

20. Linguaggio e categorie: apporti della simulazione con le reti neurali

di Alberto Greco e Angelo Cangelosi

In: Scopesi A., Zanobini M. (a cura di), Processi comunicativi e linguistici nei bambini e negli adulti: prospettive evolutive e sociali, Angeli, Milano, 1997, pp. 411-433.

Dai rapporti fra linguaggio e pensiero ai rapporti fra linguaggio e categorie

Il problema dei rapporti fra linguaggio e pensiero, uno dei classici della psicologia, oggi sembra aver perso di attualità. Già da un pezzo è stata superata l'idea secondo la quale le categorie che usiamo per interpretare il mondo provengono dalla lingua che parliamo (la famosa ipotesi whorfiana del *determinismo* linguistico); ma anche quella, meno forte, del relativismo — secondo cui il linguaggio *favorisce* un certo pensiero piuttosto che un altro — in effetti raramente si trova espressa in forma così diretta.

Questa perdita di interesse della ricerca sui rapporti fra linguaggio e pensiero è naturalmente legata all'evoluzione storica della psicologia. Dapprima c'è stata l'idea comportamentista della coincidenza fra linguaggio e pensiero, e studiare le relazioni fra processi che coincidono non avrebbe avuto senso. In tempi più recenti c'è stata l'influenza, forte nella prospettiva cognitivista, della concezione che vede il pensiero come implementato in un codice linguistico o "linguaggio del pensiero" (Fodor, 1975): infatti, se si considera il funzionamento dei processi cognitivi come una manipolazione di simboli che avviene secondo regole formali, non ha molto senso mettere in relazione il linguaggio con processi di natura diversa, non simbolica. Oggi tuttavia questa concezione è messa in discussione dall'approccio connessionista, che fornisce modelli in grado di riprodurre operazioni cognitive senza una esplicita rappresentazione simbolica.

Anche quando nella letteratura psicologica classica c'è stato un certo accordo sul fatto che il linguaggio può influenzare il pensiero, questo accordo non ha riguardato il modo in cui questo avvenga (v. Hunt e Agnoli, 1991, per una discussione).

In realtà l'espressione "rapporti fra linguaggio e pensiero" risente probabilmente della vaghezza dei concetti messi in relazione. Innanzitutto il concetto di "linguaggio" non è univoco. Esistono diversi tipi di linguaggio, diverse funzioni linguistiche e diversi livelli linguistici. Ci sono diversi tipi di linguaggio, anche se è invalso l'uso di usare lo stesso termine sia per il sistema di comunicazione corrispondente alle lingue naturali (allargandone il riferimento a ogni sistema di comunicazione, come quando si parla di "linguaggio non verbale" o di "linguaggio delle api") che per ogni sistema con funzioni anche non comunicative che fa uso di simboli manipolati sulla base di regole formali. Ci sono diverse funzioni linguistiche: se c'è un linguaggio usato per rappresentare a se stessi non è detto che corrisponda a quello usato per comunicare con gli altri; anche nell'ambito della sola funzione rappresentativa, un linguaggio adatto a rappresentare operazioni matematiche può non essere idoneo a rappresentare le motivazioni umane. Ci sono diversi livelli linguistici: come nell'informatica si distinguono il linguaggio macchina e i linguaggi ad alto livello, possono esistere sistemi di rappresentazione — codici interni — più vicini al livello sensoriale, e altri più strutturati, analoghi al linguaggio naturale o coincidenti con esso. È bene dunque chiarire che nel contesto qui più pertinente, quello più generale, per linguaggio si può intendere un sistema di simboli idoneo a rappresentare quanto (oggetti o eventi) possa essere pensato, cioè categorizzato, senza impegnarsi sul livello, alto o basso, di analisi.

Ma è soprattutto il concetto di "pensiero" che oggi tende a essere rivisto e reso più specifico, in quanto in effetti ingloba processi psichici diversi: problem solving, ragionamento, attività di categorizzazione o concettualizzazione. È possibile teoricamente considerare il rapporto del linguaggio con ciascuno di questi aspetti del pensiero: ad esempio ci si può chiedere se avere un codice per denominare gli elementi rilevanti dei problemi (stati e operatori) ne migliori la soluzione; oppure se poter denominare le ipotesi che si formulano aiuti nel ragionamento, senza contare il ruolo che il linguaggio ha nella stessa organizzazione e sviluppo di ragionamenti complessi (cfr. Mosconi, 1990). Ma fra il linguaggio e la categorizzazione c'è un

rapporto speciale, tanto che non a caso il "vecchio" problema del determinismo o relativismo linguistico era un problema di categorie.

Il problema alla base dei rapporti fra linguaggio e categorizzazione — che costituisce anche il tema del presente scritto — è se il possedere etichette (nomi) per gli stimoli possa esercitare dei vincoli sulle categorie che verranno formate o viceversa se la creazione dei nomi (denominazione) debba necessariamente essere sempre successiva alla categorizzazione e condizionata ai processi di astrazione e rappresentazione di caratteristiche.

Non si tratta naturalmente di una alternativa da intendersi in senso assoluto. È ormai chiaro che la capacità di categorizzazione è una abilità cognitiva indipendente dal linguaggio, almeno per quanto riguarda le categorie di basso livello. Sappiamo che gli organismi interagiscono con gli oggetti dell'ambiente senza alcuna mediazione linguistica, in modo tale che esso appare organizzato in categorie di oggetti. Sappiamo, tanto per fare un esempio, che possiamo distinguere e riconoscere migliaia di stimoli (come il volto di una persona nella vita quotidiana, o una particolare configurazione della scacchiera se siamo esperti di scacchi) e usare queste informazioni per pensare, ragionare o risolvere problemi senza associarle a un particolare nome.

Un esempio chiaro della categorizzazione senza linguaggio è la classificazione degli oggetti mangiabili che fanno tutti gli animali. Un animale che si ciba di funghi è in grado di distinguere i funghi commestibili da quelli velenosi. Così come gli erbivori fanno una scelta selettiva delle erbe commestibili rispetto alle altre che non sono parte della propria dieta. Eppure i funghi commestibili, come le foglie di una pianta, sono oggetti che hanno un aspetto percettivo anche molto diverso tra loro. Inoltre, anche i funghi della stessa specie hanno forme leggermente diverse.

È altrettanto noto che a volte possiamo applicare la denominazione a categorie senza nome già esistenti e che ciò apporta benefici sul piano cognitivo. Le prove sperimentali del fatto che stimoli codificabili attraverso etichette verbali vengono ricordati più facilmente sono ormai vecchie di parecchi decenni (Clark, 1965; Greenfield, Reich e Olver, 1966). Quando il linguaggio favorisce il successo nei compiti cognitivi può sembrare che ciò avvenga anche in modi non direttamente legati alla formazione di categorie, ad esempio in quanto manipolare etichette al posto di idee complesse può ridurre il carico della memoria a breve termine. Questo, come si sa, è uno dei

motivi per cui nascono gerghi specialistici e si coniano parole speciali nella scienza o nel diritto.

Tuttavia il problema in discussione è se tali fenomeni possano essere ricondotti alla categorizzazione prima che a una economia mnemonica. In particolare, se l'aiuto che il linguaggio offre in questi processi possa essere dovuto alla stessa "identificazione" degli stati, operatori, ipotesi o regole rilevanti — il che è un fenomeno prettamente categoriale — prima ancora che all'ottimizzazione della loro codifica per la successiva manipolazione. Ad esempio in matematica, dove spesso il riconoscimento di configurazioni problematiche è prerequisito per la soluzione, la possibilità di denominare in modo caratteristico le varie configurazioni potrebbe aiutare a riconoscerle (Cocking e Mestre, 1988).

Se si considera lo sviluppo del linguaggio e della formazione dei concetti in una prospettiva evolutiva, si osserva una stretta relazione fra i due processi. La capacità di categorizzazione e quella della comunicazione linguistica sono strettamente legate tra loro fin dalle prime espressioni delle abilità cognitive del bambino. La condivisione di concetti strutturalmente simili tra le categorie degli adulti e quelle dei bambini permette lo scambio di informazioni linguistiche relative al mondo concettuale condiviso. Ed è in base a questa condivisione di categorie che il bambino inizia il processo di apprendimento del linguaggio. Infatti gli adulti comunicano con i bambini utilizzando per lo più i nomi riferiti al livello di base della categorizzazione e questo facilita l'apprendimento delle parole riferite ai concetti di base. A sua volta la tendenza vincolante dei bambini ad assumere che i nomi si riferiscono ai concetti di base facilita il successivo sviluppo del lessico (Clark, 1993).

Un altro aspetto che induce a riconsiderare le relazioni fra categorizzazione e linguaggio è infine la disponibilità di nuovi modelli come quelli connessionisti, che rendono possibile riprendere in esame da angolature diverse le ipotesi che assegnano al linguaggio un ruolo importante nella categorizzazione. In queste pagine, oltre a prendere in esame alcuni punti che di recente sono emersi a favore di queste ipotesi, discuteremo appunto il sostegno che a esse deriva dai risultati di alcune simulazioni compiute con reti neurali.

Comportamento discriminativo e categorie

Così come per altri processi cognitivi, anche la categorizzazione può essere considerata a diversi livelli, a seconda che la si metta più o meno in relazione con i processi sensoriali o di "basso livello". Nel caso della categorizzazione si va dunque dalle categorie percettive a quelle concettuali. Anche se i fenomeni categoriali a basso e alto livello presentano somiglianze (Medin e Barsalou, 1987), le tradizioni e i paradigmi di ricerca sono stati abbastanza diversi, tanto che si parla di *percezione categoriale* a basso livello e di *categorie semantiche* ad alto livello.

Il più semplice fenomeno categoriale è quello che si può definire *comportamento discriminativo*, che consiste in una risposta selettiva agli stimoli basata sull'individuazione negli stessi di diversità e uguaglianze. Il comportamento discriminativo più complesso consiste in due capacità:

- a. la capacità di rispondere selettivamente a classi generali di stimoli (*generalizzazione*), basata sulla comunanza o costanza di caratteristiche fra uno stimolo e l'altro;
- b. la capacità di rispondere selettivamente non a uno stimolo completo ma solamente alla presenza di certe caratteristiche (*astrazione*).

Nell'ambito del comportamento discriminativo si può inoltre distinguere fra discriminazione relativa (la distinzione fra coppie di stimoli che si somigliano) e assoluta (che riguarda l'identificazione di uno stimolo isolato), ma — come ha notato Harnad (1987) — tale distinzione può essere fuorviante in quanto una discriminazione non è mai assoluta ma avviene sempre rispetto alle alternative presenti in un contesto.

Abbiamo adottato il termine *comportamento discriminativo* perché la discriminazione implica semplicemente una diversità di risposta a certi stimoli piuttosto che ad altri; non implica né la consapevolezza di tale risposta, né la consapevolezza delle *caratteristiche* sulla cui base sono individuabili gli stimoli o classi di stimoli a cui si risponde. In particolare non è necessariamente presente la capacità di rappresentarsi mentalmente tali stimoli o caratteristiche. L'ipotesi avanzata da chi vede l'influenza del linguaggio come determinante sulla categorizzazione è che questa capacità dipenda in maniera critica dalla possibilità di *riferirsi* agli stimoli o alle caratteristiche utilizzando etichette (simboli linguistici).

La tendenza a rispondere in modo diverso a stimoli percettivi appartenenti a categorie diverse (discriminazione fra categorie) è normalmente affiancata dalla tendenza a rispondere in modo simile o uguale agli stimoli appartenenti alla stessa categoria. Inoltre, se si chiede di giudicare le differenze fra stimoli che rientrano in categorie diverse, queste differenze vengono ritenute maggiori di quelle esistenti fra stimoli che appartengono alla stessa categoria, e viceversa, anche quando si può mostrare obiettivamente che la distanza è uguale (ad esempio la distanza fra due rossi o fra due gialli viene ritenuta minore di quella — uguale nello spettro — fra un rosso e un giallo). Non è chiaro se questo fenomeno, che è stato chiamato "percezione categoriale" (*categorical perception*), sia attribuibile all'effetto del linguaggio, come farebbero supporre le prime ricerche sulla categorizzazione dei colori influenzate dall'ipotesi del determinismo linguistico (Berlin e Kay, 1969). Tuttavia il problema evidentemente non è chiuso: anche per quanto riguarda i colori, attualmente si ritiene che la loro percezione sia legata a meccanismi genetici specie-specifici ma che possa essere influenzata dall'esperienza linguistica, specialmente nei bambini, o da categorizzazioni precedenti (Bornstein, 1987; Goldstone, 1995).

Categorizzazione, rappresentazione e denominazione

Quando dalla percezione categoriale ci si sposta alla categorizzazione di più alto livello, diventa cruciale e di estrema importanza la distinzione tra *categorie* e *caratteristiche* (che ad alto livello sono anche chiamate *attributi*). Infatti molto spesso c'è una sovrapposizione fra le dimensioni elementari che sembrano essere i mattoni con cui si costruiscono le categorie e le categorie stesse. Ciò è dovuto al fatto che nel momento in cui una caratteristica viene "identificata" e resa oggetto di rappresentazione diventa essa stessa una categoria. Questo processo tuttavia non va avanti all'infinito, in quanto non sempre tutte le dimensioni degli stimoli rilevanti per la categorizzazione vengono rappresentate distintamente in modo da poter costituire una categoria.

Dal punto di vista della relazione con la denominazione linguistica, le *caratteristiche* sono aspetti dello stimolo che possono avere tre *status* differenti:

a. contribuiscono implicitamente alla discriminazione;

b. contribuiscono alla discriminazione in maniera esplicita ma non hanno un nome;

c. contribuiscono alla discriminazione in maniera esplicita e sono anche denominate.

Questa tricotomia corrisponde grosso modo a quella delineata da Clark e Karmiloff-Smith, con i loro tre livelli di ridefinizione delle rappresentazioni (Clark e Karmiloff-Smith, 1993; Karmiloff-Smith, 1994).

Il problema che stiamo indagando può essere dunque meglio precisato come riguardante la relazione fra rappresentazione e denominazione. Fra le più importanti questioni che sono in gioco ci sono le seguenti:

- se e come siano possibili rappresentazioni non denominate, non denominabili o addirittura non simboliche;
- se le rappresentazioni contrassegnate da un nome abbiano uno *status* speciale nei processi cognitivi, in particolare nel porre vincoli sulle successive discriminazioni.

Per quanto riguarda la prima domanda, è noto che è possibile creare e rappresentarsi concetti complessi non associati a etichette linguistiche esplicite: un concetto che in una lingua naturale corrisponde a un singolo termine può essere espresso da una circonlocuzione in un'altra lingua. Si può ritenere che alcuni concetti complessi per i quali non abbiamo un nome corrispondano magari a codici interni del tipo ipotizzato dalla teoria delle reti semantiche. C'è tuttavia un'ipotesi alternativa, che esclude la necessità di codici interni: quella connessionista delle rappresentazioni non simboliche (v. Greco, 1997).

Per quanto riguarda la seconda domanda, è evidente che se si potesse dimostrare che alcune discriminazioni dipendono dalle categorie che siamo in grado di denominare sarebbe in gioco una influenza del linguaggio sul pensiero. In effetti ci sono diverse evidenze in tal senso.

a. L'idea secondo cui la categorizzazione si basa su un repertorio fisso di caratteristiche di base, che verrebbero combinate fra loro per produrre le rappresentazioni più complesse, dopo essere stata criticata sulla base di considerazioni metateoriche improntate al costruttivismo sociale (cfr. Greco, 1995, pp. 68 sgg.), viene oggi sostenuta su basi empiriche da Schyns *et al.* (1997). Questi autori sostengono la non fissità delle caratteristiche, che possono essere anche "create"

in maniera flessibile sulla base di vincoli ambientali e delle precedenti categorizzazioni. Un'idea simile per alcuni aspetti è quella di "emergenza verso il basso" (*downwards emergence*; De Vries, 1995), secondo cui non è importante solo — come si ritiene ormai comunemente nella letteratura connessionista — l'"emergere" di rappresentazioni ad alto livello da quelle di basso livello, ma anche il processo inverso di scoperta di nuove caratteristiche a basso livello sulla base delle rappresentazioni di alto livello.

b. È noto che il *commento* linguistico può avere la funzione di dirigere l'attenzione su certe caratteristiche piuttosto che su altre. Ad esempio nel mostrare a un bambino un'automobilina è possibile dire "guarda che bella, è una Ferrari rossa" in modo da fargli creare una nuova categoria (Ferrari) o associare esplicitamente un nuovo oggetto a una o all'altra categoria già esistente (degli oggetti rossi o degli oggetti belli, ecc.). Come ha fatto notare Parisi (1995), la rappresentazione di un oggetto non è mai "puramente" percettiva ma risente sempre degli stimoli concomitanti, fra i quali i commenti linguistici sono fondamentali nell'esperienza del mondo fin da bambini. Tutte le esperienze percettive del bambino sono pressoché costantemente accompagnate da stimoli linguistici, che normalmente sono arbitrari, non avendo una relazione diretta con le prime, e ciò fa sì che la rappresentazione dello stesso oggetto o evento non sia mai esattamente la stessa quando il commento linguistico è diverso.

Categorizzazione e linguaggio: l'apporto delle reti neurali

Da quanto si è esaminato finora, è chiaro che un aspetto importante della relazione tra linguaggio e categorizzazione è il valore aggiunto che il linguaggio ha nella formazione e organizzazione strutturale delle categorie, o meglio nella rappresentazione mentale del significato comune agli oggetti di una stessa categoria. In particolare è interessante chiedersi quali conseguenze abbia la disponibilità di un'etichetta linguistica nel processo di acquisizione delle categorie e la sua applicazione alla struttura stessa di una categoria e alla sua rappresentazione mentale.

In letteratura non esistono molti studi che hanno confrontato in maniera sistematica l'influenza di fattori linguistici e non linguistici sulla formazione dei concetti.

Lo studio della categorizzazione negli animali non è il più adatto per compiere indagini sperimentali sulla rappresentazione dei concetti. Anche nelle specie che usano dei sistemi di comunicazione, è difficile studiare il ruolo di questi ultimi nella formazione dei concetti. Ad esempio Cheney e Seyfarth (1990) hanno fatto delle osservazioni sull'apprendimento delle vocalizzazioni nelle giovani scimmie *vervet* (*cercopithecus*), analizzando i cambiamenti referenziali cui è soggetto l'uso delle grida di allarme per i predatori come aquile, leopardi e serpenti. Ma anche con questi animali è impossibile fare uno studio preciso sulla relazione tra le categorie senza etichette e quelle a cui si associa l'uso di un segnale di comunicazione.

Negli studi con soggetti umani, invece, la categorizzazione è quasi sempre associata all'uso di parole o etichette linguistiche. Il classico compito sperimentale di categorizzazione in laboratorio prevede infatti che il soggetto formi delle nuove categorie associando un'etichetta linguistica a un gruppo di oggetti. E anche negli studi classici con etichette "senza senso" non si può facilmente escludere l'implicazione del nostro sistema linguistico.

È arduo anche indagare sul ruolo che il linguaggio potrebbe avere sulla struttura delle rappresentazioni concettuali, in quanto ci sono dei limiti sperimentali difficili da superare. Per esempio è molto difficile pensare a una ricerca sperimentale che intenda studiare il modo in cui il linguaggio trasforma e ristrutturata la rappresentazione mentale di categorie che sono state precedentemente acquisite attraverso una modalità non linguistica.

La simulazione con le reti neurali sembra essere una buona metodologia per una analisi sistematica del ruolo specifico del linguaggio nella categorizzazione e degli effetti della denominazione sulla struttura di rappresentazione dei concetti. Le reti neurali, o modelli connessionisti, sono infatti dei modelli computazionali adatti a simulare compiti cognitivi, come appunto la categorizzazione di stimoli e compiti linguistici. Esse hanno inoltre il vantaggio di essere vincolate ad alcuni principi strutturali e funzionali del sistema nervoso, come l'organizzazione in gruppi di unità (neuroni artificiali) interconnessi, l'elaborazione in parallelo e la rappresentazione distribuita delle informazioni (Rumelhart e McClelland, 1986).

Queste due caratteristiche rendono le reti neurali particolarmente adatte per la costruzione di modelli cognitivi computazionali in cui si vogliono anche indagare alcuni aspetti della rappresentazione neurale delle categorie. Questi modelli permettono, ad esempio, degli

studi sistematici sugli aspetti quantitativi delle rappresentazioni della rete, e cioè sul pattern di attivazione delle unità interne della stessa (Hanson e Burr, 1990). Sui valori numerici di attivazione delle unità, ottenute dalla rete in risposta all'*input* di uno stimolo linguistico o percettivo, è infatti possibile effettuare delle misure quantitative per comparare tra loro le rappresentazioni di diverse reti neurali, o quelle ottenute da diversi stimoli nella stessa rete.

Inoltre la simulazione sul calcolatore permette di gestire e manipolare un gran numero di variabili. In un modello per la categorizzazione e la denominazione di stimoli possono essere manipolate, ad esempio, le variabili relative al livello di apprendimento come la durata dell'addestramento, la frequenza di presentazione di ciascuno stimolo, il tasso di apprendimento della rete. Inoltre possono essere tra loro comparate delle condizioni sperimentali in cui l'*input* visivo e linguistico degli stimoli viene sistematicamente combinato ottenendo così situazioni paragonabili di categorizzazione di stimoli visivi senza linguaggio, categorizzazione di stimoli visivi senza linguaggio, categorizzazione a partire solo dal nome degli oggetti, categorizzazione con entrambi gli *input*.

Nella simulazione connessionista dei compiti linguistici e di categorizzazione sono stati finora seguiti diversi approcci. Tuttavia la gran parte dei lavori sperimentali ha il limite di studiare isolatamente queste due capacità cognitive, utilizzando modelli che svolgono un compito di sola elaborazione linguistica o di sola classificazione degli stimoli. Un modello inteso a esplorare l'influenza del linguaggio nella formazione dei concetti dovrebbe prevedere la contemporanea simulazione dell'apprendimento dei concetti a partire dall'*input* degli stimoli percettivi e a partire dall'*input* dei nomi degli oggetti. Infatti solo la combinazione dei due compiti in una stessa rete neurale può permettere di verificare come la rappresentazione neurale dei concetti è influenzata dal linguaggio.

Prima di discutere in dettaglio alcuni dei modelli più recenti che simulano insieme i compiti di denominazione e categorizzazione degli stimoli, accenneremo brevemente alla variabilità di approcci nei modelli per le singole abilità cognitive.

Tra le diverse metodologie connessioniste per compiti linguistici le più comuni sono quelle che riguardano la lettura e scrittura e quelle riguardanti la simulazione di problemi di morfologia. Questi modelli prevedono per lo più una struttura della rete a più strati *feedforward* in cui gli stimoli linguistici di *input* e di *output* vengono codificati tenendo conto delle caratteristiche rilevanti per il compito

linguistico. Ad esempio in compiti di lettura è importante che gli stimoli di *input* siano costruiti secondo un codice che rappresenti la struttura grafemica delle lettere, mentre le parole di *output* riporteranno le informazioni fonologiche rilevanti. Gli algoritmi di apprendimento della rete neurale in questi compiti linguistici sono prevalentemente di tipo supervisionato, in cui cioè si usa un valore corretto di attivazione delle unità di *output* per correggere e aggiustare i pesi delle connessioni. Tra questi metodi quello sicuramente più diffuso è l'algoritmo di *backpropagation*.

Alcuni tra i modelli di reti neurali linguistiche più studiati e conosciuti sono quelli sulla formazione del tempo passato per i verbi inglesi e quello della lettura di parole inglesi. Nel modello di Rumelhart e McClelland (1986b) sull'acquisizione e formazione del passato è stata utilizzata una rete neurale a due strati in cui la codifica delle parole tiene conto della rappresentazione fonologica del verbo. Per poter includere nel modello gli effetti di contesto dovuti alla concatenazione sequenziale dei fonemi nella parola, sono state usate delle unità per la rappresentazione delle combinazioni di trigrammi fonetici, dette *wickelphones*. Nel modello della lettura di Seidenberg e McClelland (1989) è stata simulata una grande rete neurale a tre strati, con 400 unità ortografiche di *input*, 200 unità nascoste e 460 unità per l'*output* fonologico. Oltre alle connessioni *feedforward*, vi erano delle connessioni ricorrenti tra le unità *hidden* e quelle di *input*. La rete neurale è stata addestrata secondo l'algoritmo della *backpropagation* con un *set* di stimoli costituito da 2897 parole inglesi monosillabiche di almeno tre lettere. La frequenza di presentazione di ogni parola alla rete era proporzionale al logaritmo della frequenza di uso delle parole.

Le reti neurali sono usate anche per la categorizzazione e classificazione di stimoli; questo compito è stato anzi uno degli iniziali campi di applicazione di queste reti e rimane anche adesso uno dei più diffusi. Vi sono numerosi lavori che usano le reti neurali per la classificazione di stimoli di vario tipo (immagini visive, stimoli acustici) e per il riconoscimento di nuove immagini in base alla conoscenza estratta dagli esemplari di addestramento. Le reti neurali sono infatti molto adatte alla simulazione di compiti di classificazione e di riconoscimento di configurazioni per via della loro capacità a estrarre le caratteristiche comuni salienti dell'*input* identificando le proprietà dell'*input* visivo comuni a più stimoli.

Tra i modelli di categorizzazione che hanno studiato problemi più vicini ai temi della psicologia cognitiva vi sono il lavoro di Shanks (1991) e quello di Jacobs e Kosslyn (1994). Nel modello di Jacobs e Kosslyn, ad esempio, si utilizzano delle reti che devono classificare e riconoscere degli stimoli retinici costituiti da configurazioni casuali di punti. La rete neurale deve imparare a identificare la posizione dello stimolo e la sua categoria semantica di appartenenza.

Modelli integrati per compiti di categorizzazione e denominazione

Nei modelli che ci accingiamo qui a descrivere vedremo come la simulazione integrata dei due compiti di categorizzazione e denominazione di stimoli può servire a studiare i fenomeni di interazione tra queste due abilità cognitive.

Un primo studio che ha affrontato in maniera diretta la relazione tra categorizzazione e denominazione è il lavoro di Plunkett e Sinha (1992) che ha ripreso la situazione sperimentale della categorizzazione di configurazioni di punti (Posner e Keele, 1968). Con questo modello Plunkett e Sinha hanno inteso studiare la relazione fra la formazione di categorie, a partire da esemplari prototipici e non, e i fenomeni di comprensione-produzione linguistica. Gli autori hanno addestrato una rete a riprodurre in *output* la rappresentazione visiva della stessa configurazione di punti data in *input* (rete autoassociativa), aggiungendo delle ulteriori unità di *input* e *output* linguistico in modo da permettere di associare una etichetta alle configurazioni visive. La rete, dopo aver appreso a riprodurre separatamente gli stimoli visivi sulle unità di *output* visivo e quelli verbali sulle unità di *output* verbale, veniva addestrata ad associare etichette verbali con immagini. Diverse immagini simili potevano avere un'unica etichetta. Successivamente veniva verificata la capacità di "produrre" una denominazione linguistica sulla base di una sola immagine e viceversa la capacità di "comprendere" un nome attraverso la rappresentazione in *output* della corrispondente immagine.

Il modello di Plunkett e Sinha ha mostrato che il linguaggio facilita la formazione delle categorie in termini di velocità dell'apprendimento. Nella simulazione del modello si osservano dei fenomeni di esplosione del vocabolario (*vocabulary spurt*) in cui si ha una fase iniziale con apprendimento di poche parole e poi una

improvvisa crescita del numero di queste. Infine gli autori mostrano anche che la prestazione nel compito di comprensione linguistica è migliore che nella produzione soprattutto nelle prime fasi dell'apprendimento.

Un modello a reti neurali con il quale è stato studiato in maniera diretta il valore aggiunto del linguaggio nella rappresentazione concettuale è quello di Harnad, Hanson e Lubin (1991; 1995). Essi hanno addestrato una rete neurale a categorizzare delle linee in due gruppi (le linee lunghe e quelle corte) e a denominarle attraverso due etichette: "lunga" e "corta". Dopo che la rete apprende ad autoassociare le linee, cioè a riprodurre in *output* la stessa configurazione di *input*, essa viene addestrata ad associare in *output* le due etichette alle diverse linee. L'uso di una rete autoassociativa con uno strato di unità nascoste è stato deciso in maniera che la rete debba costruire una rappresentazione compressa di ciascuno stimolo.

Analizzando le rappresentazioni interne della rete categorizzatrice prima dell'apprendimento del compito linguistico di denominazione sono stati trovati alcuni effetti tipici della teoria della percezione categoriale. Ad esempio è stato osservato che la rappresentazione interna delle linee viene ridotta tra i membri della stessa categoria (variabilità intra-categoriale). La misura della distanza tra le rappresentazioni è ottenuta attraverso dei calcoli sulla rappresentazione euclidea delle attivazioni delle unità nascoste della rete neurale. Confrontando invece la rappresentazione categoriale degli stimoli prima e dopo l'apprendimento del compito di denominazione delle categorie si è osservato che il linguaggio ha l'effetto di ridurre maggiormente la variabilità delle distanze intra-categoriali e di aumentare quella inter-categoriale.

Risultati simili a questi sono stati osservati negli studi di Harnad e Cangelosi (in preparazione) e di Cangelosi (1997) in cui vengono utilizzati modelli di reti neurali ecologiche. Un modello prevede, ad esempio, la simulazione di una popolazione di organismi che vivono in un mondo artificiale e devono evolvere la capacità di discriminare i funghi commestibili da quelli velenosi. La rete neurale che controlla il comportamento degli organismi deve organizzarsi in maniera da imparare a riconoscere e categorizzare i funghi raccogliendo quelli commestibili e allontanandosi da quelli velenosi. Gli organismi possono inoltre comunicare attraverso dei segnali generati dalla loro rete neurale in risposta alla differente categoria di funghi.

L'analisi dell'attivazione delle unità nascoste della rete neurale mostra come l'uso dei segnali di comunicazione trasformi la rappresentazione neurale che gli organismi si costruiscono. Infatti le rappresentazioni categoriali in organismi con linguaggio sono meglio strutturate rispetto alle stesse rappresentazioni di organismi senza linguaggio. Per esempio, i valori delle distanze euclidee tra le categorie opposte aumentano con la presenza del comportamento di comunicazione linguistica.

Nello studio di Greco e Cangelosi (1996) è stato preso in considerazione il ruolo dell'*input* linguistico nella rappresentazione interna di alcuni oggetti che possono appartenere contemporaneamente a diverse categorie. La rete neurale deve apprendere a raggruppare e denominare gli oggetti secondo tre diversi criteri di categorizzazione: colore, forma e funzione. L'*input* alla rete è sia di tipo percettivo, con la codifica delle categorie salienti della forma dell'oggetto, che linguistico, con una codifica fonetica delle etichette linguistiche.

Useremo la descrizione dettagliata di quest'ultimo modello per mostrare come la relazione tra categorizzazione e linguaggio possa essere agevolmente indagata con i modelli connessionisti integrati di categorizzazione e denominazione.

Poiché lo scopo del modello è studiare come la denominazione di uno stimolo possa avere degli effetti nel tipo di concetti che sono attivati e nella rappresentazione neurale di tali concetti, vengono utilizzate tre condizioni sperimentali di confronto nei quali la rete viene esposta ai soli stimoli percettivi (Solo_Oggetto), ai soli *input* linguistici (Solo_Nome), o ad ambedue i tipi di stimoli (Oggetto+Nome). La situazione qui simulata si ispira a un compito di apprendimento dei concetti in cui il bambino impara a denominare degli oggetti che gli vengono mostrati solo visivamente, solo con il loro nome, o mostrando contemporaneamente l'oggetto e una etichetta linguistica a esso riferito.

Il compito di denominazione degli oggetti si articola in tre sottocompiti: qual è il nome dell'oggetto, quale il suo colore, quale la sua funzione. Questi tre sottocompiti corrispondono alle tre tipologie di categorie che la rete dovrà imparare a denominare: Nome, Colore e Funzione. Ad esempio, quando il sottocompito richiede di categorizzare l'oggetto in base alla tipologia del Colore, la rete dovrà denominare l'oggetto con l'etichetta della specifica classe di appartenenza (rosso o blu). Nella condizione Solo_Oggetto, poiché la richiesta del tipo di categoria da denominare sarebbe ambigua, vengono usate

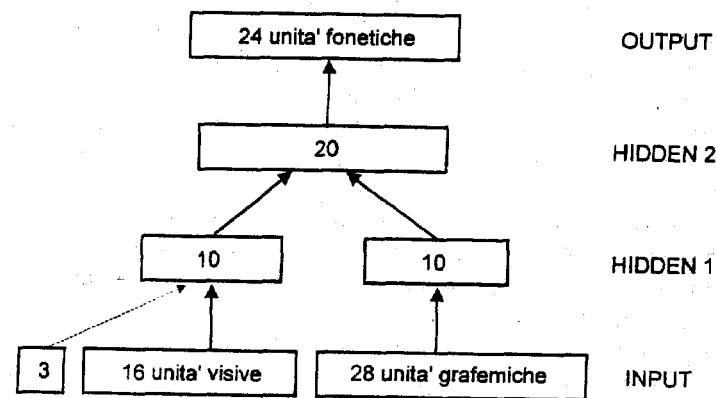
tre unità di contesto che attivano la specifica tipologia di categoria in base al quale l'oggetto deve essere denominato.

Il compito della rete neurale è quindi quello di identificare e denominare la specifica categoria di appartenenza. Il *set* di stimoli è costituito da 4 diversi oggetti: ascia, bullone, penna e inchiostro. I quattro oggetti possono appartenere a due diverse categorie funzionali. Ciascun oggetto è presentato in due diversi colori. La codifica delle caratteristiche visuo-percettive degli oggetti e delle caratteristiche fonologiche dei loro nomi è ottenuta secondo il sistema di codifica distribuito usato da Plaut e Shallice (1993).

L'architettura della rete è costituita da quattro strati *feedforward* (Fig. 1). Delle 47 unità di *input*, 16 codificano le proprietà visive dell'oggetto, 28 codificano la rappresentazione fonetica degli stimoli e 3 fungono da unità di contesto per indicare il sottocompito che la rete deve eseguire. Le unità fonetiche di *output* sono 24. Nel primo strato di unità nascoste vi sono due gruppi separati di 10 unità per l'*input* visivo e 10 per quello linguistico. Il secondo strato nascosto contiene invece 20 unità raggruppate insieme. L'algoritmo di apprendimento è la *backpropagation*. Nelle simulazioni sono state necessarie circa 1000 epoche di addestramento per arrivare a un livello di errore molto basso inferiore all'1%. Un'epoca di addestramento era costituita da 72 presentazioni di *input* (4 oggetti X 2 colori X 3 sottocompiti X 3 condizioni).

Sui valori di attivazione delle unità nascoste sono state applicate

Figura 1 - Architettura della rete neurale



delle tecniche di analisi quantitativa perché l'obiettivo della simulazione era lo studio del modo in cui le rappresentazioni neurali interne della rete neurale si organizzano in conseguenza dell'*input* linguistico e non. In particolare, utilizzando il metodo della analisi dei *cluster* è possibile vedere quanto le attivazioni delle unità nascoste sono tra loro simili nelle tre condizioni sperimentali (Solo_Oggetto, Solo_Nome, Oggetto+Nome) e per i tre sottocompiti (Nome, Colore, Funzione). Attraverso i dendrogrammi, che rappresentano le gerarchie di raggruppamento delle rappresentazioni in base alla somiglianza tra i valori di attivazione delle unità, si riesce a dare una rappresentazione grafica della somiglianza tra le attivazioni.

I principali risultati che sono stati ottenuti dall'analisi dei *cluster* delle unità degli strati nascosti sono i seguenti.

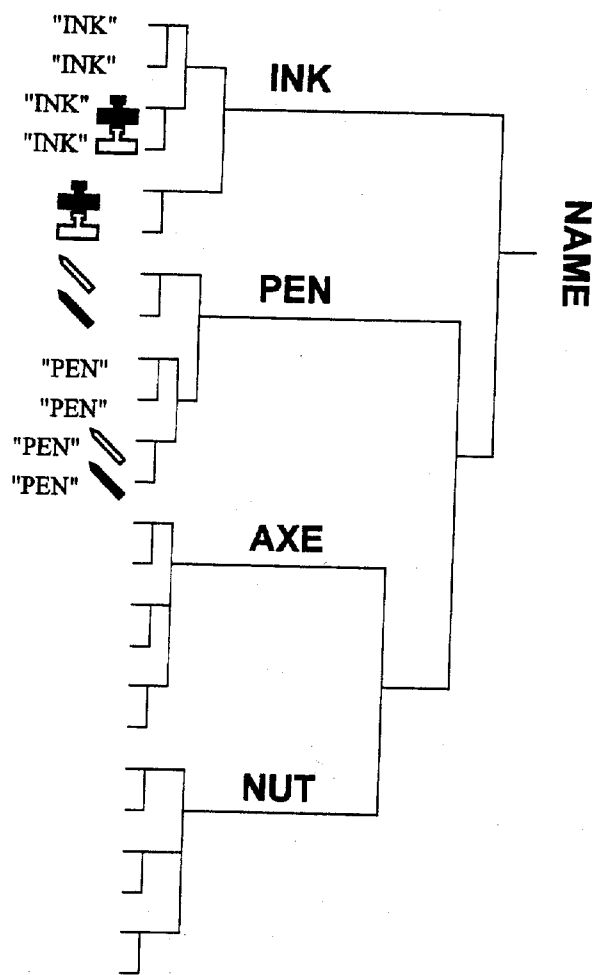
1. Il dendrogramma generale sulle attivazioni del secondo strato nascosto (riportato parzialmente nella Fig. 2) mostra come emergono delle rappresentazioni interne categoriali che riflettono la struttura semantica degli stimoli. Cioè la rete neurale mentre impara a denominare gli oggetti costruisce delle rappresentazioni specifiche per le diverse categorie.

2. Nel confrontare le attivazioni neurali delle unità interne rispetto alle tre condizioni sperimentali si vede che la rappresentazione dello stesso oggetto è molto simile in tutte le situazioni. Ad esempio, la rappresentazione della penna blu nella condizione Solo_Nome è molto simile alla stessa rappresentazione elicitata dall'*input* delle proprietà percettive nella condizione Solo_Oggetto.

3. L'analisi dei *cluster* per le unità del primo strato nascosto, quelle in cui le due informazioni linguistiche e percettive sono elaborate separatamente e non interagiscono, indica che, separatamente per ciascun gruppo di unità, ogni oggetto attiva lo stesso pattern di rappresentazione nelle diverse condizioni di *input*.

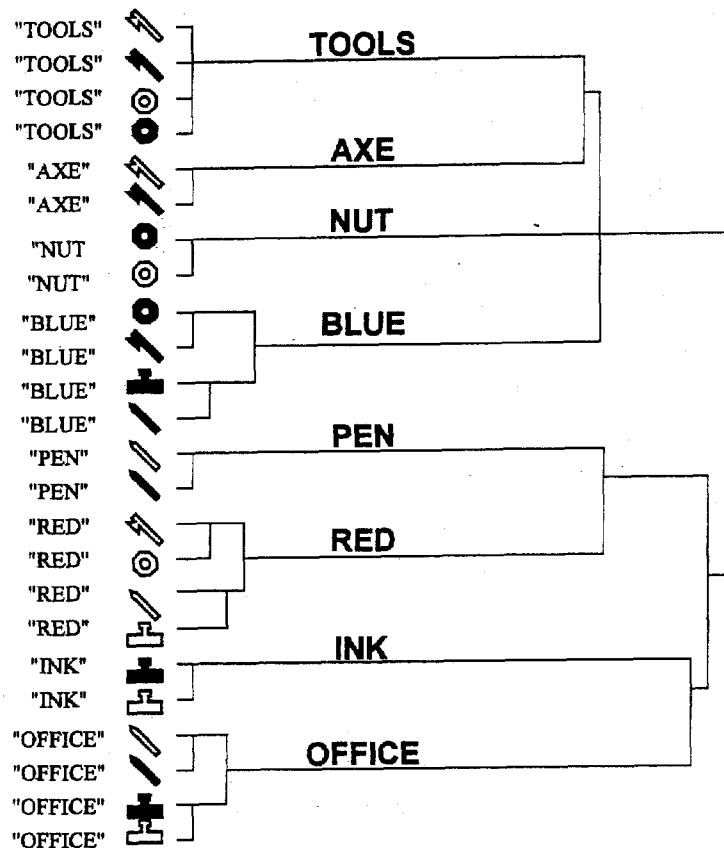
4. Lo studio del modo in cui lo stesso oggetto viene rappresentato all'interno di ciascuna condizione di *input* permette di capire come il linguaggio influenzi le differenti rappresentazioni categoriali ottenibili a partire dallo stesso *input* visivo. Il dendrogramma per la condizione Oggetto+Nome (Fig. 3) mostra che l'*input* corrispondente alle caratteristiche visive di un singolo oggetto attiva tre diverse rappresentazioni neurali a seconda del sottocompito di categorizzazione indicato dal nome. L'*input* delle caratteristiche visive della penna blu cambia a seconda dell'etichetta che è a esso associato. Nel sottocompito Colore la rappresentazione neurale della penna blu è simile

Figura 2 – Analisi dei cluster delle rappresentazioni interne (sottocompito del Nome)



a quella di tutti gli altri oggetti blu. Cioè si raggruppa con gli altri oggetti per rappresentare il concetto di "bluezza" e si distanzia dagli altri oggetti rossi appartenenti al concetto di "rossità". Lo stesso avviene nella condizione Nome e in quella Funzione.

Figura 3 – Analisi dei cluster delle rappresentazioni interne (condizione Oggetto+Nome)



Questi risultati mostrano come sia possibile studiare sistematicamente il modo in cui il linguaggio e la categorizzazione interagiscono. Il linguaggio è infatti in grado di guidare e modificare la rappresentazione categoriale degli oggetti che la rete riceve in *input*. Nel confronto delle rappresentazioni interne tramite il metodo della analisi dei *cluster* viene mostrato come l'*input* linguistico abbia un ruolo rilevante nel contenuto della rappresentazione del concetto. Infatti a seconda dell'*input* linguistico le rappresentazioni interne dei concetti conservano informazioni solamente sul colore o sulla forma o sulla funzione dell'oggetto.

Ulteriori sviluppi di questo semplice modello potrebbero permettere maggiori approfondimenti sulla relazione esistente tra categorizzazione e linguaggio. A questo scopo sarebbe necessario utilizzare un numero maggiore di categorie rispetto al modello descritto e il ricorso a una organizzazione gerarchica multilivello delle stesse categorie. Ad esempio utilizzando una rappresentazione retinica di figure geometriche si potrebbe creare una gerarchia di categorie di forme geometriche che vanno dalle forme di base, quali cerchi e quadrati, fino a categorie di ordine superiore con quello delle figure curvilinee o lineari. Associando a ciascuna di queste categorie l'etichetta linguistica di denominazione si potrebbe vedere come il linguaggio aiuta la formazione di gerarchie di concetti e come le rappresentazioni interne della rete cambiano a seconda del livello della gerarchia.

Conclusione

Il fatto che molte delle questioni che costituivano il classico tema dei rapporti fra pensiero e linguaggio siano ancora controverse (si pensi all'esistenza di problemi come quelli dei primitivi concettuali o degli universali semantici) rende tuttora interessante lo studio dei rapporti fra linguaggio e categorie. In questo articolo è stata discussa l'ipotesi secondo la quale i processi di categorizzazione possano essere vincolati dalla disponibilità di etichette linguistiche, analizzando l'apporto del metodo della simulazione con reti neurali. In particolare si è mostrato come in diversi lavori con metodologia connessionista sia possibile studiare in maniera sistematica la relazione esistente tra la rappresentazione dei concetti e la loro denominazione.

Un primo aspetto, preliminare allo studio della relazione tra linguaggio e categorizzazione, è l'analisi di rappresentazioni concettuali non denominate, cioè indipendenti da qualunque rappresentazione simbolica o proprietà linguistica. Nell'analizzare come dal comportamento discriminativo si giunga all'identificazione di caratteristiche abbiamo sottolineato quanto sia problematico considerare la categorizzazione come basata su un repertorio fisso di caratteristiche di base. L'ottica che si profila come più produttiva non è quella che concepisce i processi di basso livello come produttori del materiale a cui si applicano quelli di alto livello, ma quella che vede i due processi come complementari e circolari. In questa ottica diventa più

naturale considerare come possano esistere rappresentazioni non linguistiche o non simboliche.

Le reti neurali sono, per loro natura, sistemi a rappresentazione distribuita e non simbolica. Inoltre esse sono adatte all'estrazione, dagli stimoli percettivi di *input*, delle caratteristiche a essi comuni e rilevanti per la formazione di una categoria concettuale. Queste due proprietà delle reti neurali le rendono particolarmente appropriate per la simulazione di compiti di categorizzazione di stimoli senza ricorso a sistemi linguistici o simbolici. Durante l'addestramento in un compito di classificazione la rete neurale costruisce, nelle sue unità interne, una rappresentazione distribuita (detta anche *subsimbolica*) degli stimoli di *input* senza fare ricorso ad alcuna mediazione simbolica o linguistica. Queste rappresentazioni interne, o nascoste, acquisiscono inoltre le proprietà tipiche delle rappresentazioni categoriali. In Harnad, Hanson e Lubin (1991; 1995), ad esempio, sono state analizzate quantitativamente le caratteristiche delle rappresentazioni categoriali che presentano i fenomeni di massimizzazione delle distanze inter-categoriali e di minimizzazione della variabilità intra-categoriale. Quando una rete impara a categorizzare, cioè ad associare la stessa risposta di *output* a un gruppo di stimoli percettivi, le rappresentazioni interne corrispondenti agli stimoli di una categoria sono molto simili tra di loro e inoltre tendono a essere più distanti dalle rappresentazioni corrispondenti a stimoli di categorie diverse.

L'aspetto più strettamente inerente la relazione linguaggio-categorizzazione riguarda la differenziazione e caratterizzazione delle rappresentazioni concettuali che sono state acquisite linguisticamente, cioè di quelle categorie cui è associato un nome o una etichetta. Con le reti neurali è possibile studiare le caratteristiche quantitative e qualitative della rappresentazione di concetti linguistici e il ruolo del linguaggio nella formazione di diverse rappresentazioni concettuali di uno stesso stimolo. Abbiamo presentato alcune esemplificazioni di come ciò possa essere realizzato. Nel modello di Harnad e Cangelosi (in preparazione), ad esempio, si è mostrato che le rappresentazioni di concetti acquisiti linguisticamente sono molto meglio differenziate tra loro in quanto viene accentuata la distanza inter-categoriale tra i diversi concetti. Nel lavoro di Greco e Cangelosi (1996), qui presentato più in dettaglio, è stato invece analizzato il modo in cui il commento linguistico può determinare la formazione di diverse categorie a partire dallo stesso gruppo di stimoli. La

rete neurale impara a classificare una penna blu sia come parte della categoria degli oggetti di ufficio che come parte degli oggetti di colore blu a seconda dell'etichetta linguistica a essa associata. Inoltre le rappresentazioni interne della rete per i due concetti si strutturano in maniera tale da astrarre solo quelle proprietà dello stimolo che sono coerenti con ciascuna categoria.

I modelli qui presentati costituiscono, ovviamente, solo i primi passi verso lo sviluppo di sistemi connessionisti sempre più complessi che permettano di affrontare in maniera più efficiente lo studio del rapporto tra categorizzazione e linguaggio. Gli attuali modelli di reti neurali che simulano i compiti di categorizzazione e denominazione necessitano di un maggiore livello di complessità. Ciò vale sia per quanto riguarda l'aspetto sintattico che quello semantico. Infatti da una parte la simulazione di veri e propri sistemi linguistici, anziché di parole singole, permetterebbe di studiare in che modo possa emergere naturalmente, attraverso l'astrazione, un'organizzazione sintattica e valutare il vantaggio espressivo e forse adattivo che essa comporta. Dall'altra parte, l'estensione dei modelli a gerarchie più articolate di categorie semantiche potrebbe render conto meglio della capacità tipicamente linguistica di favorire la generalizzazione, soprattutto in contesti più naturali.

La simulazione connessionista basata su sistemi integrati per la categorizzazione e la denominazione costituisce comunque già da adesso un approccio metodologico promettente per l'esplorazione dei rapporti tra le funzioni cognitive di pensiero e quelle linguistiche.

Riferimenti bibliografici

- Berlin B., Kay P. (1969). *Basic color terms: their universality and evolution*. Berkeley: University of California Press.
- Bornstein M.H. (1987). Perceptual categories in vision and audition. In Harnad S. (Ed.), *Categorical perception* (287-300). Cambridge: Cambridge University Press.
- Cangelosi A. (1997). Verso nuovi modelli computazionali del linguaggio: test e valutazione delle reti neurali ecologiche. *Tesi di Dottorato*, Università di Genova.
- Cheney D.L., Seyfarth R.M. (1990). *How monkeys see the world: inside the mind of another species*. Chicago: Chicago University Press.

- Clark H.J. (1965). Recognition memory for random shapes as a function of complexity, association value, and delay. *Journal of Experimental Psychology*, 69, 590-595.
- Clark E.V. (1993). *The lexicon in acquisition*. Cambridge: Cambridge University Press.
- Clark E.V., Karmiloff-Smith A. (1993). The cognizer's innards: a psychological and philosophical perspective on the development of thought. *Mind & Language*, 8, 487-519.
- Cocking R.R., Mestre J.P. (Eds.) (1988). *Linguistic and cultural influences on learning mathematics*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Goldstone R.L. (1995). Effects of categorization on color perception. *Psychological Science*, 6, 298-304.
- Greco A. (1995). Il disagio del cognitivismo: un simposio immaginario. In Greco A. (a cura di) *Oltre il cognitivismo* (26-84). Milano: Franco Angeli.
- Greco A. (1997). Rappresentazione e simbolo in psicologia. In Braga Illa F. (a cura di), *Livelli di rappresentazione. Percorsi tra il naturale e l'artificiale*. Urbino: Quattroventi.
- Greco A., Cangelosi A. (1996). Language, categorization, and representation: a pilot study using neural networks. *Technical Report*. DISA-96-03, Department of Anthropological Sciences, University of Genoa, Ap.
- Greenfield P.M., Reich L. Olver R.R. (1966). On culture and equivalence (II). In Bruner J.S., Olver R.R., Greenfield P.M. *Studies in cognitive growth*. New York: Wiley.
- Hanson S.J., Burr D.J. (1990). What connectionist models learn: learning and representation in connectionist networks. *Behavioral and Brain Sciences*, 13, 471-518.
- Harnad S. (1987). Category induction and representation. In Harnad S. (Ed.), *Categorical perception* (535-565). Cambridge: Cambridge University Press.
- Harnad D., Cangelosi A. (in preparazione). Theft versus toil: the adaptive advantage of grounded symbols in category acquisition.
- Harnad S., Hanson S.J., Lubin J. (1991). Categorical perception and the evolution of supervised learning in neural nets. In D.W. Power e L. Reeker (Eds.), *Working papers of the AAI Spring Symposium on Machine Learning in Natural Language and Ontology*.
- Harnad S., Hanson S.J., Lubin J. (1995). Learned categorical perception in neural nets: implication for symbol grounding. In V. Honavar e L. Hur (Eds.), *Symbol processors and connectionist network models in artificial intelligence and cognitive modelling: Step toward principled integration*. New York: Academic Press.
- Hunt E., Agnoli E. (1991). The whorfian hypothesis: a cognitive psychology perspective. *Psychological Review*, 98, 377-389.
- Jacobs R.A., Kosslyn S.M. (1994). Encoding shape and spatial relations: the role of the receptive field size in coordinating complementary representations. *Cognitive Science*, 18, 361-386.

- Karmiloff-Smith A. (1994). Précis of Beyond modularity: a developmental perspective on cognitive science. *Behavioral and Brain Sciences*, 17, 693-745.
- Markman E. (1989). *Categorization and naming in children. Problems of induction*. Cambridge, MA: Bradford (MIT Press).
- Medin D.L., Barsalou L.W. (1987). Categorization processes and categorical perception. In Harnad S. (Ed.) *Categorical perception* (455-490). Cambridge: Cambridge University Press.
- Mosconi G. (1990). *Discorso e pensiero*. Bologna: Il Mulino.
- Parisi D. (1995). Percepire un mondo commentato: cosa cambia? *Istituto di Psicologia CNR, Roma*.
- Plaut D.C., Shallice T. (1993). Deep dyslexia: a case study of connectionist neuropsychology. *Cognitive Neuropsychology*, 10, 377-500.
- Plunkett K. e Sinha C. (1992). Connectionism and developmental theory. *British Journal of Developmental Psychology*, 10, 209-254.
- Posner M.I. e Keele S.W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, 77, 353-363.
- Rumelhart D.E., McClelland J.L. (Eds.) (1986a). *Parallel Distributed Processing: Explorations in the microstructure of cognition*. Cambridge, MA: MIT Press.
- Rumelhart D.E., McClelland J.L. (1986b). On learning the past tense of English verbs. In Rumelhart D.E., McClelland J.L. (Eds.), *Parallel Distributed Processing: Exploration in the microstructure of cognition: (vol. 1, 318-362)*. Cambridge, MA: MIT Press.
- Shanks D.R. (1991). Categorization by a connectionist network. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 17, 3, 433-443.
- Schyns P.G., Goldstone R.L., Thibaut J.P. (in stampa). The development of features in object concepts. *Behavioral and Brain Sciences*.
- Seidenberg M.S., McClelland J.L. (1989a). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96, 523-568.