

Tecnologie del Linguaggio Naturale

Traccia: Traduttore transfer sintattico IT -> EN

Amedeo Raganati 928995

Angelo Pio Sansonetti 928869

Introduzione al problema

L'obiettivo del progetto è la realizzazione di un programma che traduca delle frasi dalla lingua italiana a quella inglese.

In particolare le frasi da tradurre sono:

- *È la spada laser di tuo padre*
- *Ha fatto una mossa leale*
- *Gli ultimi avanzzi della vecchia Repubblica sono stati spazzati via*

Il traduttore è composto dai seguenti sotto-moduli:

1. Parsing delle frasi italiane
2. Trasformazione dell'albero sintattico in un sentence plan
3. Generazione della traduzione

Parsing delle frasi italiane

Per il parsing è stato necessario creare una grammatica di tipo Context free.

Di seguito sono elencate le **regole grammaticali** da noi definite:

```
S    --> VP NP | NP VP NP | NP VPP
VPP  --> AUX V V
VP   --> AUX V
VP   --> V
V    --> V V-MOD
NP   --> DET N ADJ | DET ADJ N | ADJ N | DET N N | POS N | NP PP | N
PP   --> P NP
```

Di seguito sono elencate le **regole lessicali**:

N --> 'spada' | 'mossa' | 'avanzi' | 'padre' | 'Repubblica' | 'laser'
DET --> 'la' | 'una' | 'gli'
V --> 'è' | 'fatto' | 'spazzati' | 'stati'
AUX --> 'ha' | 'sono'
ADJ --> 'leale' | 'ultimi' | 'vecchia'
P --> 'della' | 'di'
V-MOD --> 'via'
POS --> 'tuo'

S rappresenta l'elemento 'frase' nella sua interezza.

VPP rappresenta il sintagma verbale quando il verbo è nella forma passiva, mentre *VP* rappresenta il sintagma verbale nella forma attiva.

AUX rappresenta un verbo ausiliare, mentre *V* rappresenta un verbo non ausiliare, al quale può essere associato un modificatore *V-MOD*.

NP rappresenta il sintagma nominale mentre *PP* rappresenta quello preposizionale.

DET rappresenta l'articolo, *N* il sostantivo e *ADJ* l'aggettivo.

Infine *P* rappresenta una preposizione e *POS* indica un aggettivo/pronome possessivo.

Risultati

È stata utilizzata la libreria **nltk** in Python per effettuare la generazione dell'albero a costituenti, data la grammatica sopra definita. Il parsing è stato realizzato utilizzando la classe [ChartParser](#), la quale sfrutta il principio della Dynamic programming per la generazione dell'albero sintattico.

Di seguito è illustrato l'output prodotto per le tre frasi in lingua italiana.

"È la spada laser di tuo padre" -->

```
(S
  (VP (V È)
    (NP
      (NP (DET la) (N spada) (N laser))
      (PP (P di) (NP (POS tuo) (N padre))))))
```

“Ha fatto una mossa leale” -->

(S
 (VP (AUX Ha) (V fatto))
 (NP (DET una) (N mossa) (ADJ leale)))

“Gli ultimi avanzi della vecchia Repubblica sono stati spazzati via” -->

(S
 (NP
 (NP (DET Gli) (ADJ ultimi) (N avanzi))
 (PP (P della) (NP (ADJ vecchia) (N Repubblica))))
 (VPP (AUX sono) (V stati) (V (V spazzati) (V-MOD via))))

Trasformazione dell'albero sintattico in un sentence plan

L'albero a costituenti prodotto nella fase di parsing viene elaborato in seguito da un programma implementato in Java.

Per la generazione delle frasi in inglese è stata utilizzata la libreria [Simple-NLG](#). Grazie a questa libreria è stato possibile definire il sentence plan partendo dall'albero sintattico: in questa maniera le realizzazioni morfologiche delle parole dovute anche alla sintassi sono state gestite automaticamente dalla libreria.

E' importante sottolineare che nelle frasi ove il soggetto è implicito, per determinare il soggetto (obbligatorio nell'inglese) si sono fatte delle assunzioni: si è specificato il soggetto *he* per quei verbi che implicano un agente umano – come ad esempio *fare* – mentre per quei verbi che non sono associabili ad un agente umano, il soggetto specificato è *it*.

Con la definizione del sintagma VPP è stato possibile stabilire quando il verbo è definito in forma passiva o attiva.

Infine, la presenza dell'ausiliare implica che il verbo principale sia coniugato al passato.

Di seguito è mostrato lo pseudocodice dell'algoritmo ricorsivo che, dato un albero a costituenti (espresso sotto forma di un oggetto json), provvede a generare una parte del sentence plan.

getNLGElement(AlberoCostituenti albero) restituisce un sentence plan

node_type = radice dell'albero

se *node_type = "NP":*

crea il sintagma nominale resultNP

firstNoun = true

per ogni *sotto_elemento dell'albero do:*

se *sotto_elemento = "DET":*

associa sotto_elemento come articolo di resultNP

se *sotto_elemento = "N" :*

se *firstNoun = true:*

associa sotto_elemento come sostantivo di resultNP

firstNoun = false;

altrimenti:

associa sotto_elemento come modificatore di resultNP

se *sotto_elemento = "NP":*

//sotto_elemento è a sua volta un albero

se *firstNoun = true:*

*associa **getNLGElement**(sotto_elemento) come
sostantivo di resultNP*

firstNoun = false;

altrimenti:

*associa **getNLGElement**(sotto_elemento) come
modificatore di resultNP*

se *sotto_elemento = "ADJ":*

associa sotto_elemento come aggettivo di resultNP

se sotto_elemento = "POS":

associa sotto_elemento come possessivo di resultNP

se sotto_elemento = "PP":

associa **getNLGElement**(sotto_elemento) come
modificatore di resultNP

restituisce il sintagma resultNP

se node_type = "PP":

crea sintagma preposizionale resultPP

per ogni sotto_elemento **do**:

se sotto_elemento = "P":

associa sotto_elemento come preposizione di resultPP

se sotto_elemento = "NP":

associa **getNLGElement**(sotto_elemento) come oggetto di
resultPP

restituisce il sintagma resultPP

se node_type = "VP":

crea il sintagma verbale resultVP

per ogni sotto_elemento **do**:

se sotto_elemento = "V":

associa **getNLGElement**(sotto_elemento) come verbo di
resultVP

se sotto_elemento = "AUX":

imposta il tempo verbale al passato

restituisce il sintagma resultVP

se node_type = "VPP":

crea il sintagma verbale passivo resultVPP

firstVerb = **true**;

per ogni sotto_elemento **do**:

se sotto_elemento = "V":

se firstVerb = **true**:

//Il primo verbo in una VPP viene scartato

firstVerb = **false**

altrimenti:

//Verbo principale

associa **getNLGElement**(sotto_elemento) come verbo
di resultVPP

imposta il tempo verbale al passato

restituisce il sintagma resultVPP

se node_type = "V":

crea il verbo resultVerb

se albero è una foglia:

associa il contenuto dell'albero come verbo di resultVerb

determina se il verbo fa riferimento ad un agente umano o meno

altrimenti:

per ogni sotto_elemento dell'albero **do**:

se sotto_elemento = "V":

associa **getNLGElement**(sotto_elemento) come
verbo di resultVerb

se sotto_elemento = "V-MOD":

associa sotto_elemento come modificatore di
resultVerb

restituisce il verbo resultVerb

Generazione della traduzione

Per effettuare la traduzione abbiamo creato una struttura a dizionario che tiene traccia della corrispondenza tra i termini nelle due lingue:

è = be
la = the
spada = sword
laser = laser
di = of
tuo = your
padre = father
lui = he
fatto = do
una = a
mossa = move
leale = fair
gli = the
ultimi = last
avanzi = leftover (al plurale)
della = of the
vecchia = old
Repubblica = Republic
spazzati = wipe
via = out

Di seguito sono presentate le frasi tradotte in inglese:

“È la spada laser di tuo padre” --> “It is the laser sword of your father.”

“Ha fatto una mossa leale” --> “He did a fair move.”

“Gli ultimiavanzi della vecchia Repubblica sono stati spazzati via” -->

“The last leftovers of the old Republic were wiped out.”