DHQ: Digital Humanities Quarterly
Editorial

ITALMUD: a Computational Terminology of the Italian Translation of the Babylonian Talmud

Emiliano Giovannetti <emiliano_dot_giovannetti_at_ilc_dot_cnr_dot_it>, , Istituto di Linguistica Computazionale “A. Zampolli” (CNR-ILC) https://orcid.org/0000-0002-0716-1160
Andrea Bellandi <andrea_dot_bellandi_at_ilc_dot_cnr_dot_it>, , Istituto di Linguistica Computazionale “A. Zampolli” (CNR-ILC) https://orcid.org/0000-0002-1900-5616
Flavia Sciolette <flavia_dot_sciolette_at_uniroma1_dot_it>, , Postdoctoral research fellow https://orcid.org/0000-0002-7998-9768

DOI: pending

Abstract

ITALMUD is a multilingual computational terminology built from the Italian translation of the Babylonian Talmud. Terms were extracted from glossaries compiled during ten years of translation work and modeled using the OntoLex-Lemon model. The resource includes over 1,600 terms in Italian, Hebrew, and Aramaic. A key innovation that was introduced involved linking terms to the CompL-it computational lexicon of the Italian language, in order to place the terminology within a rich network of semantic relations and enable more sophisticated linguistic searches. ITALMUD is available as an open resource and conforms to the principles of Linked Open Data to ensure maximum interoperability and reuse.

1. Introduction

The Babylonian Talmud is one of the fundamental texts of the Jewish tradition, the result of a complex interweaving of legal, narrative, and sapiential discussions [Strack and Stemberger 1996]. This text presents terminological peculiarities that make it difficult to translate into other languages: it is in fact written mainly in Aramaic, but it includes Hebrew terms, linguistic borrowings, and a vast lexical repertoire with a high degree of polysemy and contextual variability [Schäfer 2007]. The exact interpretation of Talmudic terms often depends on the immediate context and the cultural and legal implications implicit in the discussion. 1

From a linguistic point of view, the main terminological difficulties are related to semantic ambiguity, widespread polysemy, variations in meaning according to contexts of use, and frequent implicit references to concepts or cultural and legal institutions specific to the rabbinic world [Satlow 2006]. 2

Translating a text of such complexity (and vastness) constitutes a challenge; from a terminological point of view, translators are required not only to have considerable linguistic skills, but also a profound knowledge of Jewish culture and norms. 3

In this context, translation cannot be reduced to the substitution of words across languages. It requires the transposition of concepts, legal categories, ritual practices, and culturally embedded references from the rabbinic tradition into another linguistic and interpretive framework. Terminological work is therefore essential, since it helps preserve the specificity of the source tradition while making its concepts accessible to readers of the target language. 4

To date, there are few complete translations of the Babylonian Talmud. One of the most renowned is the translation edited by Rabbi Adin Steinsaltz, ([Steinsaltz 1965-2010])[1] rendered in modern Hebrew with commentary. The Schottenstein Edition ([Schottenstein et al. 1990-2005]), published in English by ArtScroll between 1990 and 2005, is widely recognized for its accuracy and popularity in the English-speaking world. Moreover, ArtScroll has also published a complete translation of the Talmud in French [Safra 2003-ongoing]. Another noteworthy translation is the Goldschmidt Talmud ([Goldschmidt 1897-1936]) a German translation edited by Lazarus Goldschmidt between 1897 and 1936. 5

A complete Italian translation of the Babylonian Talmud is, at the time of writing, underway since 2012 within the framework of the “Traduzione del Talmud Babilonese” project2.[2] 6

The role of technology has proven fundamental for this highly interdisciplinary project: without specially developed computational tools, it would have been extremely difficult to coordinate dozens of translators and reviewers. Indeed, the computer-assisted translation (CAT) tool Traduco ([Giovannetti 2016]) was essential in ensuring terminological consistency, coherence, and accuracy throughout the translation process. 7

The aim of the work presented in this article is to develop ITALMUD, a multilingual terminological resource for the Babylonian Talmud — pivoting on Italian — using the Linked Data paradigm ([Bizer, Heath, and Berners-Lee 2009]) and computational lexical resources. The focus is therefore not on the vocabulary of the Babylonian Talmud as a whole, but on the set of specialised terms documented in the expert glossaries produced during the Italian translation process. The significance of ITALMUD lies not only in the creation of a specialised resource for Talmudic studies, but also in the methodological problem it addresses. Many digital humanities projects generate rich terminological knowledge in the form of glossaries, annotations, translation notes, and expert commentaries; however, such knowledge often remains difficult to query systematically or to connect with other linguistic resources. ITALMUD shows how this kind of expert knowledge can be converted into Linked Open Data and connected to a computational lexicon, thereby making implicit semantic relations explicit. This makes it possible to ask questions that go beyond simple word lookup — for example, to retrieve terms connected by broader semantic domains, to compare translation choices, or to identify conceptual patterns across a large and multilingual textual tradition. The effort is therefore worthwhile because it turns a glossary into an instrument for discovery: a resource that supports both close philological work and scalable computational exploration. 8

The rest of the article is organised as follows. A brief overview of previous works on Talmudic terminology is provided in Section 2. 9

In particular, as detailed in Section 3.1, the Linked Data approach offered a way to formalize and standardize the resource, foster its shareability ([Chiarcos et al. 2020]), and significantly aid in managing the terminological variety and complexity within the Talmudic domain. 10

The resource was built using data extracted from a series of glossaries produced during the translation process and subsequently converted into a formal representation compliant with the aforementioned standards, as described in Section 3.2. 11

It was also decided to link the resulting terminological resource to a computational lexicon of the Italian language, to be used as a lexical superstructure, in order to systematically connect and more clearly define Talmudic terms through an established lexical framework (see Section 3.3). This approach made it possible to embed the terminology within the dense network of semantic relationships that constitutes the lexicon, thereby enhancing its usability. 12

In Section 4, the resource will be described from a quantitative point of view, including some statistics on the various types of linking with the Italian lexicon; Section 5, on the other hand, will describe how to access the resource and some possible uses of it. The last section, finally, will conclude the article with some reflections on the work done and its future prospects. 13

2. Related Works

Freely available lexical resources related to the religious texts of Judaism are relatively scarce. As for the Hebrew language, the only resource that, to the best of our knowledge, is freely accessible is Jastrow's Dictionary,[3] hosted by Sefaria,[4] which includes English definitions of approximately 30,000 Hebrew and Aramaic words used in various rabbinic texts, including the Babylonian Talmud. This digital dictionary is based on the work of the scholar Marcus Jastrow, whose "A Dictionary of the Targumim, the Talmud Bavli and Yerushalmi, and the Midrashic Literature" ([Jastrow 1926]) remains an authoritative source for understanding the Aramaic lexicon in rabbinic Jewish literature. 14

We also cite for Hebrew the Historical Dictionary Project of the Academy of the Hebrew Language, initiated in the 1950s. This project aims to compile an authoritative historical dictionary encompassing the entire Hebrew lexicon, tracing the meanings and morphology of words from their earliest attestations to their most recent usage. The project has developed a comprehensive computerized database, known as Ma'agarim, which serves as a unique resource for researchers. The project is documented online by the Academy of the Hebrew Language.[5] The database is freely accessible online but cannot be downloaded. 15

For Aramaic, we cite the Comprehensive Aramaic Lexicon (CAL),[6] a digital project hosted by the Hebrew Union College-Jewish Institute of Religion. Initiated in the 1980s, CAL provides a text base of Aramaic texts across all dialects 16

from the 9th century BCE to the 13th century CE. It offers a searchable dictionary (which cannot, however, be downloaded) and access to text corpora.

17 Regarding research involving Italian lexical resources, several studies conducted within the Babylonian Talmud Translation Project into Italian are mentioned below. These studies have addressed various aspects, ranging from the computational representation of Talmudic terms to the ontological modelling of rabbinic figures mentioned in the text. Specifically, we highlight the contribution on extraction, representation, and use of the Babylonian Talmud terminology ([Giovannetti 2020]) and the definition of an ontology dedicated to the Talmudic masters [Giovannetti 2021]. Furthermore, experiments were conducted on graphical visualization of Talmudic terms to enhance accessibility for scholars and non-specialist users [Marchi 2022]. Additional works explored the application of the OntoLex-Lemon model to formally link the Talmudic text to lexical entries ([Sciolette 2024]), and structured methodologies were proposed for building integrated computational resources (terminological and conceptual) that are shareable and interoperable, derived from religious texts, including Hebrew texts [Saponaro, Giovannetti, and Sciolette 2022].

18 Taken together, these works show that Talmudic terminology constitutes not only a specialised domain of Jewish studies, but also a useful testing ground for broader questions concerning the computational representation, publication, and reuse of expert knowledge in multilingual textual traditions.

3. The Construction of the Resource

19 The terminological resource was built based on a selection of tractates from the Babylonian Talmud translated into Italian. This translation effort is part of the Babylonian Talmud Translation Project, introduced in Section 1. The project's goal is to create a digital Italian version of the Babylonian Talmud — an essential work in Jewish tradition that encompasses a wide range of knowledge, from law and science to philosophy and daily life.

20 The translation process involves a dedicated team of translators, subject matter experts, and editorial reviewers, all collaborating synergistically through the use of the aforementioned Traduco CAT tool. In this respect, the construction of ITALMUD illustrates a recurrent challenge in digital humanities projects: how to transform knowledge produced during collaborative scholarly work into structured data that can be preserved, queried, and reused beyond the immediate context in which it was created.

21 The reference corpus used as the source for the terminological data was compiled from all the tractates that had been translated and published at the time of writing this article, amounting to ten in total, namely *Berakhòt*, *Rosh haShanà*, *Ta'anit*, *Qiddushìn*, *Chaghigà*, *Betzà*, *Meghillà*, *Sukkà*, *Mo'èd Qatàn*, and *Sotà*.

22 More specifically, the terms were selected from the glossaries included in the ten tractates under consideration. During the translation process, in fact, Talmudic technical terms — as well as any other terms that the translators deemed it useful to accompany with a definition to facilitate the reading of the Talmud — were added to thematic glossaries. Of the seven glossary categories created, three were taken into account for the construction of the present resource, i.e., "Concept," "Linguistics," and "Nature." [7] These categories were selected because they contain the entries with the strongest terminological relevance for the purposes of the present resource.

23 Each glossary entry contains the following information: the category (e.g., "Concept"), the entry label (e.g., "Sinedrio"), the original Hebrew term (e.g., "סַנְהֶדְרִין"), its transliteration ("Sanhedrìn"), its translation (used for terms that were not translated into Italian but left in their transliterated form — for example, "Tevilà", translated as "immersione"), and finally, its definition.

24 Once the glossary entries from which to extract the terms had been identified, it was necessary to define a model capable of accommodating and formalising these terms. As previously mentioned, the Linked Data paradigm was adopted as a reference.

3.1 The Model of Reference

25 The paradigm of Linked Open Data (LOD) ([Berners-Lee et al. 2006]) today constitutes one of the methodological foundations for the management, publication, and interoperability of structured data within the context of the Semantic Web [Bizer, Heath, and Berners-Lee 2009]. It enables the transformation of data into knowledge that is accessible, shareable, and reusable on a global scale, fostering an open and interconnected information ecosystem. In accordance with the FAIR principles (Findable, Accessible, Interoperable, Reusable) ([Wilkinson 2016]), the LOD paradigm promotes traceability and methodological transparency, ensuring the accessibility and reusability of informational resources in domains such as scientific research, education, and technological innovation. The LOD approach, in addition to improving data quality and usability, enhances their scientific, cultural, and social significance, contributing to

making knowledge — including linguistic knowledge — an open, shared, and dynamically integrable resource within the global digital landscape. From this perspective, the Linguistic Linked Open Data[8] cloud (LLOD) represents a collaborative initiative promoted by members of the Open Linguistics Working Group ([Chiarcos et al. 2020]), with the aim of building a coherent subset of open and interconnected data specifically dedicated to linguistic resources [McCrae et al. 2016]. The adoption of the Linked Data paradigm in the linguistic domain enables mechanisms of semantic interoperability, while also promoting the integration of linguistics with other scientific and applied fields already employing this approach, such as geography ([Goodwin, Dolbear, and Hart 2008]), biomedicine ([Ashburner et al. 2000]), and public administration.[9]

In the context of terminology, the adoption of Linked Open Data (LOD) enables the description not only of terms and their definitions, but also of the full range of semantic, morphological, etymological, and pragmatic relations that connect them, as well as their usage contexts, sources of attestation, and equivalents in other languages. To this end, standard languages such as RDF (Resource Description Framework),[10] OWL (Web Ontology Language),[11] and SPARQL (a query language for RDF data)[12] are employed, along with internationally recognised and shared ontological vocabularies, such as OntoLex-Lemon ([McCrae et al. 2017]) for the representation of lexical resources, SKOS (Simple Knowledge Organization System, for the structuring of concepts),[13] and LexInfo (for the encoding of grammatical and syntactic properties).[14] The OntoLex-Lemon model has gained popularity as the de facto standard for representing lexical data to express information on the Semantic Web as Linked Data. Although originally conceived as a linguistic model for describing the lexicalisation of concepts represented in a formal ontology, over the years several methodologies have been proposed, both for applying Linked Data principles to terminologies ([Cimiano et al. 2015]; [Bosque-Gil et al. 2015]), including services that generate LD terminologies from corpora such as termItUp ([Martín-Chozas 2022]), and for converting terminologies represented in the ISO standard TBX/XML format into OntoLex-Lemon, such as Term-à-LLOD [Di Buono et al. 2020], TBX2RDF ([Cimiano et al. 2015]); ([Montiel-Ponsoda et al. 2015]), and LemonizeTBX ([Bellandi 2023]). 26

To make the modelling approach more explicit for non-specialist readers, Figure 1 presents a simplified view of the core OntoLex-Lemon components used in this work. Rather than treating a term as a simple label, the model distinguishes between the lexical entry, its written forms, and its senses. It also makes it possible to associate a lexical sense with a conceptual reference and to establish semantic links with external lexical resources. 27

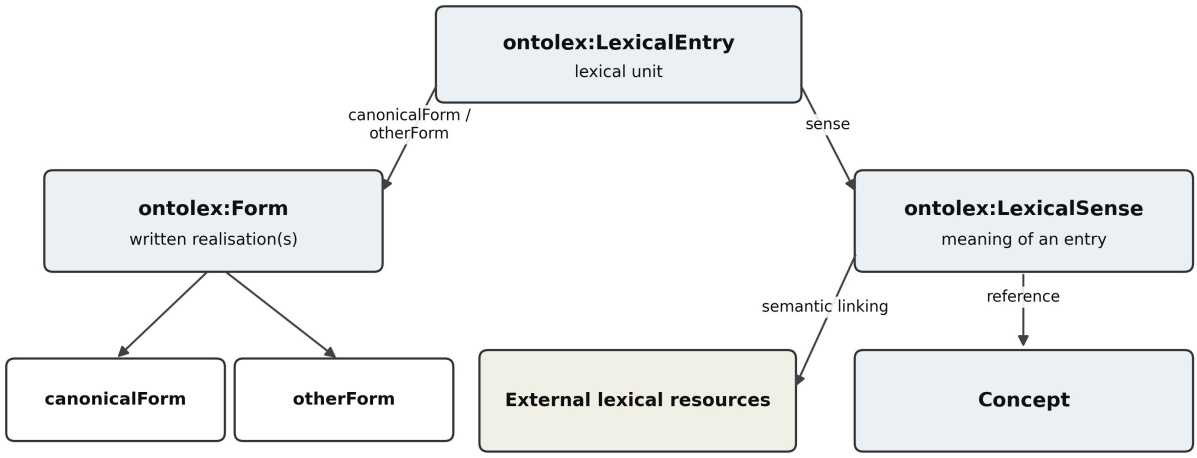

Figure 1. Simplified representation of the core OntoLex-Lemon components used in ITALMUD.

This distinction is particularly important for ITALMUD, where each glossary item may involve several layers at once: a lexical entry, one or more written forms, a culturally specific sense, and, when relevant, a conceptual reference. The same structure also provides the basis for linking ITALMUD senses to external resources such as the Italian computational lexicon CompL-it. 28

More generally, this modelling step shows how domain-specific knowledge, often embedded in glossaries or editorial notes, can be translated into a form that supports computational analysis and scholarly reuse. 29

3.2 From Glossaries to Terminology

The glossaries compiled for the translation of the Babylonian Talmud are resources manually curated by domain experts over a span of ten years. This extended time frame allowed for the development of resources that are rich in 30

information, highly accurate, and thoroughly validated. However, like any manually created dataset, they also exhibit reworkings, duplicate entries, annotation inconsistencies, and certain types of ambiguity, such as, for example, the presence of two forms — singular and plural — in the field “name” of the glossary entry, or the presence of separators, in the same field, for different forms. While such issues may be negligible in the context of human consultation, they become critical for computational processing. This situation is common in many scholarly projects: resources designed for human consultation often contain precisely the richness that makes them valuable, but they require an additional layer of modelling before they can become computationally actionable.

Accordingly, the glossaries underwent manual preprocessing to ensure data homogeneity. The primary interventions focused on the extraction and analysis of alternative forms of the lemma. Each entry of glossaries was subsequently mapped and converted into a Lexical Entry following the OntoLex-Lemon model, as described in the previous section. 31

Lexical entries were classified into two distinct categories: Word entries, representing single words, and Multiword entries, comprising compound words, fixed expressions, and complex phrases. From a theoretical perspective, the Multiword category was designed to encompass a wide range of linguistic phenomena, which were later subdivided into more specific syntactic classes. Where possible, the various forms associated with a lemma were extracted, particularly when such forms were explicitly provided by domain experts either in the entry label or within its definition. 32

In fact, both of these fields — entry name and definition — were used in the glossaries not only to convey semantic information, but also to include grammatical annotations such as plural forms. 33

Once lemmas were identified as canonical entries and the alternative forms were treated as instances of ontolex:OtherForm in accordance with the OntoLex-Lemon model, all data was tagged with Part-of-Speech (PoS) information and morphological features, specifically gender and number. Furthermore, wherever feasible, entries and forms in Aramaic were identified and distinguished. Cross-references between glossary entries were formalized using the rdfs:seeAlso property. 34

LexInfo, the reference linguistic ontology for the OntoLex-Lemon model, was adopted both for the syntactic classification of phrases and for the attribution of PoS tags and morphological traits. In cases where it was not possible to assign a specific syntactic class to a multiword unit, the entry was left underspecified and labelled simply as a multiword. 35

Regarding the handling of duplicates, the selection process was based on criteria agreed upon with domain experts: in cases of equivalent informational content, preference was given to the most recently annotated (and therefore updated) entry with the greatest number of annotated contexts. In cases where the contexts associated with duplicate candidates differed, a merging operation was performed to consolidate all contextual annotations under a single entry. 36

From this preprocessing of the 887 glossary entries, a total of 828 entries were obtained, distributed as follows: 461 entries of the “concept” type, 276 of the “linguistics” type, and 91 of the “nature” type. 37

Finally, the lexical data extracted from glossaries and mapped to OntoLex-Lemon were converted into RDF: the resulting resource, we called ITALMUD, is a multilingual, open, and extensible terminology compliant with the LOD paradigm. 38

3.3 The Linking to the Italian Lexicon

The set of terms contained in the resource essentially constitutes a list of Italian, Hebrew, and Aramaic terms, linked by translation relations. What is lacking, therefore, is a semantic network of relations among the terms themselves. 39

What characterizes a computational terminology, however, is its capacity to formally define semantic relations among the linguistic entities it contains. Such a structure enables the resource to be exploited as a representation of the underlying domain, allowing for advanced queries both on the resource itself and on the texts to which it is linked. 40

In light of this, we decided, as a first step, to address the lack of relations among the ITALMUD terms by anchoring its Italian entries to the words of an Italian computational lexicon, CompL-it, in which the words, through their respective senses, are already connected to each other by a dense network of relations. 41

3.3.1 The Computation Lexicon for Italian: CompL-it

CompL-it ([Sciolette 2024]) is a computational lexicon for contemporary Italian, developed in accordance with the OntoLex-Lemon model. The construction of this resource involved the integration of three distinct data sources: (1) M-GLF (MAGIC-Generated Lemmatized Forms), a collection of lemmatized entries enriched with morphological 42

information produced by the MAGIC morphological analyzer ([Battista and Pirrelli 1999]; [Pirrelli and Battista 2000]); (2) a selection of Italian language treebanks — including the Italian Stanford Dependency Treebank (ISDT), the Venice Italian Treebank (VIT), ParallelTut (ParTut), and ParlaMint-It — accessible via the Universal Dependencies (UD) repository; and (3) the LexicO computational lexicon ([Sciolette, Giovannetti, and Marchi 2023]), which serves as the foundational component of the resource, particularly from a modeling perspective. For a detailed account of the resource construction process and the characteristics of the underlying model, the reader is referred to the previously cited works [Sciolette 2024], [Sciolette, Giovannetti, and Marchi 2023].

Below, a sample entry from the resource is briefly described. In this example, the Italian noun *arca* (ark) is modelled as a lexical entry with two distinct senses: one meaning “wooden chest” and the other “monumental sarcophagus.” Each form of the word (*arca* and its plural *arche*) is encoded with grammatical information such as gender and number. The lexical senses are semantically enriched with definitions, usage examples, and links to two other senses, with two different relations (i.e., hyponym and usedFor). 43

```
lex:MUSarcaNOUN a ontolex:Word;
   rdfs:label "arca"@it;
   lexinfo:partOfSpeech lexinfo:noun;
   ontolex:sense lex:USem63144arca;
   ontolex:sense lex:USem63145arca;
   ontolex:canonicalForm      lex:arca_MUSarcaNOUN_noun_singular-feminine;
   ontolex:otherForm lex:arche_MUSarcaNOUN_noun_plural-feminine .
lex:arca_MUSarcaNOUN_noun_singular-feminine a ontolex:Form;
   lexinfo:number lexinfo:singular;
   lexinfo:gender lexinfo:feminine;
   lexinfo:degree lexinfo:positive;
   ontolex:writtenRep "arca"@it .
   lex:arche_MUSarcaNOUN_noun_plural-feminine a ontolex:Form;
lexinfo:number lexinfo:plural;
   lexinfo:gender lexinfo:feminine;
   lexinfo:degree lexinfo:positive;
   ontolex:writtenRep "arche"@it .
lex:USem63144arca a ontolex:LexicalSense;
   skos:definition "cassa di legno";
   lexinfo:senseExample "l'arca delle tavole della legge" .
lex:USem63145arca a ontolex:LexicalSense;
   skos:definition "sarcofago monumentale" .
lex:USem63144arca lexinfo:hyponym lex:USem2372cassa .
lex:USem63144arca compl-it:usedFor lex:USemD883contenere .
```

The CompL-it lexicon can be freely downloaded from GitHub.[15] 44

3.3.2 Linking

At the current stage of development, 227 ITALMUD terms have been linked to CompL-it. This figure does not represent the lexical coverage of the Babylonian Talmud in CompL-it, nor the proportion of Talmudic words that occur in the Italian lexicon. Rather, it refers to the subset of curated ITALMUD terminological entries for which a validated semantic connection with CompL-it has so far been established. Although this linking is still partial, it already enables ITALMUD terms to be explored through the semantic network of a general-language computational lexicon. The value of the operation therefore lies not in simple lexical overlap, but in the possibility of using CompL-it as a semantic infrastructure through which specialised Talmudic terminology can be queried, compared, and progressively enriched. 45

Specifically, the following types of linking have been identified: 46

i) linking between senses of the same word, which can be further divided into two cases:

a) the lemma form in ITALMUD and CompL-it is identical, e.g., *bietola* (chard); 48 47

49

b) the lemma form in ITALMUD does not match that in CompL-it, e.g., *vetriolo* (vitriol) in ITALMUD appears as *Qanqantòm* (a transliteration of the Hebrew word קַנְקַנְתּוֹם);

ii) linking between senses of different words, based on semantic relations (such as hyponymy);

iii) linking between a lexical entry in CompL-it and a sense in ITALMUD, which can thus constitute an additional lexical sense of the entry. This also occurs in two cases:

a) the lemma forms in both CompL-it and ITALMUD are identical, e.g., *abisso* (abyss);

b) the lemma forms differ between the two resources, e.g., *racimolo* (raceme), which in ITALMUD appears under the term *Lèqet* (the transliteration of the Hebrew word לֶקֶט).

In alignment with the principles of Linked Open Data, it was decided to link ITALMUD to CompL-it by adopting standard linking strategies. The vocabulary selected for the relations between senses with a certain degree of equivalence is SKOS, which is widely used within the LOD community and in the domain of computational terminologies. For all other types of relations (such as the hyponymy), it was decided to use the CompL-it vocabulary, which includes many properties mapped to the Lexinfo, the main lexical ontology, for OntoLex-Lemon, as well as numerous custom properties, specific to the CompL-it model, which could not be mapped to LexInfo.

To handle the first type of cases introduced earlier, relative to the linking between the senses conveyed by the entries, three different SKOS relations have been used. A first example, where the relation skos:exactMatch has been applied, is relative to term *manna*, understood as the miraculous food sent by God to the Israelites during their journey through the desert, whose sense is also present in CompL-it, albeit with a less detailed definition.

However, cases of perfect equivalence are relatively rare, due to the complexity of the subject matter, which often calls for a cautious approach in order to avoid overly strong assertions. For this reason, the relation that has been more frequently preferred was skos:closeMatch. This property allows the expression of conceptual similarity and limited interchangeability in specific practical contexts, without implying full logical identity or triggering strong semantic inferences, as would be the case with OWL properties such as owl:sameAs or owl:equivalentClass. An example of this case is the entry *Androgino* (androgynous), which in ITALMUD includes additional nuances in its definition compared to the sense described in CompL-it, although both refer to the same underlying concept.

In cases where the senses are more loosely semantically related, the skos:relatedMatch property is used. This relation expresses a meaningful semantic association between concepts, without implying interchangeability or logical equivalence, as for the case of *levita* (levite), member of the Levi tribe.

As for the second type of linking, it undoubtedly involves a highly extensive undertaking in order to find the most appropriate “anchoring” of talmudic words to the Italian lexicon. Indeed, the CompL-it vocabulary of relations is exceptionally rich and allows for a very fine-grained semantic description of entries. A large-scale project aimed at mapping and establishing links between the entries of the two resources will be the subject of future work. By way of example, a possible linking case is provided involving the entry *shofar*, defined as “a horn of a kosher animal (usually a ram) that was blown on various occasions [...].”

The sense of word *shofar*, in ITALMUD, could be linked via the hyponymy relation to the sense of the word *corno* (horn) in CompL-it, defined as “an ancient wind instrument,” and through the relation *madeOf* to the sense of *corno* as “a horn-like or bony protrusion found on the head of many animals.” It may also be connected to the sense of *suonare* (to blow/play) via the *Object of Activity* relation, which describes the action for which an object is used.

As for the third type of linking, it involves creating a new sense corresponding to the meaning of the entry in ITALMUD, which does not match any of the senses of the entry in CompL-it. An example is racimolo, which in CompL-it has only the sense of “a small bunch of grapes,” whereas in ITALMUD it also means “Ears of grain fallen in the field during harvesting, which under certain conditions are the rightful property of the poor (Lev. 19:9–10, 23:22).” Consequently, the ITALMUD sense is linked to the CompL-it entry using the sense relation, as specified in the OntoLex-Lemon model. An illustrative schematic of these cases is shown in Figure 2.

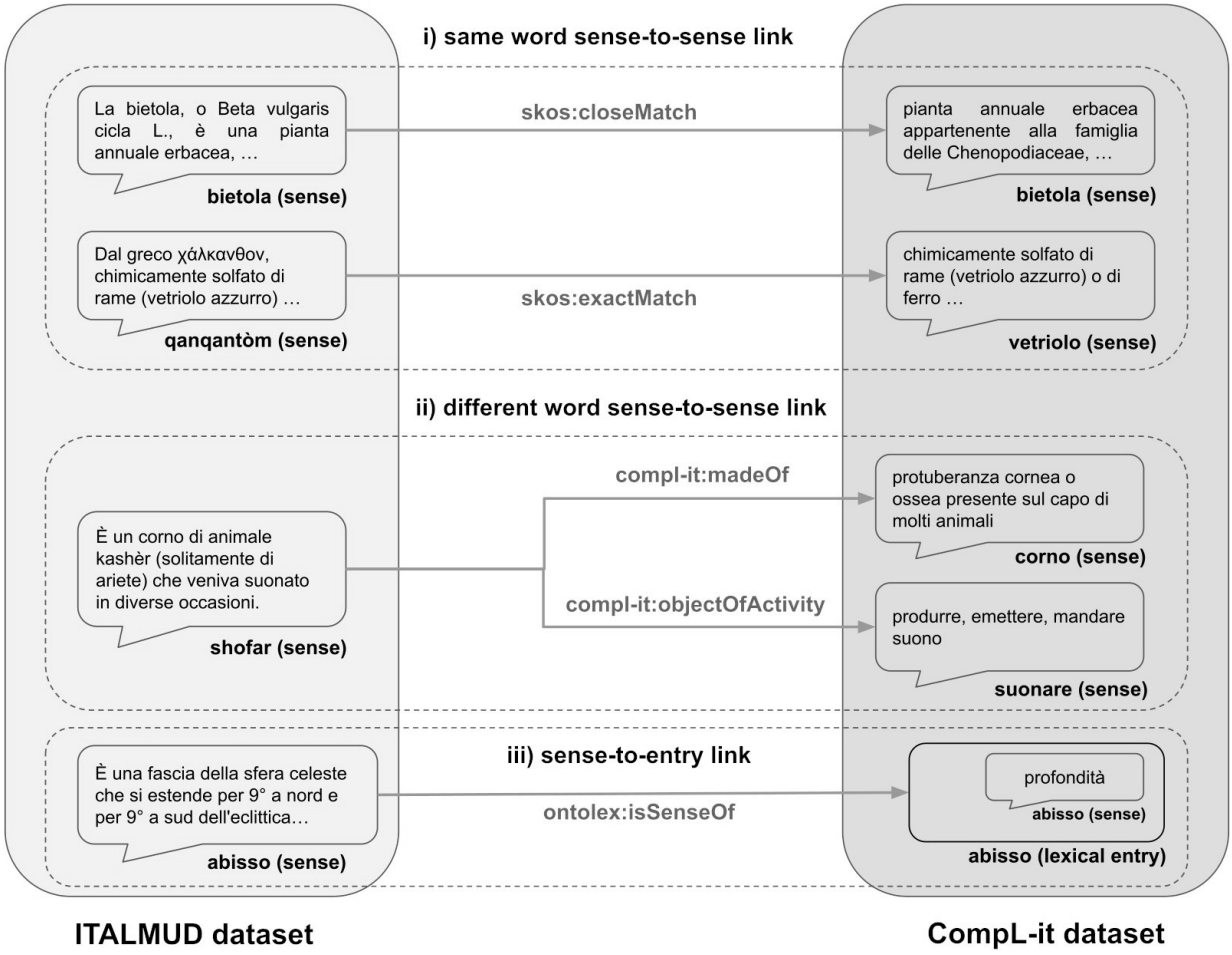

Figure 2. The three types of linking used to connect ITALMUD to CompL-it. In the first two cases, senses are linked to senses, while in the last case, a sense of ITALMUD is linked to a lexical entry of CompL-it, “abisso”, which in the lexicon has only one sense.

Each of the CompL-it words mentioned is itself connected to other words through additional relations. This enables the construction of highly sophisticated queries (see Section 5). 61

4. ITALMUD in Numbers

The trilingual terminological resource produced through this work comprises a total of 1,613 Talmudic terms, including 823 terms in Italian, 773 in Hebrew, and 17 in Aramaic. Of the Italian terms, 782 have a Hebrew translation, 17 an Aramaic one and 4 terms have both a Hebrew and an Aramaic translation. 62

From the point of view of lexical entry type, 1,242 are classified as “Word” (625 in Italian, 601 in Hebrew, and 16 in Aramaic), and 371 as “Multiword Expression” (198 in Italian, 172 in Hebrew, and 1 in Aramaic). A schematic representation of the distribution of terms by language and type is shown in Figure 3. 63

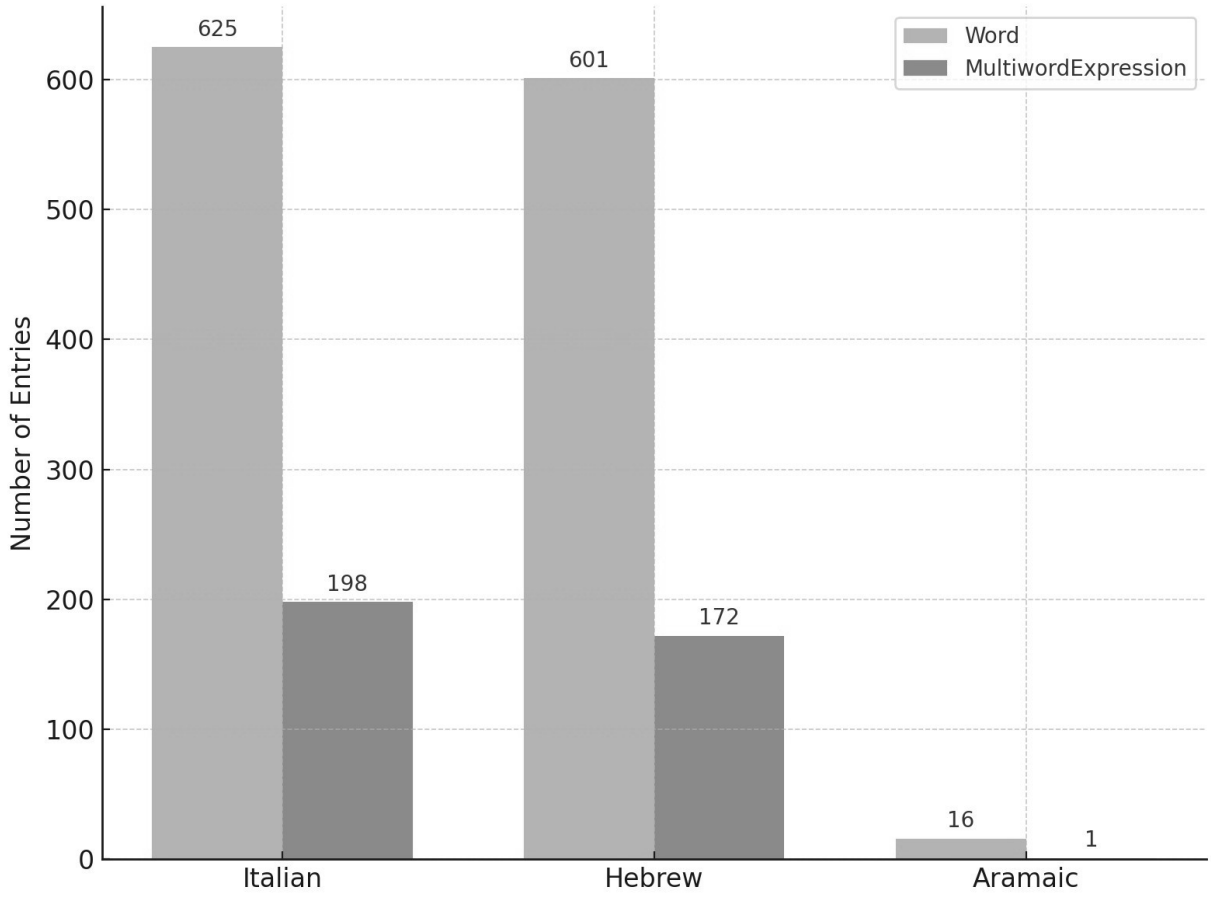

Figure 3. Distribution of lexical entries by language and type. The chart compares the number of entries classified as Word and Multiword Expression across Italian, Hebrew, and Aramaic. Italian and Hebrew show a predominance of Word entries, while Aramaic includes only a small number of terms.

As for the PoS of the Italian terms, as noted in Section 3.2, no syntactic categories were assigned to multiword expressions. Of the remaining 625 Italian terms, 606 are nouns, 12 adjectives, 4 interjections, 2 adverbs, and 1 conjunction. The PoS of the Hebrew and Aramaic terms, of course, are the same as the Italian “Word” terms linked to them by translation relations. However, there are 11 Italian multiword expressions with Hebrew or Aramaic single-word translations, all of which have been assigned the PoS “Noun”, as they are all nouns. 64

The total number of terms that have so far been linked to the CompL-it lexicon is 227. Of these, 158 terms have been connected, via their respective senses, to senses in the CompL-it lexicon through three types of SKOS relations: skos:exactMatch (68 links), skos:closeMatch (72 links), and skos:relatedMatch (18 links). The remaining 69 terms have been linked, again via their respective senses, to lexical entries in CompL-it using the ontolex:isSenseOf property. The connection between ITALMUD and CompL-it has been technically implemented through a dedicated linkset, used as a “bridge” resource. 65

5. Use and Applications

ITALMUD is an open and freely downloadable resource.[16] The linkset connecting it to the CompL-it lexicon is also available for download at the same address. The resource can additionally be queried via SPARQL through a dedicated web interface.[17] Turtle was chosen for representing the resource — a serialisation format for RDF data designed to be both readable and compact. Turtle expresses information as “subject–predicate–object” triples and allows the use of prefixes to shorten URIs,[18] making the content more readable. It is a W3C standard widely used in the context of Linked Data and terminological resources to describe knowledge in a formal, shareable, and interoperable way. 66

Below is an example of combined access to the three resources — ITALMUD, CompL-it, and the linkset connecting them — aimed at demonstrating their potential use in the study of Talmudic terminology. 67

The following query, based on the data currently available in the resources, illustrates how semantic modelling changes the scale and nature of inquiry. Instead of manually comparing isolated glossary entries, users can formulate repeatable queries over a network of lexical and terminological relations. In this case, the query concerns the search for Talmudic terminology related to the concept of “container”. A possible formulation in natural language might be as follows: 68

Retrieve all Talmudic terms that are related to containers. 69

By leveraging the connection established between ITALMUD and CompL-it, it is possible to answer this query by first searching within the language lexicon for the lexical senses of the word "*contenitore*" (container), its hyponyms (including indirect ones),[19] and/or the senses linked by a usage relation to the senses of the word "contenere" (to contain). Then, the bridge resource — the linkset — can be used to identify which of the resulting lexical senses have been linked to terms in ITALMUD. Technically, to respond to this request the following operations are required: 70

1. search in CompL-it for the senses of the word "contenitore" (there is only one) and the word "contenere" (which has three senses); 71
2. retrieve, via the lexinfo:hyponym relation, the hyponym senses of the sense of "*contenitore*" (there are 337, of which 160 are direct and 177 indirect), and the senses linked via the compl-it:usedFor relation to the three senses of "*contenere*" (there are 233), the union of which (thus excluding senses in common) amounts to 390 senses;
3. search the linkset to see whether any senses of Talmudic terms are linked — through one of the three sense relations (cf. 3.3.2) — to at least one of the 390 CompL-it senses retrieved in the previous step.

This query, available as a preset through the aforementioned query interface, returns 12 Talmudic terms related to containers — such as *tasqà* (sack), *disaqqayà* (saddlebag), and *peyalè* (cup) — along with their respective Hebrew translations, the definition, and additional information.

To emphasize the multilingual nature of the resource, we have also added to the SPARQL interface a preset query that starts from the Hebrew term "סִרְפָּד" (nettle) (which can, if desired, be replaced with another term by directly editing the query) and retrieves from the database the related terms — particularly the co-hyponyms, i.e., other types of plants, since they are hyponyms of the term "pianta" (plant), which, in CompL-it, is the hypernym of "ortica" (nettle). 72

As these examples suggest, the integration of the produced terminological resource (beyond its intrinsic value) with a computational lexicon such as CompL-it is not merely an exercise in data linking, but opens up new perspectives in the study of Talmudic terminology. Thanks to the alignment between lexical senses and Talmudic terms, it becomes possible to carry out terminological research that would otherwise require lengthy manual comparisons across heterogeneous sources, but can now be resolved through formalised and repeatable semantic querying. 73

The use of the RDF formalism and the Turtle format also ensures that each term and each relation are expressed in an interoperable way and in compliance with W3C standards, facilitating the reuse of data in international and multidisciplinary research contexts. The ability to simultaneously query the three resources — the Talmudic terminological repertoire, the computational lexicon of the Italian language, and the linkset connecting them — will make it possible, especially as the resources grow, to accurately map shared or divergent semantic areas, to monitor lexical variation across different periods and traditions, and to compare the evolution of concepts with their treatment in modern languages. 74

6. Conclusions

This article has presented ITALMUD, a multilingual terminological resource based on the glossaries developed within the framework of the Italian translation of the Babylonian Talmud project. Through the adoption of the OntoLex-Lemon model and the principles of Linked Open Data, the Talmudic terms have been formalised and additionally linked to a computational lexicon of contemporary Italian (CompL-it). The overall work included phases of normalisation, linguistic modelling, and the creation of a dedicated linkset, transforming the originally "flat" list of Talmudic terms into a resource that can be queried along semantic pathways inherited from the language lexicon. For this reason, ITALMUD should also be understood as a methodological case study in how specialised, multilingual, and culturally situated terminologies can be made computationally accessible without detaching them from the expert knowledge on which they depend. 75

In addition to the outcome achieved with the release of the ITALMUD resource, we believe that the methodology presented — anchoring a specialised terminology to a computational lexicon of general language — can be fruitfully applied in other domains as well. Linking terms to a structured lexicon not only makes implicit semantic relations explicit, but also enhances access to the texts in which such terms occur. This enables scholars and researchers to carry out targeted investigations, identify conceptual variants, disambiguate meanings, and accurately trace terminological evolution within complex textual traditions. 76

The future development of this work will primarily involve the expansion of the ITALMUD resource, which will progress in parallel with the translation of new tractates of the Talmud and the simultaneous production of additional glossary 77

entries, which will be converted into terms. Alongside this expansion of the terminology, semantic relations will also be extended — both through the linking approach already outlined in this article (further expanded to include additional semantic relations offered by CompL-it), and by defining semantic relations among the Talmudic terms themselves.

It will be possible to directly involve Talmud scholars in the expansion of the ITALMUD resource by providing them with access to the collaborative web tool Maia ([Giovannetti et al. 2024]), designed for the construction of lexical and terminological resources based on the OntoLex-Lemon model and for linking them to annotated corpora. 78

A shared platform of this kind, based on Linked Open Data, would foster collaboration among specialists, ensuring terminological consistency, traceability of translation choices, and interoperability with other resources. It would also allow for the continuous updating of terminology and semantic access to texts, facilitating interdisciplinary research and comparative approaches across languages and traditions. 79

Funding

This work was supported by the TALMUD project and carried out within the scientific collaboration between S.c.a r.l. Progetto Traduzione Talmud Babilonese and Cnr-Istituto di Linguistica Computazionale “A. Zampolli”. 80

Data Availability Statement

All data involved in this work are openly available at the following locations: i) the ITALMUD resource and the linkset to CompL-it on the dedicated GitHub repository at https://github.com/klab-ilc-cnr/ITALMUD, ii) the CompL-it lexicon in the KLAB GitHub repository at https://github.com/klab-ilc-cnr/CompL-it, iii) the CompL-it item in ILC4CLARIN repository at https://dspace-clarin-it.ilc.cnr.it/repository/xmlui/handle/20.500.11752/ILC-1007, iv) the web interface for the SPARQL querying of ITALMUD, the linkset and CompL-it at https://klab.ilc.cnr.it/CompL-it-SPARQL-interface. 81

Notes

[1] https://steinsaltz-center.org/

[2] https://www.talmud.it/en/

[3] https://www.sefaria.org/Jastrow

[4] https://www.sefaria.org/

[5] https://eng.hebrew-academy.org.il/our-work/historical-dictionary-project/

[6] http://cal.huc.edu/

[7] The classes Concept, Nature, and Linguistics represent three categories of glossary terms. They were selected because they contain the entries with the strongest terminological relevance for the construction of ITALMUD, namely terms that identify domain-specific concepts, culturally or legally significant entities, linguistic phenomena, etymological information, and names of plants and animals occurring in the Talmudic text. Concept refers to names related to domain-specific concepts, often untranslatable into Italian, such as Kohèn, the priest entrusted with performing the priestly duties in the Temple of Jerusalem. Linguistics includes entries that explore the origin of a word and its parallels in other languages, such as Dugmà, which comes from Greek and means “sample.” Nature covers entries related to plants and animals, such as Tzir'à, which refers to the wasp. The remaining glossary categories, namely, Idiom (referring to idiomatic expressions), Rabbi, Name (referring to names of historical characters), and Daily (referring to daily life), were not included at this stage because they are less directly terminological in nature or are more closely related to other types of scholarly annotation.

[8] https://linguistic-lod.org/

[9] https://data.gov.uk/

[10] https://www.w3.org/TR/rdf12-concepts/

[11] https://www.w3.org/TR/owl-features/

[12] https://www.w3.org/TR/rdf-sparql-query/

[13] https://www.w3.org/TR/skos-reference/

[14] https://lexinfo.net/

[15] https://github.com/klab-ilc-cnr/CompL-it/

[16] https://github.com/klab-ilc-cnr/ITALMUD

[17] https://klab.ilc.cnr.it/ITALMUD/

[18] A URI (Uniform Resource Identifier) is a standardised string used to uniquely identify a resource on the Web or within a structured data system. In RDF and Linked Data contexts, URIs allow for precise and shareable references to concepts, entities, or terms, thereby facilitating interoperability between different resources.

[19] Hyponymy is a transitive relation. This means that if word A is a hyponym of B, and B is a hyponym of C, then A is also a hyponym of C. In this way, a reasoner can make all hyponymy relations explicit. For example, if “sacco” (sack) is a hyponym of “recipiente” (bowl), and “recipiente” is a hyponym of “contenitore” (container), the reasoner will infer a hyponymy relation between “sacco” and “contenitore.” We will refer to this type of relation as inferred or indirect instances. Other examples of transitive relations include hypernymy and near-synonymy.

Works Cited

Ascoli 2022 Ascoli M. (ed.) (2022) *Talmud Babilonese: trattato Meghillà*. Firenze: Giuntina.

Ashburner et al. 2000 Ashburner, M. et al. (2000) “Gene ontology: tool for the unification of biology”, *Nature Genetics*, 25(1), pp. 25–29.

Battista and Pirrelli 1999 Battista, M. and Pirrelli, V. (1999) “Una piattaforma di morfologia computazionale per l’analisi e la generazione delle parole italiane”, ILC-CNR Technical Report.

Bellandi 2023 Bellandi, A. et al. (2023) “From TBX to Ontolex Lemon: issues and desiderata”, in *Proceedings of the 2nd International Conference on Multilingual Digital Terminology Today (MDTT 2023)*. Lisbon: CEUR Workshop Proceedings 3427.

Berners-Lee et al. 2006 Berners-Lee, T. et al. (2006) “A framework for web science”, *Foundations and Trends in Web Science*, 1(1), pp. 1–130.

Bizer, Heath, and Berners-Lee 2009 Bizer, C., Heath, T., and Berners-Lee, T. (2009) “Linked Data: the story so far”, *International Journal on Semantic Web and Information Systems*, 5(3), pp. 1–22.

Bosque-Gil et al. 2015 Bosque-Gil, J. et al. (2015) “Applying the OntoLex model to a multilingual terminological resource”, in *The Semantic Web: ESWC 2015 Satellite Events*. Cham: Springer International Publishing, pp. 283–294.

Chiarcos et al. 2020 Chiarcos, C., Cimiano, P., and Declerck, T. (eds.) (2020) *Linked Data in Linguistics: Representing and Connecting Language Data and Language Metadata*. Cham: Springer.

Chiarcos et. al 2020 Chiarcos, C. et al. (2020) “On the Linguistic Linked Open Data Infrastructure”, in *Proceedings of the 1st International Workshop on Language Technology Platforms*. Marseille: European Language Resources Association (ELRA), pp. 8–15.

Cimiano et al. 2015 Cimiano, P. et al. (2015) “Linked terminologies: applying linked data principles to terminological resources”, in *Proceedings of the eLex 2015 Conference*, pp. 504–517.

Di Buono et al. 2020 Di Buono, M. P. et al. (2020) “Terme-à-LLOD: simplifying the conversion and hosting of terminological resources as Linked Data”, in *Proceedings of the 7th Workshop on Linked Data in Linguistics (LDL-2020)*. Marseille: European Language Resources Association (ELRA), pp. 28–35.

Di Segni 2016 Di Segni, R. (ed.) (2016) *Talmud Babilonese: trattato Rosh haShanà*. Firenze: Giuntina.

Di Segni 2017 Di Segni, G. (ed.) (2017) *Talmud Babilonese: trattato Berakhòt*. Firenze: Giuntina.

Di Segni 2018 Di Segni, R. (ed.) (2018) *Talmud Babilonese: trattato Ta’anit*. Firenze: Giuntina.

Di Segni 2019 Di Segni, R. S. (ed.) (2019) *Talmud Babilonese: trattato Qiddushìn*. Firenze: Giuntina.

Di Segni 2020 Di Segni, R. S. (ed.) (2020) *Talmud Babilonese: trattato Chaghigà*. Firenze: Giuntina.

Di Segni 2021 Di Segni, D. (ed.) (2021) *Talmud Babilonese: trattato Betzà*. Firenze: Giuntina.

Di Segni 2022 Di Segni, R. S. (ed.) (2022) *Talmud Babilonese: trattato Sukkà*. Firenze: Giuntina.

Di Segni 2023 Di Segni, R. S. (ed.) (2023) *Talmud Babilonese: trattato Mo’ed Qatàn*. Firenze: Giuntina.

Di Segni 2024 Di Segni, R. S. (ed.) (2024) *Talmud Babilonese: trattato Sotà*. Firenze: Giuntina.

Giovannetti 2016 Giovannetti, E. et al. (2016) “Traduco: a collaborative web-based CAT environment for the interpretation and translation of texts”, *Digital Scholarship in the Humanities*, 32(suppl_1), pp. i47–i62. Available at: https://doi.org/10.1093/llc/fqw054 (Accessed: 29 August 2025).

Giovannetti 2020 Giovannetti, E. et al. (2020) “The terminology of the Babylonian Talmud: extraction, representation and use in the context of computational linguistics”, *Materia Giudaica*, XXV. Firenze: Giuntina.

Giovannetti 2021 Giovannetti, E. et al. (2021) “An ontology of masters of the Babylonian Talmud”, *Digital Scholarship in the Humanities*, fqab043. Available at: https://doi.org/10.1093/llc/fqab043 (Accessed: 29 August 2025).

Giovannetti et al. 2024 Giovannetti, E. et al. (2024) “Maia: an open collaborative platform for text annotation, e-lexicography, and lexical linking”, *Umanistica Digitale*, 8(18), pp. 27–52. Available at: https://doi.org/10.6092/issn.2532-8816/19705 (Accessed: 29 August 2025).

Goldschmidt 1897-1936 Goldschmidt, L. (ed.) (1897–1936) *Der Babylonische Talmud: Mit Einschluss der vollständigen Mischnah und Commentaren*. Berlin: Jüdischer Verlag

Goodwin, Dolbear, and Hart 2008 Goodwin, J., Dolbear, C., and Hart, G. (2008) “Geographical linked data: the administrative geography of Great Britain on the semantic web”, *Transactions in GIS*, 12, pp. 19–30.

Jastrow 1926 Jastrow, M. (1926) *A Dictionary of the Targumim, the Talmud Babli and Yerushalmi and the Midrashic Literature*. Available at: http://www.tyndalearchive.com/TABS/Jastrow/index.htm (Accessed: 29 August 2025).

Marchi 2022 Marchi, S. et al. (2022) “Un esperimento di visualizzazione grafica della terminologia del Talmud babilonese”, in *AIUCD 2022 - Proceedings*. Lecce.

Martín-Chozas 2022 Martín-Chozas, P. et al. (2022) “TermitUp: generation and enrichment of linked terminologies”, *Semantic Web*, 13(6), pp. 967–986.

McCrae et al. 2016 McCrae, J. et al. (2016) “The Open Linguistics Working Group: developing the Linguistic Linked Open Data Cloud”, in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. Paris: European Language Resources Association (ELRA), pp. 2435–2441.

McCrae et al. 2017 McCrae, J. et al. (2017) “The OntoLex-Lemon model: development and applications”, in *Electronic Lexicography in the 21st Century: Lexicography from Scratch. Proceedings of eLex 2017*. Brno, pp. 587–597.

Montiel-Ponsoda et al. 2015 Montiel-Ponsoda, E. et al. (2015) “Towards the integration of multilingual terminologies: an example of a Linked Data prototype”, in *Terminology and Artificial Intelligence (TAI)*, pp. 205–206.

Pirrelli and Battista 2000 Pirrelli, V. and Battista, M. (2000) “The paradigmatic dimension of stem allomorphy in Italian verb inflection”, *Rivista di Linguistica*, 12, pp. 307–379.

Safra 2003-ongoing Safra, E. J. (ed.) (2003–ongoing) *Talmud Bavli: Édition française Edmond J. Safra*. Brooklyn, NY: Mesorah Publications.

Saponaro, Giovannetti, and Sciolette 2022 Saponaro, D., Giovannetti, E., and Sciolette, F. (2022) “From religious sources to computational resources: approach and case study on Hebrew terms and concepts”, *Materia Giudaica*, XXVII. Firenze: Giuntina.

Satlow 2006 Satlow, M. L. (2006) *Creating Judaism: History, Tradition, Practice*. New York: Columbia University Press.

Schottenstein et al. 1990-2005 Schottenstein, M. et al. (eds.) (1990–2005) *The Schottenstein Edition of the Talmud Bavli*. Brooklyn, NY: Mesorah Publications.

Schäfer 2007 Schäfer, P. (ed.) (2007) *The Cambridge Companion to the Talmud and Rabbinic Literature*. Cambridge: Cambridge University Press.

Sciolette 2024 Sciolette, F. (2024) “Modeling linking between text and lexicon with OntoLex-Lemon: a case study of computational terminology for the Babylonian Talmud”, in *Proceedings of the 9th Workshop on Linked Data in Linguistics @ LREC-COLING 2024*. Torino: ELRA and ICCL, pp. 103–107. Available at: https://aclanthology.org/2024.ldl-1.13 (Accessed: 29 August 2025).

Sciolette et al. 2024 Sciolette, F. et al. (2024) “CompL-it: a computational lexicon of Italian language”, *AIDA Informazioni*, 42(3–4), pp. 119–148. Available at: https://doi.org/10.57574/596545646 (Accessed: 29 August 2025).

Sciolette, Giovannetti, and Marchi 2023 Sciolette, F., Giovannetti, E., and Marchi, S. (2023) “LexicO: an Italian computational lexicon derived from Parole-Simple-Clips”, *Umanistica Digitale*, 7(15), pp. 169–193. Available at: https://doi.org/10.6092/issn.2532-8816/15176 (Accessed: 29 August 2025).

Steinsaltz 1965-2010 Steinsaltz, A. (1965–2010) *Talmud Bavli im Perush Steinsaltz*. Jerusalem: Israel Institute for Talmudic Publications.

Steinsaltz 2012-2020 Steinsaltz, A. (2012–2020) *The Koren Talmud Bavli: With Commentary by Rabbi Adin Even-Israel Steinsaltz*. Jerusalem: Koren Publishers.

Strack and Stemberger 1996 Strack, H. L. and Stemberger, G. (1996) *Introduction to the Talmud and Midrash*. Translated by Markus Bockmuehl. Minneapolis: Fortress Press.

Wilkinson 2016 Wilkinson, M. D. et al. (2016) “The FAIR Guiding Principles for scientific data management and stewardship”, *Scientific Data*, 3.

This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License.