

Università Ca' Foscari di Venezia

Linguistica Informatica Mod. 1

Anno Accademico 2010 - 2011



Modellazione

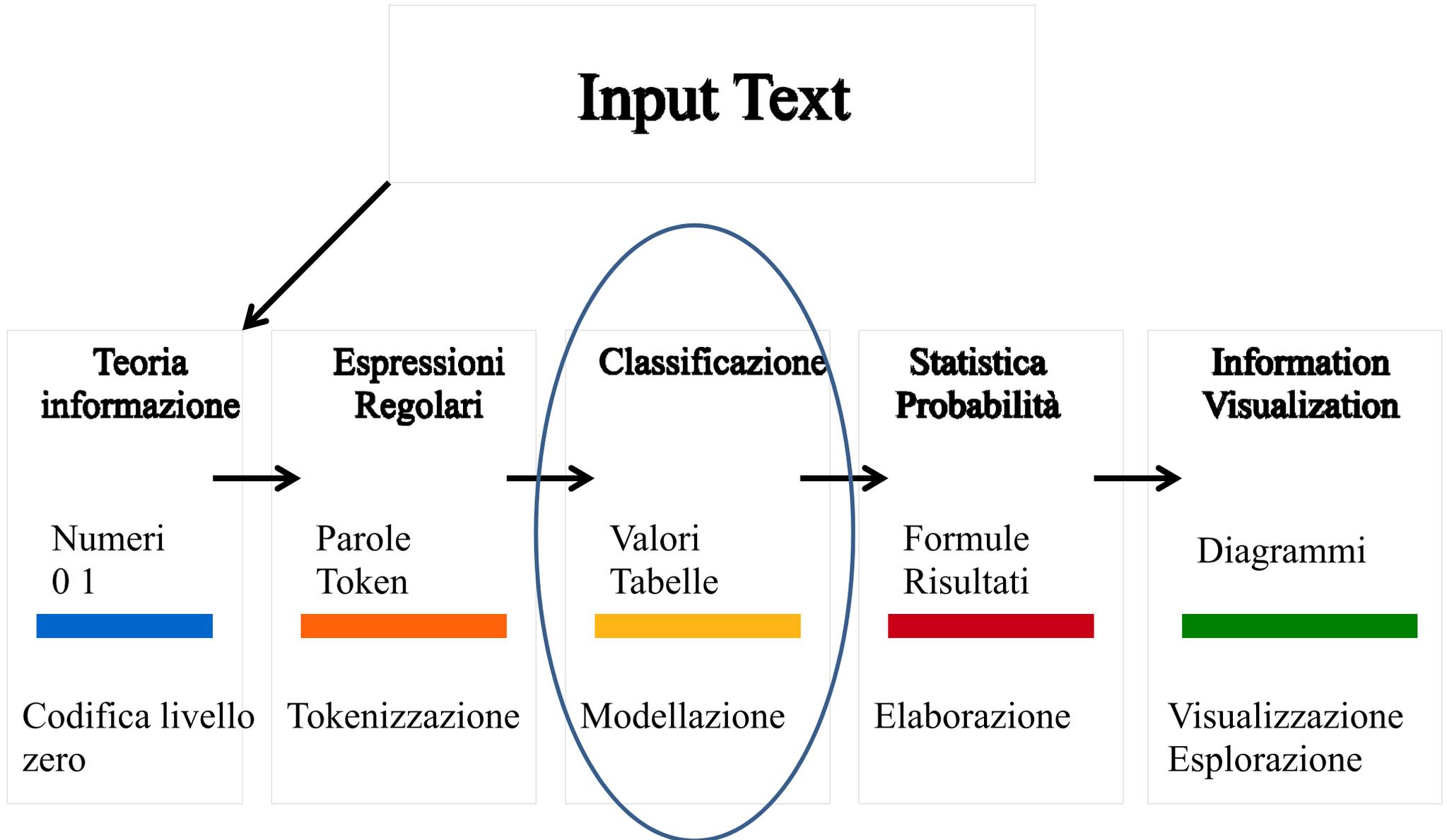
Rocco Tripodi
rocco@unive.it

Anticipazione lezione

La lezione del 23 Dicembre sarà svolta il **21 Dicembre** dalle **ore 14** alle ore 15:30 in **aula B** a Ca' Bembo

Schema

Input Text



Modellazione

Modello

struttura costruita da un particolare punto di vista. Griglia interpretativa che descrive in maniera formale gli elementi testuali e paratestuali.

Specificare ogni elemento in modo da poter essere trattato dall'elaboratore, in modo che possa simulare la lettura e la comprensione degli aspetti che ci interessa elaborare.

Classificazione

Il testo dopo la tokenizzazione si presenta all'elaboratore come una sequenza di unità minime che dovranno essere descritte in maniera non ambigua (secondo uno schema)

Verticalizzazione del testo

Conservazione della sua forma lineare e aggiunta della lettura (interpretazione) verticale (specificazione degli elementi e delle relazioni)

Testo inteso come oggetto complesso e strutturato

Organizzazione dei dati testuali su più livelli

Analisi morfo - sintattica

Forma basilare e più comune di annotazione

Assegnare ogni token ad una categoria grammaticale e specificarne le caratteristiche morfologiche (persona, genere, numero, lemma)

Consente di risolvere le ambiguità degli omografi

Esempio effettuato con Venses ([Link](#))

the	-dt
government	-nn
took	-vbd
measures	-vbz
to	-to
protect	-vbd_vbn
the	-dt
environment	-nn
.	-punto

Penn-Treebank ([Link](#))

Analisi sintattica

Annotazione di informazioni relative l'analisi sintattica della frase

Rappresentazione a costituenti: individuazione dei sintagmi, della loro costituzione e delle relazioni che intrattengono

Rappresentazione a dipendenze: annotazione delle relazioni binarie di dipendenza tra i costituenti (Es: soggetto -> oggetto) Consente

Costituenti

subj-[the-art-sn, government-n-sn]
ibarc-[took-vin-ibarc]
ibarc-[measures-v-ibarc]
vcomp-[to-p-_G29033, protect-vt-ibarc]
obj-[the-art-sn, environment-n-sn]

Dipendenze

det(environment-sn2, the).
det(government-sn1, the).
'subj-agente'(protect-cl6,
government-sn1).
'obj-result'(protect-cl6,
environment-sn2).

Categorie X-Bar: si collocano tra le categorie lessicali e quelle di frase ([Link](#))

Analisi semantica

Codifica esplicita del significato o del contenuto semantico delle parole

Classificazione delle parole all'interno di categorie semantico – concettuali

Classificazione dei ruoli tematici

Esempio estratto da Venses

George Weah is a politician of Liberia .

George Weah [umano, istituzione, legale]

politician [legale, relat, sociale, umano]

Liberia [luogo]

Si studia la lingua tramite le regole della logica e della teoria degli insiemi. I nomi comuni vengono rappresentati come classi, i determinanti come relazioni tra classi, i verbi come relazioni e le entità come riferimenti.

WordNet 1

WordNet ([Link](#))

è un database lessicale della lingua inglese sviluppato dal laboratorio di scienze cognitive della Princeton University. Al suo interno nomi, verbi, avverbi e aggettivi non sono organizzati in ordine alfabetico, ma per vicinanza semantica. Ogni raggruppamento forma un *synset* (gruppo di sinonimi).

Es: discover

Synset1: detect, observe, find, discover, notice (discover or determine the existence, presence, or fact of)

Synset2: S: learn, hear, get word, get wind, pick up, find out, get a line, discover, see (get to know or become aware of, usually accidentally)

Synset3: fall upon, strike, come upon, light upon, chance upon, come across, chance on, happen upon, attain, discover (find unexpectedly)

...

Lessemi con più significati appartengono a diversi synset.

WordNet 2

Un'altra caratteristica dei synset è che possono essere organizzati con le relazioni di

meronimia (“è parte di ...”) Es: *x is a part of y*

iponimia (“è del tipo ...”) Es: {*tree*} e {*plant*}

iperonimia (“sottoclasse di ...”) Es: {*plant*} e {*tree*}

antinomia (“non è ...”) Es: {*rise, ascend*} e {*fall, descend*}

Le parole che compongono una collocazione sono collegate da un underscore

Es: *fountain_pen*

FrameNet 1

Frame semantics

il significato di una parola è dato dalla conoscenza del contesto in cui questa viene usata. Il contesto viene strutturato dai concetti correlati alla parola.

Es: la parola “*comprare*” può essere compresa se si ha l’idea “*trasferimento commerciale*” in cui ci sono due attori (venditore e compratore) ed un oggetto da scambiare

FrameNet ([Link](#))

Gli attori sono intesi come ruoli (Frame elements) che sono evocati da parole specifiche (lexical units)

Es: Matilde fried the catfish

[Cook Matilde] fried [Food the catfish]

Nei casi più semplici il frame viene evocato da un verbo e i Frame Elements sono ad esso sintatticamente dipendenti (struttura predicato argomentale)

FrameNet 2

1	frame(TRANSPORTATION) frame_elements(MOVER(S), MEANS, PATH) scene(MOVER(S) move along PATH by MEANS)
2	frame(DRIVING) inherit(TRANSPORTATION) frame_elements(DRIVER(=MOVER), VEHICLE(=MEANS), RIDER(S)(=MOVER(S)), CARGO(=MOVER(S))) scenes(DRIVER starts VEHICLE, DRIVER controls VEHICLE, DRIVER stops VEHICLE)
3	frame(RIDING_1) inherit(TRANSPORTATION) frame_elements(RIDER(S)(=MOVER(S)), VEHICLE(=MEANS)) scenes(RIDER enters VEHICLE, VEHICLE carries RIDER along PATH, RIDER leaves VEHICLE)

Ontological Semantics 1

“una ontologia è un artefatto ingegneristico, costruito con uno specifico vocabolario usato per descrivere una certa realtà, più un insieme di assunzioni esplicite sul significato inteso delle parole del vocabolario” (Guarino, 1998).

Il significato viene inteso in senso intensionale (significato particolare)

Morning star e Evening star

Sono entrambe mappate su PLANET e VENUS

L'ambiguità è il fattore differenziante tra ontologia e linguaggio naturale. Le parole possono essere ambigue, i concetti no. Bisogna fare i conti con la polisemia e assegnare ad ogni significato di una parola un differente concetto.

Ontological Semantics 2

Una ontologia è organizzata come una serie di concetti ognuno dei quali ha una serie di proprietà collegate ad altri concetti.

Le proprietà possono descrivere sia caratteristiche fisiche degli oggetti che relazioni più complesse che riguardano oggetti immaginari o eventi. Inoltre le proprietà possono descrivere le relazioni tra i verbi e i rispettivi agenti e complementi.

Ontologia: fornisce un metalinguaggio per i concetti, le classi e i frames.

Database dei fatti: contiene le istanze dei concetti ontologici.

Lessico semantico: fornisce informazioni che specificano quale concetto e proprietà deve essere istanziata per l'entrata lessicale.

Onomasticon: le entrate dell'*onomasticon* sono collegate direttamente a quelle nel database di fatti.

Ontological Semantics 3

Es: Dresser Industries said it expects ...

Il significato del verbo SAY istanzia il concetto ontologico INFORM (che possiede una serie di slots indicanti attori e modalità).

Definition: “the event of asserting something to provide information to .. “

is-a: assertive-act

agent: human

theme: event

instrument: communication-device

beneficiary: human

Si cerca di riempire i campi con istanze della frase, riferimenti e coriferimenti.

Creazione della struttura a dipendenza semantica: *who did what to whom*.

Analisi pragmatica

Annotazione delle funzioni comunicative legate a determinate espressioni

Funzioni illocutive: Es: richiesta, domanda, promessa, invito, ecc

Funzioni retoriche: Es: introduzione, conclusione di un discorso

Essenziale nella codifica dei dialoghi

Analisi narratologica

Identificazione dei personaggi, del tema, della fabula, dell'intreccio e del genere.

Personaggio

identificato da attributi, stati e trasformazioni.

Inserimento in una base di conoscenza sotto forma di fatti (regole)

Allo stesso modo vengono create delle regole nel database che caratterizzano tutti gli oggetti del testo.

Vengono descritti tramite marche semantiche

Gli elementi possono essere considerati sia in forma sincronica (fabula) che diacronica (tema)

Cluster di marche semantiche formano i ruoli narrativi

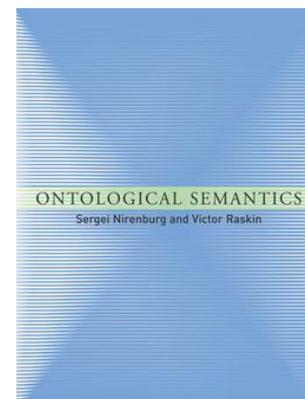
Genere

Analisi delle funzioni e delle strutture narrative

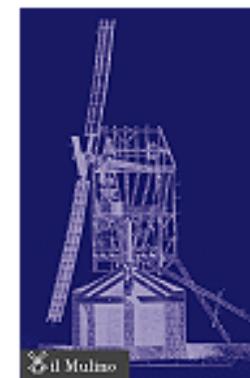
Libri consigliati

Ontological semantics

Sergei Nirenburg, Victor Raskin



Le strutture del linguaggio. Semantica
Gennaro Chierchia



L'inventario del mondo
Paolo Valore

