

INTELLIGENZA ARTIFICIALE

**A N N O A C C A D E M I C O
2 0 0 8 - 2 0 0 9**

**Introduzione al
Natural Language Processing**

SOMMARIO

- Introduzione al NLP
- Analisi Sintattica, Semantica, Pragmatica
- Applicazioni del NLP
- Statistical NLP

INTRODUZIONE AL NATURAL LANGUAGE PROCESSING

- ✿ Al di fuori e prima dell'IA, il linguaggio umano è stato studiato da specialisti di svariate discipline: linguistica, filosofia, logica, psicologia.
- ✿ Quindi non è possibile tracciare un panorama dei diversi punti di vista, dei metodi e dei risultati di questi studi.
- ✿ Ci limiteremo ad introdurre brevemente alcuni concetti, indispensabili per capire il funzionamento dei sistemi per l'elaborazione del linguaggio.

INTRODUZIONE AL NATURAL LANGUAGE PROCESSING

- ✿ Nell'elaborazione del linguaggio naturale è utile distinguere tre diversi livelli:
 - ✿ Livello Sintattico
 - ✿ Livello Semantico
 - ✿ Livello Pragmatico

IL LIVELLO SINTATTICO

- ✿ Il **livello sintattico** ha a che vedere con la grammatica di una lingua specifica. Ad esempio, il fatto che la frase italiana:

“Emanuela è prima di cordata” (1)

sia corretta, mentre la frase:

“Emanuela sono prima cordata di” (2)

non lo sia, riguarda la grammatica dell'italiano.

IL LIVELLO SINTATTICO

- ✱ Si noti che una frase può essere scorretta e restare perfettamente comprensibile, sia pure con un po' di sforzo, come la (2).
- ✱ D'altra parte, una frase del tutto priva di senso e incomprensibile, come:

“Il tavolo travalica la costituzionalità”

può essere corretta dal punto di vista grammaticale.

IL LIVELLO SEMANTICO

- ✿ Il **livello semantico** concerne il significato delle parole e delle frasi. Ad esempio, la frase (1) ha lo stesso significato della frase:

“Emanuela tira la via”

(comprensibile peraltro ai soli cultori della roccia) pur essendo costituita da parole diverse.

- ✿ Ciò che le due frasi hanno in comune è l'idea espressa, o la situazione descritta.

IL LIVELLO PRAGMATICO

- ✱ Quando due esseri umani si parlano, scambiano fra loro messaggi in una lingua di comune conoscenza.
- ✱ La comunicazione, però, va al di là del semplice significato delle frasi. Ad esempio, la frase:

“ho freddo”

può di volta in volta costituire una semplice affermazione, o una richiesta implicita che qualcuno chiuda la finestra, o altro ancora.

IL LIVELLO PRAGMATICO

- ✿ Per capire compiutamente una frase non è quindi sufficiente tener conto del significato delle parole.
- ✿ Occorre anche considerare la situazione in cui la frase viene pronunciata, e la relazione fra chi parla e chi ascolta.
- ✿ Questi problemi investono il **livello pragmatico** del linguaggio.

PROBLEMI DELL'ELABORAZIONE AUTOMATICA (ANALISI SINTATTICA)

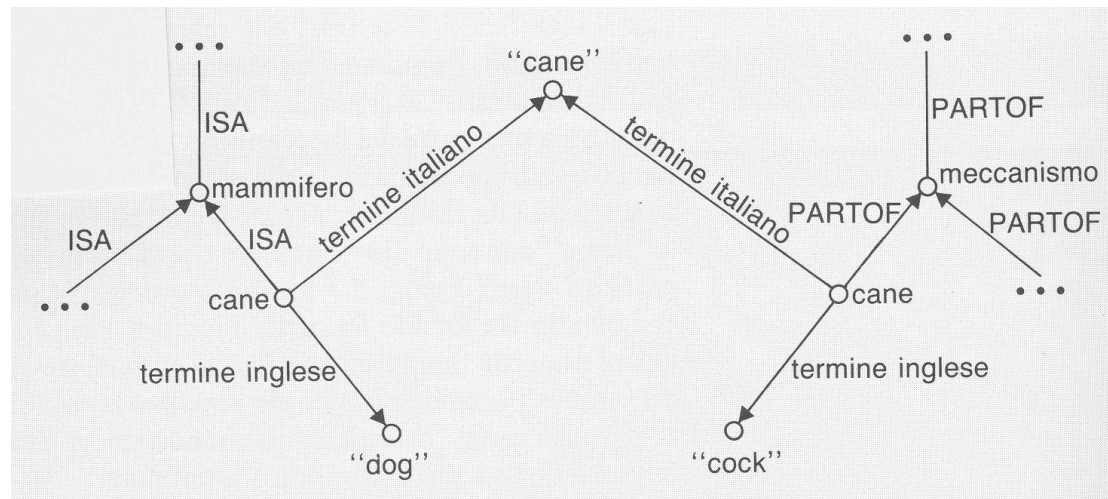
- ✻ Dal punto di vista dell'elaborazione automatica, il livello sintattico è il meno problematico.
- ✻ E' possibile analizzare frasi anche assai complesse, riconoscere la loro correttezza grammaticale, e ricostruire la loro struttura sintattica.

PROBLEMI DELL'ELABORAZIONE AUTOMATICA (ANALISI SEMANTICA)

- ✿ Molto più arduo si presenta invece il livello semantico.
- ✿ Nei casi più semplici si può procedere così:
 - ✿ il significato di una frase si ottiene dal significato delle singole parole, una volta che sia nota la struttura sintattica della frase;
 - ✿ il significato delle singole parole è definito in un **dizionario** accessibile automaticamente.

PROBLEMI DELL'ELABORAZIONE AUTOMATICA (ANALISI SEMANTICA)

- ✿ Ci sono però diversi problemi.
- ✿ Innanzitutto, la stessa parola può avere diversi significati in contesti diversi (vedi figura).



PROBLEMI DELL'ELABORAZIONE AUTOMATICA (ANALISI SEMANTICA)

- ✿ Poi, la struttura sintattica di una frase può essere **ambigua**.
- ✿ In altri termini, alla stessa successione di parole possono a volte essere attribuite diverse strutture.
- ✿ Infine, il significato di certe frasi idiomatiche è diverso dal significato letterale che si ottiene componendo i significati delle singole parole.
- ✿ Si pensi ad esempio ad espressioni come “fare l'indiano” o “gettare la spugna”.
- ✿ L'IA tenta di risolvere questi problemi utilizzando un'opportuna base di conoscenza a cui il programma di elaborazione del linguaggio è in grado di accedere.

PROBLEMI DELL'ELABORAZIONE AUTOMATICA (ANALISI PRAGMATICA)

- ✿ Più difficile ancora si presenta il **livello pragmatico**.
- ✿ Il problema principale è dovuto al fatto che quando discorriamo siamo in grado di rappresentarci certi stati mentali dei nostri interlocutori.
- ✿ Ad esempio, alla domanda: “sai che ora è?” non rispondiamo “sì” ma, poniamo, “sono le dieci e un quarto”; e questo perché riusciamo a capire che chi parla desidera conoscere l’ora, e non informarsi se possediamo un orologio.
- ✿ In altre parole, per dialogare correttamente è necessario rappresentarsi le intenzioni degli interlocutori, che sono solo parzialmente rispecchiate nelle loro parole.

ANALISI SINTATTICA (PRINCIPI GENERALI UTILIZZATI)

- ✿ Come già accennato, lo scopo dell'analisi sintattica di una frase è quello di riconoscerne la struttura sintattica.
- ✿ Al riguardo esistono diversi modelli teorici che definiscono il concetto di struttura in modo differente.
- ✿ Negli anni Cinquanta il linguista Noam Chomsky ha proposto il concetto di **grammatica a struttura di frase** (**phrase-structure grammar**).
- ✿ Una grammatica di questo tipo è costituita da un insieme di regole di produzione che consentono di associare ad ogni frase corretta del linguaggio una struttura ad albero, detta **albero sintattico**, che rappresenta i rapporti sintattici esistenti fra le varie parti della frase.

ANALISI SINTATTICA (PRINCIPI GENERALI UTILIZZATI)

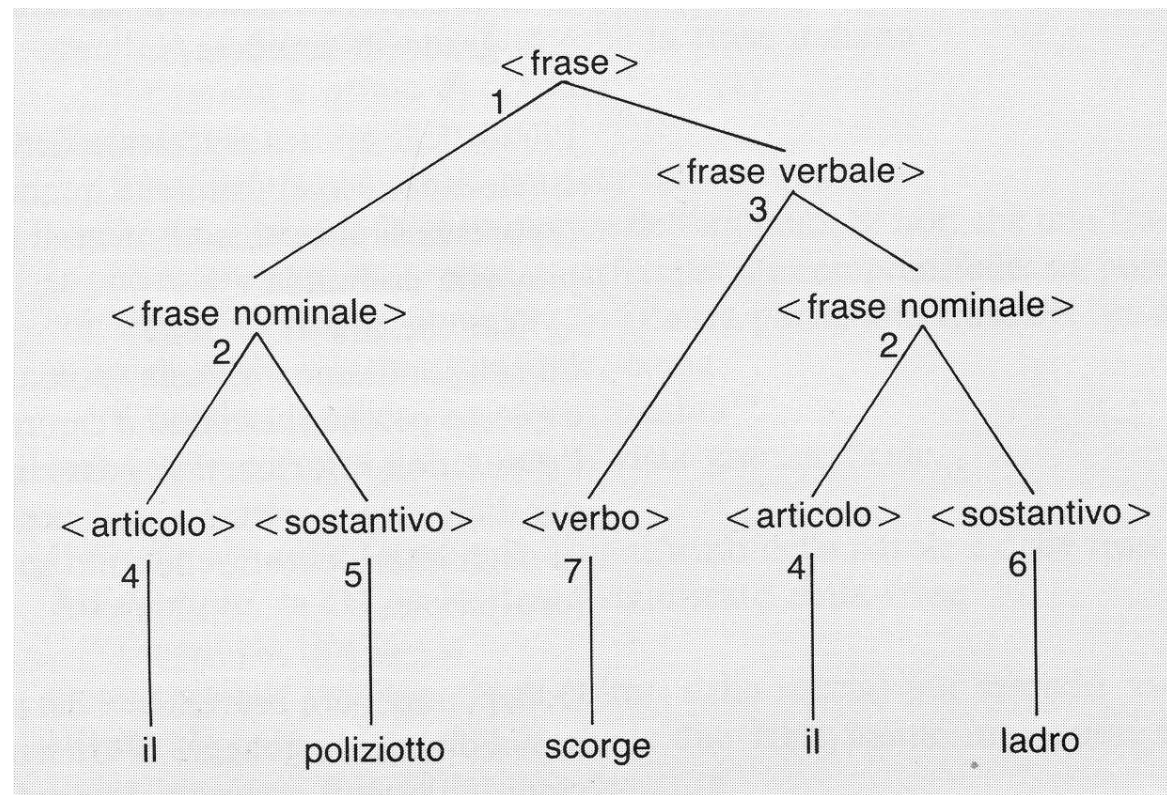
Una Grammatica Chomskiana:

1. <frase> → <frase nominale> <frase verbale>
2. <frase nominale> → <articolo> <sostantivo>
3. <frase verbale> → <verbo> <frase nominale>
4. <articolo> → il
5. <sostantivo> → poliziotto
6. <sostantivo> → ladro
7. <verbo> → scorge

- I simboli racchiusi fra < > rappresentano **categorie sintattiche**, ovvero componenti della frase.
- Gli altri simboli sono simboli terminali, e rappresentano le parole della frase.
- Una regola del tipo: <x> → <y> <z> asserisce che la categoria sintattica <x> è costituita dalla semplice successione delle categorie <y> e <z> nella frase.
- Una regola del tipo <x> → y asserisce che la parola y corrisponde alla categoria sintattica <x>.

ANALISI SINTATTICA (PRINCIPI GENERALI UTILIZZATI)

Mediante le 7 regole precedenti si può attribuire una **struttura sintattica** ad una frase:



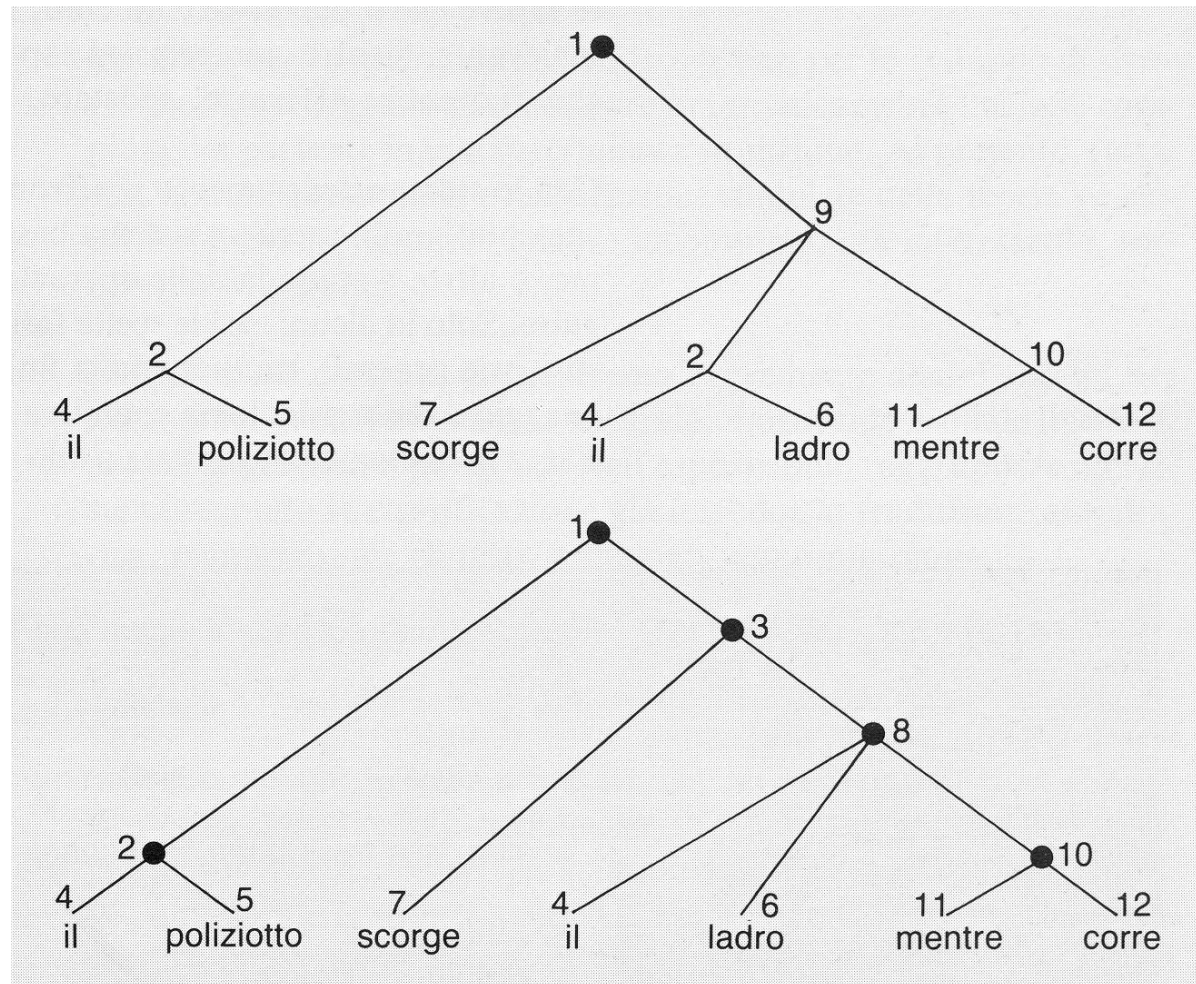
ANALISI SINTATTICA (PRINCIPI GENERALI UTILIZZATI)

Aggiungiamo le regole seguenti alla grammatica vista prima:

8. <frase nominale> → <articolo> <sostantivo> <frase attributiva>
9. <frase verbale> → <verbo> <frase nominale> <frase attributiva>
10. <frase attributiva> → <avverbio> <verbo>
11. <avverbio> → mentre
12. <verbo> → corre

Nella figura che segue sono descritte due analisi possibili della stessa frase, la cui struttura sintattica risulta quindi ambigua.

ANALISI SINTATTICA (PRINCIPI GENERALI UTILIZZATI)



ANALISI SINTATTICA (PRINCIPI GENERALI UTILIZZATI)

- ✿ Oltre alle grammatiche chomskiane, sono stati proposti altri modi di rappresentare la struttura delle frasi.
- ✿ Accenniamo qui alle grammatiche a casi di Fillmore (Fillmore, C. "The case for case". In: Bach E., Harms R. (a cura di) *Universals in linguistic theory*, Holt, Reinhart and Winston, New York, 1968.).
- ✿ Secondo questo approccio l'analisi di una frase richiede l'identificazione di un componente centrale (il verbo) e di un certo numero di parti della frase che si pongono con il verbo in certi rapporti, detti casi.
- ✿ Tra i possibili casi elencati da Fillmore ci sono l'agente, l'oggetto, lo strumento, il risultato, la provenienza, la destinazione.

ANALISI SINTATTICA (PRINCIPI GENERALI UTILIZZATI)

esempio di una analisi a casi:

Frase: “Bruno accende il sigaro con un fiammifero”

Al verbo “accendere” è associato uno schema di casi:

accendere → [agente oggetto strumento]

L’analisi della frase porta al seguente risultato:

accendere

agente:	Bruno
oggetto:	sigaro
strumento:	fiammifero

ANALISI SINTATTICA (PRINCIPI GENERALI UTILIZZATI)

- ✿ Un problema dell'analisi sintattica è che spesso è ambigua, ovvero porta ad assegnare alla stessa frase diverse strutture possibili.
- ✿ In genere, l'ambiguità può essere risolta soltanto tenendo conto del significato delle parole e del contesto in cui si trova la frase.
- ✿ Per questo motivo l'IA tende oggi a sviluppare analizzatori semantici che contengono tra l'altro una limitata componente di analisi sintattica, piuttosto che affrontare i due tipi di analisi separatamente.

ANALISI SEMANTICA (PRINCIPI GENERALI UTILIZZATI)

- ✿ Il primo problema che si incontra nella costruzione di analizzatori semantici è la definizione di che cosa si intenda precisamente per **significato**.
- ✿ Ad esempio, potremmo a seconda dei casi considerare come significato della parola “peperone”:
 - uno specifico peperone (come in “il peperone preparato da Caterina”);
 - l’insieme di tutti i peperoni esistenti (come in “il peperone è un ortaggio molto saporito”);
 - il concetto di peperone (descritto, poniamo, da una rete semantica).

ANALISI SEMANTICA (PRINCIPI GENERALI UTILIZZATI)

- ✱ In molte applicazioni dell'Intelligenza Artificiale si sceglie come significato di un termine un concetto, descritto secondo un formalismo di rappresentazione adeguato.
- ✱ Ma questo non risolve il problema!
- ✱ Quali caratteristiche dobbiamo scegliere per rappresentare il concetto di peperone? La forma, il colore, il sapore? Oppure la sua funzione di cibo? E quando abbiamo scelto le caratteristiche rilevanti, come possiamo rappresentarle nel calcolatore?

ANALISI SEMANTICA (PRINCIPI GENERALI UTILIZZATI)

- ✱ Alle domande precedenti l'IA tende a fornire risposte motivate dalle esigenze applicative.
- ✱ Le caratteristiche di un oggetto che consideriamo rilevanti per la sua rappresentazione vengono così a dipendere dagli scopi per cui il programma è sviluppato:
 - In certi casi sarà importante sapere che il peperone è un cibo.
 - In altri casi, questo dato può essere irrilevante, mentre ci interesserà sapere che il peperone è un oggetto che può essere afferrato, spostato, tagliato.

ANALISI SEMANTICA (PRINCIPI GENERALI UTILIZZATI)

- ✿ Con lo stesso atteggiamento si affronta la scelta del formalismo rappresentativo.
- ✿ Si userà un sistema basato su reti semantiche, o su frames, o altro, a seconda delle operazioni che il programma deve eseguire.
- ✿ Solo in alcuni casi le scelte fatte dai ricercatori di IA si basano su teorie generali, mutate dalla linguistica, dalla filosofia del linguaggio o dalla psicologia cognitiva.
- ✿ Più spesso un sistema fa caso a sé, e la sua validità è limitata a uno specifico campo di applicazione.

ANALISI PRAGMATICA (PRINCIPI GENERALI UTILIZZATI)

- ✿ A tutt'oggi, i tentativi di inserire un livello di **analisi pragmatica** nei sistemi di elaborazione automatica del linguaggio non sono numerosi, e spesso hanno un valore puramente esplorativo.
- ✿ Alcuni di questi tentativi si basano su un'interessante teoria sviluppata dai filosofi del linguaggio: la teoria degli **atti linguistici** (*speech acts*).
- ✿ Secondo questa teoria, per comprendere una frase è necessario capire che tipo di azione compie chi parla nell'emettere il suo messaggio. Azioni possibili sono l'affermare, il richiedere, il domandare, il promettere, ecc.

ANALISI PRAGMATICA (PRINCIPI GENERALI UTILIZZATI)

- ✿ Riconsideriamo la frase:

“Sai che ora è?”

- ✿ Letteralmente, si tratta di una domanda sullo stato delle conoscenze dell'ascoltatore.
- ✿ Questo però è in grado di capire che chi parla sta in realtà compiendo una richiesta: appunto, la richiesta che venga comunicata l'ora.
- ✿ Per arrivare a questa conclusione, l'ascoltatore si basa sulla sua conoscenza della situazione e sulla valutazione dei probabili scopi del parlante.

ANALISI PRAGMATICA (PRINCIPI GENERALI UTILIZZATI)

- ✿ Scartata come implausibile l'eventualità che lo scopo del parlante sia di scoprire se l'ascoltatore possiede un orologio (ma è facile immaginare una situazione in cui la frase in questione venga pronunciata proprio a questo scopo), è possibile ricostruire lo scopo effettivo (cioè: conoscere l'ora) seguendo i passi logici descritti da Searle (Searle J.R., *Expression and meaning*. Cambridge University Press, Cambridge, 1979).
- ✿ Questo processo di ricostruzione richiede fra l'altro che l'ascoltatore faccia alcune ipotesi su ciò che è plausibile attendersi da chi parla in determinate situazioni.
- ✿ In altre parole, l'ascoltatore deve farsi una rappresentazione di certi **stati mentali** del parlante, fra cui i suoi scopi. Naturalmente, se la rappresentazione è scorretta la frase verrà male interpretata.

APPLICAZIONI: LA TRADUZIONE AUTOMATICA

- ✿ Il problema di tradurre automaticamente un testo da una lingua all'altra è stato affrontato ancora prima della nascita dell'IA.
- ✿ Negli anni Cinquanta diversi ricercatori, tra cui l'italiano Silvio Ceccato, si sono occupati della questione, senza poter raggiungere risultati significativi.
- ✿ Negli anni Settanta il problema è stato affrontato di nuovo utilizzando le tecniche di Intelligenza Artificiale.
- ✿ Tuttavia, anche se sono stati raggiunti vari risultati apprezzabili, si può dire che l'obiettivo di tradurre automaticamente testi complessi è ancora lontano.

APPLICAZIONI: LA TRADUZIONE AUTOMATICA

- ✿ Uno dei sistemi di traduzione più interessanti è stato realizzato da Wilks a partire dal 1973.
- ✿ Si tratta di un sistema che produce una buona traduzione francese di brevi periodi in lingua inglese.
- ✿ Questo, nelle sue linee essenziali, il funzionamento:
 - Il testo inglese in ingresso viene convertito, sulla base delle conoscenze contenute in un dizionario, in una forma interna che utilizza un numero limitato di **termini primitivi**. I termini primitivi, che in tutto sono un'ottantina, costituiscono per Wilks una specie di vocabolario essenziale, che consente di definire ogni altra parola con sufficiente precisione. Queste definizioni sono contenute in un dizionario.

APPLICAZIONI: LA TRADUZIONE AUTOMATICA

- Quando una parola è ambigua, cioè può essere ridotta a diversi termini primitivi in diversi contesti, il sistema sceglie la riduzione più adeguata sulla base di una **semantica preferenziale** (preference semantics). Per fare un esempio in italiano, nella frase “**il cane corre**” si sceglie di ridurre la parola “**cane**” ai termini che rappresentano l’animale e non la parte di un’arma da fuoco, sulla base del fatto che il verbo “**correre**” regge preferenzialmente un soggetto animato.
- La forma interna della frase inglese viene convertita, sempre sulla base del dizionario, in una frase francese.

APPLICAZIONI: LA TRADUZIONE AUTOMATICA

- ✿ Il sistema di Wilks è in grado di utilizzare anche conoscenze linguistiche non contenute nel dizionario, come le regole per risolvere i riferimenti pronominali.
- ✿ Un aspetto interessante, che ritroveremo in altri sistemi, è che l'aspetto sintattico del linguaggio è ridotto a giocare un ruolo marginale.
- ✿ Il sistema si basa essenzialmente su un'analisi semantica diretta della frase in ingresso, giacché i termini primitivi contenuti nel dizionario forniscono una rappresentazione del significato delle parole.

APPLICAZIONI: LA TRADUZIONE AUTOMATICA

- ✱ In realtà è difficile considerare l'approccio di Wilks come una teoria del significato dotata di una sufficiente generalità.
- ✱ In particolare, la scelta dei termini primitivi è completamente empirica, e finalizzata unicamente alla traduzione.
- ✱ Non è chiaro inoltre se lo stesso insieme di termini primitivi sarebbe adeguato alla traduzione fra lingue diverse da quelle scelte da Wilks, come ad esempio il russo e il tedesco; oppure, invertendo la direzione, dal francese all'inglese.

APPLICAZIONI: LA COMPrensIONE DEI TESTI

- ✱ Comprendere un testo richiede, come minimo, di comprendere le singole frasi che lo compongono. In genere, però, ciò non è sufficiente!
- ✱ Consideriamo ad esempio una “storia” brevissima, costituita da due semplici frasi:

“Giuliano ha sete” (1)

“Giuliano decide di entrare in un bar, ma si accorge di non avere denaro con sé” (2)

APPLICAZIONI: LA COMPrensIONE DEI TESTI

- ✿ Prese isolatamente, le due frasi hanno un significato facile da rappresentare.
- ✿ La frase (1) esprime un bisogno fisiologico di Giuliano.
- ✿ La frase (2) esprime due eventi che concernono lo stato mentale di Giuliano (**decidere** e **accorgersi**).

APPLICAZIONI: LA COMPrensIONE DEI TESTI

- ✱ Dal testo, tuttavia, un lettore umano capisce molto di più:
 - Per prima cosa, si può collegare la sete con la decisione di entrare in un bar, perché si sa che nei bar c'è da bere, che bevendo passa la sete, che la sete è uno stato spiacevole e che pertanto si cerca di eliminarla.
 - In secondo luogo, possiamo collegare le due parti della (2), perché sappiamo che nei bar ci vuole denaro per ottenere una bevanda.
- ✱ Solo dopo aver compiuto questi collegamenti possiamo dire di aver capito veramente il testo.

APPLICAZIONI: LA COMPrensIONE DEI TESTI

- ✿ Se riflettiamo un attimo, vediamo che la conoscenza utilizzata per porre i collegamenti è di due tipi:
 - conoscenza relativa a situazioni stereotipate che si verificano usualmente: nei bar si beve e si paga la consumazione;
 - conoscenza relativa agli obiettivi ed ai piani di comportamento dei personaggi della storia: la sete genera l'obiettivo di bere, cui segue un piano per raggiungere l'obiettivo.

APPLICAZIONI: LA COMPrensIONE DEI TESTI

- ✿ Non è pensabile che le conoscenze di questo tipo facciano parte di un dizionario del linguaggio, dove sono semplicemente definiti i significati delle parole: il fatto che bevendo passi la sete fa parte più della nostra cultura spicciola che delle nostre conoscenze linguistiche.
- ✿ Possiamo dire che per comprendere un testo non basta utilizzare un dizionario, ma occorre un'enciclopedia, contenente le conoscenze necessarie.
- ✿ Il processo di comprensione deve essere in grado, partendo dalle frasi, di utilizzare l'enciclopedia per costruire le connessioni dirette fra i significati delle parole.

APPLICAZIONI: LA COMPrensIONE DEI TESTI

- ✿ Ci si potrebbe chiedere come mai i testi siano così lacunosi, e obblighino il lettore ad uno sforzo deduttivo per integrare le informazioni per così dire sottintese.
- ✿ Ma questa è una caratteristica costante del linguaggio: le informazioni ovvie, che ogni lettore è in grado di ricostruire da sé, vengono di solito taciute.
- ✿ Ciò corrisponde ad un principio di economia: se ogni dettaglio andasse specificato, i testi raggiungerebbero inutilmente dimensioni impraticabili.
- ✿ Naturalmente, ciò costituisce una notevole difficoltà per la comprensione automatica perché ci obbliga a fornire al calcolatore una conoscenza enciclopedica, per lo meno all'interno del dominio applicativo di nostro interesse.

APPLICAZIONI: L'INTERAZIONE UOMO-MACCHINA

- ✿ Nell'interazione uomo-macchina in linguaggio naturale, l'utente sottopone alla macchina frasi che esprimono solitamente **richieste**.
- ✿ Ad esempio, una frase può esprimere un comando diretto ad un robot industriale, o una domanda intorno a dati contenuti in un archivio.
- ✿ Il calcolatore deve essere in grado di comprendere la frase e di soddisfare la richiesta dell'utente.

APPLICAZIONI: L'INTERAZIONE UOMO-MACCHINA

- ✿ In genere, l'interpretazione della frase e l'esecuzione della prestazione richiesta sono due processi distinti, gestiti da programmi diversi.
- ✿ Un primo programma, detto **interfaccia per il linguaggio naturale** (**natural language interface** o **front-end**) accetta la frase e la analizza trasformandola in una rappresentazione simbolica equivalente.
- ✿ Questa rappresentazione costituisce l'ingresso per il programma che deve eseguire la richiesta (**esecutore**).

APPLICAZIONI: L'INTERAZIONE UOMO-MACCHINA

- ✱ La maggior parte delle interfacce per il linguaggio naturale esistenti non comprende completamente le frasi analizzate.
- ✱ In genere è sufficiente estrarre dalla frase le informazioni che costituiscono la rappresentazione formale della richiesta.
- ✱ A questo scopo, l'interfaccia utilizza di solito due tipi di conoscenza:
 - un dizionario;
 - una rappresentazione delle funzioni espletate dall'esecutore, utilizzata per estrarre dalla frase le informazioni rilevanti.

ELIZA

- ✿ Realizzato da Weizenbaum nel 1966, **Eliza** è uno dei più noti programmi per l'interazione uomo-macchina in linguaggio naturale.
- ✿ In realtà, non si può considerare Eliza come un vero e proprio sistema per l'elaborazione del linguaggio.
- ✿ E' piuttosto un intelligente giocattolo, che merita di essere citato più che altro per motivi storici.

ELIZA

- ✿ Eliza simula uno psicoterapeuta.
- ✿ L'utente inserisce frasi in inglese del tutto arbitrarie, e il sistema risponde con un'altra frase, dando l'impressione che si stia svolgendo un vero proprio dialogo.
- ✿ Esempio:

Utente: Oggi mi sento un po' depresso.

Eliza: MI DISPIACE. C'E' QUALCOSA IN PARTICOLARE CHE TI DEPRIME?

Utente: Forse un insuccesso sul lavoro:

Eliza: COME VIVI GLI INSUCCESSI?

Utente: Penso sempre a che cosa mi direbbe mio padre.

Eliza: PARLAMI ANCORA DELLA MIA FAMIGLIA.

ELIZA

- ✿ Il metodo utilizzato da Eliza è molto elementare.
- ✿ Semplificando al massimo, le frasi in ingresso vengono confrontate con un archivio di schemi di frasi.
- ✿ Quando si trova uno schema corrispondente alla frase d'ingresso, si stampa una risposta costruita a partire da uno schema parimenti memorizzato nell'archivio.
- ✿ Quando la frase in ingresso non corrisponde a nessuno schema noto, si stampa una risposta generica, abbastanza vaga da essere adatta praticamente ad ogni situazione.

ELIZA

- ✱ In realtà Eliza non capisce nulla di quanto viene detto, perché non costruisce nessuna rappresentazione del significato delle frasi.
- ✱ Ciò nonostante, il dialogo ha spesso un aspetto piuttosto realistico.
- ✱ Questo è dovuto da un lato ad alcuni trucchi ingegnosi utilizzati nella costruzione della risposta, e d'altro lato alla scelta dl dominio applicativo, che non è certo casuale.

ELIZA

- ✿ In un dialogo psicoterapeutico, infatti, nessuno si stupisce se il terapeuta non risponde direttamente alle domande, o se salta all'improvviso da un argomento all'altro.
- ✿ Uno psicoterapeuta umano si comporterebbe in questo modo seguendo una strategia dettata da una tecnica del colloquio.
- ✿ Eliza non segue nessuna strategia, ma non è facile accorgersene.

SHRDLU

- ✿ **Shrdlu** è stato realizzato da Winograd nel 1972.
- ✿ Viene considerato la prima realizzazione significativa nel campo dell'elaborazione automatica del linguaggio.
- ✿ Il sistema opera all'interno di un dominio applicativo giocattolo, comprendente un robot dotato di un braccio in grado di spostare alcuni oggetti (cubi, piramidi, scatole) posti su di un tavolo.
- ✿ Il braccio, il tavolo e gli oggetti non esistono fisicamente, ma sono simulati da un programma che visualizza su monitor la disposizione degli oggetti e la posizione del braccio ad ogni istante.

SHRDLU

- ✿ L'utente può richiedere al sistema di spostare certi oggetti mediante il braccio, sottoponendo una frase in inglese.
- ✿ Shrdlu interpreta la frase, controlla l'eseguibilità della richiesta e ne simula l'esecuzione, visualizzando infine il risultato dell'operazione.
- ✿ Shrdlu non utilizza schemi di rappresentazione della conoscenza come quelli che abbiamo discusso nelle lezioni precedenti.
- ✿ La conoscenza è codificata mediante procedure scritte in un'estensione del linguaggio di programmazione Lisp.

SHRDLU

- ✱ Se le frasi in ingresso non sono troppo complesse, Shrdlu è in grado di eseguire le richieste correttamente, fornendo una prestazione che non manca di colpire.
- ✱ Tuttavia, non sembra che questo schema possa essere considerato come un modello generale del processo di comprensione del linguaggio.
- ✱ Infatti, le sue capacità derivano essenzialmente da un'intelligente codifica delle procedure di risoluzione automatica di problemi relative al dominio applicativo, che non sono facilmente generalizzabili a domini più ampi o qualitativamente diversi.

STATISTICAL NATURAL LANGUAGE PROCESSING

- ✱ L'interesse per questo tema scaturisce dalle esigenze di poter trattare enormi insiemi di documenti disponibili in rete per le più varie applicazioni.
- ✱ D'altro canto, la disponibilità di estesi corpora testuali ha cambiato l'approccio scientifico allo studio del linguaggio in linguistica e in psicologia cognitiva.
- ✱ Per Statistical NLP intendiamo tutti gli approcci "quantitativi" all'elaborazione del linguaggio naturale (modellazione probabilistica, information theory, linear algebra).

STATISTICAL NATURAL LANGUAGE PROCESSING

- ✿ Nell'approccio basato su corpus alla comprensione del linguaggio possiamo considerare tre attività specifiche:
 - il recupero delle informazioni (**information retrieval**)
 - l'estrazione delle informazioni (**information extraction**)
 - traduzione automatica

INFORMATION RETRIEVAL

- ✿ L'**Information Retrieval** consiste nel trovare i documenti rilevanti per le necessità informative dell'utente (**information needs**).
- ✿ I sistemi più noti sono probabilmente i motori di ricerca per il Web: un utente può inserire una **query** e ottenere una lista di **pagine rilevanti**.

INFORMATION RETRIEVAL

- ✻ Un sistema di Information Retrieval può essere caratterizzato in base ai seguenti componenti:
 - Una collezione di documenti.
 - Una query posta in un linguaggio di interrogazione.
 - Un insieme di risultati.
 - Una presentazione dell'insieme di risultati.

INFORMATION RETRIEVAL

- ✿ Si potrebbe pensare che per costruire un sistema di Information Retrieval sia sufficiente eseguire l'analisi sintattica di un insieme di documenti e riempire una base di conoscenza utilizzando un certo formalismo di rappresentazione per poi interpretare ogni query e chiedere la risposta alla base di conoscenza.
- ✿ Purtroppo nessuno è riuscito a sviluppare un sistema di questo tipo su larga scala.
- ✿ Costruire un lessico e una grammatica in grado di coprire una collezione di documenti così vasta è semplicemente troppo difficile, per cui tutti i sistemi di Information retrieval utilizzano modelli di linguaggio più semplici.
- ✿ Nelle prossime lezioni approfondiremo tali metodi.