

NLP - LEZIONE 10

DEL 04/11/2019

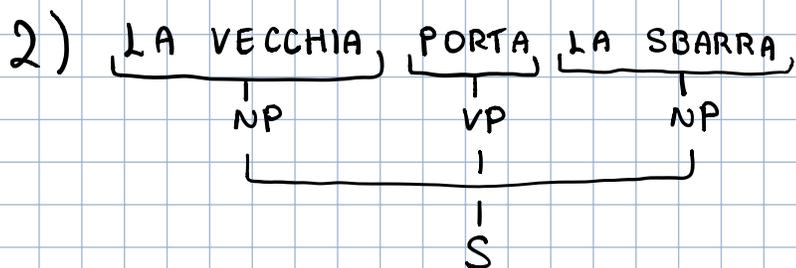
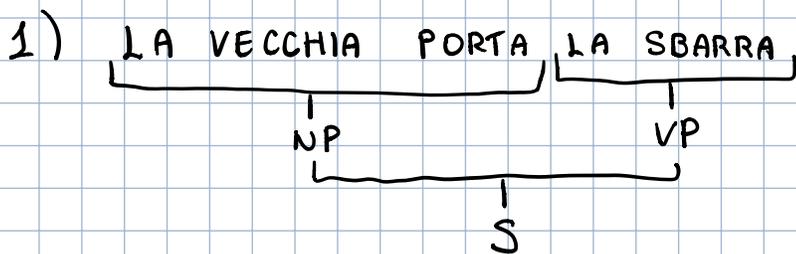
SINTASSI III

FEATURE STRUCTURE

Consideriamo la frase

LA VECCHIA PORTA LA SBARRA

abbiamo due possibili interpretazioni sintattiche e semantiche per questa frase:



Notiamo che se al posto di VECCHIA mettiamo VECCHIO, l'interpretazione corretta è era solo



Un primo approccio è quello di codificare le informazioni sul GENERE all'interno delle gramm. G tramite l'introduzione di NON-TERMINALI come segue

NOME M \rightarrow VECCHIO

NPM \rightarrow ART M NOME M

NOME F \rightarrow VECCHIA

NPF \rightarrow ART F NOME F

ART M \rightarrow IL

⋮

ART F \rightarrow LA

ADS M \rightarrow VECCHIO

S \rightarrow NPM VPM

ADS F \rightarrow VECCHIA

S \rightarrow NPF VPF

Notiamo subito nei limiti di questo approccio:

i) Si aumentano il # di regole della grammatica in modo significativo.

ii) Non si è comunque in grado di gestire le seguenti frasi

MARIO VA A ROMA

MARIA VA A ROMA

I BAMBINI VANNO A ROMA

Coni facendo infatti per ogni caratteristica saremmo costretti a introdurre nuovi terminali.

Le caratteristiche più importanti che ci vengono in mente sono GENERE (maschile o femminile), e NUMERO (singolare o plurale).

Un altro approccio per codificare queste info è tramite una ETICHETTATURA delle regole della grammatica. A seguire un possibile esempio...

NP_gen:F_num:S → ART_gen:F_num:S NOM_gen:F_num:S

NP_gen:M_num:S → ART_gen:M_num:S NOM_gen:M_num:S

Volevo ASTRARRE vediamo che le regole seguono lo schema STRUTTURA, e quindi possiamo completare la scrittura come segue

NP_gen:[X]_num:[Y] → ART_gen:[X]_num:[Y] NOM_gen:[X]_num:[Y]

Adesso ci dobbiamo porre i seguenti problemi:

i) Come codifichiamo queste etichette?

ii) Come operiamo sulla struttura dati utilizzate per codificare le etichette?

Per quanto riguarda i) possiamo utilizzare un DIZIONARIO key-value. In PYTHON avremmo quindi

{ CAT: NOME, gen: F, num: S }

nomini come anche la CLASSE GRAMMATICALE e' diventata una caratteristica.

In generale ad ogni nome associamo delle FEATURES. Possiamo avere casi, come per i verbi, a cui non associamo NESSUN valore per determinate features come nomi avere casi in cui era parole che più di un valore per una data feature.

ESEMPIO : Consideriamo la parola

PORTACCIA

nomi aggiunge questa parola come valore della caratteristica "DIMINUTIVI" associate alla parola PORTA.

Non tutte le features devono essere CONFRONTATE tra loro, ma tutte le features devono essere portate avanti nelle varie fasi.

Per semplificarci le vite, esumiamo di lavorare con DIZIONARIO STATICO. Questo ci permette di scrivere la grammatica basandoci solamente sui termini presenti nel dizionario. Per complicare le cose nomi anche pensare di utilizzare un DIZIONARIO GENERATIVO utilizzando delle info. MORFOLOGICHE.

OSS: Le grammatiche HPSG e LFG hanno provato a gestire sia la sintassi che la semantica.

ESEMPIO: Per la parola PORTA nominato avere le seguenti features

{
FORMA: PORTA,
LEMA: PORTA,
CAT: VERBO,
NUM: SING,
PERSONA: 3^o,
TEMPO: PRESENTE,
MODO: INDICATIVO}

Per quanto riguarda il punto ii) l'idea è quella di vedere le variabili come VARIABILI DICHIARATIVE, e non PROCEDURALE. Così facendo siamo in grado di UNIFICARE e PROPAGARE il valore di una variabile in tutti i luoghi in cui la variabile appare.

OSS: Nei linguaggi dichiarativi lo statement

$$2x = x + 1$$

implica che il valore di x è 1.

Questo valore nei non cambia nel tempo.

Nei linguaggi PROCEDURALI il concetto di VARIABILE UNIFICATA deve essere implementato, e non è supportato nativamente.

Notiamo che durante il controllo devo stare attento a non considerare le features il cui valore non ci interessa per capire se possiamo accordare i vari componenti. In altre parole, vogliamo UNIFICARE le informazioni là dove non unificabili. Per fare questo utilizzeremo l'operatore di SUSSUNZIONE, che ci permette di UNIFICATE STRUTTURE RICORSIVE.

Intuitivamente la sussumzione su due sequenze di features F_1 e F_2 è rappresentata con $F_1 \sqsubseteq F_2$ ed è definita come segue

F_1 è unificabile e F_2 se non unificabili tutte le features in comune tra F_1 e F_2 .

Andiamo adesso a FORMALIZZARE quanto descritto...

FORMALIZZAZIONE ~ (01:23:00 min)

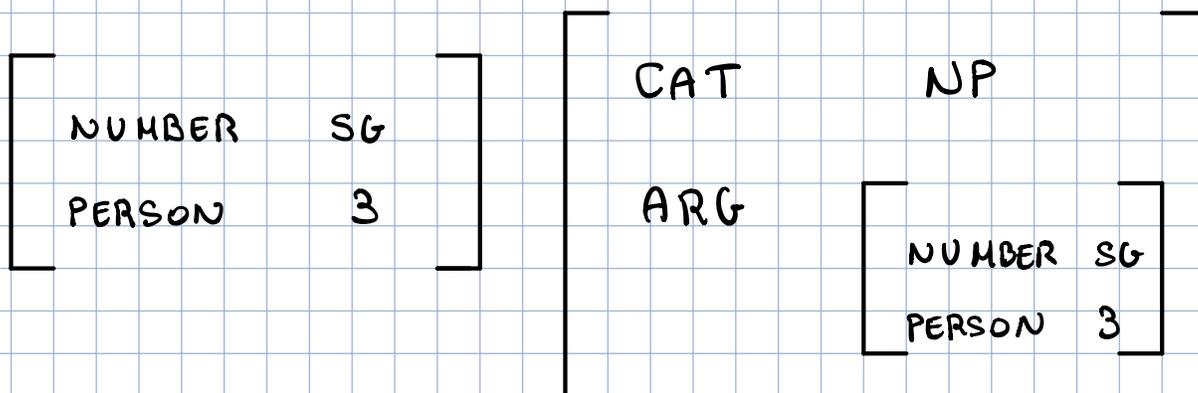
- DEF. FEATURE STRUCTURE

Una FEATURE STRUCTURE fs è una struttura ricorsiva di caratteristiche che può essere

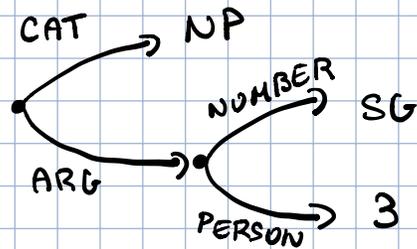
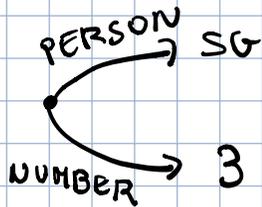
- i) ATOMICA, che non può essere scomposta ulteriormente
- ii) INSIEME DI COPPIE $\langle \text{FEATURE}, \text{VALUE} \rangle$, dove VALUE è a sua volta una feature structure.

Notiamo quindi che una FS può avere una forma arbitrariamente complessa. Abbiamo due modi per rappresentare una FS:

i) MATRICIALE

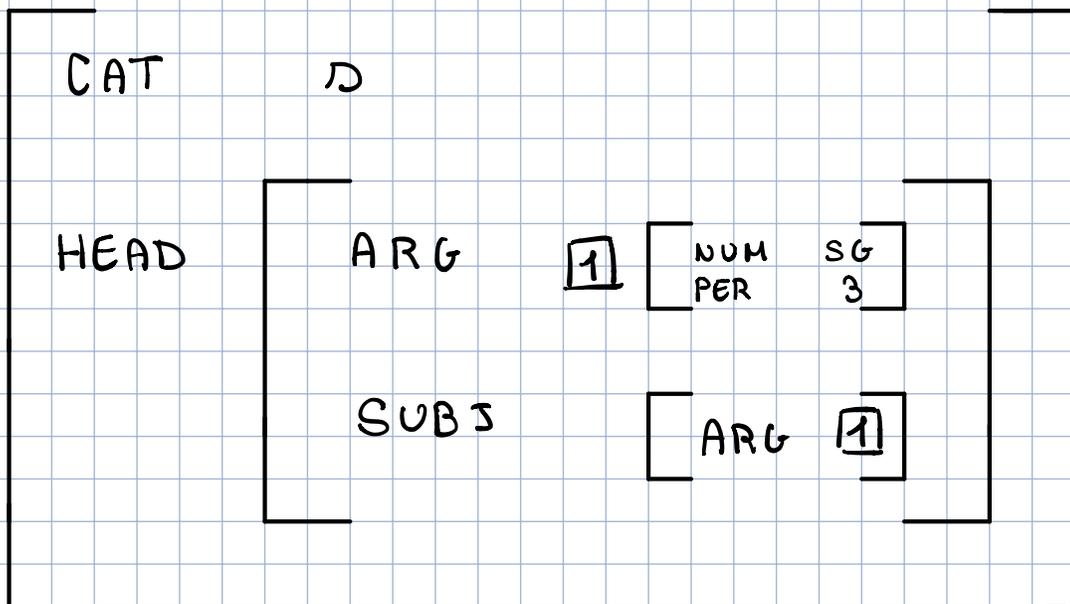


ii) Con GRAFO,



Tramite questo tipo di rapp. possiamo utilizzare degli algoritmi già esistenti di matching su grafi diretti e ciclici per implementare l'UNIFICAZIONE.

Le variabili vengono indicate con un QUADRATINO [1] come segue



OPERAZIONI SU FEATURE STRUCTURE

Abbiamo due operazioni, la SUSSUNZIONE e l'UNIFICAZIONE. Queste non def. come segue

- F_1 SUSSUME F_2 se e solo se tutte le informazioni contenute in F_1 sono anche contenute in F_2 . Formalmente,

Se F_1 e F_2 non valori atomici, allora

$$F_1 \sqsubseteq F_2 \Leftrightarrow F_1 = F_2$$

Se $F_1 = \{ m_1: v_1, \dots, m_k: v_k \}$, e
 $F_2 = \{ m_1: v_2, \dots, m_s: v_s \}$, allora

$$F_1 \sqsubseteq F_2 \Leftrightarrow \forall (p, v) \in F_1: \exists (p, v') \in F_2: v \sqsubseteq v'$$

- L'UNIFICAZIONE di F_1 e F_2 è definita come $F_1 \sqcup F_2$ ed è la PIÙ PICCOLA feature structure che è sommanza sia da F_1 che da F_2 .

Se F_1 e F_2 hanno valori diversi nella stessa feature, allora $F_1 \sqcup F_2$ NON È DEFINITO.

In altre parole, l'operazione di unificazione è PARZIALMENTE DEFINITA.

Per il calcolo effettivo di $F_1 \cup F_2$ procediamo come segue

- 1) F contiene (p, σ) se p è in F_1 ma non in F_2 o viceversa;
- 2) F contiene (p, σ) se (p, σ_1) è in F_1 , (p, σ_2) è in F_2 , e σ_1 è unificabile e σ_2 .

Se nei σ_1 e σ_2 non sono ATOMICI, F_1 è unificabile e F_2 se $\sigma_1 = \sigma_2$

COMMENTI FINALI ~ (2:07:00 min)

Queste feature strutturali unite al concetto di unificazione ci permettono di mettere insieme le informazioni dei tre livelli di interpretazione del linguaggio: MORFOLOGIA, SINTASSI e SEMANTICO.

Esempi di questo si vedono nelle HPSG e LFG.

Le grammatiche HPSG e LFG non solamente COMPLESSE che quando noi vengono applicate hanno troppi VINCOLI e dunque modulare con pochi simboli.

La difficoltà è quindi capire quali non le REGOLE e allo stesso tempo capire quali non le ECCEZIONI alle regole e quando bisogna chiamare i simboli.

SELECTIONAL RESTRICTIONS: vengono utilizzati per selezionare l'informazione annotata da inserire nelle feature structure.