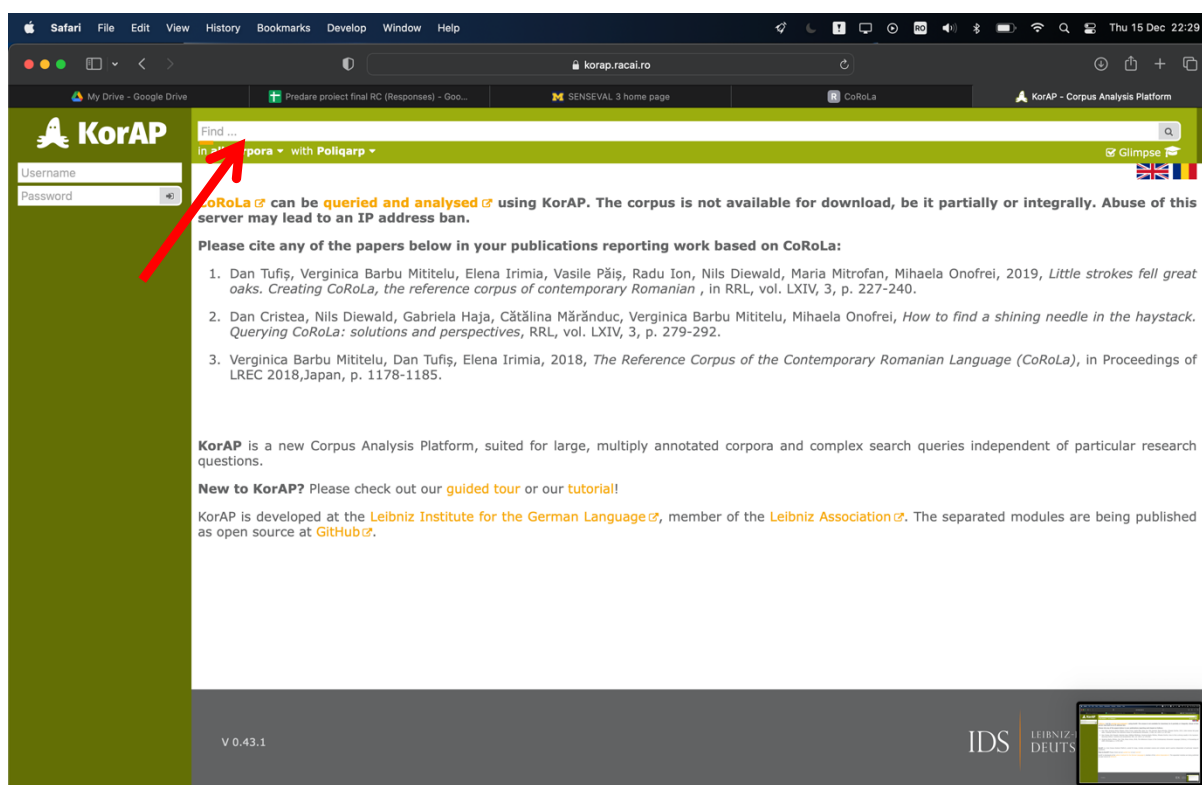


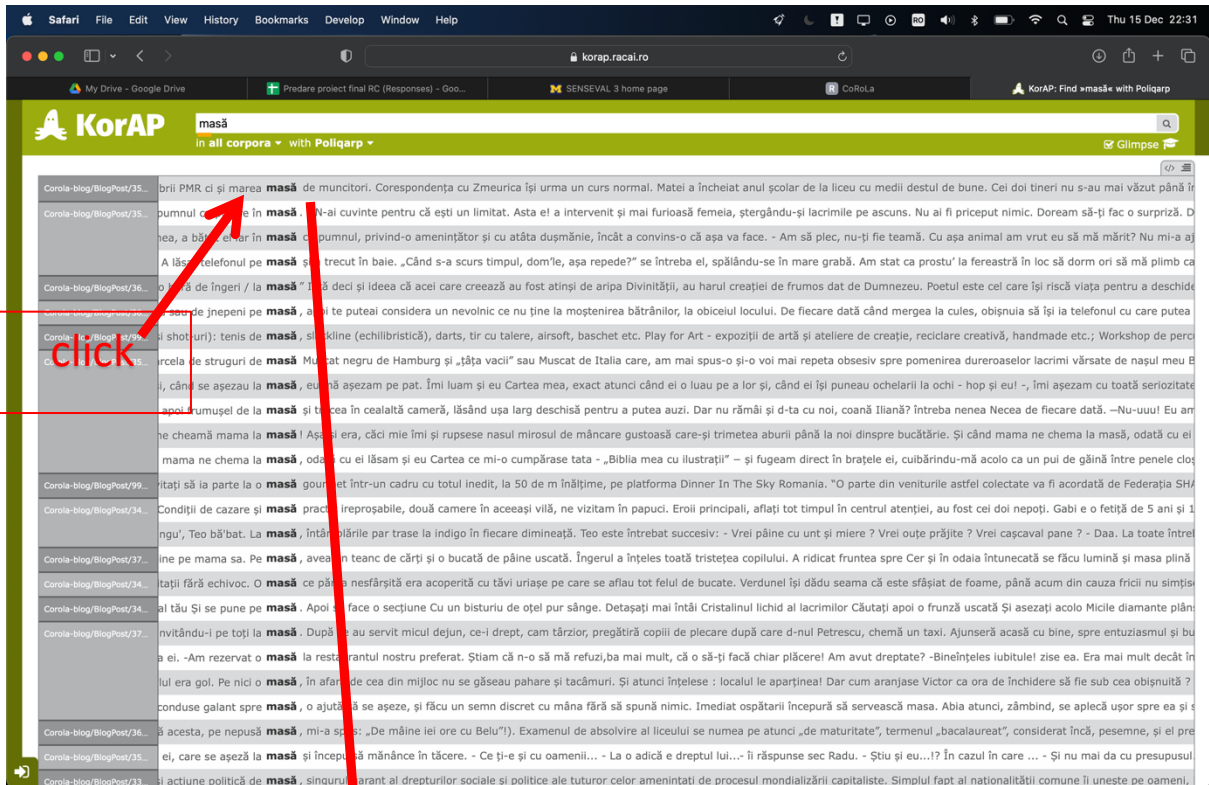
Manual adnotare

Plecând de la lista de cuvinte care au mai multe sensuri, căutam în corpusul CoRoLa aceste intrări:

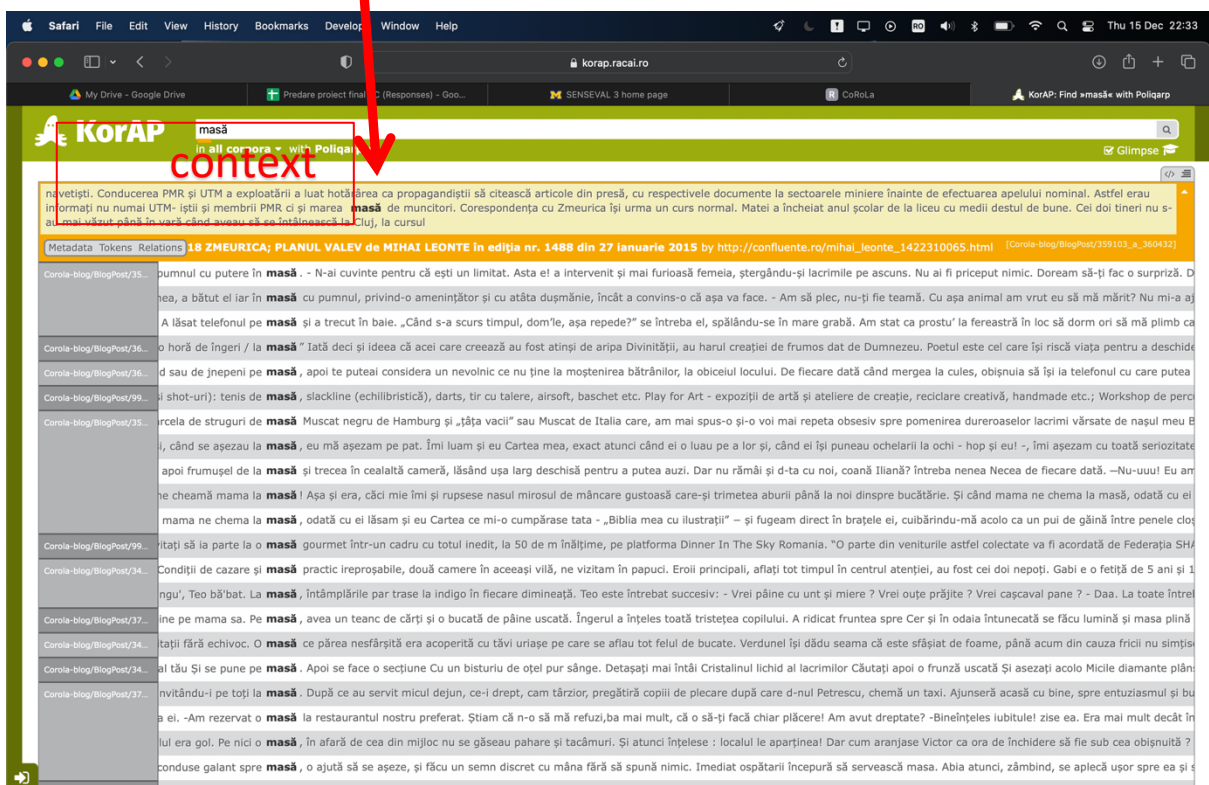
- vom utiliza interfața KorAP (<https://korap.racai.ro/>)



- căutare ne va genera o lista cu toate contextele unde apare cuvântul căutat:



- în momentul în care se face click pe context se deschide o fereastră cu tot contextul:

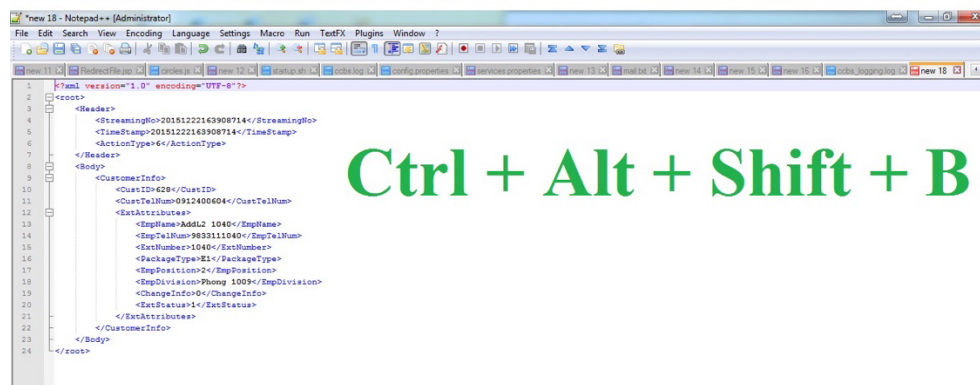
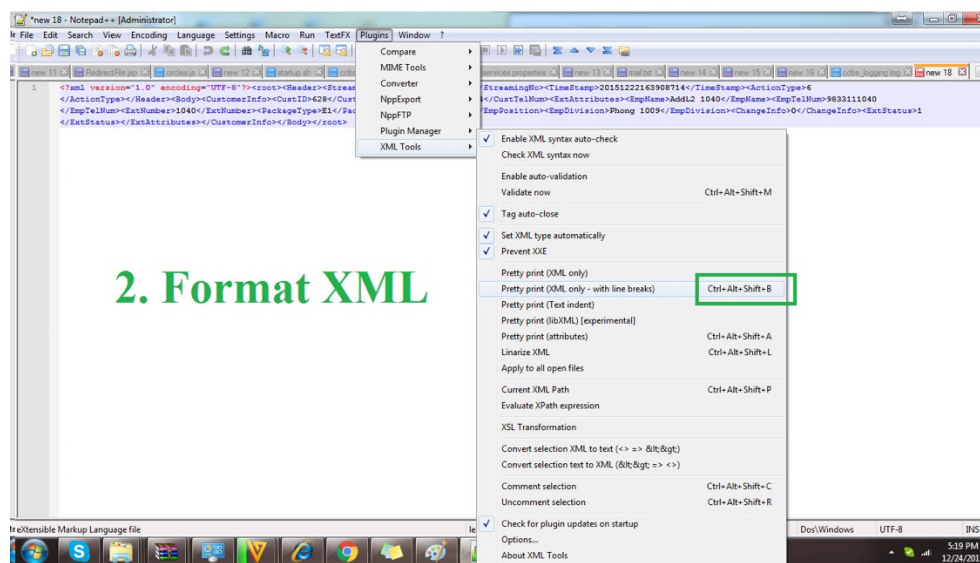
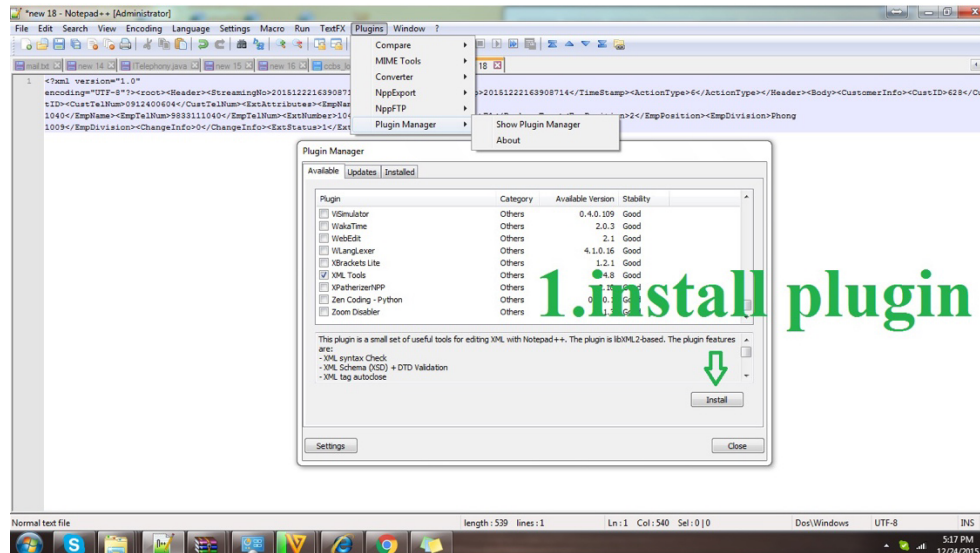


- acest context se copiază în fișierul unde adnotăm cuvântul (în exemplul de față: masă)

Pentru adnotare este recomandat să se instaleze utilitarul Notepad++
(<https://notepad-plus-plus.org/downloads/>) și adițional XML Tools pentru
formatare

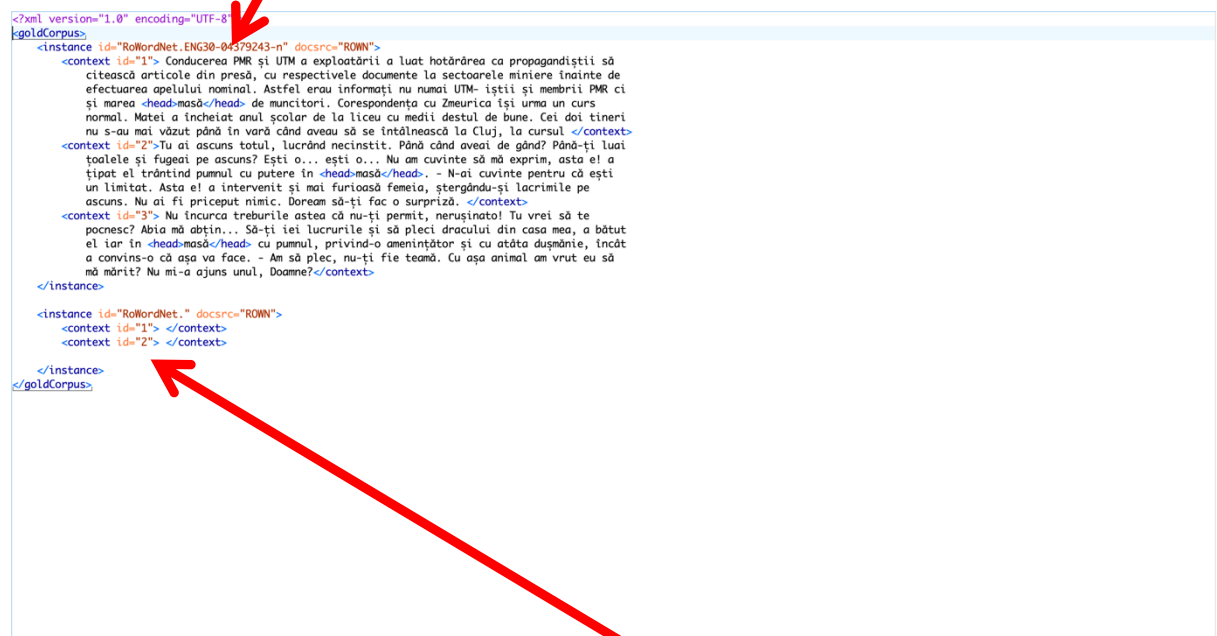
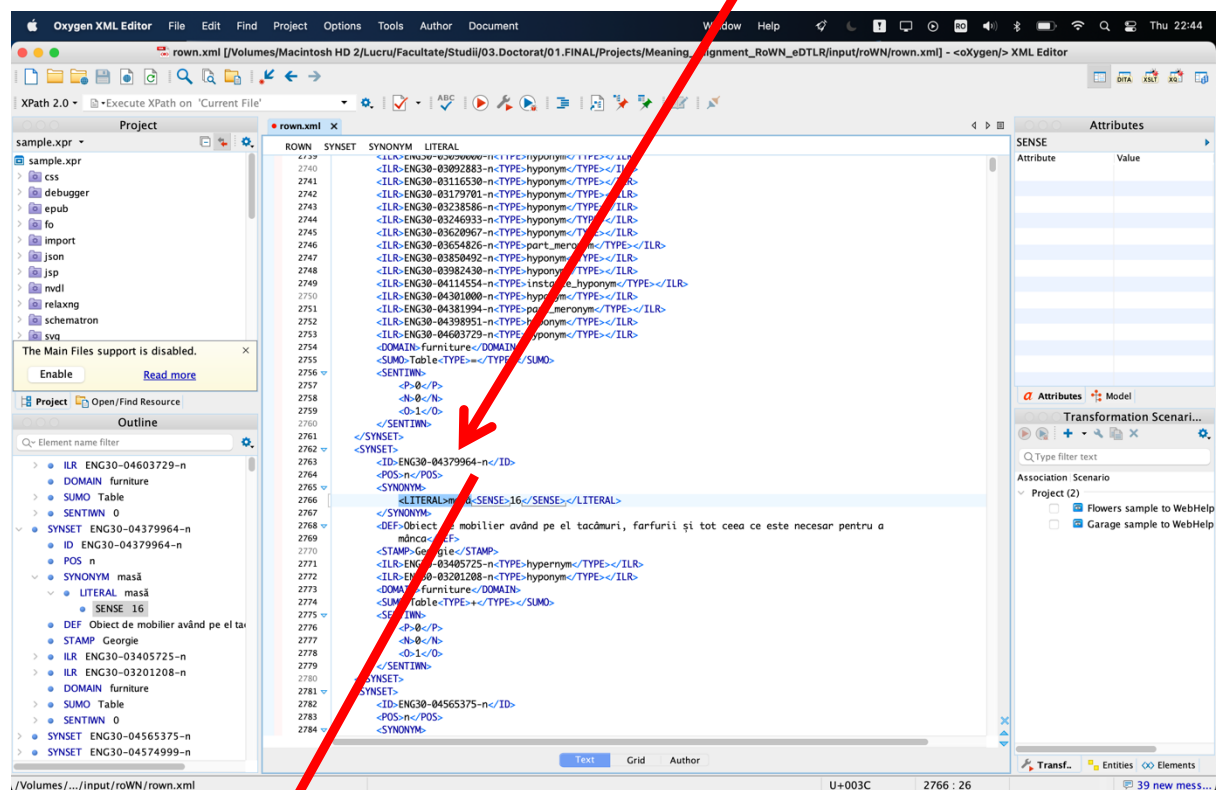
Step 1: Install XML Tools plugin

Step 2: Format ...completed



După instalarea acestui utilitar adnotarea se va face după cum urmează:

- deschidem fișierul corpusGold.xml
- în tagul instance – vom introduce la id -> id-ul care se găsește în RoWordNet:



- contextul îl vom prelua din interfața KorAP (descrișă mai sus)
- fiecare context va avea un identificator unic (id="n")
- dacă există mai multe contexte vom adăuga taguri de context de forma:
`<context id="n"> text text text </context>`

- pentru a adăuga o nouă instanță (adică un cuvânt nou) vom folosi tagul de forma:

```
<instance id="RoWordNet.id" docsrc="ROWN">
```

în interiorul instanțelor vom adăuga tagurile de forma

```
<context id="..."> .... </context>
```

```
</instance>
```