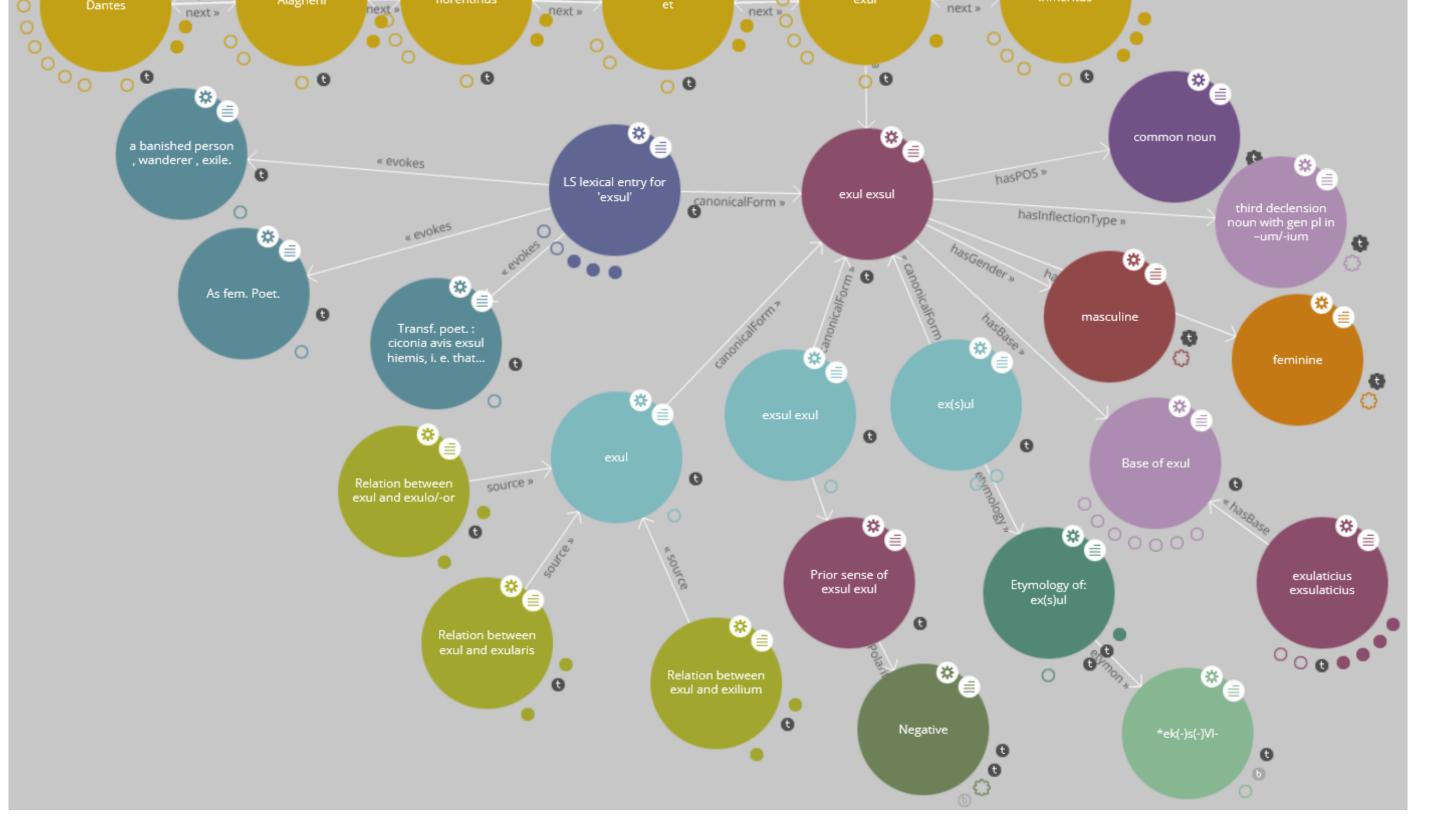
LINKING THE LEWIS & SHORT DICTIONARY TO THE LILA KNOWLEDGE BASE OF INTEROPERABLE RESOURCES FOR LATIN F. Mambrini, E. Litta, M. Passarotti and P. Ruffolo CIRCSE Research Centre, Università Cattolica del Sacro Cuore, Milan

LiLa	Lewis & Short
 The LiLa KB is a collection of interlinked textual and lexical resources for Latin described with the same vocabulary of knowledge description that uses common data categories and ontologies Interoperability is attained by linking all the entries in lexical resources and tokens in corpora that point to the same lemma to the corresponding lemmas in the Lemma Bank The Lemma Bank is a large collection of Latin lemmas, currently comprising of almost 200,000 canonical forms (approx. 130,000 lexical items) 	 The Latin Dictionary, curated by Ch. T. Lewis and Ch. Short (1879) Encoded in TEI xml by the Perseus Project and available on a series of desktop and web applications XML structure : 1. Entry: the headword, encoded within the TEI element <entryfree>. 51,596 in total</entryfree> 2. Information about inflection 3. Etymological or derivational information, encoded within the same element <etym></etym>
Image: Second secon	4. Sense(s): containers where the meaning of the word is matched with citations from Classical Latin sources (e.g. "Cic. Sen. 8, 26").



5. "sub-entries", words that are not given a record of their own, but are discussed within another entry, encoded within the **<sense>** element and followed by the same type of information structured as the main entries

<entryFree id="n23298" type="main" key="infrequens" opt="n">
 <orth extent="full" lang="la" opt="n">in-frĕquens</orth>,
 <itype opt="n">tis</itype>,
 <pos opt="n">adj.</pos>,
 <sense id="n23298.0" n="I" level="1" opt="n">
 <hi rend="ital">that does not often</hi> or <hi rend="ital">frequently take place</hi>,
 <hi rend="ital">that does not often</hi> or <hi rend="ital">frequently take place</hi>,
 </sense>
 <sense id="n23298.1" n="I" level="1" opt="n">

 <hi rend="ital">that does not often</hi> or <hi rend="ital">frequently take place</hi>,
 <hi rend="ital">that does not often</hi> or <hi rend="ital">frequently take place</hi>,
 <hi rend="ital">that does not often</hi>

 <li

Linking the L&S entries

1. Spelling normalisation

2. Mapping of part-of-speech and inflectional information

Modelling

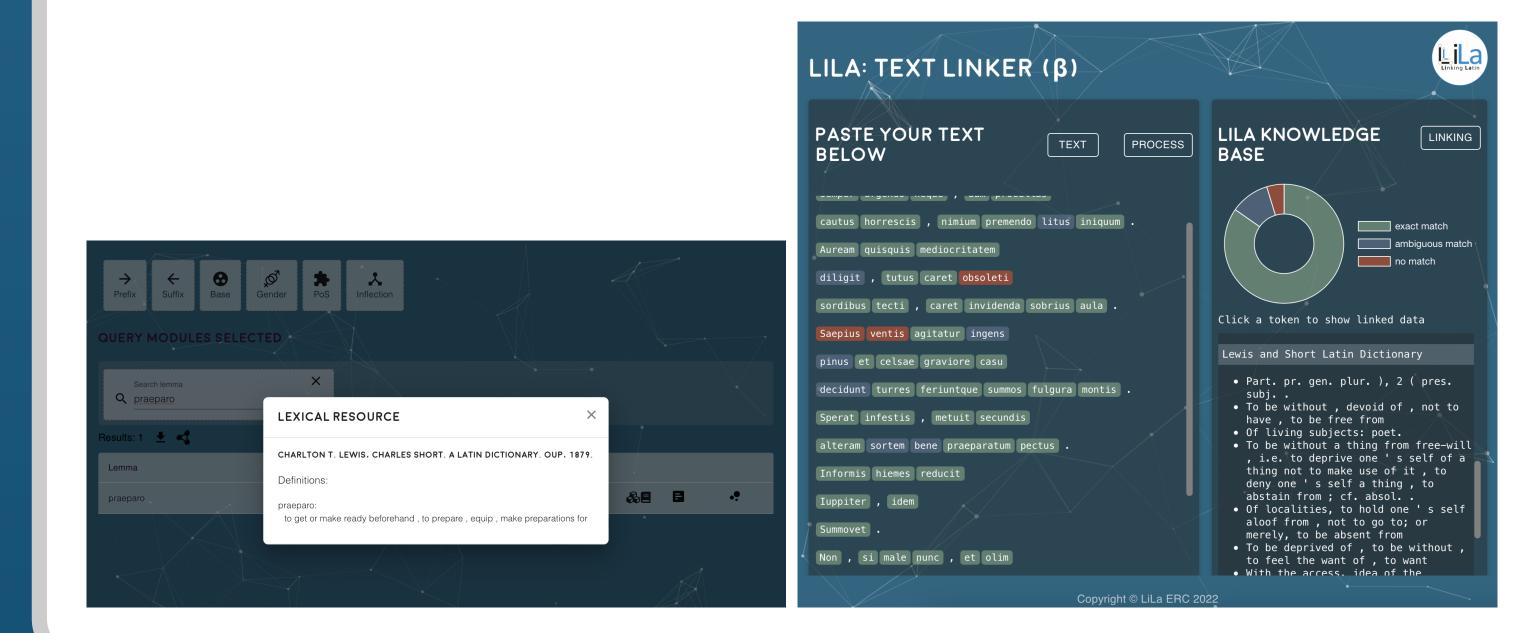
• We use the "OntoLex lexicography module" or **lexicog**, to capture the structural information expressed in a lexicographic resource

3.31,142 1:1 matches, 2,998 1:N matches and 4,553 1:0 matches

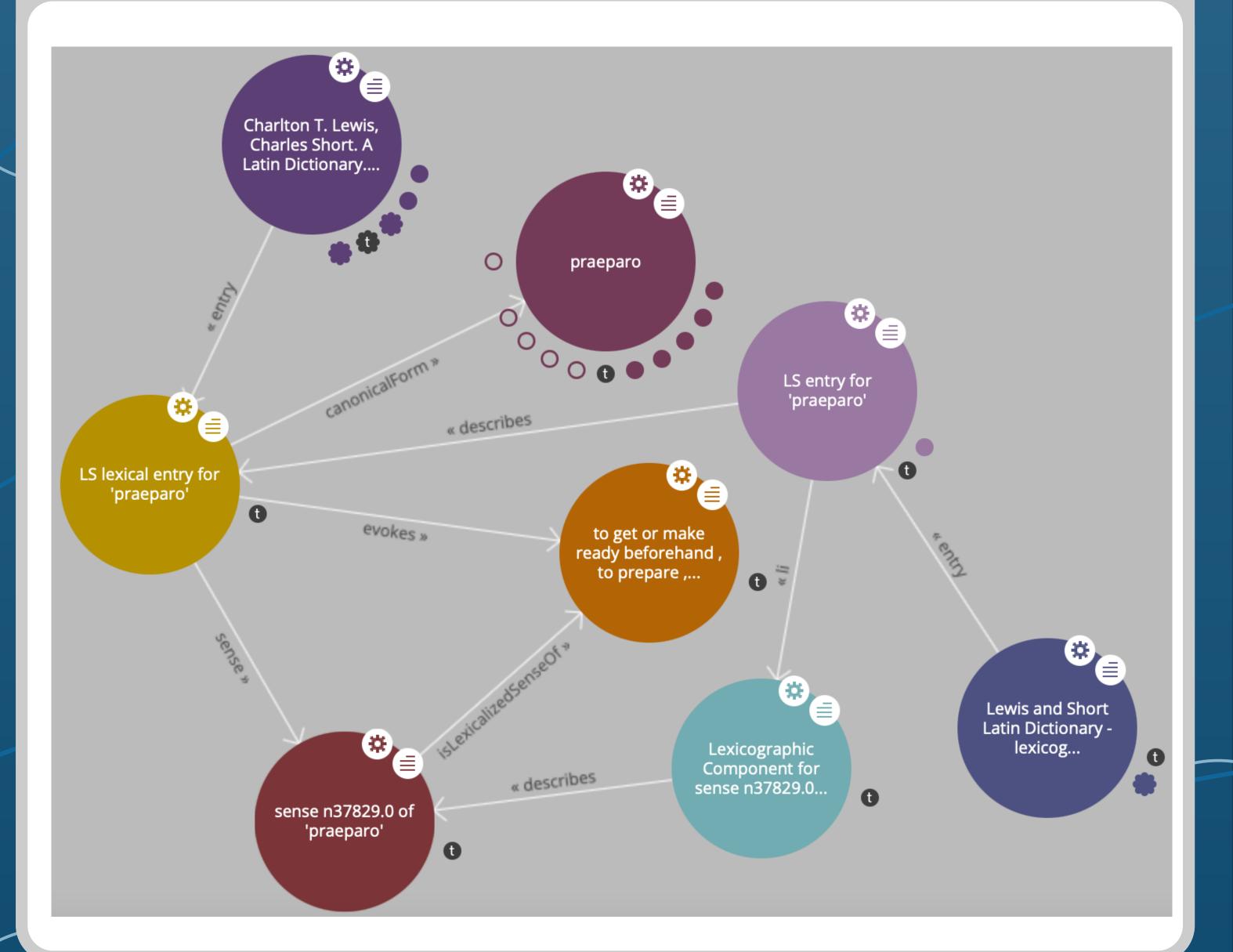
- 4. 1,289 unmatched entries successfully analysed by the morphological analyser Lemlat5. Multi-word spellings and graphical variants
- 6.1:N leftover matches manually disambiguated and matched to the relevant lemmas
- 7. Definitely unmatched entries have been added to the LiLa Lemma Bank
- The most important structural element in lexicog is the Lexicographic Entry
- Lexicographic entries are a special subset of Lexicographic Components. Components can also be used to represent senses, sense groups or subentries
- Components are structural units; all linguistic information must be expressed using Ontolex. The property lexicog:describes provides a link between the two dimensions: a lexicographic entry can be said to *describe* a lexical entry

Services and use cases

- L&S can be accessed from the main lemma query interface access point, together with the other lexical resources currently linked to LiLa: https://lila-erc.eu/query/
- L&S can be used when linking textual material to LiLa during the lemmatisation process for quick check in disambiguation processes



Modeling: praeparo



₽́Т_ЕХ TikZ**poster**

erc This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, Grant Agreement No. 769994