

The Challenge of the Talmud in Italian: Tools and Computational Linguistics in the Service of Translation

Mafalda Papini

Cnr-Istituto di Linguistica Computazionale, Italia

Emiliano Giovannetti

Cnr-Istituto di Linguistica Computazionale, Italia

Abstract The translation of the Babylonian Talmud into Italian constitutes a true challenge. To address the numerous issues posed by translating such a vast and complex text, the Traduco system was developed to support translators through collaborative features, a translation memory, distributional semantics, glossary management, and annotation tools. Additionally, recent experiments have explored the integration of large language models to assist in the translation and the construction of a Talmud Knowledge Base designed to provide a formal representation of the terminological and conceptual data identified in the Talmud.

Keywords Computer-Assisted Translation. Translation Memory. Babylonian Talmud. Traduco. Large Language Models.

Summary 1 Introduction. – 2 Translating the Talmud: One Text, Many Complexities. – 2.1 Notes on the Nature of the Talmud. – 2.2 Structural Complexity and Dimension. – 2.3 Linguistic Complexity. – 2.4 Thematic Intricacy and Variety. – 3 From Translation Issues to Technological Solutions. – 3.1 Handling Corpus Size. – 3.2 Solutions to the Management of Structural Complexity. – 3.3 Dealing with Linguistic Complexity. – 3.4 Addressing Thematic Intricacy and Variety. – 4 Work in Progress. – 4.1 The Talmud Knowledge Base. – 4.2 LLMs and TM Integration. – 5 Conclusions.



Peer review

Submitted 2025-07-14
Accepted 2025-09-09
Published 2025-11-24



Open access

© 2025 Papini, Giovannetti | 4.0



Citation Papini, Mafalda; Giovannetti, Emiliano (2025). "The Challenge of the Talmud in Italian: Tools and Computational Linguistics in the Service of Translation". *TranScript*, 4, 135-154.

DOI 10.30687/TranScript/2785-5708/2025/01/005

Happy is the man who finds wisdom, the man who
attains understanding.

(*Proverbs*, 3:13)

1 Introduction

This quote from the *Tanakh*, like many others, highlights the importance of knowledge and the transmission of wisdom, and the crucial role they can play in the very fabric of human history. In an increasingly interconnected world, the sharing of traditions and cultures becomes essential for our collective enrichment.

In this context, the project to translate the Babylonian Talmud into Italian¹ emerges as a bridge – accessible to all – towards understanding and appreciating Jewish culture. Making the wisdom of the Talmud² accessible through its translation into Italian opens the door to broader participation and a deeper comprehension of the rich Jewish tradition.

Currently, initiatives aimed at translating the Babylonian Talmud around the world represent a significant commitment to preserving and disseminating the rich heritage of Jewish culture. These efforts span multiple languages and communities, underlining the Talmud's universal importance and the widespread desire to make it accessible to a broader audience.

In the United States, the *Soncino Press* project has served as a landmark for English translations of the Talmud (Epstein 1935-52). Launched in the 1930s, this project produced a complete translation of the Babylonian Talmud with commentary, offering English-speaking scholars a key resource for the study and understanding of the text. Among the most well-known English translations of the Talmud is also the Steinsaltz edition, originally conceived as a translation into Modern Hebrew (Steinsaltz 1965-2010), and later developed into an English version published by Koren Publishers Jerusalem (Steinsaltz 2012-20). Another widely used translation – available in English, French, and Modern Hebrew – is the Schottenstein edition of the Babylonian Talmud, published between 1990 and 2005 by ArtScroll (Scherman, Zlotowitz 1990-2005).

Beyond English, the Talmud has been translated – or is in the process of being translated – into many other languages. Among the

1 Progetto traduzione Talmud Babilonese. <https://www.talmud.it/en/>.

2 The Talmud is one of the main sacred texts of the Jewish tradition and is essentially a commentary on the *Mishnah*, a collection of laws and interpretations of the Torah that were initially transmitted orally (see §2.1).

Spanish translations, a notable example is the *Proyecto TaShema*,³ launched in 2012 and carried out in Jerusalem within the framework of the *yeshiva* directed by Rav Yaakov Benaim. In the French-speaking world, the most advanced translation is undoubtedly the *Édition Edmond J. Safra*, which offers a French version of ArtScroll's Schottenstein edition. This translation, started in 2003, had, at the time of writing, completed and published nine tractates.⁴ A partial Arabic translation of the Babylonian Talmud has also been produced. At the Middle East Studies Center in Jordan, more than 90 scholars – both Muslim and Christian – worked for six years on an Arabic version, resulting in the 2012 publication of 20 volumes (Jeremy 2012). Another noteworthy translation is the Goldschmidt Talmud (Goldschmidt 1897-1936), a German translation edited by Lazarus Goldschmidt over a span of forty years.

In this landscape, the Italian translation of the Babylonian Talmud stands as a significant addition to the international panorama of Talmudic translation. Italian, with its rich cultural and literary history, offers a unique context for interpreting and understanding the Talmudic text, opening up new perspectives and insights.

However, the undertaking of translating the Talmud into Italian would not have been possible without the support of digital technologies. In particular, the collaborative CAT tool Traduco (Giovannetti et al. 2016), specifically developed for the translation of the Talmud into Italian, has played a fundamental role, facilitating cooperation among translators and accelerating the translation process. Yet it is not only human collaboration that made this achievement possible: a range of language technologies has contributed to making the translation smoother and more accurate, allowing the essence and richness of the original text to be captured.

This article will explore the significance and impact of the Italian translation of the Talmud, with particular focus on the crucial role that technology has played in this ambitious translation project. We will also present the latest technological developments in the project, relating both to a complete redesign of the interface and to new lines of research (see §4) involving, on the one hand, the creation of a dedicated knowledge base and, on the other, the first applications of large language models (LLMs) to support translation.

3 *Proyecto Tashema: Unidos por el Talmud*. Tashema. <https://www.tashema.es/el-proyecto>.

4 *Édition Edmond J. Safra du Talmud Bavli – Le Talmud en français*. https://www.talmud-en-francais.fr/#dernieres_nouvelles.

2 Translating the Talmud: One Text, Many Complexities

Stating that the Babylonian Talmud is a complex text comes as no surprise. As Rabbi Riccardo Shemuel Di Segni writes about the Talmud in the introduction of the Italian translation of the *Berakhot* tractate:

La complessità del testo talmudico ne rende praticamente impossibile lo studio senza l'aiuto di guide e di opere di commento.

The complexity of the Talmudic text makes it virtually impossible to study without the aid of guides and commentaries. (Di Segni 2017, XIX; Author's translation)

Not by chance, the Italian translation of the Talmud was conceived both as a domesticating and foreignizing translation⁵ which “si ripropone di trasportare il lettore all'interno del mondo del Talmud” (seeks to transport the reader into the world of the Talmud; Di Segni 2017, XXXI), with the precise aim of being understandable even to those who are not scholars of the Talmud.

More in general, every act of translation inevitably involves interpretation. The illusion of a perfect equivalence between languages quickly fades when we realize that, in most cases, a literal translation fails to convey the full and profound meaning of the source text. This issue becomes dramatically more pronounced in the case of ancient and layered texts such as the Babylonian Talmud, whose dense, elliptical, and allusive discursive form renders literal translation not only inadequate but often misleading or incomprehensible.

Translating the Talmud is not merely about transferring words from one language to another; it is above all about reconstructing a conceptual and cultural universe that is profoundly distant from the contemporary one. For this reason, an effective translation requires deep knowledge of both the historical-cultural context in which the original text was produced and the cultural framework of the modern reader. Only through this dual expertise can the meaning of the text be made accessible without betraying its essence.

In the remainder of this section, the main features that make the Talmud a challenging text to be translated will be introduced. In 3, the techniques (and the related technologies) adopted to address these issues will be described.

⁵ This refers to the two translation strategies described by Lawrence Venuti (1995), according to whom ‘domestication’ occurs when the translator adapts the text to the value system of the target language in order to reduce the reader’s sense of alienation, while ‘foreignization’ occurs when the target values are forced in order to highlight the linguistic and cultural differences of the translated text.

2.1 Notes on the Nature of the Talmud

Alongside the *Written Torah* (the Pentateuch), Judaism developed a parallel tradition of authoritative and sacred rules and interpretations, known as the *Oral Torah*, which was transmitted orally for many centuries. At the end of the second century of the Common Era, due to the loss of Jewish political independence and the dispersion into the Diaspora, the Sages (called *Tannaim*) decided to organize and write down this oral heritage, giving rise to the *Mishnah*. The work of study did not end with the *Mishnah*, and centuries of this ongoing scholarly activity gave birth to the Talmud.

The Talmud is essentially a commentary on the *Mishnah* and is composed of two parts – the *Mishnah* (divided article by article) and the *Gemara* (the commentary on each article). The *Gemara* includes traditions from various schools, explains the rules in the *Mishnah* (their sources, reasons, and the meanings of words), compares the *Mishnah* with other *Tannaitic* traditions, and discusses new cases (Di Segni 2016, XIV). The discussions in the *Gemara* expand to various topics through associations of ideas and analogies. A significant portion of these ‘extensions’ does not have strictly legal implications and is referred to as *Aggadah* (narrative), which includes biblical exegesis, stories, and moral teachings.

The Talmud is an extraordinarily complex work that differs significantly from traditional encyclopedic structures, presenting itself instead as the ultimate ‘hypertext’. Its complexity stems not only from the vastness of its content but also from the unique methodology of thought and composition.

2.2 Structural Complexity and Dimension

Since the seminal printed edition by Daniel Bomberg (1519-23), the layout of the Babylonian Talmud has been standardized: the Talmudic text appears in the center, Rashi’s commentary on one side, and the *Tosafot* on the other, using different typefaces (‘square’ script for the main text and ‘Rashi’ script for the commentaries). Citations are made by indicating the tractate name, folio number, and side (e.g., *Berakhot* 11a).

The structure is based on continuous interaction between the *Mishnah* and the *Gemara*. The *Mishnah* is a code of legal norms divided into six ‘orders’ (*sedarim*), each subdivided into tractates (*masekhtot*), chapters, and individual articles (*mishnayot*). The *Gemara* is the commentary and analysis of these articles, characterized by a continuous flow of questions, answers, objections, and refutations. This style creates an ideal ‘symposium’ in which sages of various generations engage in dialogue. The text does not merely present

conclusions; it also includes all the hypotheses and possibilities formulated during the discussions, making it a document of ongoing creative process. Despite its precise lexical editing, the Talmud is built upon free associations and mental connections, appearing disorganized but highly meticulous in detail and language.

In the context of this translation project, we chose to maintain a structure of the text as close to the tradition as possible, providing for a subdivision of the work into tractates, chapters, blocks and logical units. Specifically, the block (*sughyà*) describes a portion of the text dedicated to a specific theme or a narrow aspect of the general topic, according to the subdivision identified in the edition by Rabbi Adin Steinsaltz. Logical units, on the other hand, are traditionally referred to by the Talmudic Masters as *chalaqim* and serve to account for the dialogical structure that the commentary on the *Mishnah* assumes, as seen in the example in fig. 1, where objections (*obiezioni*) and responses (*risposte*) alternate. Another structural feature that is marked is the numbering of the paragraphs, included to allow easier orientation in the original Hebrew-Aramaic text, where the distribution can be complex for those not well-versed in the subject.

Regarding the size of the Babylonian Talmud, the total number of folios (*dappim*) is 2.711, collected in a total of 37 tractates. The dimensions of the Talmudic tractates vary greatly, ranging from the 8 *dappim* of Tractate Tamid to the 176 *dappim* of Tractate Bava Batra.

It is important to note that each folio (*daf*) includes both the recto and verso sides (each of which is known as *amud*). Classical editions of the Babylonian Talmud are printed on large folio-sized pages, and this pagination has remained unchanged in subsequent editions, with citations referring to the tractate name, the folio number, and the side (a-b) (Di Segni 2016, XXX).

2.3 Linguistic Complexity

Another distinctive feature of the complexity of the Talmud is its language. While the Aramaic-Hebrew dialect was commonly spoken among Babylonian Jews for generations, it was unknown elsewhere, and, even in Babylonia, Aramaic became less widespread. The Talmud features various linguistic varieties, ranging from different stages of the Hebrew language (biblical, Mishnaic, and Amoraic) and Aramaic (Babylonian and Palestinian) to direct borrowings from Akkadian, Ancient Greek, Latin, Middle Persian (Pahlavi), Syriac, and Arabic. This linguistic richness brings with it certain challenges, particularly considering the skill set required of individual translators.

Additionally, the text is known for its extreme conciseness and sometimes cryptic formulaic language. This results in the presence

of repeated formulas,⁶ especially for introducing specific types of logical units, while the extremely stringent form of the discourse presupposes a thorough knowledge of the context and themes in order to compensate for the implied elements within the text.

2.4 Thematic Intricacy and Variety

The inclusion of heterogeneous materials heightens the complexity of the Talmud: in addition to the Mishnah's rules, the Gemara incorporates *baraitôt* (Tannaitic teachings excluded from the Mishnah) and *Aggadah* (which includes biblical exegesis, narratives, and moral teachings). This thematic variety means the Talmud addresses virtually every aspect of reality, from jurisprudence to religious history, from exegesis to philosophy, and even medical advice and historical anecdotes.

The difficulty of studying the Talmud also lies in its inherent purpose: it is not a schematic or orderly study book, but a reproduction of 'life itself'. The Talmudic sages were not solely concerned with practical consequences but with the 'intrinsic problematics' of issues. They often explored hypothetical or 'bizarre' problems seemingly distant from reality (like the 'tower floating in the air' or the man with an animal's head), which serve as models for exploring broader principles and demonstrative methods. This focus on debate and deep analysis, rather than mere practical utility, is a defining characteristic.

Finally, the Talmud was never 'completed'. It demands active student participation, inviting readers to become part of the discussion, pose questions, and raise doubts. The possibility of questioning and challenging – even with full awareness of the text's sanctity – is a core principle that makes Talmudic study a continuous process of *renewal* and discovery of new perspectives.

3 From Translation Issues to Technological Solutions

According to what emerges from the previous section, a computational system designed to support the translation of complex texts like the Babylonian Talmud must therefore go beyond simply accelerating the translation process. It must be capable of identifying difficult passages and suggesting interpretative – albeit not always definitive – alternatives. Furthermore, it should provide tools for representing

⁶ For example, the word תָּנִי is usually translated with the phrase "è stato insegnato" (it has been taught).

and handling contextual information – historical, linguistic, and conceptual – that can assist the translator in faithfully reconstructing the original meaning in a living and accessible language. Only in this way can we hope to bridge, at least in part, the gap that separates the modern reader from the great texts of tradition.

3.1 Handling Corpus Size

The Talmud, as previously emphasized, constitutes a textual corpus of considerable size. To undertake the translation and subsequent publication of such a text, the collaboration of a large team of experts with a wide variety of skills and levels of training has been and continues to be necessary. At the time of writing, there are around 70 active experts, including translators, revisors, and editors. Contributors were selected in Italy and abroad on the basis of the skills required. In particular, translators were required to be native speakers of Italian, have knowledge of Hebrew and Aramaic, and be familiar with the Talmud. In the case of revisors and editors, in addition to more strictly linguistic skills, a very thorough knowledge of the Talmud was required. This distribution of personnel across space and time brings with it great richness, but also a series of challenges regarding both the coordination of work and the synchronization of data. To support such an endeavor, it was decided to develop Traduco, the collaborative web platform dedicated to assisting in the translation of complex texts, as mentioned in the introduction (see 1).

The platform, developed since 2012 with numerous versions, is the result of an incremental refinement process. Starting from the initial specifications, various features have been added to Traduco based on both the emerging needs of its users and advancements in linguistic research. Furthermore, given the diversity of the target audience, the usability of the platform has been validated in the field, integrating continuous feedback from users into the development process.

Traduco was created as a CAT software that is directly accessible via a browser. Developed since 2012 through various versions, this platform has allowed for overcoming the typical issues of desktop systems related to installation difficulties and compatibility with different operating systems. Even more significant is its nature as a natively collaborative platform, which is expressed through user management. Each user on Traduco is characterized by a specific role that defines their activities and access and modification permissions, ranging from a simple Viewer to an Administrator. Work, whether it involves translation or revision, is distributed through a resource assignment feature, where each resource is one of the text portions defined within the project. The combination of the role and the assigned resources allows the system to identify the permissions of

individual users, particularly regarding the ability to activate editing mode in the translation environment.

However, the collaborative aspect of the platform is not limited to the simple assignment of roles and resources. First, a system for managing concurrent modifications has been implemented to ensure that users do not simultaneously intervene on the same data, thus guaranteeing data consistency. Secondly, both for the translated text and for glossaries and other contextual data, a version history feature is available that allows for the viewing and potential restoration of previous versions.

3.2 Solutions to the Management of Structural Complexity

The Talmudic corpus, in addition to its physical structure (initially organized into tractates), is characterized by a complex logical structuring, particularly regarding the commentary on the *Mishnah*, which takes the form of an ongoing debate (see §2.2). The edition of the Italian translation of the Talmud has set two important objectives in relation to this complexity: on one hand, to preserve the traditional physical structure of the text (which is also necessary for using the standard citation format), and on the other hand, to simplify its understanding and reading in parallel with the original text. To meet these needs, Traduco allows for the management of various data related to the structuring of the resource.

The screenshot displays the Traduco platform interface. On the left is a 'Risorse' (Resources) navigation tree. It is structured as follows:

- ▼ 1 Berakhōt
 - > 1 Meematai – Da quando
 - Capitolo 2. Hayà
 - ▼ 2 qorè – Stava leggendo
 - > 1 MISHNÀ
 - ▼ 2 GHEMARÀ Intenzione nell'applicazione delle mitzvot
 - 1 [OBIIEZIONE]
 - 2 [RISPOSTA]
 - 3 [OBIIEZIONE]
 - 4 [RISPOSTA]
 - > 3 Regole riguardanti la lettura dello Shemà'
 - > 4 Intenzione e lettura dello Shemà'

On the right is a translation table for the selected resource 'GHEMARÀ Intenzione nell'applicazione delle mitzvot'. The table has columns for a reference number (13a), a circular icon with a number (5), the original Hebrew text, the Italian translation, and the user responsible for the entry.

GHEMARÀ Intenzione nell'applicazione delle mitzvot				
[OBIIEZIONE]				
13a	5	שָׁמַע מִיָּדָה	Deducit de eis , ossia dalla mishnà che dice all'inizio SE AVEVA INTENZIONE DI USCIRE DALL'OBLIGO DELLA LETTURA DELLO SHEMA' È USCITO D'OBLIGO.	usavni
13a	5	מִמַּת צִוְּיֹת קוֹרֵא	che le (mitzvot) richiedono intenzione di uscire d'obbligo?	decachenco
[RISPOSTA]				
13a	5	כִּי אִם כִּן לֹא –	Cosa significa SE AVEVA INTENZIONE quando ha letto lo Shemà'?	jottes-ber
13a	5	לִרְאוֹת.	Significa che aveva intenzione di leggere la Torà, senza aver avuto l'intenzione di uscire d'obbligo dalla mitzvà dello Shemà'.	jottes-ber
[OBIIEZIONE]				
13a	5	לִרְאוֹת?	Con'è possibile che l'espressione aveva INTENZIONE si riferisca all'intenzione di leggere?	jottes-ber
13a	5	וְכִי אִם קוֹרֵא	Stava appunto leggendo! Perciò sicuramente aveva intenzione di leggere.	usavni

Figure 1 Resource navigation tree and translation table

In fig. 1, we can see part of the Traduco translation interface. In this figure, on the left, we find the quick navigation system for resources, which effectively illustrates the hierarchy of the text, while the rest of the screen consists of the translation table. Regarding the physical structure of the Talmud, the main hierarchical level is represented by the division of the text into tractates (in this case, we are within the tractate *Berakhòt*), while the data related to the division into individual *amudim* is included in the translation table at the level of individual text segments, as the end of a *daf* does not necessarily correspond to the closure of one of the logical divisions described in §2.2.

In Traduco, the possibility of arbitrarily dividing the text into multiple nested logical groupings (the chapters, blocks, and logical units introduced earlier) has been provided, each with its own title and potentially accompanied by descriptions and attachments. This logical structuring can also be seen represented in the left column of fig. 1, where the groupings are hierarchically organized within the individual tractates.

The different types of divisions incorporated into the system are, as mentioned, functional for managing the printed aspect of the translated text. Users can gain insight into the final effect of their work by consulting a section of editorial preview, of which we present an excerpt in fig. 2. As we can see, the titles of the different logical levels of the text have a descriptive function and guide the reader for better understanding. An example is the titles of the logical units, which, directly integrated into the text, allow for a clear distinction of the sequence of responses and objections that would otherwise flow continuously.

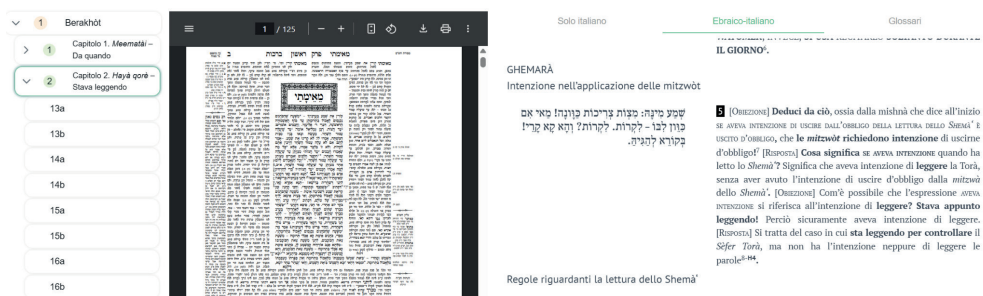


Figure 2 Editorial preview side-by-side with the original text

3.3 Dealing with Linguistic Complexity

The language with which Talmud translators engage is, as described in §2.3, complex from various perspectives. The Hebrew-Aramaic linguistic variety is just the first hurdle in the translation process, as the language used in the Talmud is characterized by its cryptic nature. The constructions are extremely concise, and the text features the use of formulas and implications whose understanding presupposes a full mastery of the subject. In such a context, considering a literal translation would result in a text of limited readability, likely intended for a narrow audience of experts, which is entirely at odds with the project's objectives.

In the project of translating the Babylonian Talmud, the text is rendered into Italian through an interpretative translation, unfolding the formulas and thus expanding the text, also thanks to the inclusion of contextual elements that we will discuss in more detail in §3.4. However, it remains necessary to preserve a distinction between the original text and the explanatory additions, a distinction that must also be reflected in the printed text. To this end, Traduco has developed a dedicated text editor that allows for the use of a predefined markup scheme to differentiate translation portions, as well as to include additional information such as notes and glossaries. An example of this is shown in the excerpt in fig. 3, where we find three different types of markup. In particular, the entire portion of literal translation is in bold, while explanatory insertions are in italics; furthermore, citations from the *Mishnah* are distinctly marked using small caps.

תנא ה"א קא' דקתני "מאימתי?"	Il tannà (autore della nostra mishnà) a quale norma si riferisce quando dice "DA QUANDO"? ^③
-----------------------------	--

Figure 3 Extract of the translation table with marking of the translated text

As emphasized from the beginning, Traduco was created as a CAT platform and, like all systems of this kind, features a suggestion system based on a translation memory (TM). In practice, for each string to be translated, the platform searches among the already translated strings (which constitute the TM) for Hebrew-Aramaic strings that correspond, even partially, and proposes them to the user as translation suggestions, accompanied by a similarity score. As of the writing of this article, Traduco's memory has reached a considerable size, with over 300.000 pairs of translated strings capable of providing an exact translation suggestion (excluding any contextual additions) in over 50% of cases. This fundamental mechanism of CAT systems has been further refined in Traduco with

two specific strategies: vowel removal from the text and the adoption of distributional semantics techniques.

First, the search for similar translated strings is performed on the unvocalized Hebrew-Aramaic strings. Indeed, since there is no lemmatizer for the idioms of the Talmud, using strings without vowels allows for a sort of approximate lemmatization and yields useful matches even in the case of inflected forms.

The second type of strategy applied is the integration of distributional semantics approaches into Traduco (Bellandi et al. 2016). The distributional hypothesis states that, given two words represented as n-dimensional vectors where each dimension records the number of times a word appears in a certain context, the smaller the distance between the vectors in the distributional space, the greater their semantic similarity (Sahlgren 2008). For instance, if we consider the words ‘miracles’ and ‘wonders’, these, while being distinct lemmas, are semantically related, and consequently, the translation of one can be significant for the other [fig. 4]. Considering the distributional hypothesis in Traduco means that distinct words in two sentences can have a weight greater than 0 in the calculation of the similarity percentage if they are semantically close.



Figure 4 Applying the distributional hypothesis, the ‘cost’ required to go, for example, from the word מִסִּימָנוֹת (miracles) to the word מִפְּלִאוֹת (wonders) is reduced because they are recognized as synonyms, consequently their similarity increases significantly

The suggestion system has thus been enhanced to meet two specific needs. On one hand, it speeds up the translation process (especially concerning standardized formulas that often recur across different tractates), and, on the other hand, it serves as a stylistic normalizer. As mentioned, the translation of the Talmud involves numerous translators who are variously distributed across space and time. This broad collaboration, while allowing for the tackling of such an extensive and complex text, complicates the production of a homogeneous Italian text. Despite the establishment of strict editorial guidelines, the risk of stylistically divergent translations is high and adds an additional workload during the revision phase. The

translation memory suggestions address this aspect. When selecting a string to translate on the platform, the system displays a list of related suggestions, and a double-click on one of them is sufficient to apply its translation to the string being worked on. This promotes consistent lexical and stylistic choices at every point in the work, reducing the need for subsequent interventions.

3.4 Addressing Thematic Intricacy and Variety

The Talmud, as we have mentioned, is a reproduction of ‘life itself’, reflecting a variety of topics and themes. This extreme heterogeneity of content has significant implications for the entire process of translation, revision, and consumption of the text.

First, translators and reviewers cannot be perfect experts of all the knowledge present in the Talmud, which spans historical facts to philosophical and religious concepts. Therefore, linguistic competence is not always accompanied by a deep understanding of the themes addressed in a specific section, making the role of so-called ‘content reviewers’ necessary. These reviewers verify not only the quality and coherence of the translation but also the accuracy of the concepts conveyed. As shown in fig. 1, translators work with a division of the text at roughly the level of sentences, a method that, however, makes continuous engagement with the text more challenging. The system then supports the reviewers by providing them with various preview views, both as pop-ups within the translation section and as a standalone section [fig. 2]. This type of engagement allows for a reading experience as close as possible to that of the final reader, simplifying the revision process and providing a clear idea of the final editorial product.

However, managing the thematic variety of the Talmudic work cannot be solely entrusted to *a posteriori* revision processes. It is essential that the working platform provides elements to support translators, offering contextual information that can guide them in accurately rendering concepts.

Traduco allows enriching the translation with a range of contextual information. The first elements are the glossaries, descriptive entries that can be created and manipulated either directly from the text (highlighting a portion of translation being worked on) or from a dedicated section of the interface. Glossary entries are typed information, used to distinguish different reference domains (among which we have the types Concept and Linguistics),⁷ which are linked

7 Glossary entries typified as Linguistics are also used to mark those portions of the text pertaining to languages other than Hebrew and Aramaic, marking both borrowings

to inline annotations of the translated text. Some types of glosses can be further expanded by linking them to attachments. The system, through a dedicated section, allows connecting a glossary entry to an extended biography (usually the lives of rabbis mentioned in the text) or to a file, such as an image, useful for fully describing an object or a plant.

Within the translated text, it is also possible to insert typed notes, some functional to the translation work (e.g., system notes), and others useful additions of external content. For example, *Halakha*-type notes refer to religious norms and contextualize the references present in the text.

It is essential to note that all these contextual elements added to accompany the translation not only help translators and reviewers in their work but also significantly assist future readers. Indeed, glossaries, notes, and attachments find their place in the printed edited text, narrating the socio-cultural panorama in which the Talmud is set and which, often, is far from the environment and knowledge of the reader, especially in a work that aims to spread this important element of Jewish tradition throughout the Italian community.

4 Work in Progress

The Talmud Project, in addition to being a translation project, is also a research initiative. For this reason, numerous activities are underway aimed at experimentation and innovation, both in the field of Talmudic studies and in the area of computer-assisted translation. Below, two specific areas are outlined, relating to the development of the Talmud Knowledge Base and the experimentation with artificial intelligence (AI) techniques to support the translation.

4.1 The Talmud Knowledge Base

The Talmud Knowledge Base (TKB) is an onto-terminological resource, currently under development at the time of writing this article, designed to provide a formal representation of the most relevant information elements identified in the Babylonian Talmud.

In general terms, an onto-terminological resource integrates ontological (i.e., conceptual structuring) and terminological (i.e., sets of terms and definitions) dimensions with the aim of constructing an

and any etymologies. This is to account for the variety of languages that make up the Talmudic work, for which see §2.3.

organized representation of knowledge within a specific disciplinary domain (Roche 2003; 2012).

The essential components of such a resource include, on the one hand, the ontology, which defines the concepts, relationships, and properties of a domain, offering a formal representation of knowledge, and on the other hand, the terminology, which collects the terms, definitions, and linguistic variants associated with the concepts identified in the ontology.

The terminological section of the TKB includes a collection of Hebrew, Aramaic, and Italian terms that are unique to the Talmudic field. These terms come from a manually validated set selected by domain experts during the translation phase and are part of the glossaries found in both the Traduco platform and the Talmud Project. However, the term set is expected to be expanded by integrating the results of an automatic extraction process applied to the Babylonian Talmud, as documented in Giovannetti et al. 2020.

Out of the seven glossary types⁸ developed within the project, three were utilized for building the TKB: Concept, Linguistics, and Nature. Each entry in the glossary provides the following elements: the category (e.g., Concept), the term label (e.g., *Giubileo*), the corresponding Hebrew term (e.g., גִּיּוּבִּיל), its transliteration (*yovèl*), the translation (included for terms left untranslated in Italian but presented in transliterated form, such as *Simchà*, meaning ‘gioia’), and a definition. Furthermore, each glossary entry is linked to every instance in the text where translators annotated words with that specific semantic reference.

As for the formalization of the resource, the OntoLex-Lemon model (McCrae et al. 2017) was adopted for the terminological component, and OWL⁹ was used for the conceptual component. At the current stage, most of the ontological part of the TKB consists of OWL ‘stub’ classes (i.e., classes that will be described in more detail as the work progresses), which are formally linked to the lexical senses of the corresponding denoting terms via the lemon:reference relation. However, two more developed ontological components are already in place: the first concerns the Talmudic Masters (Giovannetti et al. 2021), and the second – currently under development – is designed to describe the Second Temple of Jerusalem, including formal representations of events, places, figures, offerings, rituals, and so on. Quantitatively, as of the time of writing this article, the TKB consists of nearly 900 terms and approximately 1.000 concepts.

⁸ Namely: Concept, Linguistics, Idiom (referring to idiomatic expressions), Rabbi, Nature, Name, and Daily (referring to daily life).

⁹ OWL (Web Ontology Language) Overview. <https://www.w3.org/TR/owl-features/>.

The availability of a resource of this kind, developed starting from the terms actually present in a text and the concepts they express, makes it possible to implement advanced methods for accessing the text itself, thus facilitating its analysis and in-depth study, especially during the translation process.

In particular, the integration of a structured terminological component with a formally defined ontology allows translators to disambiguate meanings, trace the recurrence and semantic function of key terms across the text, and ensure consistency in their choices. Moreover, the ability to semantically annotate and link specific textual occurrences to shared conceptual representations fosters a deeper understanding of the underlying knowledge system of the Talmud. This not only enhances the quality of translation, but also contributes to its reproducibility, pedagogical usability, and alignment with digital scholarship standards.

4.2 LLMs and TM Integration

The rapid expansion of LLMs in all areas of knowledge has led to evaluating their potential application in supporting the translation of the Talmud.

A generative LLM enables the generation of natural language based on user input. Considering the most well-known example, ChatGPT, it is based on LLMs that interpret the user's request (or 'prompt') and construct a response based on that interpretation. Using this mechanism, models can also produce automatic translations, whose quality largely depends on the training data provided and their linguistic coverage.

Our hypothesis is that, given the linguistic and stylistic peculiarities of the Talmud, as well as the specific translation requirements of the project (stylistic homogeneity, explanatory expansions of the text, etc.), a generalist LLM is not capable of automatically providing an adequate translation in context. Considering the technical difficulties connected to specializing a model through training or fine-tuning,¹⁰ we drew inspiration from two well-known techniques: Retrieval Augmented Generation (RAG) (Lewis et al. 2020) and few-shot prompting (Brown et al. 2020) to define our approach.

As previously mentioned, Traduco is already equipped with a translation suggestion system based on a translation memory that, at the time of writing this article, consists of approximately 303.000

10 Fine tuning refers to the supervised adjustment of a pre-trained language model's parameters on a specific domain or task, typically requiring labeled data and significant computational resources.

pairs of Aramaic/Hebrew-Italian strings. Our methodology involves leveraging the translation memory to extend the LLM's query prompt. Specifically, the workflow we hypothesized involves invoking the model's support to supplement the translation memory in the absence of 'perfect' suggestions, i.e., with a 100% similarity. If a string to be translated presents at least three translation suggestions from the TM, with a similarity percentage between 50% and 99%, then the Aramaic/Hebrew-Italian pairs of suggestions are provided to the model as translation examples integrated into the prompt (Papini et al. 2025), which will combine with its innate translational capabilities.

We tested the methodology on 300 randomly extracted strings from the Talmud, all of which had a revised translation to use as a reference. Each string was also subjected to a direct translation by a model to compare the generated texts. Three standard machine translation metrics were considered for the evaluation: SacreBLEU (Post 2018), which evaluates lexical correspondence; BERTscore (Zhang et al. 2020), which considers semantic similarity; and Meteor (Lavie et al. 2007), which accounts for synonymy and morphological inflection. For each string translated by the model (LLM) or by the model combined with the TM (TM+LLM), the metrics were calculated in comparison with the string translated by the expert, yielding promising results. In particular, we calculated the difference between the metrics of the TM+LLM version and those of the LLM version, considering as positive those cases in which there was an increase in all values. The result is that 79% of the sample considered shows an increase in translation quality using the hybrid approach.¹¹ Let's consider, for example, the sentence "כִּי־נֶאֱמַר לָהּ הִנֵּה פְּרוּטָה", translated by the experts as "Dal momento che le ha detto: 'Mezza perutà' ha interrotto l'azione". We therefore produced a first translation by directly querying the model: "Poiché disse a lei mezza peruta, la interrompe". A second translation was then obtained by providing TM suggestions as context to the LLM: "Dal momento che le dice: 'Mezza perutà' ha interrotto l'azione". Even at first glance, the improvement achieved by hybridizing TM and LLM is noticeable, starting with the correct translation of *perutà* with its accented form, in line with the editorial standards of the project.¹²

We therefore engaged an expert to perform a qualitative assessment of the remaining 21%, i.e., 63 pairs of strings, obtaining the following breakdown: in 16 cases, the TM+LLM translation was the best; in 4 cases, the LLM translation was the best; in 18 cases,

¹¹ For further details, see Papini et al. 2025.

¹² The test results are available on a dedicated repository: <https://github.com/klab-ilc-cnr/TalmudAI-AIUCD2025>.

both translations were acceptable; while in the remaining 25 cases, neither was acceptable.

Some critical issues remain open, particularly regarding the explanatory expansion of the text. The model, even using the provided examples for selecting terms and constructions to use, proposes a literal translation.¹³ This is not an unexpected result, as expansions are usually the result of domain knowledge on the part of the translator or based on the content of the surrounding context of the single string, of which the model is not aware.

5 Conclusions

The translation of the Babylonian Talmud into Italian represents a true challenge: the Talmud is not merely a text, but a complex stratification of languages, references, discursive structures, and exegetical practices that require deep interpretive expertise. In this context, the role of digital tools and language technologies has proven essential. Platforms such as Traduco – with its collaborative features, translation memory enriched by distributional semantics, glossary and annotation management, and full workflow supervision – effectively support human translators, revisors, and editors in producing a comprehensive translation that is also accessible to non-experts in Talmudic studies.

In the future, a crucial role will also be played by the Talmud Knowledge Base, which, by integrating terminology and ontology, will enable a structured representation of the knowledge embedded in the text, enhancing both consistency and accessibility of the translation. At the same time, the tests conducted with LLMs have revealed promising potential, especially when combined with techniques such as Retrieval-Augmented Generation (RAG) and few-shot prompting. These approaches have shown improvements in the quality of translation suggestions, though they still fall short in generating explanatory expansions – an aspect that remains tied to the translator’s domain expertise.

Looking ahead, the project opens up new opportunities for interdisciplinary research, demonstrating how translation practices, software engineering, AI, and cultural studies can converge within a productive and scientifically innovative framework. Future developments will include expanding the talmudic ontology, refining

13 For example, the string, “בִּיחָדָה”, which the model translates as “voti”, was translated in the Talmud (and considered as the reference segment in the experiment) with the phrase “per quanto riguarda le offerte di voto”. However, the model with the translation examples provides a better translation (*nedarim*).

the semantic quality of automated suggestions, formalizing the adopted translation strategies, and integrating the project with user interfaces and educational resources to provide broader and more pedagogically effective access to the Talmudic heritage.

Bibliography

- Bellandi, A.; Benotto, G.; Di Segni, G.; Giovannetti, E. (2016). "Investigating the Application and Evaluation of Distributional Semantics in the Translation of Humanistic Texts: A Case Study". Orăsan, C.; Parra, C.; Barbu, E.; Feder, M. (eds), *2nd Workshop on Natural Language Processing for Translation Memories (NLP4TM 2016) = Workshop Proceedings* (Portorož, 28 May 2016). Paris: European Language Resources Association ELRA, 6-11.
- Bomberg, D. (ed.) (1519-53). *Talmud Bavli*. 1-12 vols. Venice: Daniel Bomberg.
- Brown, T.B. et al. (2020). "Language Models are Few-Shot Learners". Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M.F.; Lin, H. (eds), *Advances in Neural Information Processing Systems 33 (NeurIPS 2020) = Conference Proceedings* (virtual, 6-12 December 2020), 1877-901.
- Di Segni, R.S. (ed.) (2016). *Talmud Babilonese – Trattato Rosh haShanah*. Firenze: Giuntina.
- Di Segni, D.G. (ed.) (2017). *Talmud babilonese – Trattato Berakhot*. Firenze: Giuntina.
- Epstein, I. (ed.) (1935-52). *The Babylonian Talmud (English translation)*. 18 vols. London: Soncino Press.
- Giovannetti, E.; Albanesi, D.; Bellandi, A.; Benotto, G. (2016). "Traduco: A Collaborative Web-Based CAT Environment for the Interpretation and Translation of Texts". *Digital Scholarship in the Humanities*, 32(suppl. 1), 47-62. <https://doi.org/10.1093/llc/fqw054>
- Giovannetti, E.; Bellandi, A.; Del Grosso, A.M.; Marchi, S.; Piccini, S. (2020). "The Terminology of the Babylonian Talmud: Extraction, Representation and Use in the Context of Computational Linguistics". *Materia Giudaica*, XXV, 61-74.
- Giovannetti, E.; Albanesi, D.; Bellandi, A.; Dattilo, D.; Del Grosso, A.M.; Marchi, S. (2021). "An Ontology of Masters of the Babylonian Talmud". *Digital Scholarship in the Humanities*, 37, 3, 725-37. <https://doi.org/10.1093/llc/fqab043>
- Goldschmidt, L. (1897-1936). *Der Babylonische Talmud: Übersetzt und erläutert von Lazarus Goldschmidt*. 12 vols. Berlin: Jüdischer Verlag.
- Jeremy, S. (2012). *Jordan Group Translates Babylonian Talmud to Arabic*. March 28, 2012. <https://www.jpost.com/jewish-world/jewish-news/jordan-group-translates-babylonian-talmud-to-arabic>
- Lavie, A. et al. (2007). "METEOR: An Automatic Metric for MT Evaluation with High Levels of Correlation with Human Judgments". Callison-Burch, C.; Koehn, P.; Fordyce, C.S.; Monz, C. (eds), *Proceedings of the Second Workshop on Statistical Machine Translation* (Prague, 23 June 2007), 228-31. <https://aclanthology.org/W07-0734/>
- Lewis, P. et al. (2020). "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks". Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M.F.; Lin, H. (eds), *Proceedings of the 34th International Conference on Neural Information Processing Systems* (Vancouver, 6-12 December 2020), 9459-74.
- McCrae, J.P.; Bosque-Gil, J.; Gracia, J.; Buitelaar, P.; Cimiano, P. (2017). "The Ontolex-Lemon Model: Development and Applications". Kosem, I. et al. (eds), *Proceedings*

- of the Electronic Lexicography of the 21st Century* (Leiden, 19-21 September 2017), 587-97.
- Papini, M.; Albanesi, D.; Dattilo, D.; Giovannetti, E.; Marchi, S. (2025). "Experiments on the Use of LLMs for the Translation of the Babylonian Talmud". Rebor, S.; Rospocher, M.; Bazzaco, S. (eds), *Diversità, equità e inclusione: Sfide e opportunità per l'informatica umanistica nell'era dell'intelligenza artificiale*. = *Proceedings del XIV Convegno Annuale AIUCD2025* (Verona, 11-13 June 2025), 363-7. <https://doi.org/10.6092/unibo/amsacta/8380>
- Post, M. (2018). "A Call for Clarity in Reporting BLEU Scores". Bojar, O. et al. (eds), *Proceedings of the Third Conference on Machine Translation: Research Papers = Conference Proceedings* (Brussels, 31 October-1 November 2018), 186-91. <https://doi.org/10.18653/v1/W18-6319>
- Roche, C. (2003). "Ontology: A survey". *8th IFAC Symposium on Automated Systems Based on Human Skill and Knowledge 2003 = Conference Proceedings* (Göteborg, 22-24 September 2003), 187-92. [https://doi.org/10.1016/S1474-6670\(17\)37715-7](https://doi.org/10.1016/S1474-6670(17)37715-7)
- Roche, C. (2012). "Ontotermology: How to Unify Terminology and Ontology into a Single Paradigm". Calzolari, N. et al. (eds), *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)* (Istanbul, 21-27 May 2012), 2626-30. http://www.lrec-conf.org/proceedings/lrec2012/pdf/567_Paper.pdf
- Sahlgren, M. (2008). "The Distributional Hypothesis". *Italian Journal of Linguistics / Rivista di Linguistica*, 20, 1, 33-53.
- Scherman, N.; Zlotowitz, M. (eds) (1990-2005). *The Babylonian Talmud: The Schottenstein Edition*. 73 vols. Brooklyn, NY: ArtScroll / Mesorah Publications.
- Steinsaltz, A. (1965-2010). *Talmud Bavli Steinsaltz*. 45 vols. Jerusalem: Israel Institute for Talmudic Publications.
- Steinsaltz, A. (2012-20). *The Koren Talmud Bavli*. 41 vols. Jerusalem: Koren Publishers Jerusalem.
- Venuti, L. (1995). *The Translator's Invisibility*. London; New York: Routledge.
- Zhang, T.; Kishore, V.; Wu, F.; Weinberger, K.Q.; Artzi, Y. (2020). "BERTScore: Evaluating Text Generation with BERT". *International Conference on Learning Representations*. <https://openreview.net/forum?id=SkeHuCVFDr>