

***Grounding transfer* di simboli con reti neurali**

A. Greco, T. Riga, A. Cangelosi

Università di Genova
Università di Plymouth

Paper presentato al **Congresso dell'Associazione Italiana di Scienze Cognitive (AISC)**
Rovereto 20-21 settembre 2002.

1. Introduzione

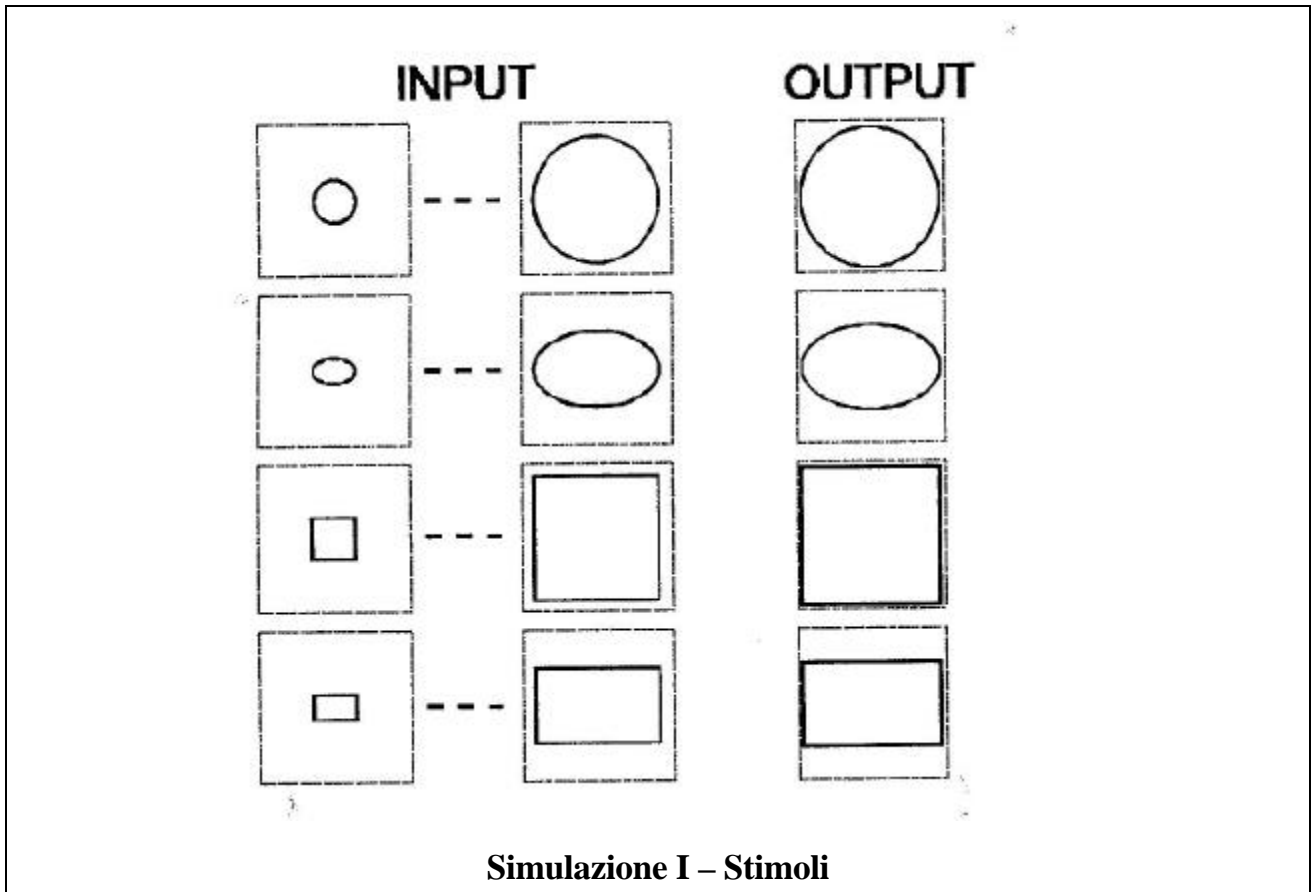
Il *symbol grounding* è uno dei principali problemi incontrati dalla metodologia simulativa in psicologia. Nelle simulazioni di tipo tradizionale, il significato dei simboli non è intrinseco al sistema ma dipende da una interpretazione esterna ad esso. I programmi, infatti, sono insiemi di regole o algoritmi per manipolare simboli, che non hanno di per sé significato ma che vengono sistematicamente interpretati come significativi. Lo stesso problema si può riscontrare nelle simulazioni con reti neurali, quando agli input e output vengano attribuiti significati psicologici che dipendono in realtà dall'interpretazione di chi le costruisce o usa.

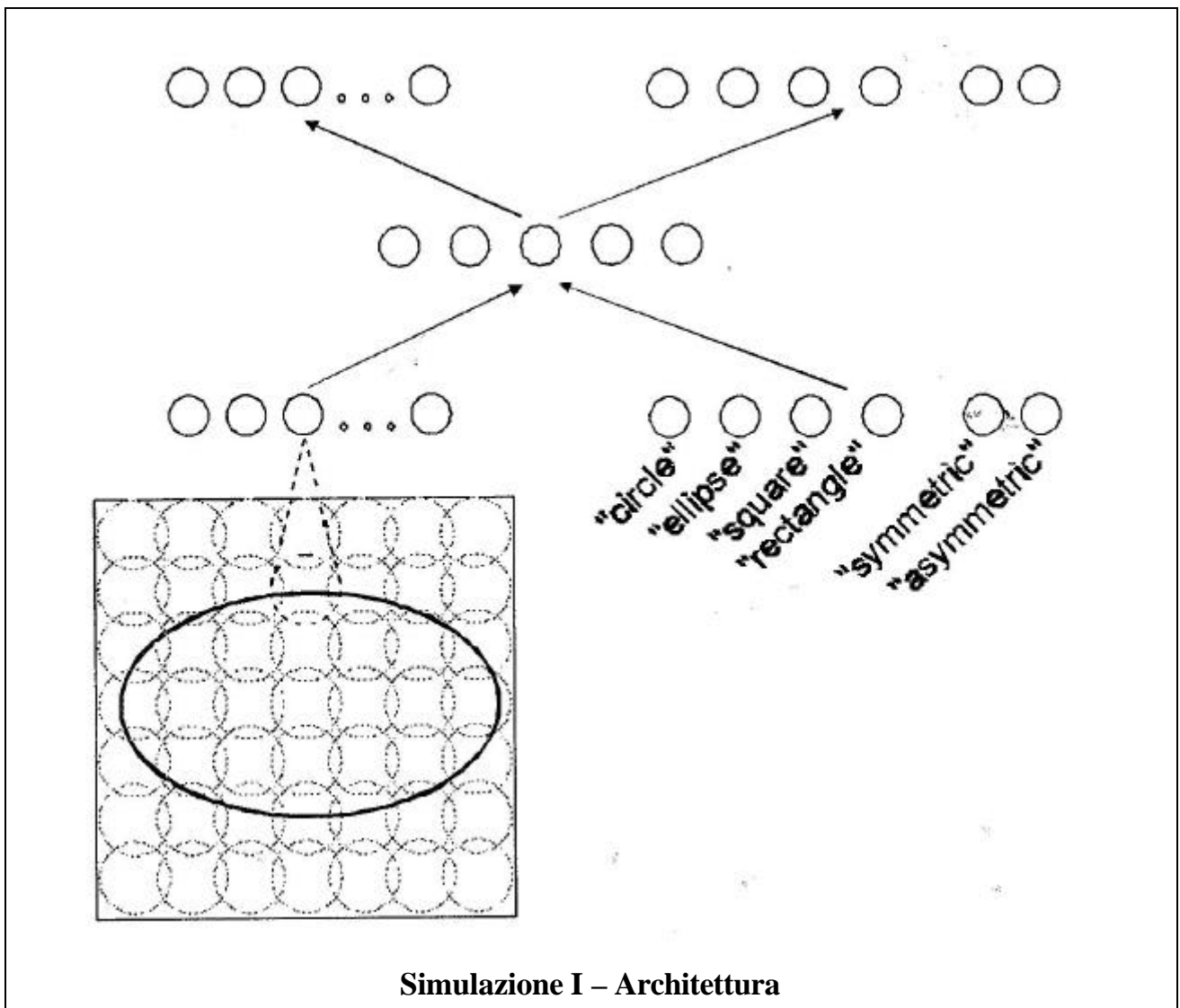
Lo scopo del presente lavoro è mostrare come questo problema possa essere affrontato con la metodologia connessionista. È possibile costruire reti neurali che categorizzano e associano etichette "linguistiche" a tali categorie (Plunkett, Sinha, Moller & Strandsby, 1992; Greco & Cangelosi, 1999). Ciò che è più importante, tuttavia, è la possibilità di combinare i simboli fra loro in proposizioni, in modo da esprimere significati più complessi che vanno al di là del dominio della percezione immediata. Se non ci fossero simboli con un fondamento in rappresentazioni di origine sensomotoria, anche gli altri sarebbero privi di senso, essendo rimandati all'infinito da ogni definizione ad altri simboli non definiti (Harnad, 1990).

Per affrontare questo problema, abbiamo sviluppato alcuni modelli che imparano dapprima a categorizzare stimoli visivi attraverso un training basato esclusivamente sulla percezione (apprendimento per prove ed errori, supervisionato attraverso un feedback correttivo); poi le rappresentazioni di queste categorie sono associate a nomi, che possono essere successivamente usati come simboli per comporre descrizioni linguistiche più complesse, utili per categorizzare stimoli mai incontrati in precedenza e acquisire categorie di livello più alto. Questa acquisizione deriva dall'associazione di nomi "fondati" con nuovi nomi arbitrari. Per questo motivo, si può dire che il *grounding* viene trasferito dai primi ai secondi e questo processo è stato chiamato *grounding transfer*.

2. Simulazione I

Stimoli. In questo modello, dieci reti neurali sono state addestrate a categorizzare e denominare immagini di forme geometriche (cerchi, ellissi, quadrati e rettangoli).

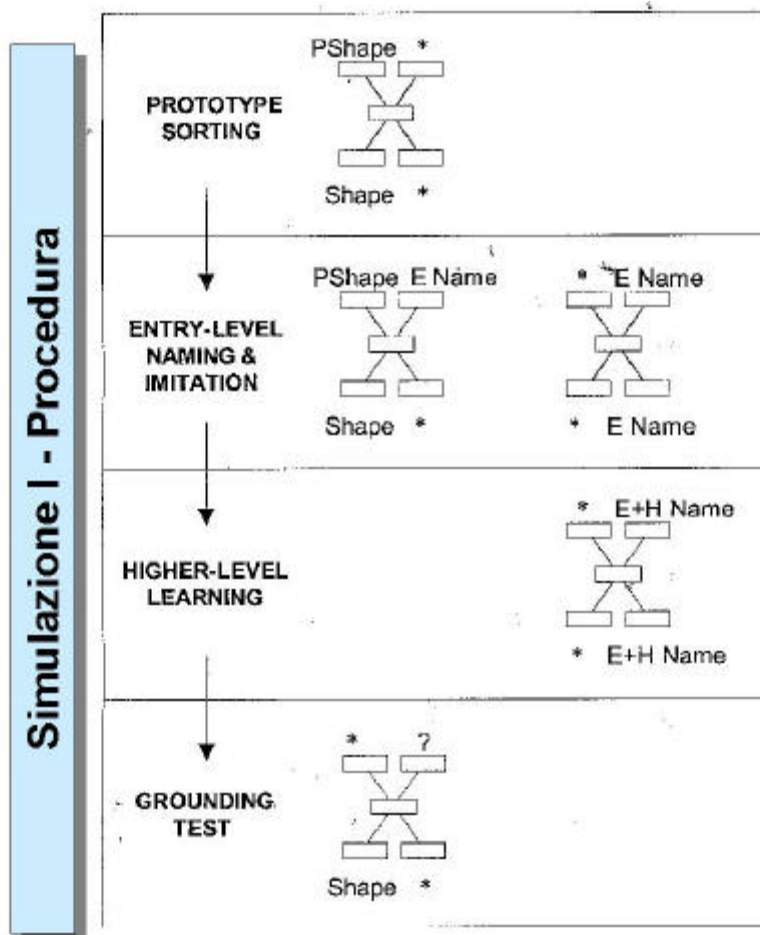




Architettura. Sono state usate 10 reti a tre strati, di tipo feed-forward, con i pesi iniziali casuali. Lo strato di input conteneva due gruppi di unità: 49 unità che simulavano la retina e 6 unità linguistiche (una per ciascun nome di categoria: "cerchio", "ellissi", "quadrato", "retangolo", "simmetrico", "asimmetrico"). Lo strato nascosto era formato da 5 unità connesse a entrambi i gruppi di unità di input. Lo strato di output aveva la stessa organizzazione di quello di input, con 49 unità corrispondenti a quelle retiniche più 6 unità per i nomi.

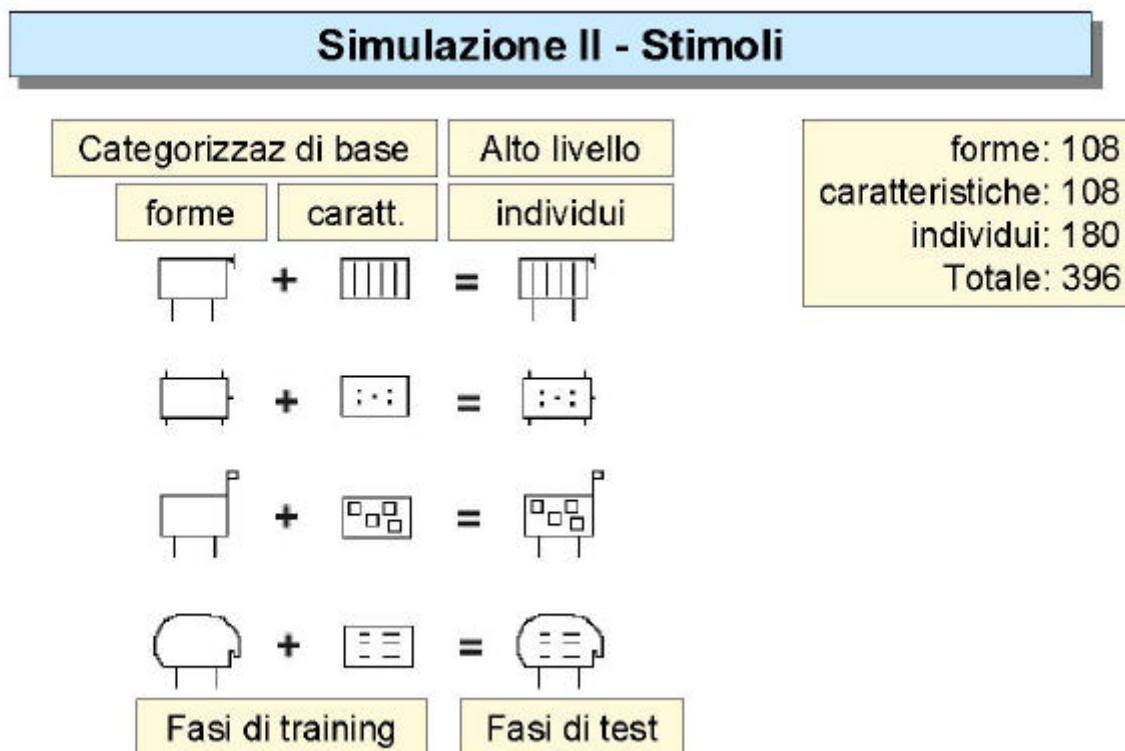
Procedura. La procedura di training era costituita da tre fasi di apprendimento supervisionato mediante la tecnica di *backpropagation*. Nel primo compito (*prototype sorting*), le reti imparavano a categorizzare le figure: avendo in input le forme, producevano in output il prototipo della categoria. Nella seconda fase (*entry-level naming*), le reti imparavano ad associare ciascuna forma con il corrispondente nome. La parte più interessante della simulazione era il terzo compito (*higher-level*), in cui alle reti venivano presentati i nomi già fondati (ad esempio "quadrato" o "rettangolo") in congiunzione con nuovi nomi arbitrari ("simmetrico", "asimmetrico"). Questo equivale in un certo senso ad insegnare che "cerchio e quadrato [sono figure] simmetriche" e che "ellissi e rettangolo [sono figure] asimmetriche". E' stata così insegnata alle reti una categoria di ordine superiore, associando le nuove stringhe a quelle relative ai simboli fondati in precedenza.

Nella successiva fase di test, sono state presentate alle reti le forme come input retinici e sono state controllate le risposte nell'output linguistico. Le reti sono state in grado nell' 80% dei casi di produrre per ciascun input sia il nome della forma che quello della categoria superiore (simmetrico o asimmetrico). È stato così verificato il *grounding transfer*.



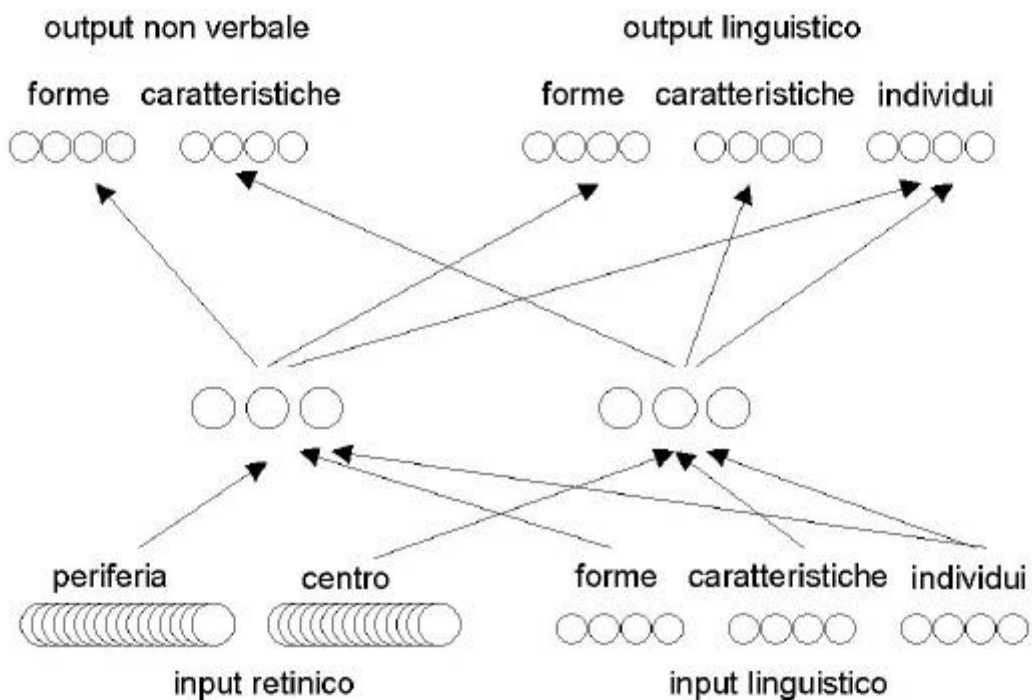
3. Simulazione II

La prima simulazione ha mostrato che è possibile trasferire il fondamento posseduto da simboli elementari che si riferiscono a *individui* (es. quadrato) a categorie di livello superiore (es. simmetrico). Tale modello è stato replicato ed esteso ad una congiunzione *intensionale*, cioè basata sulla combinazione di *caratteristiche*. Abbiamo addestrato dieci reti a categorizzare e denominare immagini di forme animali (es. cavallo, tartaruga) e di loro caratteristiche (es. strisce, puntini). È stata quindi insegnata alle reti una categoria di ordine superiore basata sulla combinazione di caratteristiche per definire un nuovo animale (es. zebra).



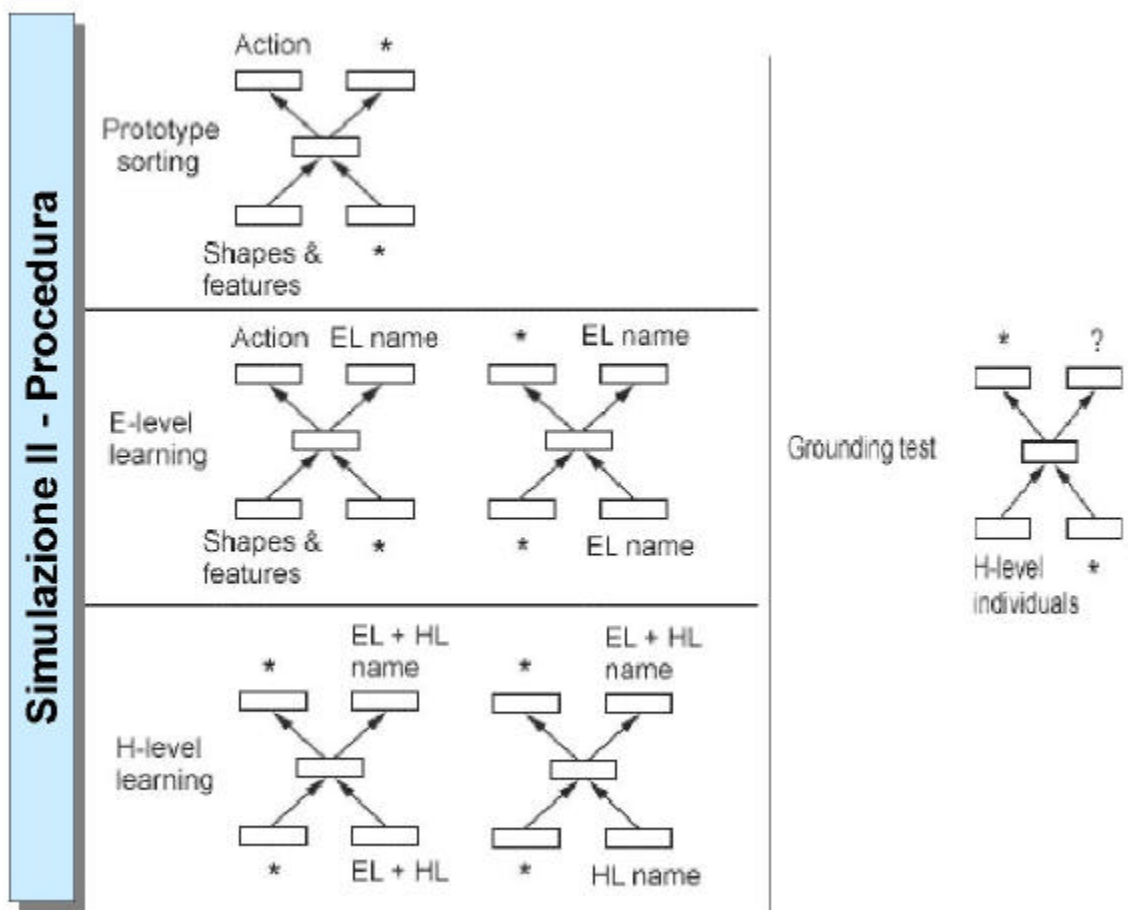
Architettura. Sono state usate 10 reti simili alle precedenti. Lo strato di input era anche in questo caso diviso in un gruppo di unità retiniche e un gruppo di unità linguistiche. A differenza del precedente modello, le unità retiniche sono state suddivise in una parte centrale e una periferica. Lo strato nascosto, costituito da 6 unità, era connesso a quello di input in modo che le prime tre unità nascoste ricevessero input dalla periferia della retina e le altre tre unità dalla parte centrale della retina stessa. In questo modo si consentiva alle unità nascoste di specializzarsi nel riconoscere le forme o le trame. Per l'apprendimento, in questa simulazione, non veniva richiesto in output il prototipo delle categorie, ma una risposta considerata di tipo sensomotorio. Lo strato di output era suddiviso in due gruppi di unità come quello di input: il primo conteneva 8 unità che rappresentavano l'output sensomotorio e il secondo conteneva 12 unità linguistiche strutturate in maniera simile a quelle di input. Lo strato di output era connesso con quello nascosto in modo particolare per ottenere una struttura modulare, in cui tre unità si specializzavano nel rappresentare forme animali e le altre tre nel rappresentare loro caratteristiche.

Simulazione II - Architettura della rete



Procedura. La procedura è stata del tutto simile a quella adottata nella prima simulazione. Dopo le fasi di categorizzazione e di denominazione delle categorie di base, in cui era sempre presente l'input visivo, nella terza fase (apprendimento di alto livello) sono state presentate alle reti soltanto stringhe di simboli, costituite dalla combinazione di nomi di forme animali e di loro caratteristiche (ad es. cavallo, strisce; zebra), che definivano 4 categorie di alto livello.

Nella fase di test sono stati presentati 180 nuovi stimoli visivi, che combinavano insieme forme e *texture* per formare immagini di nuovi animali (es. cavallo+strisce=zebra). Anche se questi stimoli non erano mai stati presentati alle reti, queste sono state in grado di denominarli correttamente, mostrando che anche in questo caso il *grounding* era stato trasferito dai simboli di base a quelli di ordine superiore.



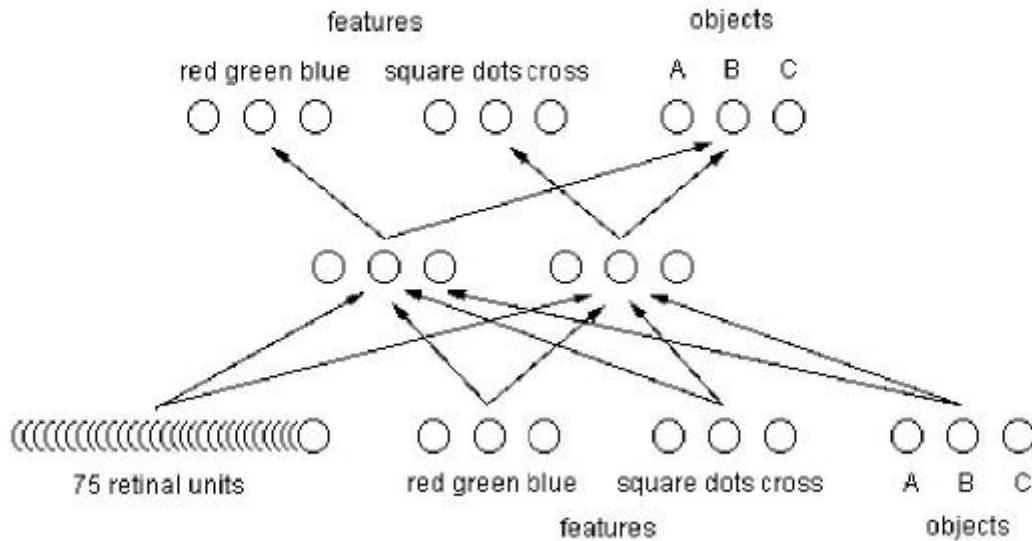
4. Simulazione III

I risultati della precedente simulazione, basata su una architettura in cui le parti periferica e centrale dello stimolo retinico erano separate, potevano dare adito al dubbio che la corretta performance fosse indotta da un artificio. Infatti gli stimoli erano costruiti in modo tale che le forme occupassero sempre la zona periferica della retina e le *texture* in esse contenute si trovassero sempre in quella centrale. Nella terza simulazione sono stati usati stimoli completamente diversi e la retina non è stata più divisa. Un secondo problema, condiviso da entrambe le prime simulazioni, era il fatto che non fosse del tutto chiaro quale output le reti dovessero fornire durante la prima fase dell'apprendimento, quella riguardante le categorie di base. Nella prima simulazione l'output era una sorta di duplicazione della retina e veniva richiesto di fornire il prototipo della categoria che veniva appresa. Nella seconda simulazione, l'output era di tipo localistico e veniva interpretato come una sorta di risposta sensomotoria. La plausibilità di queste interpretazioni non era tuttavia del tutto soddisfacente. La terza simulazione ha ovviato al problema lasciando solo l'output simbolico.



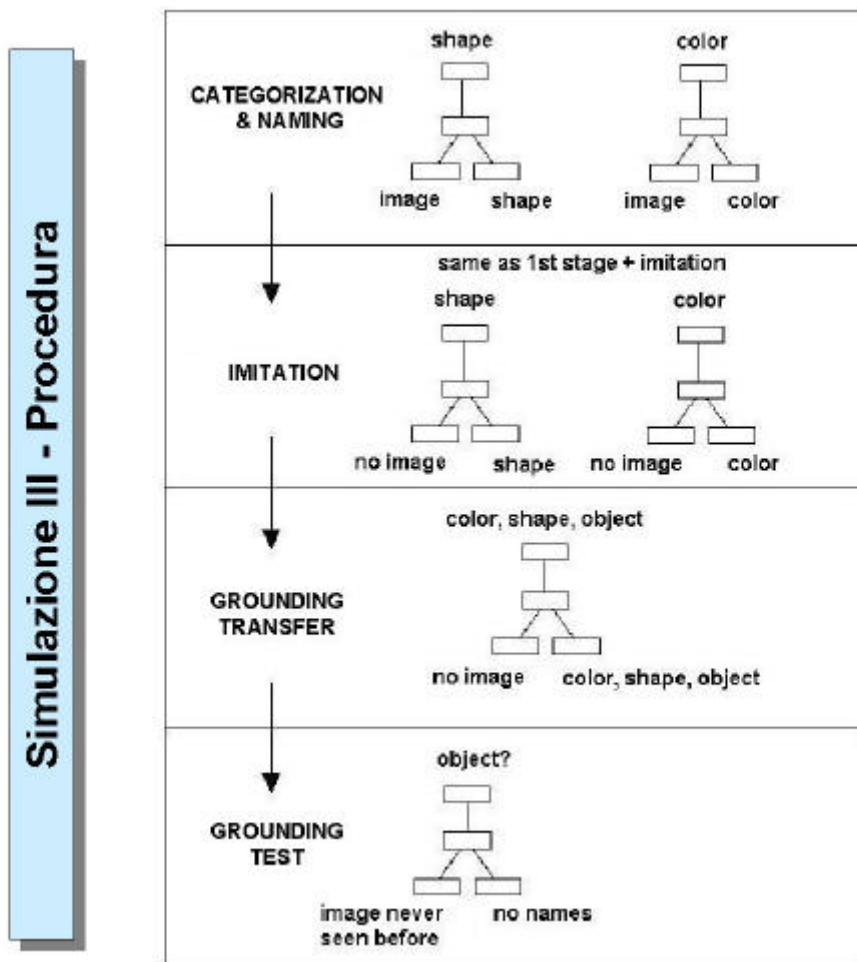
Stimoli. Sono stati usati quali stimoli 3 forme geometriche (quadrato, croce, punti) e 3 colori (rosso, verde, blu). In totale sono stati presentati 81 stimoli (derivanti dalla combinazione delle 3 forme con i 3 colori e ponendo ogni forma in 9 posizioni diverse), di cui 54 presentati nella fase di training e i rimanenti 27 nella fase di test. Ogni pixel dell'immagine occupava tre unità nella retina, una per ogni colore primario.

Simulazione III - Architettura della rete



Architettura. Sono state usate 30 reti feed-forward con tre strati (input, output, intermedio). Per quanto riguarda l'input, la retina conteneva 75 unità e le unità simboliche erano 9. Lo strato nascosto era costituito da 6 unità completamente interconnesse con le unità di input. Lo strato di output comprendeva 3 unità per i colori, 3 per le forme e 3 per gli oggetti composti presentati durante il test. Abbiamo ottenuto una specializzazione delle unità nascoste attraverso la loro connessione selettiva con le unità di output: tre unità erano connesse alle unità di output per il colore e tre alle unità per la forma. Tutte le unità hidden erano invece connesse alle unità per gli oggetti composti.

Procedura. La procedura di apprendimento è stata quella delle precedenti simulazioni, con la differenza che durante la prima fase le reti rispondevano agli stimoli visivi corrispondenti a forme di un certo colore (es. quadrato rosso) direttamente con un output simbolico, lo stesso che poi è stato associato nella seconda fase ai nomi delle forme e dei colori. Nella terza fase le reti hanno imparato a denominare figure non ancora viste attraverso descrizioni simboliche costituite dai nomi fondati nella fase precedente associati a nuove etichette arbitrarie (ad es. croce verde = DAX).



Nella fase di test sono state presentate per la prima volta sulla retina le figure definite solo simbolicamente nella terza fase di training e le reti hanno fornito in output nell' 85,3% dei casi l'etichetta ad esse arbitrariamente associata, mostrando ancora una volta di aver appreso il *grounding transfer*.

Risultati

Le reti categorizzano e denominano correttamente gli stimoli retinici.
(**sensorimotor toil**)

Le reti acquisiscono nuove categorie sulla base di descrizioni linguistiche.
(**symbolic theft learning**)

Le reti denominano correttamente gli input retinici di stimoli di alto livello mai visti prima.
(**grounding transfer**)

5. Discussione

Nel loro complesso, le simulazioni qui presentate hanno la caratteristica di far uso di simboli il cui significato è fondato in esperienze sensomotorie fatte dal sistema stesso, attraverso un apprendimento diretto per prove ed errori, supervisionato dalle conseguenze di una corretta o errata categorizzazione. Si è mostrato come possano essere appresi simboli di ordine superiore, attraverso un meccanismo, definito "furto simbolico" (Cangelosi, Greco & Harnad, 2000), di trasferimento del significato dei simboli di base a categorie più complesse senza più bisogno di un faticoso apprendimento diretto. Ciò è quanto avviene normalmente attraverso l'uso simbolico del linguaggio, al di là degli aspetti di coordinazione di schemi sensomotori che anch'esso possiede.

Certamente queste simulazioni costituiscono un mondo giocattolo; gli aspetti su cui lavorare al fine di avere simulazioni più realistiche sono innumerevoli. Di particolare interesse potrebbe essere l'uso di un meccanismo non supervisionato nella fase di apprendimento delle categorie di base. Inoltre è opportuno studiare in maniera più approfondita le rappresentazioni interne delle reti al fine di evidenziare meglio in che modo esse siano tradotte nel codice simbolico e rese riutilizzabili in contesti di apprendimento diversi.

Conclusioni

Le reti modellano sistemi cognitivi **autonomi**, in cui i collegamenti tra simboli e i loro significati sono diretti e **intrinseci** al sistema, senza bisogno della mediazione di un interprete esterno.

Si può ottenere l'apprendimento di nuove categorie dall'interazione sensomotoria e da descrizioni simboliche in un modello **interamente connessionista**, senza bisogno di altre tecniche di elaborazione.

L' **organizzazione modulare** delle unità nascoste suggerisce che il grounding sensomotorio per diverse chiavi di classificazione dev'essere separato.

Le tre simulazioni forniscono risultati soddisfacenti pur implementando modelli progressivamente più plausibili e compiti più difficili

Confronto fra le tre simulazioni

Prima simulazione:

- **stimoli per il test:** gli stessi usati per l'apprendimento (forme geometriche);
- rete non modulare;
- retina non divisa
- **output** retinico e linguistico

Seconda simulazione:

- **stimoli per il test:** diversi dall'apprendimento (figure animali mai viste);
- rete modulare;
- retina divisa (centro-periferia)
- **output** non verbale e linguistico

Terza simulazione:

- **stimoli per il test:** diversi dall'apprendimento (forme colorate mai viste);
- rete modulare;
- retina non divisa
- **output** solo linguistico

Riferimenti bibliografici

Cangelosi A., Greco A., & Harnad S. (2000). From robotic toil to symbolic theft: Grounding transfer from entry-level to higher-level categories. *Connection Science*, 12(2), 143-162.

Greco A., Cangelosi A. (1999). Language and the acquisition of implicit and explicit knowledge: a pilot study using neural networks. *Cognitive Systems*, 5-2, 148-165.

Harnad S. (1990). The Symbol Grounding Problem. *Physica D*, 42, 335-346

Plunkett K., Sinha C., Møller M.F., & Strandsby O. (1992). Symbol grounding or the emergence of symbols? Vocabulary growth in children and a connectionist net. *Connection Science*, 4, 293-312.