

## Idee e spunti di discussione per il trattamento automatico dei linguaggi artificiali

Maurizio Atzori

## Sommario

- Biologia: cervello e linguaggio
- Reti neurali: un modello diretto di quello biologico
- Basi di conoscenza
- Alcune idee
  - Tipi con proprietà modificabili e pesate
  - Fuzzy Pattern matching
  - Ricerca Breath-First

## Biologia: cervello e linguaggio

- Studio delle basi neurofisiologiche del linguaggio e della memoria: Antonio & Hanna Damasio
- Linguaggio: *capacità di usare parole e di combinarle in frasi in modo che i concetti della nostra mente possano essere trasmessi ad altri*

## Biologia: cervello e linguaggio

- Funzioni del linguaggio:
  - Mezzo di comunicazione
  - Mezzo di compressione cognitiva
    - Serve a categorizzare il mondo (ridurre la complessità ad una scala gestibile)
    - Esempio: “cacciavite” (oppure “democrazia”)
      - Descrizione movimento
      - L’uso a cui è destinato
      - Il gesto della mano, ecc.

## Biologia: cervello e linguaggio

- 3 sistemi neurali:
  - Uno sensitivo
    - è il più vasto (emisfero sx e parte dx)
    - Riconosce colori, suoni, ecc.
    - Classifica sia in base al tipo di stimolo (colore, temperatura, ecc) sia ad un livello più alto
      - Oggetti, eventi, relazioni
    - Stratificazione delle classificazioni

## Biologia: cervello e linguaggio

- Uno più piccolo (emisfero sx)
  - Rappresenta fonemi
  - Regole sintattiche per la combinazione delle parole e la costruzione delle frasi
  - Eseguono l'elaborazione iniziale dei segnali linguistici visivi ed uditivi

## Biologia: cervello e linguaggio

- Sistema neurale di mediazione (SX)
  - Viene stimolato dalle parole, producendo concetti
  - Svolge anche la funzione inversa, ovvero dato un concetto, stimola le zone neurali capace di produrre parole (e suoni)

## Biologia: cervello e linguaggio

- La tripartizione è avvalorata da studi indipendenti (psicolinguistici)
- Le zone del cervello e del cervelletto sono state classificate per “funzione”
  - Hanno nomi V1, V2, V3...
  - La funzione è stabilita in base alle disfunzioni nei soggetti non sani

## Biologia: cervello e linguaggio

- Esempio: *l'esperienza del colore*
  - Partecipano retina e nucleo genicolato laterale
    - A livello di segnale visivo (elaborazione iniziale)
  - Corteccia visiva primaria e zone V2 e V4
    - Lesioni porzioni occipitale giri linguali (Acromatopsia)
      - Visione a toni di grigio
      - Nessun difetto concettuale

## Biologia: cervello e linguaggio

- Lesioni alla corteccia temporale post sx
  - Concetti corretti
  - Perdita completa della capacità di produrre parole morfologicamente corrette
    - Blu -> BU
- Anomia di colore
  - Sbagliano completamente l'associazione tra colore e termine
    - Rosso al posto di blu
  - Concetti corretti e corretto abbinamento dei colori
    - Gettone giallo e banana

## Biologia: cervello e linguaggio

- Componenti di un concetto
  - I concetti sono archiviati nel cervello sotto forma di *registrazioni quiescenti*
  - Quando vengono attivate, ricreano sensazioni e azioni associate all'ente (o categoria di enti) attivante



## Biologia: cervello e linguaggio



- Rappresentazioni visive e tattili (aspetto statico)
  - Forma, colore, materiale (tazza)
  - Profumo, gusto (bevanda)
- Azioni (aspetto dinamico)
  - Traiettorie compiute da mano e braccio per portarla alle labbra

## Reti neurali: un modello diretto

- Le reti neurali sono un modello matematico che approssima in modo diretto le computazioni che avvengono nel cervello
- Negli ultimi due decenni hanno avuto molti sviluppi ed applicazioni pratiche
  - Presenti anche in prodotti commerciali

## Reti neurali: un modello diretto

- Pregi:
  - Capacità di categorizzare e generalizzare
  - Alto parallelismo
  - Logica semplice
  - Adattabilità
- Difetti:
  - Dipendenza dal formato input
  - Alto numero di unità per compiti complessi
  - Impossibilità di prevedere il risultato
  - Difficoltà nel valutare le performance

## Reti neurali: un modello diretto

- Le reti neurali risultano essere una “copia approssimata” del sistema biologico
- I meccanismi neuro-fisici sono conosciuti solo in parte
- Non sono state ancora utilizzate per la comprensione automatica dei linguaggi naturali

## Reti neurali: un modello diretto

- La strada delle reti neurali non è ancora adatta ai nostri scopi ☹
- Anziché un approccio diretto basato sul funzionamento “fisico”, un approccio “astratto”
  - basato sull’ “emulazione del risultato”

## Basi di conoscenza: idee

- L’idea base è di puntare sulla conoscenza, sui concetti strutturati in modo da essere facilmente accessibili e correlati
- Basate sulla struttura delle frasi e sulle regole seguite nel produrre frasi di senso compiuto (grammatiche)

## Basi di conoscenza: idee

- Oggi ampiamente diffuse
  - su domini ristretti
- Attualmente si utilizza un modello della conoscenza che fa uso dei frame
- Reti semantiche

## ELIZA

- Uno dei primi sistemi in grado di “simulare un discorso” con un interlocutore umano
- In realtà “stupido”, con nessuna conoscenza (o quasi) del mondo e del significato delle frasi
- Il goal non era la comprensione, ma il dialogo

## ELIZA

- **Ragazza:** Gli uomini sono tutti uguali.  
**Eliza:** *In che senso*  
**Ragazza:** Trovano sempre il modo di tormentarci.  
**Eliza:** *Può farmi un esempio?*  
**Ragazza:** Beh è il mio ragazzo che mi ha fatto venire qui.  
**Eliza:** *È il suo ragazzo che l'ha fatta venire*  
**Ragazza:** Dice che sono depressa.  
**Eliza:** *Mi dispiace sentire che è depressa.*  
**Ragazza:** È vero. Sono infelice.  
**Eliza:** *Pensa che il fatto di venire qui l'aiuterà a non essere infelice?*

## ELIZA

### ELIZA - a friend you could never have before

Eliza: Hello. I am ELIZA. How can I help you?

[My brother is dangerous

Invia query

### ELIZA - a friend you could never have before

You: My brother is dangerous

Eliza: Who else in your family is dangerous?

[My sister]

Invia query

### ELIZA - a friend you could never have before

You: My sister

Eliza: Who else in your family []?

[nobody]

Invia query

### ELIZA - a friend you could never have before

You: nobody

Eliza: Surely not everyone?

[yes

Invia query

## La conoscenza del mondo

- La conoscenza del mondo è fondamentale
  - Fabio era stato invitato al compleanno di Ugo, ma quando agitò il suo salvadanaio non sentì nessun rumore

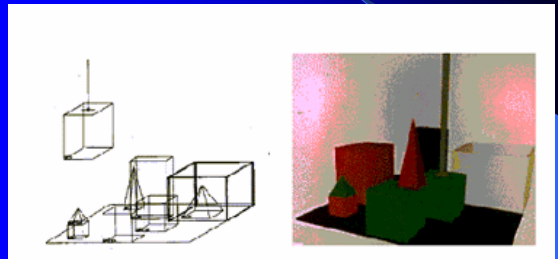
## La conoscenza del mondo

- Dall'idea di rendere completa la conoscenza del mondo, nascono negli anni '70 i micro-mondi
- Tutt'ora un risultato considerevole
  - Un grosso passo in avanti rispetto a ELIZA

## I micromondi e SHRDLU

- SHRDLU è un esempio di comprensione del linguaggio naturale basato su micromondi
- Realizzato da Terry Winograd
- È noto come un mondo di blocchi
  - Tavolo con piramidi e scatole colorate

## I micromondi e SHRDLU



## I micromondi e SHRDLU

**Persona:** how many things are on top of green cubes? [quante cose ci sono sopra i cubi verdi?]  
**SHRDLU:** i'm not sure what you mean by "on top of" in the phrase "on top of green cubes". Do you mean: 1 - directly on the surface 2 - anywhere on top of? [non sono sicuro su cosa intendi con 'sopra' nell'espressione 'sopra i cubi verdi'. Intendi 1- direttamente sulla superficie 2- ovunque sopra ai cubi?]  
**Persona:** 2  
**SHRDLU:** three of them [ci sono tre cose]  
**Persona:** had you touched any pyramid before you put the green one on the little cube? [avevi toccato qualche piramide prima di mettere quella verde sul cubo piccolo?]  
**SHRDLU:** yes, the green one [sì quella verde]  
**Persona:** when did you pick it up? [quando l'hai presa?]  
**Persona:** when did you pick it up? [quando l'hai presa?]

## I micromondi e SHRDLU

– **SHRDLU:** while i was stacking up the red cube, a large red block and a large green cube [mentre stavo mettendo l'uno sopra l'altro il cubo rosso, un blocco rosso grande e un cubo verde grande]  
**Persona:** why? [perché?]  
**SHRDLU:** to get rid of it [per toglierlo di mezzo]  
**Persona:** why did you do that? [perché l'hai fatto?]  
**SHRDLU:** to clean off the red cube [per liberare il cubo rosso]  
**Persona:** why did you clear off that cube? [perché hai liberato il cubo rosso]  
**SHRDLU:** to put it on a large green cube [per metterlo su un cubo verde grande]  
**Persona:** why did you do that? [perché l'hai fatto?]  
**SHRDLU:** to stack up the red cube, a large red block and a large green cube [per mettere uno sopra l'altro il cubo rosso, un blocco rosso grande e un cubo verde grande]  
**Persona:** why did you do that? [perché l'hai fatto?]  
**SHRDLU:** because you asked me to [perché me lo hai chiesto]

## Script Applier Mechanism

- SAM è un software ideato da Schank e da un gruppo di suoi collaboratori
- Si basa essenzialmente sulla sequenza temporale tipica delle azioni che si compiono
- Si tratta di un vero e proprio “copione”, una sceneggiatura.

## Script Applier Mechanism

- La sequenza di azioni tipiche è una forma di rappresentazione della conoscenza del mondo (delle azioni)
- Permette di comprendere azioni anche in assenza di alcune informazioni

*Fabio andò in un ristorante. Ordinò alla cameriera una bistecca ai ferri. Pagò il conto e se ne andò*

- cosa mangiò Fabio?

## Script Applier Mechanism

- Il formalismo di Shank si è rivelato incredibilmente funzionale
- Poteva riassumere in inglese dei testi e rispondere alle domande

## MindPixel: talking to GAC

- Il Digital Mind Project sta sviluppando il GAC
- "Generic Artificial Consciousness" è una personalità artificiale realizzata da MindPixel, in collaborazione con oltre 40 mila utenti della Rete e 200 paesi del mondo.
- La vera novità sta nel reperimento delle informazioni per riempire il database semantico



## MindPixel: talking to GAC

I think the answer to: **summer is hot** is:

FALSE

Respond to the Potential Mindpixels below either True or False and rate their subjective Quality. Respond the way you think most people will respond, even if you think that the majority response is wrong. Do not skip any items. Do not use bots. Read the FAQ.

True	False	Quality	Potential Mindpixel
<input type="radio"/>	<input type="radio"/>	Average	Was Allen Ginsberg a poet?
<input type="radio"/>	<input type="radio"/>	Average	Computers use the binary system of counting
<input type="radio"/>	<input type="radio"/>	Average	Is humanity supposed to survive
<input type="radio"/>	<input type="radio"/>	Average	Do friends sometimes disagree?
<input type="radio"/>	<input type="radio"/>	Average	Do you want others to feel hope?
<input type="radio"/>	<input type="radio"/>	Average	Is show hard?
<input type="radio"/>	<input type="radio"/>	Average	Do you feel proud when you think of your childhood?

## FrameNet

- Progetto che consiste nella creazione di un database di voci del lessico inglese utilizzabile a livello semantico
  - Suddiviso in domini semantici
  - I nomi ed i verbi sono legati in base al significato
  - Ai verbi sono associate le parti interne (soggetto) ed esterne (complementi) del discorso
  - Alle parole viene associato un *tag* che ne specifica il ruolo all'interno della frase (costituisce la fondazione del dizionario)

## FrameNet

### Frame: Communication

#### Lexemes

*communicate.v*

#### Frame Elements (FEs)

FE	Tag	Example (in italics)
Communicator	Com	<i>Pat communicated the message to me.</i>
Addressee	Add	<i>Pat communicated the message to me.</i>
Message	Msg	<i>Pat communicated the message to me.</i>
Topic	Top	<i>Pat communicated with Kim about the festival.</i>
Medium	Medium	<i>Pat communicated with Kim by letter.</i>

## FrameNet

- Di ogni Frame Element viene data anche una specifica descrittiva
- In sostanza, si cerca un matching della frase con quelle nel database:
  - In base al verbo (si trova il giusto Frame)
  - Un secondo matching permette di definire quelli che sono i Frame Element
    - Informazione esterna al verbo (supplementare)

## Modello esteso della conoscenza

- È un modello orientato agli oggetti
- Classe: insieme di oggetti che condividono le stesse proprietà (rappresenta i concetti, non le parole)
- Sono gerarchicamente organizzate tramite il meccanismo di ereditarietà multipla (C++ like), col collegamenti Is-A

## Modello esteso della conoscenza

- Sono descritte in termini di slot (attributi)
- Ciascuno slot può contenere valori o anche insiemi
- Esso è descritto da:
  - Nome
  - Dominio
  - Tipo,
  - Vincoli (detti anche *Facet*)

## Modello esteso della conoscenza

- I facet possono essere di diversi tipi:
- Domain: dominio dei valori
- Defining values: valore dello slot (anche un sottoinsieme di Domain)
- Value descriptor: prototypical *tipico*, inherited *ereditato*, distinguishing *si distingue dagli altri*, value *nessuno dei precedenti*
- Exceptions: dominio delle eccezioni (o dei valori improbabili)
- Modality: 1 *Tutti*, 7 *Nessuno*
  - Es: Classe:Uccelli, Slot:SaperVolare, Facet:Modality=3
- Change frequency: Regular, once only, volatile, never
  - Es: Classe:Persona, Slot:Età, Facet:=Regular (l'età varia regolarmente)
- Event: (E,S,V, R) E=evento, S=stato iniziale (attributo-valore), V=quando l'evento può essere applicato, R=reversibilità

## Modello esteso della conoscenza: Esempio

- C: Circulatory system
  - S: Blood pressure
    - F: Domain: [(0,0)-(300,200)]
    - F: Defining: [(90,60)-(130,85)]
    - F: Value descriptor: prototypical
    - F: Exceptions: [(0,0)-(89,59)] U [(131,86)-(300,200)]
    - F: Modality: 3
    - F: Change frequency: Volatile

## Modello esteso della conoscenza: Esempio

- F: Event: (Age=60,[(0,0)-(89,59)]) U [(131,86)-(300-200)],after,I)
- F: Event: (shock,[(0,0)-(89,59)],after,R)
- F: Event: (haemorrhageI,[(0,0)-(89,59)],after,R)

## Alcune idee

- Tipi con proprietà modificabili e pesate
  - Per l'adattamento e l'apprendimento
- Fuzzy Pattern matching
  - Per la gestione dell'eccezioni e delle categorie
- Sintassi con termini non ordinati
- Ricerca Breath-First
  - Per rendere efficiente il sistema di link
- In sostanza cercare di recuperare i vantaggi dell'adattabilità nei sistemi *frame-based* ©