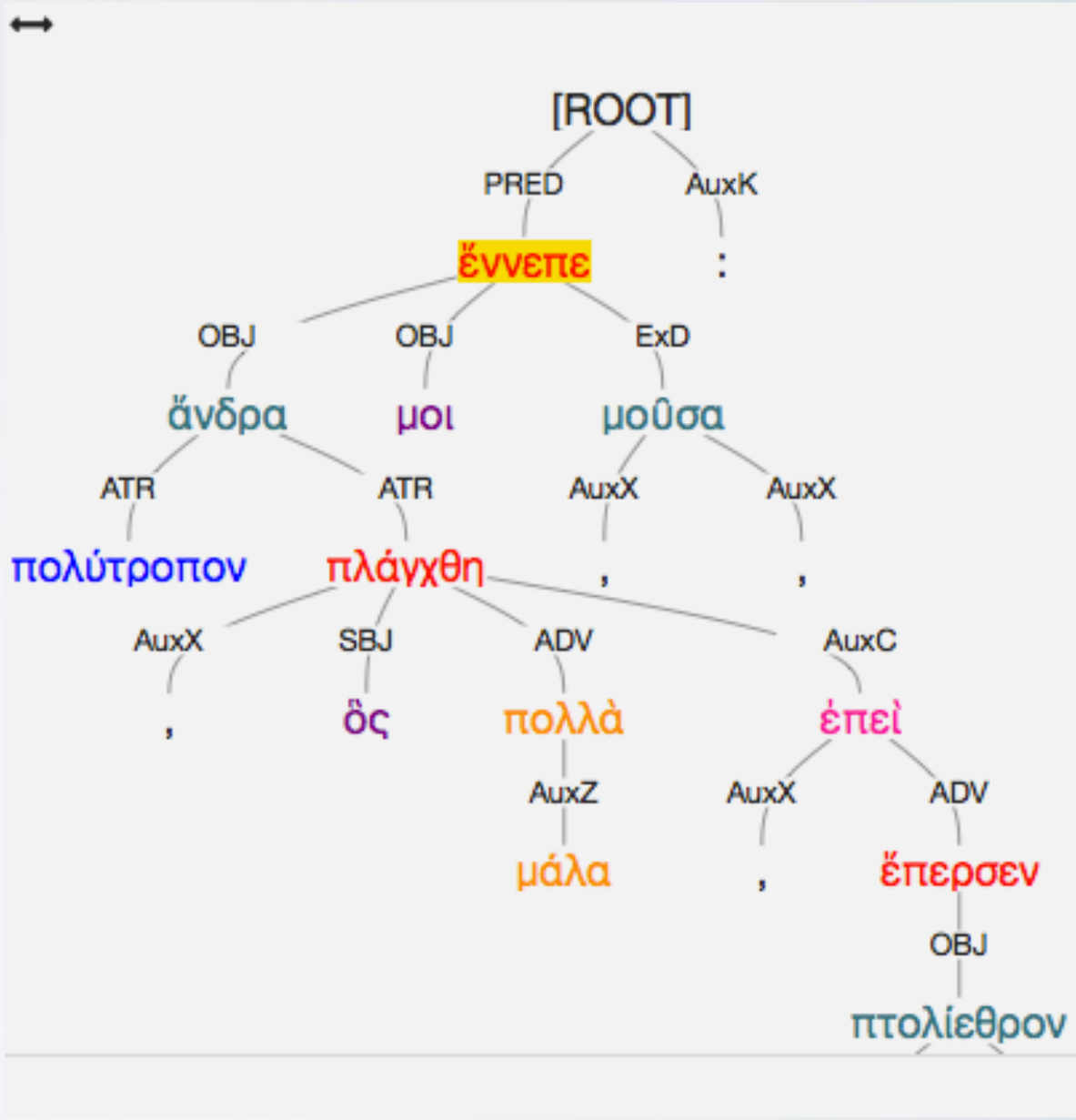
QUERYING THE AGLDT

Francesco Mambrini
Sunokisis-DC-2017-2018

AGLDT

- Started and maintained by the Perseus Project (Website)
- Greek: more 750k annotated words, including:
  - lots of poetry (Homer, tragedy)
  - some prose (bk 1 of Hdt. and Thuc.)
- Latin: not very developed…



Morphology

```
<sentence subdoc="4-5" id="2386547" document_id="urn:cts:greekLit:tlg0011.tlg004.pe
  <annotator>FrancescoM</annotator>
  <word id="1" form="πόλις" lemma="πόλις" postag="n-s---fn-" head="6" relation="SBJ
  <word id="2" form="θ'" lemma="θέ" postag="g--------" head="6" relation="AuxY" cit
  <word id="3" form="ὁμοῦ" lemma="ὁμοῦ" postag="d--------" head="4" relation="AuxZ"
  <word id="4" form="μὲν" lemma="μέν" postag="g--------" head="9" relation="AuxY" c
  <word id="5" form="θυμιαμάτων" lemma="θυμίαμα" postag="n-p---ng-" head="9" relati
  <word id="6" form="γέμει" lemma="γέμω" postag="v3spia---" head="0" relation="PRED
  <word id="7" form="," lemma="," postag="u--------" head="9" relation="AuxX" cite=
  <word id="8" form="ὁμοῦ" lemma="ὁμοῦ" postag="d--------" head="9" relation="AuxZ"
  <word id="9" form="δὲ" lemma="δέ" postag="g--------" head="6" relation="COORD" ci
  <word id="10" form="παιάνων" lemma="Παιάν" postag="n-p---mg-" head="12" relation=
  <word id="11" form="τε" lemma="τε" postag="g--------" head="12" relation="AuxY" c
  <word id="12" form="καὶ" lemma="καί" postag="c--------" head="9" relation="COORD"
  <word id="13" form="στεναγμάτων" lemma="στέναγμα" postag="n-p---ng-" head="12" re
  <word id="14" form="." lemma="." postag="u--------" head="0" relation="AuxK" cite
</sentence>
```

TEXT

Syntax

WHAT ABOUT QUERYING?

- No software that work natively with the AGDT format
- Some "3rd-party" tools are linked on the website:
  - Structural Search (by N. Kallen)
- or… you can **hack it yourself**, by:
  - querying the AGDT XML files (refer to G. Celano's tutorial on using XQUERY)
  - using a custom DB: e.g. the gAGDT

THE REQUIREMENTS

- **Intuitive!** i.e. easy to learn

- **Powerful!**
  - it must be able to support all the stuff that INESS does (regexp, constituent order, complex dependency chains…)

- **in a word**: users must have no problem translating their research questions into the formal language for querying!

THE REQUIREMENTS 2

- must be fast!
- support output as:
  - visualized trees
  - tabular form and/or statistics
- synergize well with the digital editions of the texts
- be extensible (new texts and annotation types)

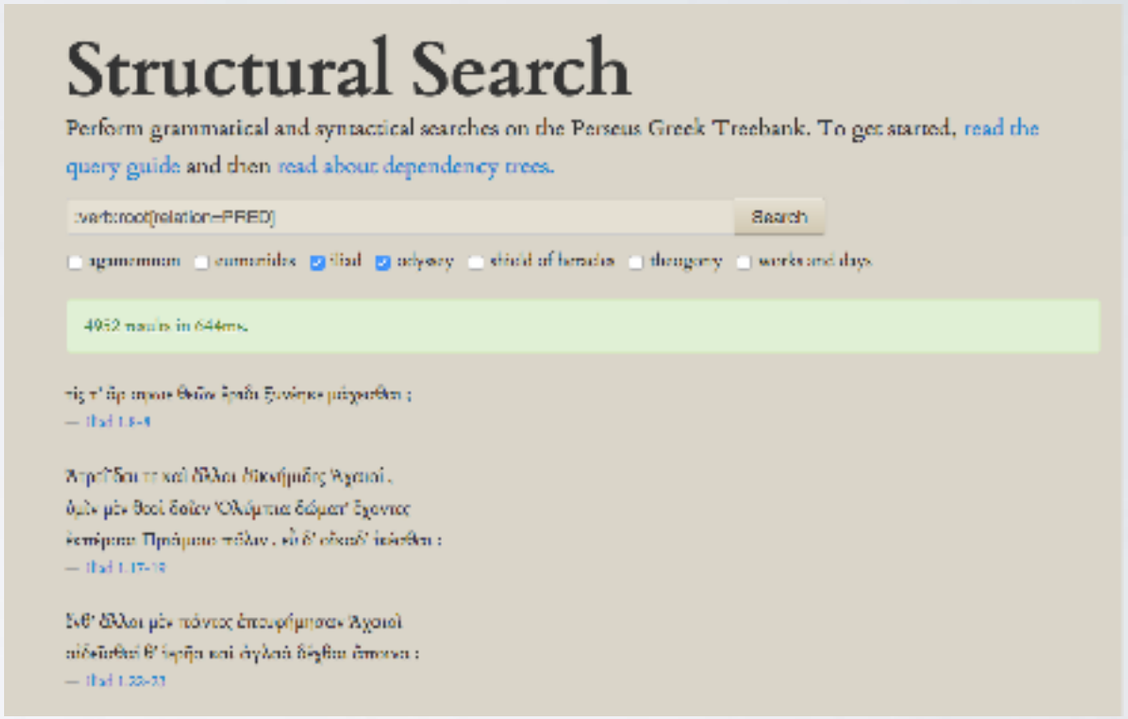
```
<milestone n="933" unit="card"/>
<div type="textpart" subtype="strophe" n="4">
  <sp>
    <speaker>Χορός</speaker>
    <l n="933">ὁμόσποροι δῆτα καὶ πανώλεθροι,</l>
    <l n="934">διατομαῖς οὐ φίλοις,</l>
    <l n="935">ἔριδι μαινομένᾳ,</l>
    <l n="936">νείκεος ἐν τελευτᾷ.</l>
  </sp>
</div>
```

I'd like to inspect all the "genitive absolutes" or accusative + inf. in Homer

I'd like to know what are the most frequent compounds of "*syn*" used by each characters in Sophocles

STRUCTURAL SEARCH

A simple but very powerful solution to query (some of) the AGDT, based on CSS3 selector syntax



http://www.iliados.com/

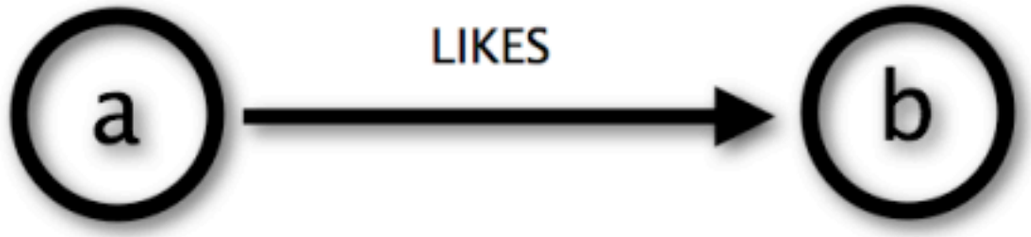
gAGDT

A graph database (based on Neo4j) that includes the AGDT data in a graph structure. Still experimental!

https://github.com/francescomambrini/gAGDT

- it's fast and scalable
- it allows powerful semantic queries
- extensible at will
    - add new types of nodes and relations

Cypher using relationship 'likes'



Cypher

```
(a) -[:LIKES]-> (b)
```

CYPHER

a query language inspired by SQL and based on ASCII art… ;-)

(node {property: “value”})

(head)-[:Dependecy]->(dep)