

## PARSING, DISAMBIGUAZIONE, IR E SOTTOCATEGORIZZAZIONE VERBALE

### Costruzione di grammatiche

- (1) Collegarsi all'indirizzo <http://www.ciscl.unisi.it/lab> (più un codice numerico comunicato durante l'esercitazione)
- (2) aprire la finestra per l'editing delle grammatiche (**edit grammar**)
- (3) inserire le seguenti regole rispettivamente sotto le colonne "rules" e "lexicon"

| Rules     | Lexicon    |
|-----------|------------|
| S > DP VP | D > il     |
| DP > D N  | D > lo     |
| VP > V DP | D > la     |
|           | D > i      |
|           | D > gli    |
|           | D > le     |
|           | D > uno    |
|           | D > un     |
|           | D > una    |
|           | D > l      |
|           | N > mela   |
|           | N > topo   |
|           | N > gatto  |
|           | N > uomo   |
|           | V > vede   |
|           | V > mangia |

- (4) scrivere, una per volta, le seguenti frasi (**step 1 – write a sentence**) e provare il parsing (**step 3 – parse it!**) con vari algoritmi (**step 2 - select a parsing algorithm**):
  1. l'uomo vede la mela
  2. l'uomo mangia la mela
  3. il gatto mangia il topo
  4. un gatto mangia un topo
  5. un topo mangia un uomo
  6. un topo mangia il uomo
  7. l'topo vede l'gatto
  8. il topo vede gatto
  9. la gatto vede la uomo
  10. ...
- (5) modificare il lessico in modo da catturare l'accordo
- (6) introdurre ambiguità nel lessico (D > la; pro > la... S > DP VP, S > VP) e provare diversi algoritmi di parsing
- (7) costruire una grammatica per parseggiare (in tutti i suoi sensi possibili) le classiche frasi:
  1. la vecchia legge la regola
  2. Gianni vede un uomo nel parco con il cannocchiale

### Dump database del lessico (nome del gatabase: "grammar")

```
CREATE DATABASE `grammar`;
USE grammar;
```

```
#
# Struttura della tabella `v`
#
```

```
CREATE TABLE `v` (
  `id` tinyint(4) NOT NULL auto_increment,
  `left` varchar(12) NOT NULL default "",
  `right` varchar(12) NOT NULL default "",
  PRIMARY KEY (`id`),
  KEY `left` (`left`,`right`)
) TYPE=MyISAM AUTO_INCREMENT=47 ;
```

```
#
# Dump dei dati per la tabella `v`
#
```

```
INSERT INTO `v` VALUES (46, 'PP', 'P DP');
INSERT INTO `v` VALUES (40, 'VP', 'V DP PP');
INSERT INTO `v` VALUES (37, 'DP', 'D N');
INSERT INTO `v` VALUES (34, 'VP', 'V DP');
INSERT INTO `v` VALUES (41, 'VP', 'cl V');
INSERT INTO `v` VALUES (32, 'S', 'DP VP');
INSERT INTO `v` VALUES (42, 'DP', 'D A N');
INSERT INTO `v` VALUES (43, 'DP', 'N');
INSERT INTO `v` VALUES (44, 'VP', 'aux V DP');
INSERT INTO `v` VALUES (45, 'DP', 'D N PP');
```

```
#
# Struttura della tabella `vt`
#
```

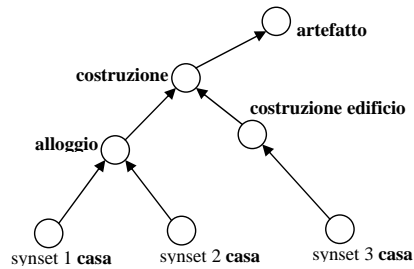
```
CREATE TABLE `vt` (
  `id` tinyint(4) NOT NULL auto_increment,
  `left` varchar(5) default NULL,
  `right` varchar(12) default NULL,
  PRIMARY KEY (`id`),
  KEY `left` (`left`),
  KEY `right` (`right`)
) TYPE=MyISAM PACK_KEYS=0 AUTO_INCREMENT=62 ;
```

```
#
# Dump dei dati per la tabella `vt`
#
```

```
INSERT INTO `vt` VALUES (4, 'D', 'il');
INSERT INTO `vt` VALUES (5, 'D', 'lo');
INSERT INTO `vt` VALUES (6, 'D', 'la');
INSERT INTO `vt` VALUES (7, 'D', 'i');
INSERT INTO `vt` VALUES (8, 'D', 'gli');
INSERT INTO `vt` VALUES (9, 'D', 'le');
INSERT INTO `vt` VALUES (10, 'P', 'di');
INSERT INTO `vt` VALUES (11, 'P', 'a');
INSERT INTO `vt` VALUES (12, 'P', 'da');
INSERT INTO `vt` VALUES (13, 'P', 'in');
INSERT INTO `vt` VALUES (14, 'P', 'con');
INSERT INTO `vt` VALUES (15, 'P', 'su');
INSERT INTO `vt` VALUES (16, 'P', 'per');
INSERT INTO `vt` VALUES (17, 'P', 'tra');
INSERT INTO `vt` VALUES (18, 'P', 'fra');
INSERT INTO `vt` VALUES (19, 'A', 'bella');
INSERT INTO `vt` VALUES (20, 'A', 'grande');
INSERT INTO `vt` VALUES (21, 'N', 'casa');
INSERT INTO `vt` VALUES (22, 'N', 'bambina');
INSERT INTO `vt` VALUES (23, 'N', 'gatto); ...
```

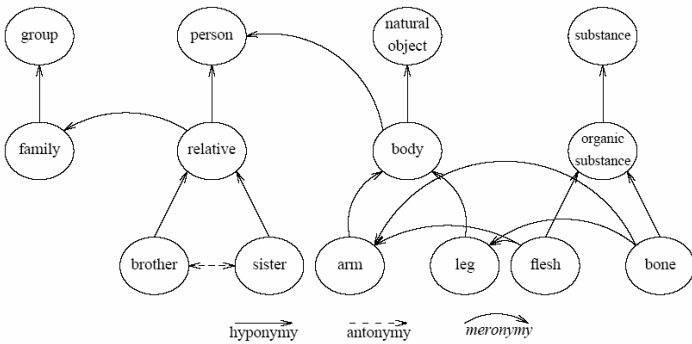
**Esplorazione di Wordnet**

- (8) Collegarsi al sito web di MultiWordNet (<http://tcc.itc.it/projects/multiwordnet/multiwordnet.php>)
- (9) Selezionare "Italian" dal menù delle lingue > scrivere "piatto" nella casella testuale > premere "search"
- (10) Visualizzare le definizioni di ogni synset premendo su ">"
- (11) Scegliere, in ordine, di visualizzare gli **iponimi**, gli **iperonimi**, i **termini coordinati** ecc. dal menù sotto "**Noun**"
- (12) Digitare la parola **casa**, visualizzare gli iperonimi e provare a disegnare un albero delle relazioni presenti (-> indica un iperonimo, quindi un nodo padre, => è un iponimo, cioè un nodo figlio):



- (13) Provare con **verbi, avverbi, aggettivi**

es. di visualizzazione delle relazioni di iponimia, antinomia, meronimia fornito da Miller:



**Disambiguazione, IR**

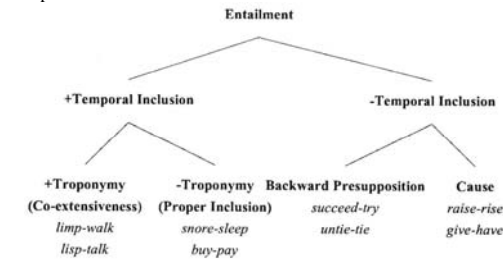
- (14) cercare alcune parole ambigue (es. cuore, aereo, seguire...) su **Wordnet**

- (15) una volta identificata almeno una coppia di sensi chiaramente distinguibili, provare a fare una ricerca con il termine ambiguo su **Google** e annotarsi i primi 5, 10 documenti trovati, classificandoli rispetto al diverso senso individuato
- (16) usare alcuni **iperonimi** (1 per volta, partendo da quello immediatamente precedente) in aggiunta al termine ambiguo e rilanciare la ricerca su **Google**
- (17) usare alcuni **iponimi** (1 per volta, partendo da quello immediatamente precedente) in aggiunta al termine ambiguo e rilanciare la ricerca su **Google**
- (18) usare **meronimi, olonimi, termini coordinati...** e registrare i risultati della ricerca in termini di sensi individuati

**Sottocategorizzazione verbale**

- (19) Classi verbali (eventi, azioni) semantiche in Wordnet (gli stati sono rappresentati in un'altra categoria):  
**cura e funzioni corporali – cambiamento – cognizione – comunicazione – competizione – consumo – contatto – creazione – emozione – movimento – percezione – possesso - interazione sociale – tempo**

Implicazioni lessicali:



- (20) verificare se le classi di Wordnet sono in qualche modo conformi alle classi di alternanza (es. Levin 93)
  - a. provare con verbi come **correre, cadere, baciare, chiedere, piacere, sorridere, pensare, mangiare, mettere...**
  - b. per ogni verbo selezionare prima l'**iperonimo**, quindi i **troponimi** dell'iperonimo e vedere se gli argomenti richiesti dai verbi individuati sono coerenti

**Riferimenti**

**PAB** si può liberamente utilizzare collegandosi al seguente indirizzo web:  
<http://www.ciscl.unisi.it/progetti/pab/>

**Wordnet** (Princeton University)  
<http://www.cogsci.princeton.edu/~wn/>

**MultiWordnet** (IRST, Trento)  
<http://tcc.itc.it/projects/multiwordnet/multiwordnet.php>