**ALBANESI — Translation**

<3 figures>

**Translation, Annotation, and Knowledge Modelling of the Babylonian Talmud: The *Traduco* System**

Albanesi, D., Bellandi, A., Benotto, G. and Giovannetti, E.

In this work we are going to present the *Traduco* System, a collaborative web-based application for the translation of the Babylonian Talmud (BT) into Italian. The system has been designed around a computer-assisted translation (CAT) component (Gordon, 1996; Huerta, 2011), constituting its core. However, *Traduco* is not limited to assisting the translation process and providing printing functionalities. In fact, it allows linguistic and semantic annotations and advanced searches, paving the way to the construction of a Talmudic knowledge base. In order to achieve these results, the *Traduco* development process abided by a model that took into account aspects of Natural Language Processing and Knowledge Engineering. The component-based architectural structure was implemented using the object-oriented Java 2 Enterprise Edition framework.

In particular, each *Traduco* component implements specific functionalities targeted at different types of users:

1. **Translators** and **revisors** are supported by the use of CAT technologies, including a Translation Memory (TM) designed to ‘remember’ every translated portion of text. The system takes as input the Hebrew text segment to be translated, queries the TM, and suggests the Italian translations relative to the Hebrew text segments recognized to be more similar to the one that has to be translated (Bellandi et al., 2014a), as illustrated in Figure 1.

2. **Philologists** and **linguists** can insert notes, comments, and bibliographical references (HaCohen-Kerner, 2010).

3. **Domain experts** are allowed to structure relevant terms into glossaries (for example, proper names, plants, measures, concepts) and, potentially, domain ontologies (Guarino, 1998) represented in Simple Knowledge Organization System (SKOS) or Ontology Web Language (OWL), by using a graphical ontology editor (see Figure 2), which is currently under development (Bellandi et al., 2014b).

4. **Researchers** and **scholars** can carry out complex searches on both a linguistic and semantic basis. In more detail, we are developing two morphological analysers, one for Italian and one for Mishnaic Hebrew. These instruments should allow the creation of lexical indices, where each entry (‘lemma’ or ‘root’) will be associated to all its morphologically correlated inflected forms and to all the contexts in which they occur.

5. **Editors** can easily produce the printed edition of the translation of the BT by arranging translations and notes in standard formats for desktop publishing software (typically XML-based, see Figure 3).

Needing a common and shared platform for the translation, revision, and editing of the BT, we went to the Web and the related technologies, providing users the ability to work on the same data collaboratively. Furthermore, it allows the supervisors to keep track, in real time, of the work done by the translators on the portions of the Talmud they were assigned to.

*Traduco* can be used for translating other texts by adapting its linguistic analysis and semantic annotation components to different languages and domains with relative ease. In particular, the approach employed by *Traduco* for the processing of Italian and Hebrew is based on extensively tested and state-of-the-art machine learning technologies (see Bar-Haim et al., 2005; Itai, 2006), built on highly flexible supervised models that can be trained using pre-annotated texts. Concerning the knowledge engineering components, the model allows for the definition of arbitrary semantic classes, thus enabling users to construct specific domain ontologies starting from the annotated terms.

**Acknowledgement**

This work has been conducted in the context of the research project TALMUD and the scientific partnership between S.c.a r.l. ‘Progetto Traduzione del Talmud Babilonese’ and ILC-CNR, and on the basis of the regulations stated in the ‘Protocollo d’Intesa’ (memorandum of understanding) between the Italian Presidency of the Council of Ministers; the Italian Ministry of Education, Universities, and Research; the Union of Italian Jewish Communities; the Italian Rabbinical College; and the Italian National Research Council (21 January 2011).

**References**

**Bar-Haim, R., Sima’an, K. and Winter, Y.** (2005). Choosing an Optimal Architecture for Segmentation and POS-Tagging of Modern Hebrew. *Proceedings of the Association for Computational Linguistics Workshop on Computational Approaches to Semitic Languages*, Michigan, June 2005.

**Bellandi, A., Bellusci, A. and Giovannetti, E.** (2014a). Computer Assisted Translation of Ancient Texts: The Babylonian Talmud Case Study. *Proceedings of the 11th International Workshop on Natural Language Processing and Cognitive Science*, Venice.

**Bellandi, A., Bellusci, A., Giovannetti, E. and Carniani, E.** (2014b). Content Elicitation: Towards a New Paradigm for the Analysis and Interpretation of Text. *Proceedings of the 13th International Conference on Informatics*, Innsbruck, February 2014.

**Gordon, I.** (1996). Letting the CAT out of the Bag—or Was It MT? *Proceedings of the 8th International Conference on Translating and the Computer*, London.

**Guarino, N.** (1998). Formal Ontology in Information Systems. *Proceedings of the First International Conference (FIOS’98)*, Vol. 46. June. Trento: IOS Press.

**HaCohen-Kerner, Y., Schweitzer, N. and Shoham, Y.** (2010). Automatic Identification of Biblical Quotations in Hebrew-Aramaic Documents. *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval*, Valencia, October 2010.

**Huerta, J. M.** (2011). Towards Efficient Translation Memory Search Based on Multiple Sentence Signatures, Speech and Language Technologies. DOI:10.5772/16327.

**Itai, A.** (2006). Knowledge Center for Processing Hebrew. *Proceedings of the 5th International Conference on Language Resources and Evaluation, Workshop ‘Towards a Research Infrastructure for Language Resources*. Genoa, May 2006.

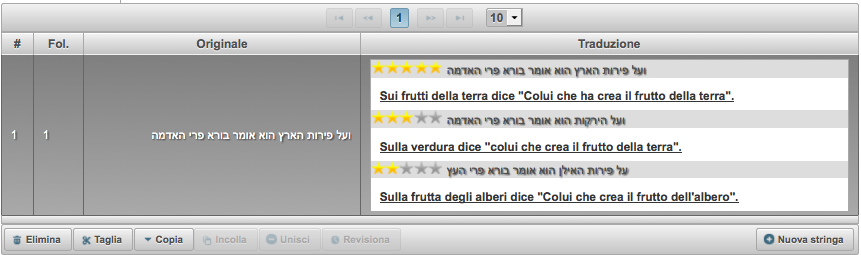


Figure 1. Translation Suggestion Component (five stars indicate an exact match). The translations in English, starting from the first, are: (i) *On the fruits of the earth he says, ‘He who owns, creates the fruit of the earth’*; (ii) *On vegetables he says, ‘He who creates the fruit of the earth’*; (iii) *On the fruits of the trees he says, ‘He who creates the fruit of the tree’*.

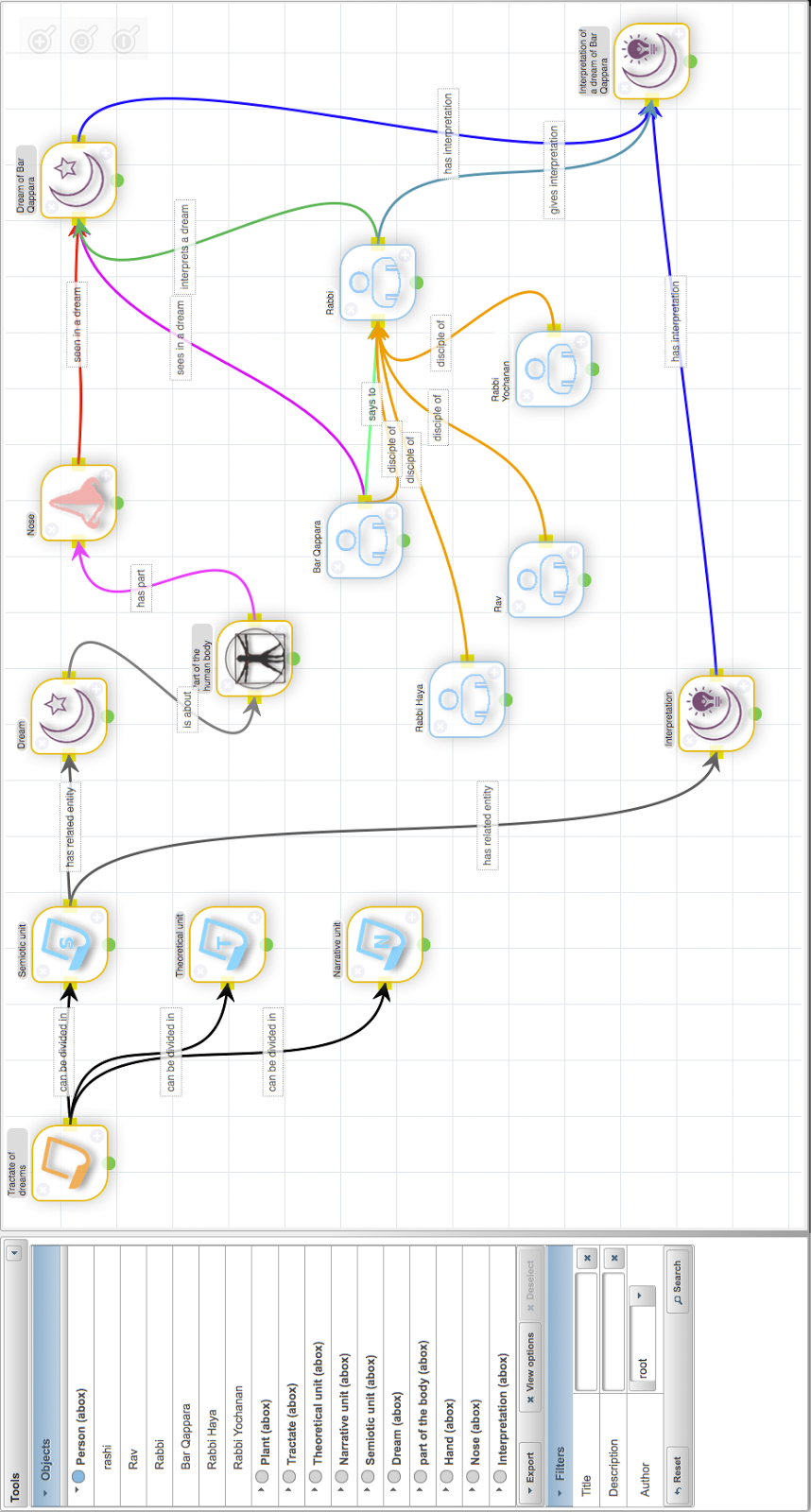


Figure 2. GUI for building the Talmudic knowledge base starting from domain terms (under development).

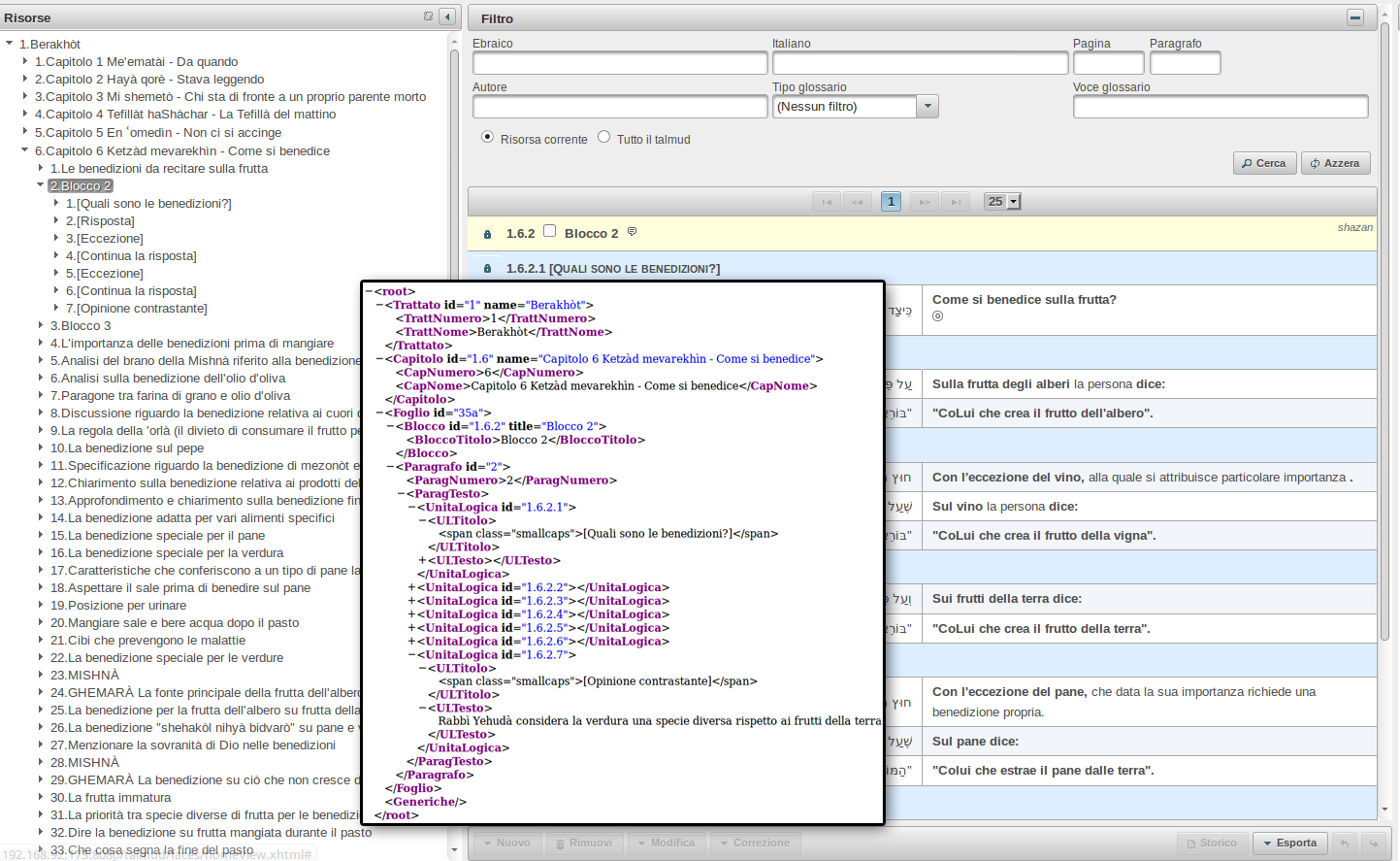


Figure 3. Publishing software export functionality. Example related to the Adobe InDesign export.