pylyshyn, Z. (1984), *Computation and Cognition: Toward a Foundation for Cognitive Science*, Cambridge, MA, MIT Press.
- Quillian, R. (1969), "Semantic memory", in M. Minsky (ed.), *Semantic Information Processing*, Cambridge, MA, MIT Press.
- Reiter, R. (1980), "A logic for default reasoning", *Artificial Intelligence*, 13, pp. 81-132.
- Rich, E., Knight, K. (1991), *Artificial Intelligence*, 2nd ed., New York, McGraw-Hill, 1991.
- Rosch, E. (1975), "Cognitive representations of semantic categories", *Journal of Experimental Psychology: General*, 104, pp. 192-233.
- Rosch, E., Mervis, C., Gray, W., Johnson, D., Boyes-Braem, P. (1976), "Basic objects in natural categories", *Cognitive Psychology*, 8, pp. 382-439.
- Rosenbloom, P. S., Laird, J. E., Newell, A. (1992), *The Soar Papers: Research on Integrated Intelligence*, vol. 1 and 2, Cambridge, MA, MIT Press.
- Roy, D. (2005), "Grounding words in perception and action: computational insights", *Trends Cognitive Science*, 9(8), pp. 389-396.
- Roy, D., yuh Hsiao, K., Mavridis, N., Gorniak, P. (2006), "Ripley, hand me the cup: Sensorimotor representations for grounding word meaning", in *Int. Conf. of Automatic Speech Recognition and Understanding*.
- Rumelhart, D.E., McClelland, J.L., and the PDP Research Group (1986), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, 2 voll., Cambridge, MA, MIT Press.
- Rumelhart, D.E., Ortony, A. (1977), *Schooling and the Acquisition of Knowledge*, Hillsdale, NJ, Erlbaum.
- Saussure, F. de (1986), *Corso di linguistica generale*, Roma-Bari, Laterza (introduzione, traduzione e commento di T. De Mauro).
- Schank, R.C., Abelson, R.P. (1977), *Scripts, Plans, Goals and Understanding: an Inquiry into Human Knowledge Structures*, Hillsdale, NJ, Erlbaum.
- Sloman, A., Chappell, J. (2005), "The altricial-precocial spectrum for Robots", in *Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI 2005)*.
- Smith, E.E., Medin, D. (1981), *Categories and concepts*, Cambridge, MA, Harvard University Press.
- Smolensky, P. (1988), "On the proper treatment of connectionism", *Behavior and Brain Sciences*, 11(1), pp. 1-74.
- Steels, L. (1996), "Synthesising the origins of language and meaning using coevolution, self-organisation and level formation", in C.K. Hurford, J. and M. Studdert-Kennedy, (eds.), *Evolution of Human Language*, Edinburgh, Edinburgh Univ. Press.
- Taylor, C. (1971), "What is involved in a genetic psychology?", in T. Mischel (ed.), *Cognitive development and epistemology*, New York, Academic Press, pp. 393-416.

- Wittgenstein, L. (1953), *Philosophical Investigations*, Oxford, Basil Blackwell.
- Wooldridge, M., Jennings, N.R. (1995), "Intelligent agents: Theory and Practice", *Knowledge Engineering Review*, 10, 2, pp. 115-152.
- Zadeh, L. (1965), "Fuzzy sets", *Journal of Information and Control*, 8, pp. 338-353.
- Zadeh, L. (1975), "Calculus of fuzzy restrictions", in L.A. Zadeh, K.-S. Fu, K. Tanaka, M. Shimura, (eds.), *Fuzzy Sets and Their Applications to Cognitive and Decision Processes*, New York, Academic Press.

Alcuni manuali di riferimento

- Artificial Intelligence: Structures and Strategies for Complex problem Solving*, Second Edition, Redwood City, CA, Benjamin Cummings Publishing Company, 1993.
- R.J. Brachman, H.J. Levesque (eds.), *Readings in Knowledge Representation*, Los Altos, CA, Morgan Kaufmann, 1985.
- T. Dean, J. Allen, Y. Aloimonos, *Artificial Intelligence: Theory and Practice*, Redwood City, CA, Benjamin Cummings Publishing Company, 1995.
- M.R. Genesereth, N.J. Nilsson, *Logical Foundations of Artificial Intelligence*, Los Altos, CA, Morgan Kaufmann, 1987.
- M. Ginsberg, *Essentials of Artificial Intelligence*, San Mateo, CA, Morgan Kaufmann, 1993.
- H. Reichgelt, *Knowledge Representation: An AI Perspective*, Norwood, NJ, Ablex Publishing, 1991.
- E. Rich, K. Knight, *Artificial Intelligence*, Second Edition, New York, McGraw Hill, 1991.
- D.E. Rumelhart, J.L. McClelland, and the PDP Research Group, *Parallel Distributed Processing*, Cambridge, MA, MIT Press, 1986.
- S.J. Russell, P. Norvig, *Artificial Intelligence: A Modern Approach*, Englewood Cliffs, NJ, Prentice Hall, 1995.
- S.C. Shapiro, *The Encyclopedia of Artificial Intelligence*, Second Edition, New York, Wiley, 1992.
- B.C. Smith, Prologue to "Reflection and Semantics in a Procedural Language", in R.J. Brachman, H.J. Levesque (eds.), *Readings in Knowledge Representation*, Los Altos, CA, Morgan Kaufmann, 1985.
- S.L. Tanimoto, *The Elements of Artificial Intelligence Using Common Lisp*, 2nd Edition, New York, Computer Science Press, 1995.
- P.H. Winston, *Artificial Intelligence*, Third Edition, Reading, MA, Addison Wesley, 1992.