

NLP - LEZIONE 14

DEL 25/11/2019

SEMANTICA I

⚠️ Persa primi 15 min!

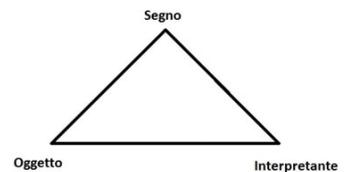
## TRIANGOLO DI PIERCE

Il TRIANGOLO DI PIERCE è uno schema che può essere utilizzato per spiegare la teoria portata avanti da PIERCE (1839-1914), considerato uno dei padri fondatori della SEMIOTICA.

Ogni lato del triangolo di Pierce deriva da

CATEGORIA di cose. In particolare troviamo:

- Il SEGNO, inteso come qualsiasi fenomeno che incontriamo;
- L'OGGETTO, ciò a cui il segno si riferisce.
- L'INTERPRETANTE, ovvero il significato che attribuiamo al segno.



Nel contesto dell'informatica il PROGRAMMA è il segno e l'oggetto è ciò che fa il programma.

Notiamo quindi che esiste una differenza sostanziale tra il SEGNO e il CONCETTO, anche se molto meno peccano l'etica a ricordarci queste differenze, e potremmo scambiare il segno per il concetto e viceversa il segno fa riferimento.

---

Consideriamo la parola MACCHINA. Questo segno può far riferimento ad almeno due insiemi di significati: "MACCHINARIO" e "AUTOMOBILE", dove utilizzando gli amici intendiamo di descrivere il concetto stesso.

Per togliere queste ambiguità semantiche è necessario l'utilizzo di un altro linguaggio per raccontare la semantica. Mai esseri umani abbiano utilizzato la TERMINOLOGIA per organizzare la semantica.

Per descrivere la SEMANTICA utilizzeremo un altro linguaggio, che è, naturalmente, PRIVO DI AMBIGUITÀ.

## PRINCIPIO DI COMPOSIZIONALITÀ ~ (09:00 min)

Il principio di composizionalità ci dice che il significato di una ESPRESSIONE LINGUISTICA è determinato dal significato delle ESPRESSIONI COSTITUENTI del linguaggio e dalle modalità in cui queste sono state COMBinate.

In altre parole l'idea è quella di utilizzare il SIGNIFICATO DELLE PARTI per capire il SIGNIFICATO DEL TUTTO.

A seconda di come COMBINIAMO le cose possiamo avere varie tipologie di COMPOSIZIONALITÀ.

Formalmente, dati due segni  $w_1$  e  $w_2$ , e date due interpretazioni  $f(w_1)$  e  $f(w_2)$  l'interpretazione del segno  $w_1 w_2$  è data da

$$f(w_1, w_2) = F(f(w_1), f(w_2))$$

dove  $F(\cdot)$  è la funzione che combina le interpretazioni di  $w_1$  e  $w_2$ .

OSS: Le TAVOLE DI VERITÀ codificano in quale modo i valori di verità delle VARIABILI vengono COMBINATI per ottenere i valori di verità di ESPRESSIONI BOOLEANE.

La COMPOSIZIONALITÀ CONCATENATIVA consiste nel GIUSTAPPORRE e interpretazione dei vari segni tra loro.

Nel contesto del NLP il principio di componizionalità è utile in quanto ci permette di tirar fuori il significato della frase basandoci sul significato delle parole. Senza componizionalità dovremmo assegnare un significato ad ogni singola parola.

---

La componizionalità risulta problematica quando un aggregato di parole comincia ad assumere un concetto diverso da quello ottenuto prendendo e combinando i significati delle singole parole.

**ESEMPIO:** Le parole HARD DISK originariamente venivano utilizzate perché il FLOPPY DISK era MOSCIO. Col passare del tempo è diventata sinonimo di "dispositivo di memoria esterna", PERDENDO la proprietà di componizionalità.

Un altro esempio problematico è il seguente:

MELA ROSSA → Qui cambia il colore  
ESTERNO del frutto.

ARANCIA ROSSA → Qui cambia il colore  
INTERNO del frutto.

---

## DESCRIVERE LA SEMANTICA ~ (33:00 min)

Vogliamo eliminare le ambiguità della semantica utilizzando un linguaggio di arrivo per descriverla.

A ricordo della specifico TASK da risolvere quindi saremo interessati a catturare vari aspetti della semantica. In generale più "sintetico" e, e più dovremo lavorare con la componizionalità.

Ci dobbiamo quindi chiedere

- i) Quali sono gli ELEMENTI chiave da CATTURARE?
- ii) Come possiamo combinare gli elementi?

Smizziamo con i). Possiamo avere i seguenti elementi

PAROLE ITALIANO	PAROLE CONCETTUALI
RAGGIUNGERE	RAGG
IO	EGO
MILANO	"SEGNAPOSTO SU MAPPA"
ARRIVARE	ARR

La frase "io arrivo a Milano" può quindi essere tradotta come ARR (EGO, I (MILANO)).

La parole IRRAGGIUNGIBILE invece significa l'immobilità di raggiungere qualcosa. Per sentire queste parole l'idea può essere quella di dare del significato ad un # di limitate di elementi, come IR e RAGGIUNGERE per poi combinarli per poter interpretare varie frasi.

# SCRIPT THEORY

~ (52:00 min)

SHANK nel 1975 ha sviluppato la SCRIPT THEORY per cercare quali non gli ELEMENTI PRIMITIVI necessari per dare senso ai verbi.

Gli ELEMENTI PRIMITIVI trovati da SHANK erano:

- ATRANS, Transfer in contact object
- SPEAK, Utter a sound
- INGEST, Actor ingesting an object
- ...

Il tentativo di SHANK non sortì i grandi risultati in quanto

- i) C'è una PERDITA DI INFORMAZIONI nell'utilizzo di queste primitive per catturare il significato dei verbi.
- ii) Le primitive scelte non ARBITRARIE, o comunque è difficile capire se le primitive scelte sono corrette oppure no.

OSS: L'esistenza di queste PRIMITIVE che danno significato alle parole è stato in parte dimostrato. In particolare si è visto che non primitive non contenute in un frase e non è facile RICORDARE la frase.

---

SE Plus che stiamo analizzando è quello Chomskiano, ed è il seguente:

$P \longrightarrow T = \text{PARSE\_TREE}(P) \longrightarrow \text{SEMANTIC}(T)$

Un altro approccio che non viene utilizzare consiste nell'assegnare ad ogni parola una CLASSE o recorde del LESSICO della parola stessa.

OSS: Nel modello E/R le RELAZIONI vengono rappresentate dai VERBI, mentre le ENTITÀ vengono rapp. dai NOMI.

Consideriamo la frase

MARIO INCONTRA UN AMICO A MILANO

non viene procedere nel seguente modo

INCONTRA  $\rightarrow$  INCONTRARE (·, ·, ·)

MILANO  $\rightarrow$  SEGNAPOSTO  $\downarrow$  MILANO

MARIO  $\rightarrow$   $\begin{matrix} P \\ \downarrow \\ M \end{matrix}$

Motivato che è difficile tradurre "UN AMICO", in quanto non sappiamo se rappresentarlo come CONCETTO generale oppure come ISTANZA specifica.

(!) NON MOLTO CHIARO

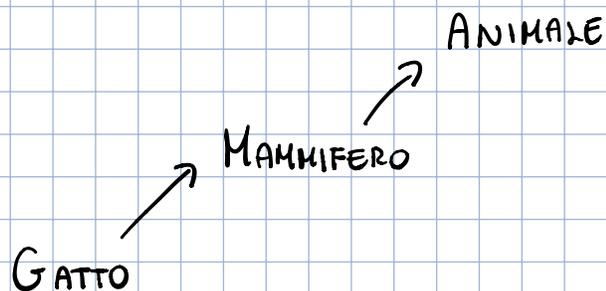
La frase viene quindi interpretata come segue

INCONTRARE (UN AMICO, MILANO)

La domanda odora diventa la seguente: quali e quanti sono le CLASSI (i CONCETTI) che utilizzo per derivare il significato delle frasi? Da dove le loro parole?

L'idea è quella di utilizzare delle strutture dati simili a DIZIONARI.

Quando cerchiamo una parola in un dizionario troviamo la SUPER-CLASSE a cui appartiene la parola insieme alla SPECIFICAZIONE associata alla parola.



Il PRIMO dizionario DIGITALE fu chiamato WORDNET.

## WORD NET ~ (1:30:00 min)

Wordnet funziona molto bene per i NOMI e gli AGGETTIVI, e si trova nella ricerca degli IPERONIMI e degli IPONIMI.

CANE è un IPONIMO di ANIMALE.

MOBILE è un IPERONIMO di SEDIE.

Inizialmente si è tentato di utilizzare la struttura LESSICOGRAFICA contenuta da WORDNET per IMPORCI la relazione LOGICA IS-A nota nel paradigma OBJECT-ORIENTED. Motivando però che questo tentativo di ONTOLOGIZZAZIONE non è stato seminale.

In WORDNET grandi cerchi sono anche le versioni presentate da serie di SYNSETS. Un SYNSET contiene le parole che secondo WORDNET hanno lo stesso significato.

Il SENSO in WORDNET non viene quindi descritto in modo ESPlicito, ma viene descritto in modo IMPlicito tramite i SYNSETS.

La stessa PAROLA può essere presente in diversi SYNSETS, a seconda dei vari significati associabili alla PAROLA.

Ciascun SYNSET ha un particolare ID.

OSS: Inizialmente è ID associato al SYNSET  
e è OFFSET all'interno di un FILE GIGANTE  
in cui sono memorizzati i vari SYNSETS.  
Ogni volta che si modifica il dizionario,  
l'ID cambia.

Ad un certo punto l'associazione

SYNSET ↔ ID

è stata resa PERMANENTE.

Notiamo che in alcuni casi notiamo avere una  
eccessiva GRANULARITÀ delle parole che porta alla  
creazione di SYNSETS troppo specifici.

WORD - SENSE DISAMBIGUATION: Task che prende una PAROLA  
e utilizza uno dei synsets presenti in WORDNET per  
interpretare la parola.