
Paradigmi e linguaggi di programmazione

Linguaggi formali e naturali e modelli per l'informatica

Paradigmi e Linguaggi di Programmazione

Cenni di Natural Language Processing (NLP)

Natural Language Processing (NLP)

- ❑ **Analisi morfologica e POS tagging**
 - ❑ **Analisi semantica**
-

Natural Language Processing (NLP)

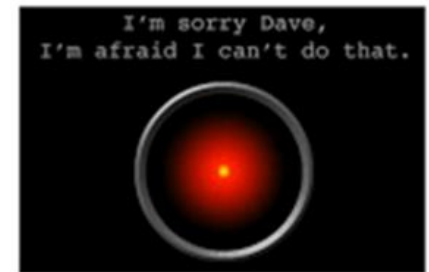
- “la linguistica computazionale e l’intelligenza artificiale sono nati nello stesso periodo e e cresciuti insieme, mescolandosi nella disciplina denominata “linguistica computazionale”
 - Russell, Stuart, Peter Norvig, and Artificial Intelligence. "A modern approach." *Artificial Intelligence*. Prentice-Hall, 1995.
- http://www1.unipa.it/sorce/didattica/sei1213/SEI1213_01_Linguistica_Computazionale_intro.pdf
- Elaborazione dei linguaggi naturali: automi e morfologia (Maria Teresa Pazienza- ART (Artificial Intelligence Research at Tor Vergata)
 - <http://slideplayer.it/slide/1006278/>

1968, Kubrick, 2001
odissea nello spazio

HAL-9000

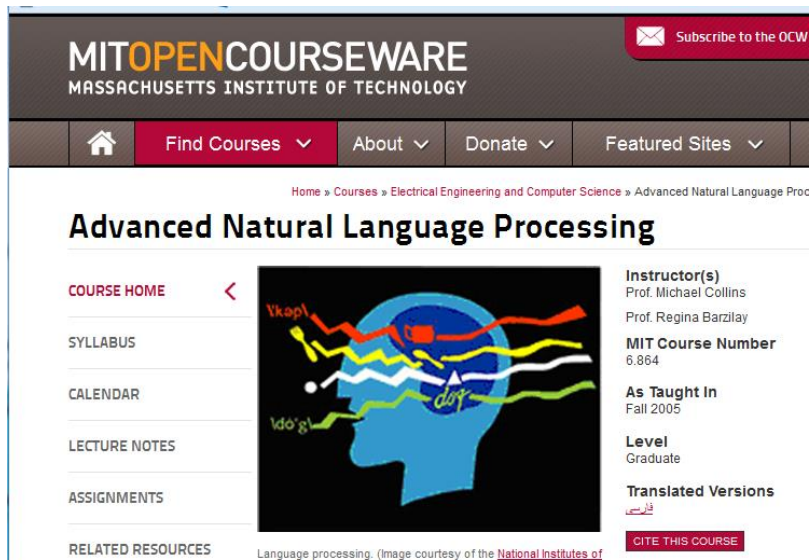


Heuristic ALgorithmic



Natural Language Processing (NLP)

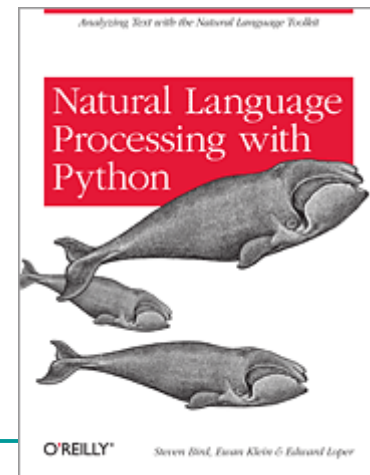
- <https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-864-advanced-natural-language-processing-fall-2005/related-resources/>



The screenshot shows the MIT OpenCourseWare website interface. At the top, there is a navigation bar with the MIT OpenCourseWare logo and a 'Subscribe to the OCW' button. Below the navigation bar, there are several menu items: 'Home', 'Find Courses', 'About', 'Donate', and 'Featured Sites'. The main content area is titled 'Advanced Natural Language Processing' and includes a sidebar with links for 'COURSE HOME', 'SYLLABUS', 'CALENDAR', 'LECTURE NOTES', 'ASSIGNMENTS', and 'RELATED RESOURCES'. The main content area features a blue silhouette of a human head with colorful lines representing neural activity and the words 'Ykopl', 'Ido'g', and 'dog' written in various colors. To the right of the image, there is a list of course details: 'Instructor(s) Prof. Michael Collins, Prof. Regina Barzilay', 'MIT Course Number 6.864', 'As Taught In Fall 2005', 'Level Graduate', and 'Translated Versions' with a small icon. A 'CITE THIS COURSE' button is located at the bottom right of the course details section.

Steven Bird, Ewan Klein, Edward Loper, *Natural Language Processing with Python Analyzing Text with the Natural Language Toolkit* - Publisher: O'Reilly Media. 2009

► <http://www.nltk.org/>



Natural Language Processing (NLP)

- Prende in esame delle stringhe di parole (*frasi*) come input ed produce come output una **rappresentazione strutturata del significato** di queste stringhe
 - Il primo tentativo (fallito) di usare un calcolatore per NLP risale al 1950 per un traduttore Russo-Inglese
 - Si traduceva ogni singola parola priva dal contesto del discorso
-

Natural Language Processing (NLP)

Problema complesso

- Complessità del problema dovuto a:
 1. Livello di ambiguità presente nei linguaggi naturali
 2. Complessità della informazione semantica contenuta anche in semplici frasi
 - «Una vecchia legge la regola.»

Qual è il soggetto? e il verbo?

La frase allude:

-a un'anziana signora che scorre con gli occhi il testo di una norma
-oppure di una «fattispecie», indicata nella frase dal pronome *la*,
regolamentata da un'antica legge?

Alcune applicazioni di NLP

- ✓ Correttori ortografici, grammaticali, ecc.
- ✓ Riconoscimento automatico del parlato
 - ✓ Automatic Speech Recognition (ASR)
- ✓ Sintesi automatica della voce
 - ✓ Text-To-Speech (TTS)
- ✓ Estrazione automatica di informazione da testi
 - ✓ Information Extraction (IE)
- ✓ Interrogazione documenti con domande in linguaggio naturale
 - ✓ Question Answering (QA)
- ✓ Traduzione (semi)-automatica di testi
 - ✓ Machine Translation
- ✓ Interazione (conversazione) uomo-macchina multimodale
 - ✓ Robot

Natural Language Processing (NLP)

Quali conoscenze linguistiche deve possedere il computer?

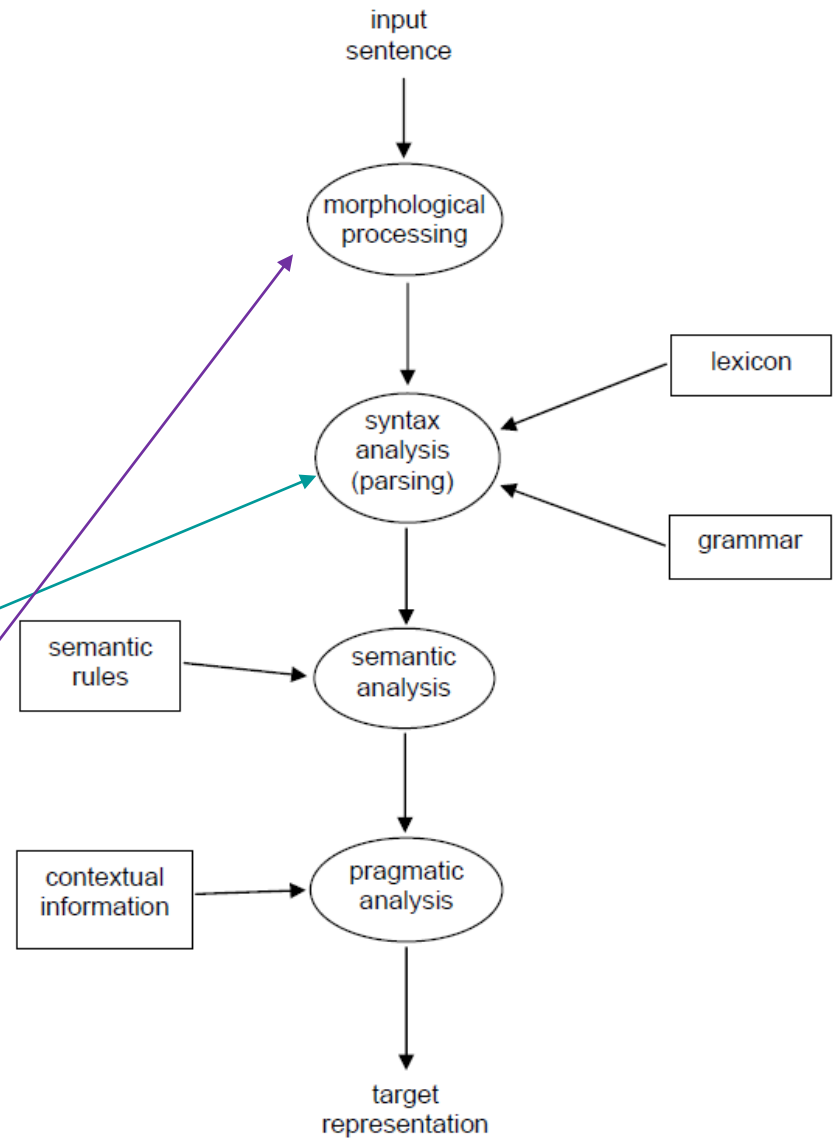
- articolare e decodificare i suoni di una lingua
 - *Fonetica articolatoria e acustica Fonologia ecc.*
 - conoscere le parole di una lingua, la loro struttura e la loro organizzazione
 - *Lessico e morfologia*
 - comporre le parole in espressioni linguistiche complesse (sintagmi, frasi, ecc.)
 - *Sintassi*
 - assegnare significati alle espressioni linguistiche semplici e complesse
 - *Semantica (lessicale e composizionale)*
 - usare le frasi nei contesti, situazioni e modi appropriati agli scopi comunicativi
 - *Pragmatica*
-

I passi logici in Natural Language Processing

Raramente nei sistemi reali queste fasi sono svolte in modo strettamente separato e sequenziale

La sintassi italiana e quella inglese sembrano essere Context-Free e per eseguire questa parte si usano le stesse tecniche implementate per la traduzione di L.d.P.

La morfologia, invece, sembra essere ancora più semplice: può essere infatti rappresentata da *Grammatiche Regolari*



Morfologia (dal greco, *morphé* "forma" e *lògos* studio ")

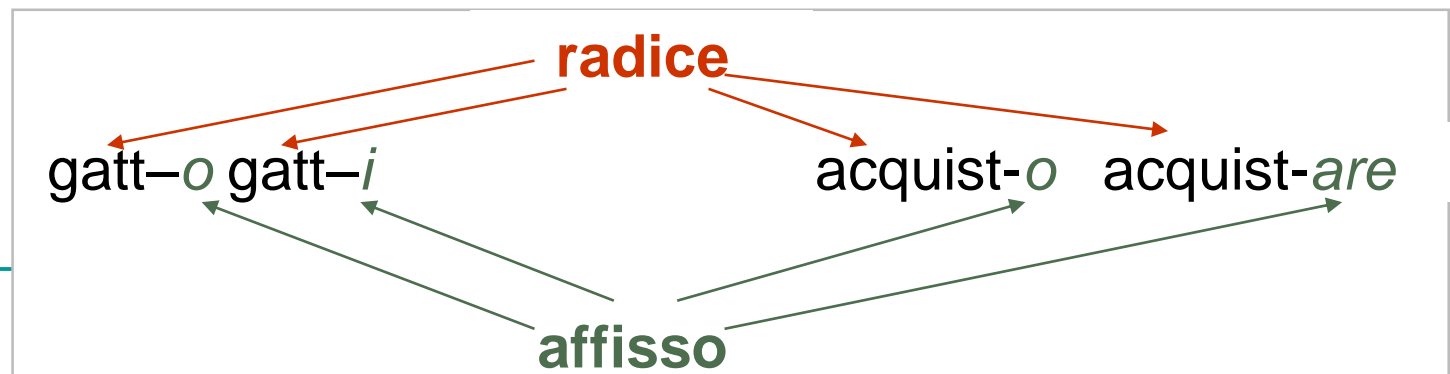
La **morfologia** è lo studio di come le parole sono costruite a partire da unità atomiche dette *morfemi*.

I **morfemi** sono le più piccole unità linguistiche che possiedono un significato.

-Come il *fonema* è l'unità elementare nel sistema dei suoni

Ne esistono due classi:

- **Radice** → il morfema che dà il significato principale alla parola
- **Affisso** → il morfema che dà significato aggiuntivo alla parola



Analisi Morfologica: Automi a Stati Finiti

Strumenti per l'analisi morfologica :

- Automi a Stati Finiti (**FSA**) → Riconoscimento
- Finite State Transducers (**FST**) → Parsing

RICONOSCIMENTO : indica se una data parola in input è morfologicamente corretta o no (ad esempio *gatti* è corretta, *gattare* è scorretta)

PARSING : produce un'analisi morfologica della parola in input (ad esempio *gatti* → *gatto* _{N PL})

*Sia gli FTA che i gli FTP sono un formalismo di tipo 3 nella gerarchia di Chomsky: l'analisi morfologica può essere quindi portata a termine con **Espressioni Regolari***

Analisi Morfologica: qualche esempio

Un analizzatore morfologico completo dovrebbe essere in grado di riconoscere la classe (nomi, verbi, ecc.) delle parole e la loro morfologia:

house

house+N+SG

houses

house+N+PL

went

go+V+PastTense+3-PL

play

play+V+Present

played

played+A+VPass

miaow

miaow+Onomatop

Analisi Morfologica

La *morfologia* può essere divisa in due parti principali

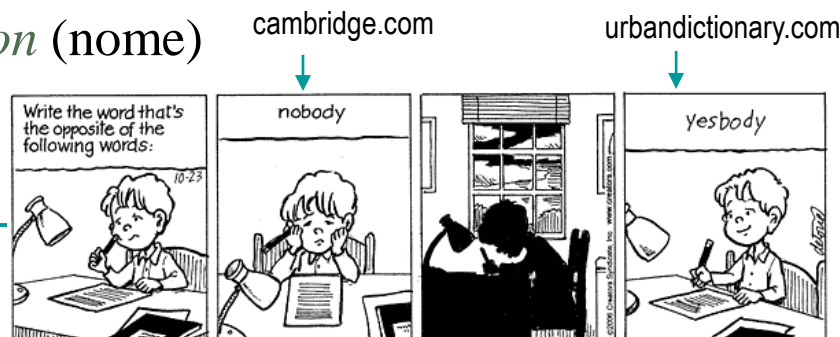
1. **Inflectional Morphology** combinazione di una radice con un affisso che risulta in una parola (*forma flessa*) della *stessa classe* (nome, verbo, aggettivo, ecc..) con una funzione grammaticale specifica

- *cat* (nome sing) → *cat-s* (nome plur)
- *cut* (verbo base) → *cut-**ting*** (verbo progressivo)



2. **Derivational Morphology** combinazione di una radice con un affisso che risulta in una parola di una *classe diversa*. Il significato della nuova parola non è facilmente prevedibile

- *trasporto* (nome) → *trasport-**abile*** (aggettivo)
- *computerize* (verbo) → *computeriz-**ation*** (nome)



Analisi Morfologica

Ogni linguaggio naturale ha una sua morfologia (affissi, regole, ecc.)

Due classi principali:

- **Concatenative morphology**: una parola è composta da morfemi concatenati insieme

- Italiano, Inglese: *gatt-o*, *cat-s*, *buy-er*

- **Non-concatenative morphology**: le parole sono composte in maniera complessa

- Ebraico: *lamad* (radice: *lmd*; affissi: *a-a* pass.rem.attivo)
-

Lingue agglutinanti: le parole sono formate da molti affissi

- Turco: *uygarlaştiramadiklarimizdanmişsinizcasına*
 (“comportarsi come se fossi tra quelli che non possono civilizzare”)
-

Analisi Morfologica

Un FSA può essere utilizzato per riconoscere se una parola è ammissibile in una lingua

Cosa serve per realizzarlo:

1. **Lessico** lista di radici ed affissi della lingua
 - Esempio: [cat,dog,cut,go,...,-s,-ed,-ation,-able,...,un-,dis-]
 2. **Regole Morfologiche** le regole di costruzione dei morfemi
 - Esempio: Plurale inglese: radice + -s
 3. **Regole Ortografiche** cambiamenti che occorrono in una parola quando due morfemi si combinano
 - Esempio: *city* → *cities*
-

Analisi Morfologica: applicazioni

- *Stemming* in Information Retrieval

- Data una parola della query, cercare le pagine che contengano anche le sue forme flesse

- *Spell Checking*

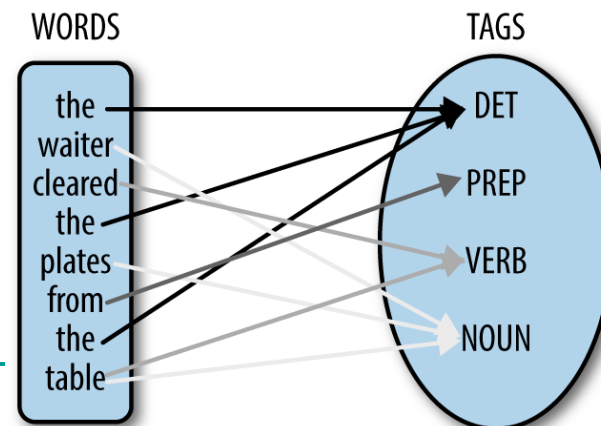
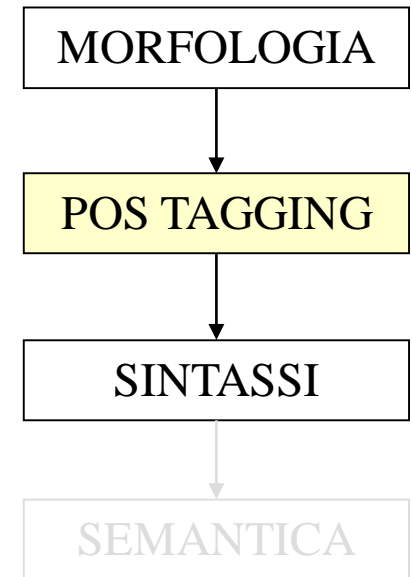
- Riconoscere quali forme flesse sono ammissibili in una lingua e quali no (ad esempio *gatt-o* e *are-gatt*)
- ***La classe (sapere se è un nome o un verbo) di una parola può aiutare a predire la struttura sintattica di una frase***

- Traduzione Automatica

- Ricondurre parole diverse a una stessa radice e quindi alla stessa traduzione (ad esempio *amatore, amare* → *love*)
-

Analisi Morfologica e POS Tagging

- **POS** (Part of Speech - classi morfologiche)
 - Categorie morfo-sintattiche a cui una parola appartiene
- Categorie principali
 - Nomi, verbi, aggettivi, avverbi, articoli, pronomi, congiunzioni ...
- **POS Tagging** assegnazione delle parole di documento fatta da modelli basati su:
 - Rule-bales
 - Stochastic
 - Misto



Analisi Morfologica e POS Tagging

ANALISI MORFOLOGICA

- Data una parola, trovare le sue interpretazioni morfologiche
- Può essere presente **ambiguità**:
ES: talks → talk+s → talk V 3PS
→ talk N PL

POS TAGGING

- Data una parola, trovare la sua unica interpretazione morfologica
 - Analisi morfologica + **disambiguazione**
 - → metodi dell'analisi morfologica (es. FST) + algoritmi di disambiguazione
-

Analisi Morfologica e POS Tagging

In genere, la definizione di una POS è basata su caratteristiche:

- **Morfologiche** gli affissi che compongono una parola
- **Sintattiche** il contesto sintattico in cui si trova la parola

Non sono considerate cioè le **caratteristiche semantiche**, sebbene le classi presentino solitamente un buon grado di coerenza semantica.

ESEMPIO (nomi)

I *nomi* in inglese hanno generalmente una forma singolare e una plurale (affisso –s)

I *nomi* in inglese sono solitamente preceduti da articoli o hanno una forma possessiva (es, *IBM's revenues*)

I *nomi* in inglese possono esprimere diverse categorie semantiche: persone, cose, astrazioni (es, *relationship*), termini simil-verbali (es, *reading*)

Analisi Morfologica e POS Tagging

Possono essere identificate due categorie principali:

- **CLASSI APERTE:**

- Classi a cui vengono spesso aggiunte nuove parole
 - in Inglese ed Italiano sono quattro: *nomi, verbi, aggettivi, avverbi*

- **CLASSI CHIUSE:**

- Classi cui appartengono un insieme relativamente statico di parole
 - *articoli, preposizioni, pronomi, congiunzioni, verbi ausiliari*

Sono le **Function Words** parole grammaticalmente significative, generalmente molto corte e frequenti nel linguaggio

- ES: *of, and, or, you ...*
-

Analisi Morfologica e POS Tagging

NOMI (N): cose, persone, luoghi...

- Nomi Comuni (NN) (*house, dog, ...*)
- Nomi Propri (NNP) (*Gino, Pino, ...*)
- Mass vs Count
 - *Count nouns*: hanno il plurale, sono enumerabili (*due panini*)
 - *Mass nouns*: gruppi omogenei, non enumerabili (*neve, sale ...*)

VERBI (V): azioni, processi ...

- Diverse classificazioni in base a: morfologia, sintassi.

AGGETTIVI (JJ): proprietà, qualità...

- Base, Superlativi, Comparativi

AVVERBI (RB): modificatori di altre classi

- Direzionali/Locativi (*here*), Temporali (*now*), Modali (*slowly*), di Gradazione (*very*)...
-

Analisi Morfologica e POS Tagging

PREPOSIZIONI : *relazioni temporali, spaziali...*

- Precedono i nomi (*of, in, for, ...*)
- Molto comuni

PARTICLE:

- Si combinano con i verbi, formando *phrasal verbs*
- Es: *go on, take off, ...*
- Si distinguono dalle preposizioni solo per caratteristiche sintattiche

ALTRE CLASSI:

- Articoli: *the, a, an*
- Congiunzioni: *and, or, but*
- Pronomi: personali (*I, you, us....*) e possessivi (*mine, yours, ...*)
- Ausiliari: copulativi (*be, do, have*) e modali (*should, must, can*)
- Interiezioni, numerali, negazioni, greetings....

ESEMPIO dal Brown Corpus(De Rose, 1988)

Tag	Description	Example	Tag	Description	Example
CC	Coordin. Conjunction	<i>and, but, or</i>	SYM	Symbol	<i>+, %, &</i>
CD	Cardinal number	<i>one, two, three</i>	TO	"to"	<i>to</i>
DT	Determiner	<i>a, the</i>	UH	Interjection	<i>ah, oops</i>
EX	Existential "there"	<i>there</i>	VB	Verb, base form	<i>eat</i>
FW	Foreign word	<i>mea culpa</i>	VBD	Verb, past tense	<i>ate</i>
IN	Preposition/sub-conj	<i>of, in, by</i>	VBG	Verb, gerund	<i>eating</i>
JJ	Adjective	<i>yellow</i>	VBN	Verb, past participle	<i>eaten</i>
JJR	Adj., comparative	<i>btgger</i>	VBP	Verb, non-3sg pres	<i>eat</i>
JJS	Adj., superlative	<i>wildest</i>	VBZ	Verb, 3sg pres	<i>eats</i>
LS	List item marker	<i>1, 2, One</i>	WDT	Wh-determiner	<i>whitch, that</i>
MD	Modal	<i>can, should</i>	WP	Wh-pronoun	<i>what, who</i>
NN	Noun, sing. or mass	<i>llama</i>	WPS	Possessive wh-	<i>whose</i>
NNS	Noun, plural	<i>llamas</i>	WRB	Wh-adverb	<i>how, where</i>
NNP	Proper noun, singular	<i>IBM</i>	\$	Dollar sign	<i>\$</i>
NNPS	Proper noun, plural	<i>Carolmas</i>	#	Pound sign	<i>#</i>
PDT	Predeterminer	<i>all, both</i>	"	Left quote	<i>(" or ")</i>
POS	Possessive ending	<i>'s</i>	"	Right quote	<i>(' or ")</i>
PRP	Personal pronoun	<i>I, you, he</i>	(Left parenthesis	<i>([({ <</i>
PRP\$	Possessive pronoun	<i>your, one's</i>)	Right parenthesis	<i>(]) } ></i>
RB	Adverb	<i>quickly, never</i>	,	Comma	<i>,</i>
RBR	Adverb, comparative	<i>faster</i>	.	Sentence-final punc	<i>(. ! ?)</i>
RBS	Adverb, superlative	<i>fastest</i>	:	Mid-sentence punc	<i>(: ; ... - -)</i>
RP	Particle	<i>up, off</i>			

Brown Corpus

- E' il primo corpus elettronico di inglese “taggato” con i POS
 - Il Brown University Standard Corpus of **Present-Day American English** è stato compilato negli anni sessanta presso la Brown University.
 - Contiene 500 estratti di testo in lingua inglese ottenuti da lavori pubblicati negli Stati Uniti nel 1961, per un totale di circa un milione di parole.
 - L'inserimento originale dei dati fu effettuato utilizzando macchine a schede perforate.
-

Analisi Morfologica e POS Tagging

Il POS tagging dovrebbe assegnare una classe ad ogni parola di un documento

- Molte parole sono **ambigue** (quindi con più POS tag possibili)
- Un *POS tagger* deve **disambiguare**, restituendo se possibile un solo tag:
 - Utilizzando evidenze contestuali
 - Utilizzando evidenze probabilistiche da corpora annotati

Esempio per
la parola
back

–The **back/JJ** door

–On my **back/NN**

–Promised to **back/VB**

Analisi Morfologica e POS Tagging

Quanto sono ambigue le parole inglesi ?

Dal Brown Corpus (De Rose, 1988)

2 tags	3,760
3 tags	264
4 tags	61
5 tags	12
6 tags	2
7 tags	1

NON AMBIGUE (1 tag): 35,340 (88,5%)
AMBIGUE (2-7 tag): 4,100 (11,5%)

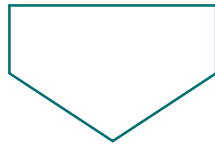
Analisi Morfologica e POS Tagging

Caratteristiche generali dei POS tagger:

METODI

INPUT

- *Tagset*
- *Dizionario con tag*
- *Frase da annotare*



OUTPUT

- *Frase annotata*

- Rule-based: database di regole di disambiguazione
- Stochastic: risolvono le ambiguità con evidenze probabilistiche estratte da un corpus annotato (*HMM, Markov models...*)
- Misti: utilizzano regole di disambiguazione ricavate con metodi stocastici

Analisi Morfologica e POS Tagging

I *sistemi rule-based* prevedono due fasi

1. Assegnazione ad una parola di tutti i possibili POS tag, utilizzando un **dizionario**
2. Applicazione di **regole di disambiguazione** create manualmente, per rimuovere tag ambigui.

ESEMPIO (fase 1)

DIZIONARIO

•she:	<i>PRP</i>
•promised:	<i>VCN , VBD</i>
•to	<i>TO</i>
•back:	<i>VB, JJ, RB, NN</i>
•the:	<i>DT</i>
•bill:	<i>NN, VB</i>

(Verb Past Participles)

(Verb Past Tense)



			NN		
			RB		
	VCN		JJ		VB
PRP	VBD	TO	VB	DT	NN
She	promised	to	back	the	bill

Analisi Morfologica e POS Tagging

ESEMPIO (fase 2)

Regola 1: Rimuovere *VBN* se è in alternativa a *VBD* e se segue “<inizio frase>*PRP*”

Regola 2: Rimuovere *VB* se è in alternativa a *NN* e se segue *DT*

	VBN								
PRP	VBD	TO	VB	DT	NN				
<i>She</i>	<i>promised</i>	<i>to</i>	<i>back</i>	<i>the</i>	<i>bill</i>				

.....

1. **VBN:** *has or have + the past participle*

be + the past participle (*be* sometimes is omitted)

2. **Modal verbs as auxiliary verbs:** *could / may/ should / might/ would / must + have + VBN*

3. **VBNs** often used as participial adjectives, whereas **VBD** often follow a subjective.

Analisi Morfologica e POS Tagging

L'*approccio stocastico* differisce dal precedente nella fase di disambiguazione:

Il tag corretto viene selezionato in base ad evidenze statistiche e alla teoria della probabilità

- **Approcci semplici:** *Most Frequent Tag*
- **Approcci complessi:** *HMM (Hidden Markov Model), Transformation-based tagging*

METODO MOST FREQUENT TAG

Le parole ambigue utilizzano un tag più spesso di altri.

- **Metodo:**
 1. Creare un dizionario e annotare manualmente un corpus
 2. Per ogni parola ambigua in un nuovo testo calcolare probabilità di annotazione
 3. Assegnare il tag più probabile
-

Analisi Morfologica e POS Tagging

ESEMPIO

I/PP give/VB you/PP **a/?** pen/NN
Section/NN 381/CD **a/?**

Cardinal Number

Possibili tag per a:

DT

NN

FW

Foreign word

Qual è il più probabile ?

Per esempio attraverso l'utilizzo del Brown corpus

- Corpus: insieme di documenti annotato manualmente con tag non ambigui
- Calcolare occorrenze di *a* con i diversi tag:

a/DT 21,830

a/NN 6

a/FW 3

Analisi Morfologica e POS Tagging

e l'assegnazione quindi del tag più probabile...

I/PP give/VB you/PP **a/DT** pen/NN

Section/NN 381/CD **a/DT**

← **ERRATO (NN)**

- ✓ Annota bene in molti casi
- ✓ Nei casi più rari sbaglia sempre
- ✓ Per aumentare le prestazioni è necessario prendere in considerazione altre informazioni.
 - Ad esempio: guardare il tag della parola precedente e successiva risolverebbe l'esempio

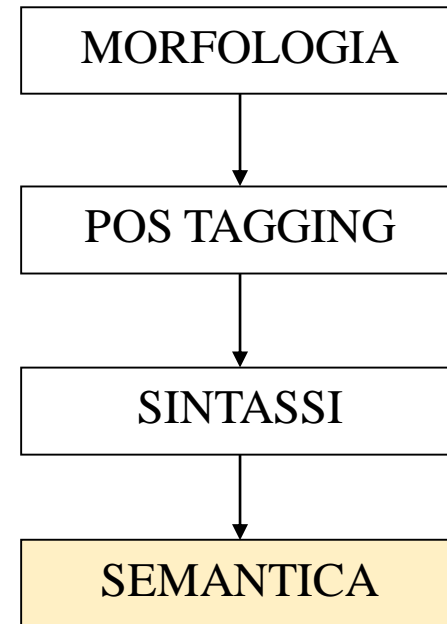
Analisi Morfologica e POS Tagging

METODO MISTO (rule based + stochastic)

- RULE BASED: La scelta del tag è effettuata usando *regole contestuali (contextual rules)* e *regole lessicali (lexical rules)*
 - STOCHASTIC: Le regole sono apprese automaticamente con tecniche di Machine Learning
 - CONOSCENZA PREGRESSA NECESSARIA:
 - *Lessico*: per assegnare i tag
 - *Corpus annotato*: per l'apprendimento delle regole
-

Analisi semantica

- *Capire il significato di una frase* serve poi per inserirla correttamente nel contesto del discorso
- *Questo compito è eseguito dall'analisi semantica* che si basa sulla rappresentazione formale del significato della frase.



Analisi semantica

- Un approccio sistematico e formale serve a collegare la rappresentazione del significato della frase con tutti i fatti e gli eventi del dominio in questione (o, almeno quelli che sono rappresentati nella base di conoscenza relativa)
- La rappresentazione del significato può avvenire in *linguaggi formali tipici dell'Intelligenza Artificiale* quali il calcolo dei predicati, reti semantiche, frames, etc..
- Un sistema potrà allora **fare inferenza**, ovvero:
 - ragionare a partire dalla frase
 - fare deduzioni valide rispetto sia alla rappresentazione formale del significato della frase in ingresso sia alla conoscenza di base del sistema stesso.

Analisi semantica

- Data la ricchezza espressiva del linguaggio naturale, si possono avere più espressioni linguistiche (frasi diverse) che esprimono lo stesso fatto e di conseguenza si avranno rappresentazioni formali diverse per lo stesso concetto

- Esempio:

- Il calcio è uno sport popolare in Italia*

- Gli italiani amano il calcio*

- Il calcio è molto seguito in Italia*

- E per capire il significato di una frase si deve ricorrere a processi di concettualizzazione (un metalivello rispetto a quello linguistico) che sono possibili solo se il linguaggio di rappresentazione del significato supporta l'uso di variabili per i concetti (termini di una ontologia)

Dato: *Gli europei seguono il calcio*

Domanda: *Gli italiani amano il calcio?*

Analisi semantica guidata dalla sintassi

- Per ricavare il corretto significato di una frase può essere utile analizzare le relazioni che esistono tra i concetti (associati a diverse parole) e le componenti frasali di ciascuna espressione linguistica
 - *Identificare, cioè, il significato completo di una frase come risultato della composizione del significato dei singoli elementi linguistici in ingresso*
- Es: i verbi (**predicati**) assumono un ruolo rilevante imponendo vincoli sulla struttura grammaticale e sulla posizione di altri elementi (**argomenti**) che devono accompagnarli nella struttura sintattica di una frase

Gli italiani amano il calcio

- Il predicato ha due argomenti di tipo nominale
- Il primo precede il verbo ed assume il ruolo di soggetto
- Il secondo segue il verbo ed assume il ruolo di oggetto

Analisi semantica guidata dalla sintassi

Quindi si può dire che il significato di una frase

- non è basato esclusivamente sulle parole che la compongono,
- ma deve tener conto dell'ordine con cui compaiono, dei raggruppamenti esistenti, e delle interrelazioni;

➤ nell'analisi semantica basata sulla sintassi la composizionalità dei significati terrà quindi conto delle diverse componenti sintattiche e delle relazioni esistenti.

Analisi semantica guidata dalla sintassi

Ovviamente esistono molti casi in cui non ha senso applicare il principio della composizionalità

Esempi

- ❑ Idiomi, giochi di parole, espressioni ironiche o sarcastiche, metafore, richieste indirette, ...
- ❑ *Occhio di bue, di palo in frasca, due piccioni con una fava, cadere dalle nuvole,, la punta di un iceberg etc...*

In molti di questi casi il significato dell'espressione come unico evento è a seconda dei casi:

- totalmente scorrelato dal significato delle singole parti
 - correlato in qualche modo oscuro e molto vago
-

Analisi semantica: il ruolo del significato delle parole

- Il significato delle frasi viene ricavato non solo dalla conoscenza del dominio e del contesto e dai significati associati alle strutture sintattiche, ma anche dai significati associati alle parole (***semantica lessicale***) e di come esse si combinano per formare il significato delle frasi (***semantica della frase***)
 - A differenza dell'analisi morfologica che studia le componenti significative di una parola, l'**analisi del lessico** studia il significato delle parole e delle relazioni lessicali (sinonimia, iperonimia, meronimia, antinomia, causa,...).
-

Analisi semantica: il ruolo del significato delle parole

- **Applicazioni tipiche:**
 - Costruzione di Thesaurus
 - Information Extraction
 - Text Summarization



Analisi del lessico

- Il *lessico* è una struttura linguistica che determina ciò che le parole possono significare e come possono essere usate;
 - la struttura consiste sia di relazioni tra parole e del loro significato, che della struttura interna di ogni parola.
 - **Lexeme/Lessema**
 - L'unità di significato in un linguaggio,
 - Può essere una sola parola o un gruppo di parole
 - Ad ogni lessema è associato un significato preciso.
 - **Lexicon/Lessico** : una collezione di lessemi
-

Relazioni tra lessemi: esempi

Homonymy/omonimia:

- Lessemi diversi che assumono significati totalmente diversi, ma condividono una stessa forma fonologica, ortografica o entrambe

- Esempio :
 - piano (progetto) vs
 - piano (piano di un edificio) vs
 - piano (pianoforte)

Relazioni tra lessemi: esempi

Polysemy/polisemia:

- ❑ Lo stesso lessema che assume più significati tra loro collegati
- ❑ Esempio: banca :
 - istituto bancario, banca dati, banca del sangue, banca del tempo

Synonymy/sinonimia:

Lessemi diversi che assumono lo stesso significato

Esempio: grande vs grosso

Analisi semantica: analisi del lessico

Che metodologie utilizzare ?

- **metodologie distribuzionali, approcci statistici non supervisionati**
 - fortemente basate su studi statistico-distribuzionali delle parole
 - adattabili *no-cost* a differenti lingue
 - non garantiscono una analisi semantica approfondita
 - **metodologie basate su *conoscenza***
 - approcci con analisi di strutture ontologiche o reti semantiche (es, *WordNet*)
 - uso di misure di distanza all'interno della rete
 - non portabili a differenti lingue se non esiste una rete per essa
 - garantiscono un'analisi semantica approfondita e precisa tanto quanto la rete è semanticamente espressiva
-

Analisi Semantica: Analisi del discorso

- Il discorso consiste di più frasi tra loro coerentemente correlate in uno schema di comunicazione
 - Cosa rende coerente un discorso?
 - Come riconoscerlo?
 - Come analizzarlo?
-

Analisi Semantica: Analisi del discorso

Cosa rende coerente un testo?

- Un uso *appropriato* di relazioni di coerenza tra parti del discorso (*strutture retoriche*)
 - ❑ *Esempi*: uso corretto dei tempi o di parole chiave: perché, ma, in aggiunta...
 - ❑ *Problemi*: Sono strutture indipendenti dalla lingua? Quante sono?
 - ❑ Non modellano completamente la struttura del discorso.
 - Una sequenza *appropriata* di parti del discorso (*struttura del discorso, articolazione degli argomenti*)
 - ❑ necessario utilizzare conoscenza di supporto del contesto e degli obiettivi del discorso
-