

Creazione e uso di *Treebanks* per lo studio di testi letterari dell'antichità classica

Federico Boschetti

federico_boschetti@yahoo.com

Università di Venezia

Treebanks su testi non letterari

- Analisi sistematica e coerente di tutte le unità testuali del *corpus*
- Possibilità di studi statistici
- Modelli sintattici adatti a descrivere un impiego ridotto di figure sintattiche (iperbato, ellissi, etc.)
- Assenza di varianti testuali
- Preferenza per le interpretazioni univoche
- Anonimato delle interpretazioni accolte

Studi sintattici su testi letterari

- Analisi condotta soltanto su luoghi notevoli
- Impossibilità di eseguire statistiche
- Studio approfondito delle figure sintattiche
- Interpretazione di eventuali varianti
- Molteplicità di interpretazioni
- Indicazioni bibliografiche delle interpretazioni proposte

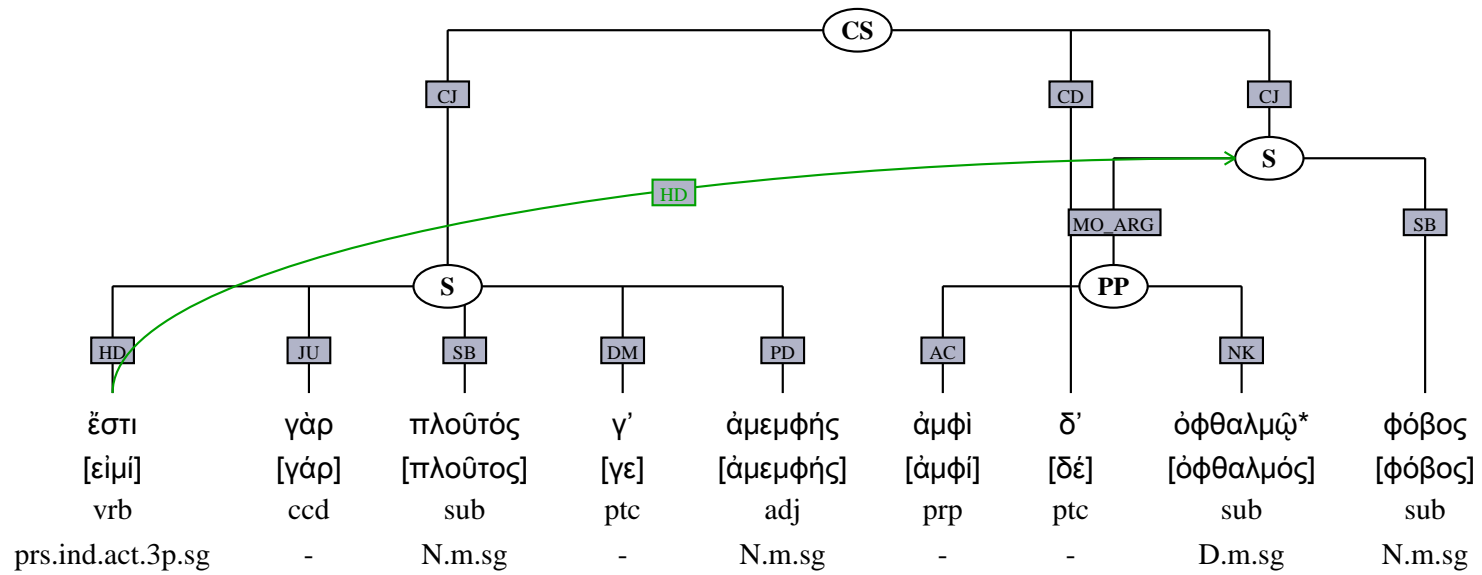
Trebanks per lo studio di testi letterari

- Analisi condotta in modo sistematico su tutte le unità testuali del *corpus*
- Possibilità di eseguire statistiche
- Modelli adeguati per rappresentare l'uso frequente di figure (iperbato, ellissi, etc.)
- Molteplicità di interpretazioni documentabili con indicazioni bibliografiche
- Interpretazione sintattica (o molteplici interpretazioni) di varianti testuali e congetture

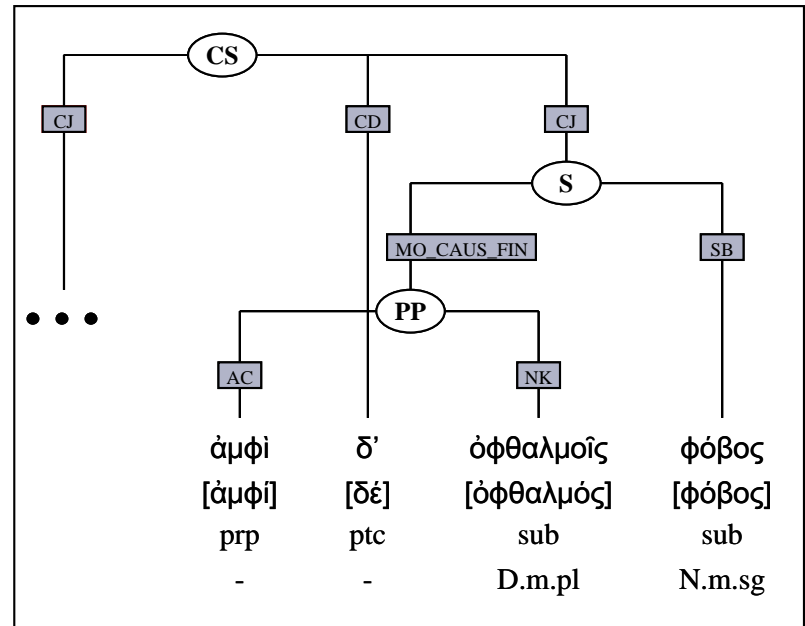
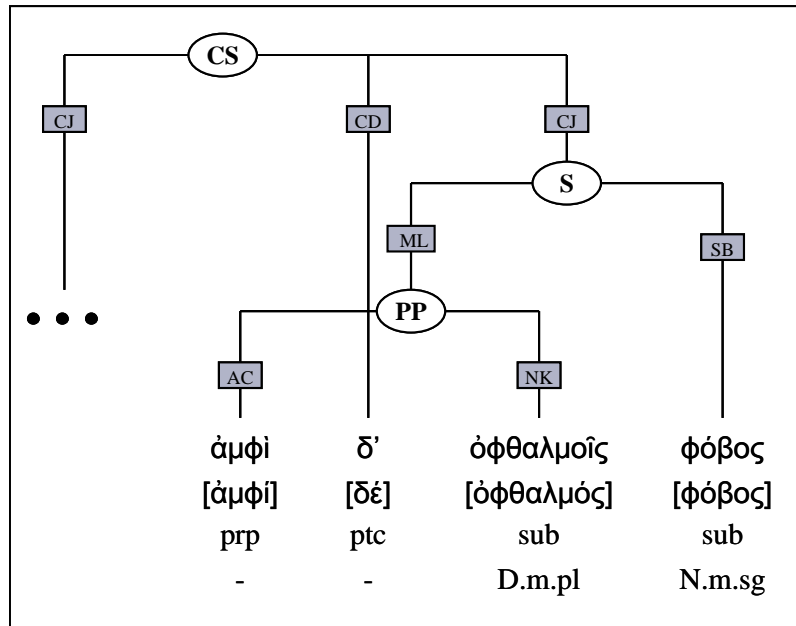
La scelta di TigerSearch

<http://www.ims.uni-stuttgart.de/projekte/TIGER/TIGERSearch>

- Implementazione in Java (garanzia di portabilità) e sviluppo di APIs *open source*
- Codifica Unicode e implementazione di *Input methods* per le lingue orientali e le lingue classiche
- Input/Output in formato XML (esportazione dei grafici in formato SVG)
- Associazione di ciascuna unità testuale (parola, sintagma, periodo) ad un identificatore univoco
- Possibilità di estendere i livelli di analisi previsti (lemmatizzazione, POS, tratti morfologici) con nuovi livelli di analisi (struttura metrica, sottocategorizzazione semantica, etc.)
- Possibilità di creare proprie etichette dei costituenti e delle funzioni sintattiche
- Modello di rappresentazione che permette *crossing edges* e *secondary edges*, molto adeguato per la rappresentazione di figure sintattiche
- linguaggio di interrogazione del *treebank* molto espressivo



Pers. 168 (editori moderni)



Due interpretazioni sintattiche di *Pers.* 168 (paradosi)

```

1 <s id="s646">
2 <graph root="x646_1">
3 <terminals>
4 <t id="x646" word="ἔσται" lemma="[ἐμί]" pos="vrb"
5 morph="prs.ind.act.3p.sg">
6 <secedge label="HD" idref="x646_3"/>
7 </t>
8 <t id="x647" word="γάρ" lemma="[γάρ]" pos="ccd" morph="-"/>
9 <t id="x648" word="πλοῦτος" lemma="[πλοῦτος]" pos="sub"
10 morph="N.m.sg"/>
11 <t id="x649" word="γ'" lemma="[γε]" pos="ptc" morph="-"/>
12 <t id="x650" word="ἀμεμφής" lemma="[ἀμεμφής]" pos="adj"
13 morph="N.m.sg"/>
14 <t id="x651" word="ἀμφί" lemma="[ἀμφί]" pos="prp" morph="-"/>
15 <t id="x652" word="δέ" lemma="[δέ]" pos="ptc" morph="-"/>
16 <t id="x653" word="ὀφθαλμοί" lemma="[ὀφθαλμός]" pos="sub"
17 morph="D.m.sg"/>
18 <t id="x654" word="φόβος" lemma="[φόβος]" pos="sub"
19 morph="N.m.sg"/>
20 </terminals>
21 <nonterminals>
22 <nt id="x646_1" cat="CS" >
23 <edge label="CJ" idref="x646_2"/>
24 <edge label="CD" idref="x652"/>
25 <edge label="CJ" idref="x646_3"/>
26 </nt>
27 <nt id="x646_2" cat="S" >
28 <edge label="HD" idref="x646"/>
29 <edge label="JU" idref="x647"/>
30 <edge label="SB" idref="x648"/>
31 <edge label="DM" idref="x649"/>
32 <edge label="PD" idref="x650"/>
33 </nt>
34 <nt id="x646_3" cat="S" >
35 <edge label="MO_ARG" idref="x646_4"/>
36 <edge label="SB" idref="x654"/>
37 </nt>
38 <nt id="x646_4" cat="PP" >
39 <edge label="AC" idref="x651"/>
40 <edge label="NK" idref="x653"/>
41 </nt>
42 </nonterminals>
43 </graph>
44 </s>
1 <sa id="s646_1" ref="s646">
2 <bibrefs>
3 <bibref>
4 <ptr xptr="id(x653)/@word">
5 <ptr xptr="id(x653)/@morph">
6 <bref target="heimsoeth:3" loc="72"/>
7 </bibref>
8 <bibref>
9 <ptr xptr="id(x646_3)/edge[1]/@label"/>
10 <bref target="matino" loc="71"/>
11 </bibref>
12 </bibrefs>
13 </sa>
1 <sa id="s646_2" ref="s646">
2 <bibrefs>
3 <bibref>
4 <ptr xptr="id(x653:2-3)/@word"/>
5 <ptr xptr="id(x653:2-3)/@morph"/>
6 <bref type="c" target="codd. praeter Q"/>
7 </bibref>
8 <bibref>
9 <ptr xptr="id(s646_2)//edge/@label"/>
10 <bref target="matino" loc="71"/>
11 <bref target="lalin" loc="20"/>
12 </bibref>
13 </bibrefs>
14 <graph>
15 <terminals>
16 <t id="x653:2-3" ref="x653" word="ὀφθαλμοί" morph="D.m.pl"/>
17 </terminals>
18 <nonterminals>
19 <nt><edge ref="id(x646_3)/edge[1]" label="ML"/></nt>
20 </nonterminals>
21 </graph>
22 </sa>
1 <sa id="s646_3" ref="s646_2">
2 <bibref>
3 <ptr xptr="id(s646_3)//edge/@label"/>
4 <bref target="matino" loc="71"/>
5 <bref target="bile" loc="101"/>
6 </bibref>
7 <graph>
8 <nonterminals>
9 <nt><edge ref="id(x646_3)/edge[1]" label="MO_CAUS_FIN"/></nt>
10 </nonterminals>
11 </graph>
12 </sa>

```