Computer Assisted Translation of Ancient Texts: the Babylonian Talmud Case Study

Andrea Bellandi, Alessia Bellusci, and Emiliano Giovannetti

Istituto di Linguistica Computazionale, Consiglio Nazionale delle Ricerche
via G. Moruzzi 1, 56124 - Pisa, Italy
{name.surname}@ilc.cnr.it

Abstract. In this paper, we introduce some of the features characterizing the Computer Assisted Translation web application developed to support the translation of the Babylonian Talmud (BT) in Italian. The BT is a late antique Jewish anthological corpus, which, as other ancient texts, presents a number of hurdles related to its intrinsic linguistic and philological nature. In this work, we illustrate the solutions we adopted in the system, with particular emphasis on the Translation Memory and the translation suggestion component.

1 Introduction

Computer-Assisted Translation (CAT) tools are conceived to aid in the translation of a text. At the very heart of a CAT tool there is always a Translation Memory (TM), a repository that allows translators to consult and reuse past translations, primarily developed to speed up the translation process. As the use of TMs spread during the 1990s, the market quickly shifted from the reduction of translation time to the reduction of costs, Gordon (1996). For example, phrases having an exact correspondence inside the TM were charged at lower rates. However, considering the nature of the texts we are working on (mainly ancient texts of historical, religious and cultural relevance), It is mandatory to safeguard the quality of the translation, beside optimizing its speed.

To translate an ancient text there is the need of (at least) two kinds of competences, both on the languages (source and target) and the "content". In this perspective users shall wear at the same time the hat of the translator and the scholar. Hence, a system developed to support the translation of ancient texts shall go beyond the standard set of functionalities offered by a CAT tool.

In this paper we discuss the experience matured in the context of the "Progetto Traduzione del Talmud Babilonese", monitored by the Italian Presidency of the Council of Ministers and coordinated by the Union of Italian Jewish Communities and the Italian Rabbinical College. In the context of the Project we are in charge of developing a collaborative web-based application to support the translation of the Babylonian Talmud (BT) into Italian by a group of authorized translators.

It is not the purpose of this paper to describe the full CAT system we are working on, which includes, among other things, semantic annotation features. Instead, we will focus here on the development of the TM and on the suggestion component, both designed for a community of translators working, in a collaborative way, on ancient texts such as the BT , whose plain and literal translation shall be enhanced to be intelligible. In this paper we will emphasize the features of our system w.r.t. those of the available open source CAT tools (*OpenTM*, *OmegaT*, *Transolution*, *Olanto*).

In Section 2 we discuss the peculiarities of the BT and how they fit for the BT translation by means of a TM. Section 3 describes the construction process of the TM, and presents all the features of our CAT system. Section 4 provides a preliminary evaluation of its performance. Finally, Section 5 outlines conclusions and future works.

2 The Babylonian Talmud

In the next sections, we provide a brief historical and linguistic analysis of the BT and discuss the relevant features of this rich textual corpus in relation to the development of a TM aimed at supporting its translation in modern languages.

2.1 Historical and Content Analysis

The Babylonian Talmud (BT) represents the most important legal source for Orthodox Judaism and the foundation for all the successive developments of *Halakhah* and *Aggadah*. Compiled in the Talmudic academies of late antique Babylonia (contemporary Iraq), the BT is based on sources from different epochs and geographical areas and maintained a fluid form at least until the sixth century CE. The BT corresponds to the effort of late antique scholars ("*Amoraim*") to provide an exegesis of the *Mishnah*, an earlier rabbinic legal compilation, divided in six "orders" ("*sedarim*") corresponding to different categories of Jewish law, with a total of 63 tractates ("*massekhtaot*"). Although following the inner structure of the *Mishnah*, the BT discusses only 37 tractates, with a total of 2711 double sided folia in the printed edition (Vilna, XIX century). The BT deals with ethics, jurisprudence, liturgy, ritual, philosophy, trade, medicine, astronomy, magic and so much more.

To this conspicuous volume of textual material and topics corresponds an intricate complex of heterogeneous sources. The BT includes (i.) quotations from the *Mishnah* ["*Mishnah*"], (ii.) long *amoraic* discussions of *mishnaic* passages aimed at clarifying the positions and lexicon adopted by the *Tannaim* ["*Gemarah*"], and (iii.) external *tannaitic* material not incorporated in the canonical *Mishnah* ["*Baraytot*"]. Intrinsically related to the *Mishnah*, the *Gemarah* develops in the form of a dialectical exchange between numerous rabbinic authorities belonging to different generations, epochs (*Tannaim* and *Amora'im*), and geographical areas (Graeco-Roman Palestine and Sassanid Babylonia). The thorough exegetical work on the *Mishnah* operated in the *Gemarah* implies discussing the words and phrases adopted in the *Mishnah*, speculating on the principles behind the *Mishnah* case laws, linking the different rules stated in the *tannaitic* text to the Bible, and harmonizing the contradictions occurring in the *Mishnah*. The *sugya'* represents the basic literary unit of the *Gemarah*, which displays the relevant textual material in a succession of questions and answers. To each *mishnah* ["statement"/ "law" from the *Mishnah*] corresponds one or more *sugyot*. Although constructed according to specific literary conventions, the succession of *sugyot* in the BT gives the perception of a fluent and live debate between rabbinic authorities. The BT attests to an extremely vast body of legal ("*halakhic*") and narrative ("*aggadic*") knowledge, which had been continuously transmitted and re-interpreted in late antique Judaism.

2.2 Linguistic and Stylistic Analysis

To the content and philological complexity of the BT, we ought to add the linguistic richness presented by this textual corpus, which inevitably affected our choices for the development of the web application for supporting its translation. In its extant form, in fact, the BT attests to (i.) different linguistic stages of Hebrew (Biblical Hebrew, Mishnaic Hebrew, Amoraic Hebrew), (ii.) different variants of Jewish Aramaic (Babylonian Aramaic and Palestinian Aramaic), and (iii.) several loanwords from Akkadian, ancient Greek, Latin, Pahlavi, Syriac and Arabic. In addition, all the different languages and dialectal variants are often alternated in the text, thus making difficult, if not impossible, to develop an automatic (statistical or pattern-based) sentence splitter (see 3.2.1).

To date, there are no available Natural Language Processing (NLP) tools suitable for processing ancient North-western Semitic languages, such as the different Aramaic idioms attested to in the BT, and for detecting the historical variants of the Hebrew language as used in the Talmudic text. The only existing NLP tools for Jewish languages (see, Bar-Haim et al. (2005), Itai (2006), *HebMorph*[1]) are specifically implemented for Modern Hebrew, a language which has been artificially revitalized from the end of the XIX century and which does not correspond to the idioms recurring in the BT. In its multifaceted form the "language" of the BT is unique and attested to only in few other writings. In addition, only few scholars have a full knowledge of the linguistic peculiarities of the BT and even fewer experts in Talmudic Studies are interested in collaborating to the creation of computational technologies for this textual corpus. Therefore, the development of NLP tools for BT would require a huge and very difficult effort, which would probably not be justified by the subsequent use of the new technologies developed. Also for these reasons, we opted to a translation system based on a TM.

More important, the literary style of the BT itself concurred to orientate us through the choice of the TM. Composed in a dialogical form, the BT is characterized by formulaic language and by the recurrence throughout the different tractates of fixed expressions used to connect the different textual units. As an exegetical text, the BT contains innumerable quotations from the *Mishnah*, from other *tannaitic* sources and even from *amoraitic* statements discussed in other passages of the BT itself. In addition, since one of the major efforts of the *Amora'im* was finding the relevant Biblical passages on which to anchor the rules of the *Mishnah*, the BT turns out to be a constellation of quotations from the Scripture (see 3.2.2). The exceptional volume of citations and recurrent fixed expressions in the BT makes the TM particularly effective and fit for this writing.

2.3 Statistics and Comparison with Other Texts

In the course of our study we carried out also some statistic analyses on the language and style of the BT, which corroborated the choice of the TM approach. Particularly,

[1] Morphological Analyser and Disambiguator for Hebrew Language `http://code972.com/hebmorph`

Table 1. Text statistics

Text	Tokens	Types	Average TTR	RR
Rosh Hashana	23,660	5,758	0.72	0.13
Berakhot	70,622	13,355	0.70	0.10
Qiddushin	59,586	10,171	0.70	0.11
Shabbat	114,082	19,096	0.70	0.11
Bava Metzia	84,675	12,774	0.70	0.10
Liqra't Medinah	4,772	2,444	0.84	0.39
Bi-Vedidut	11,552	7,230	0.90	0.46
Arieh Baal Guf	12,653	6,417	0.89	0.45
Sefer Ha-Kuzari	49,617	20,680	0.87	0.30
Sefer Ha-Bahir	15,759	5,546	0.79	0.24
Sefer Ha-Razim	7,067	3,545	0.87	0.37

the evaluation of the lexical variation within five of the BT tractates (*Berakhot*, *Shabbat*, *Rosh Ha-Shanah*, *Qiddushin*, and *Bava Metziah*) suggested that its lexicon is relatively poor when compared to other Jewish texts (see, Table 1).

The Type-Token Ratio (TTR), expressing the lexical variety, was estimated for every tractate and was computed as the ratio between the number of the distinct word tokens (ignoring the repetitions) and the number of all tokens in the analyzed text. A high TTR indicates, in fact, a high degree of lexical variations, while a low TTR demonstrates relatively low lexical variation. In our comparison, we computed TTR regardless of the text length, by taking a moving average type token ratio (MATTR) Covington and McFall (2010). By using a 100-words sized window, we computed TTRs for each window in the text and averaged them. According to the results reported in the column "Average TTR" in Table 1 (grey part), all the tractates present a poor vocabulary, containing highly repetitive words/phrases. Furthermore, the ratio between the number of hapax legomena and the number of all tokens (RR) is lower in the BT tractates, suggesting a high degree of lexical repetition.

2.4 Relevant Aspects of the BT for the CAT System

The complexity and richness of the BT makes this textual collection an exceptional object of study for many aspects (textual content; different languages attested to; history of the composition of the text; sources employed; passage from oral transmission to textual redaction; philology of the text), and, yet highly increases the hardship of analysing and translating it.

An ancient anthological writing such as the BT cannot be treated and translated as a modern text, since a plain and literal translation of this text would not be intelligible to a modern reader. The Talmudic text is exceptionally concise and several passages remain unclear even for expert Talmudists. Therefore, a good translation of the BT requires the addition of explicative integrations within the translation and a re-elaboration of the literary tradition. For these reasons we had to develop a system suitable for translators, who have a solid background in Talmudic Studies, and to create additional tools capable of expressing all the features of a translation, which is not merely a translation, but, to a certain extent, an interpretative commentary itself. Particularly, we had to enhance the system with tools that enable to distinguish the literal part of the translation (indicated in bold) from the explicative additions of the translators/scholars (see 3.1). Furthermore, due to the complexity of the inner structure of the BT, we had to provide translators with the possibility to organize their translations in specific units and subunits (see 3.2.3).

In addition, the peculiar literary style of the BT (formulaic language, standard expressions, quotations from the Bible and *tannaitic* sources, quotations from other sections of the BT itself) together with the lack of NLP tools for the languages attested to in the BT, oriented us, in the first place, to base our translation system on a TM instead of Machine Translation approaches.

3 Features of the CAT System

In the following sections we outline the most relevant features developed to face the translation of textual corpora with complex philological and linguistic peculiarities, such as the BT.

3.1 The Translation Memory

Our TM is organized at the segment level. A segment is a portion of original text having an arbitrary length. We formally define the TM $M_{BT} = \{s_i, T_i, A_i, c_i\}$ with i ranging from 1 to n, as a set of n tuples, where each tuple is defined by:

- s_i, the source segment;
- $T_i = \{t_i^1, \ldots, t_i^k\}$, the set of translations of s_i with $k \geq 1$, where each t_i^j includes a literal part $\overline{t}_i^j$ exactly corresponding to the source segment, and an explicative addition, from now on referred to as contextual information $\widetilde{t}_i^j$, with $1 \leq j \leq k$;
- $A_i = \{a_i^1, \ldots, a_i^k\}$, the set of translators' identifiers of each translation of s_i in T_i with $k \geq 1$;
- c_i, the context of s_i referring to the tractate to which s_i belongs.

As stated above, the translation of each segment can be done by differentiating the "literal" translation (using the bold style) from explicative additions, i.e. "contextual information". Segments having the same literal part can then differ by their contextual information. The tractate of the source segment of each translation is called "context".

3.1.1 The Translation Suggestion Component. The system acquires the segment to be translated, queries the TM, and suggests the Italian translations related to the most similar strings. The translation procedure involves three main aspects:

- how the system allows translators to choose a source segment (segmentation process, see 3.2.1);
- how the system retrieves pertinent translation suggestions (similarity function, see 3.1.2);

– how the system presents the suggestions to each translator (suggestions presentation, see 3.1.3).

In the following subsections we will analyze these aspects, emphasizing all the characteristics that we consider peculiar for ancient texts and which, to our knowledge, differentiate our system from state of the art open source CAT tools.

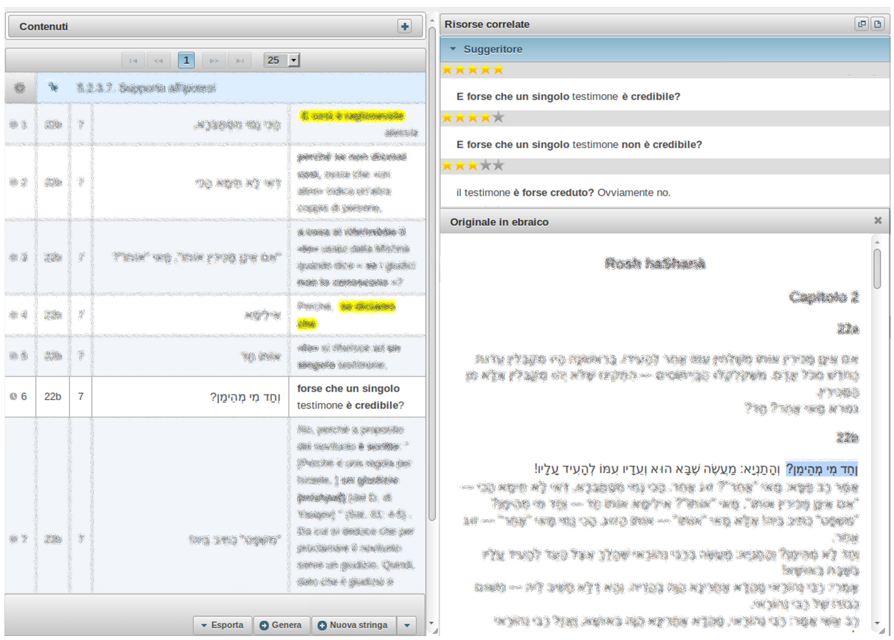

Fig. 1. The system GUI (data, authors, parts of the original Babylonian Talmud text have been clouded for privacy and rights reasons).

3.1.2 Similarity Function. For the reasons stated in Section 2.2, it has not been possible, so far, to include either grammatical or syntactic information in the similarity search algorithm. So, we took account of adopting similarity measures based on edit distance, $ED(s_1, s_2)$, by considering two segments to be more similar when the same terms tend to appear in the same order. The novelty we introduce here consists in the way we rank suggestions, based on external information, namely i) the authors of translations and ii) the context (the tractate of reference). The information about i) the author of the translation and of ii) the tractate of reference is useful both for translators and revisors. On the one hand, translators are enabled to evaluate the reliability of the suggested translations on the basis of the authority and expertise of the relative translators. On the other hand, revisors can exploit both kinds of information to ensure a more homogeneous and fluent translation.

Given a segment s_q of length $|s_q|$, and a distance error δ, our similarity function allows to:

- retrieve all segments s in the TM (called suggestions) such that $ED(s_q, s) \leq round(\delta * |s_q|)$;
- rank suggestions, not only on the basis of the ED outcome, but also on both the current context and the suggestion author.

Searching for segments within δ errors, where δ is independent of the length of the query, could be little meaningful, as described in Mandreoli et al. (2002). For this reason, we considered δ as the percentage of admitted errors w.r.t. the sentences to be translated, multiplying it by the length of the query segment. In collaboration with the translators' team, we have experimentally tuned δ to 0.7. Our algorithm is based on dynamic programming, and its implementation refers to Navarro (2001). In order to compute $ED(s_1, s_2)$, it builds a matrix $M_{(0..|s_1|,0..|s_2|)}$, where each element $m_{i,j}$ represents the minimum number of token mutations required to transform $s_1(1..i)$ in $s_2(1..j)$. The computation process is the following:

$$m_{i,j} = \begin{cases} i \text{ if } j = 0 \\ j \text{ if } i = 0 \\ m_{(i-1,j-1)} \text{ if } s_1(i) = s_2(j) \\ 1 + \mu \text{ if } s_1(i) \neq s_2(j) \end{cases}$$

where $\mu = min(m_{(i-1,j)}, m_{(i,j-1)}, m_{(i-1,j-1)})$, and the final cost is represented by $m_{(|s_1|,|s_2|)}$. The system returns those strings translated in Italian that have the lower costs. Basically, given a segment to translate, many other source segments can be exactly equal to it and each of them can be paired with multiple translations in the TM. Differently from what happens with other CAT, our similarity function immediately provides the translator with more accurate suggestions, by ranking all of them on the basis of the current context, showing first those already provided by the most authoritative translators. This aspect positively affects also the revision activity. It possible, for example, to easily retrieve translations of equal segments having the same context provided by different translators in order to analyze and, in case, uniform stylistic differences. It is noteworthy that the translations' revision activity, performed by human reviewers, guarantees that the TM contains only high-quality translations.

3.1.3 Suggestions Presentation. The system user interface shows each suggestion marked by a number of stars, as shown in Figure 1. That number is assigned by the system on the basis of how fuzzy the match between source segments is: five stars are regarded as a perfect suggestion (exact match); four stars indicate few corrections required to improve the suggestion (fuzzy match); three stars refer, in most of the cases, to acceptable suggestions (weak fuzzy match). As stated in the previous subsection (3.1.2), while dealing with suggestions ranked with the same number of stars, the system orders them by context and author.

Every suggestion can be shown with or without its contextual information; literal translation is specified by bold font, w.r.t. the contextual information. Each translator, for example, can then approve as correct the literal translation $\overline{t}_i^j$, and modify only $\widetilde{t}_i^j$. Then, each new correct translation is added to the TM, thus, increasing the pool of available translations. In this way, our system leaves human translators in control of the actual translation process, relieving them from routine work. Our system allows also to maintain the translation process as creative as necessary, especially in those cases requiring human linguistic resourcefulness.

Figure 1 shows a simple example of translation suggestions[2]. The related set of suggestions are provided and scored[3] by the system, as depicted in Figure 1:

- t_1 = "**And may a signle** witness **be credible ?**", with $\widetilde{t}_1 = \{\text{witness}\}$
- t_2 = "**And may a single** witness **not being credible ?**", with $\widetilde{t}_1 = \{\text{witness}\}$
- t_3 = "**Is the** witness **believed ?** Obviously not", with $\widetilde{t}_1 = \{\text{witness, obviously not}\}$

In the translation session of Figure 1, the translator chose t_1, and changed its literal part by deleting the word "**And**"[4].

In our view, these requirements are much more fundamental for the translation of ancient texts than for modern texts. To the best of our knowledge, there are no state of the art CAT tools taking into account these needs.

3.2 Additional Tools

3.2.1 Segmentation Process. In all CAT systems, the first step of the translation process is the segmentation of the source text. As stated in 2.2, the BT does not exhibit a linguistic continuity, thus preventing an automatic splitting into sentences. To face this uncommon situation we opted for a manual segmentation taking into account also the structure of the text (see 3.2.3). Each translator selects the segment to translate from the original text contained in a specific window (Figure 1). This process reveals to have a positive outcome: translators, being forced to manually detect the segments and organize them in a hierarchical structure, acquire a deeper awareness of the text they are about to translate. Clearly, the manual segmentation implies the engagement of the translators in a deep cognitive process aimed to establish the exact borders of a segment. The intellectual process involved in the textual segmentation deeply affects also the final translation, by orienting the content and nature of the TM. To the best of our knowledge this feature is not present in any state of the art CAT system.

3.2.2 Quotations. As outlined above, the BT is based on various sources and contains quotations of different Jewish texts (Bible, *Mishnah*, *Toseftah*, Babylonian Talmud, Palestinian Talmud). For this reason, the system provides precompiled lists containing all these texts, so that users can select the relevant source related to their translation.

---

[2] Clearly, the suggestions are in Italian. We are attempting to provide English translations as semantically similar to the original Italian sentences as possible.

[3] For privacy reasons we omitted both context and authors.

[4] Note that $\overline{t}_1$, $\overline{t}_2$, $\overline{t}_3$ are not grammatically correct in English, while they are perfectly well-formed in Italian, since Italian is a null-subject language.

3.2.3 Structure. Since the translation has to be carried out in respect of the inner structure of the BT, we had to implement a component to organize the translation following a hierarchic subdivision: tractates, chapters, blocks (sugyot), logical units (e.g. thesis, hypothesis, objections, questions, etc.), and strings, the latter representing the translation segments.

4 Evaluation

With our system not only we aim to increase the translation speed, but also and especially to support the translation process, offering the translators a collaborative environment, in which they can translate similar segments without losing contextual information on the translations. Under these premises it is quite hard to exactly evaluate the overall performance of the proposed system. In order to accomplish this task, we tried to measure both the number of similar segments stored within the TM, (i.e., the rate of TM redundancy), and how good the suggestions provided by the system are.

4.1 Translation Memory Analysis

For privacy reasons, we cannot publish the number of translated sentence present, at the moment, inside the TM. Let us say we have n translated sentences, with an average length of five tokens. Figure 2 shows that each translation is approximately between

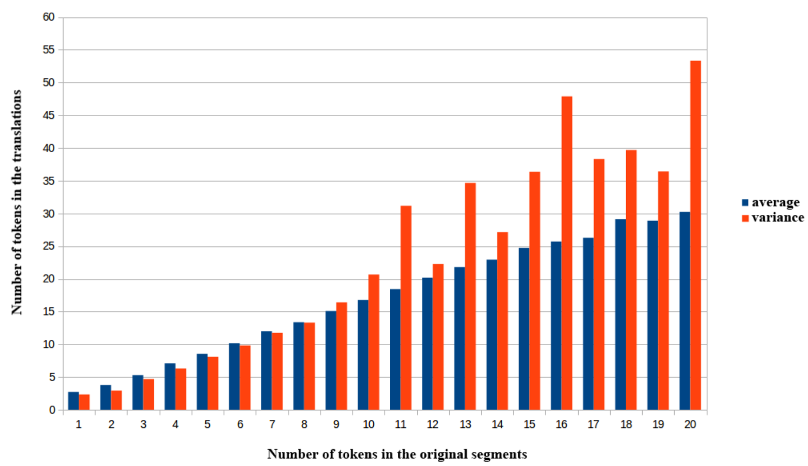

Fig. 2. Ratio between the length of the original segments and the related translations' length.

70% and 40% longer than its original text segment (these percentages are computed as the ratio between each s_i length and the average length of all translations belonging to the related T_i). This could be partially ascribed to the fact that semitic languages are agglutinated, implying that every token is made up by the union of more morphemes. Also, the high degree of variance should be taken into account; starting from eight tokens long source segments, the variance is always higher than the average, and sudden peaks can be observed examining source segments longer than eleven tokens. This means that a significant part of the obtained translations are exclusively literal, while another significant part is enriched with contextual information, derived from translators' knowledge and scholarly background, as well as from their personal interpretations of the text.

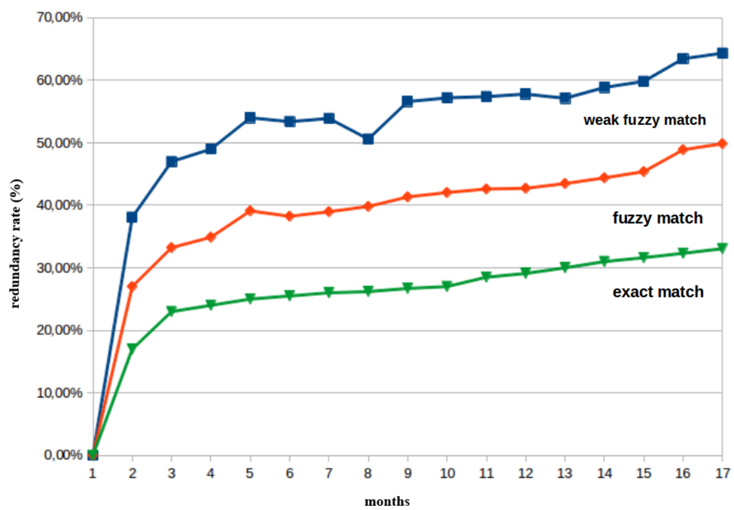

Fig. 3. TM performance.

To study the redundancy, we conducted a jackknife experiment to roughly estimate the TM performance, as in Wu (1986). We partitioned the TM in groups of one thousand, then, used one set as input and translated the sentences in this set using the rest of the TM. This was repeated n times. Intuitively, the resampling method we adopted uses the variability in the test sets to draw conclusions about statistical significance. Figure 3 shows the redundancy rate of the TM from the beginning of the project until today. Redundancy curves are drawn by considering the ranking of the similarity function. The percentage of source segments found in the TM, grows logarithmically with time (and consequently with the size of the memory). Clearly the larger the memory will be, the better the performance. For instance, if we suppose that exact matches provide at least one perfect suggestion, then each translator would save 30% of their work.

4.2 Translation Suggestions Goodness

In order to measure the goodness of a translation suggestion, we used machine translation evaluation metrics. These metrics measure the similarity between a reference trans-

Table 2. Translation suggestions evaluation by means of MT metrics. For each T_i, max represents the average of the best matches, and min means the average of the fuzzier ones

MT measure		**with** contextual information	**without** contextual information
WER	max	0.78	0.76
	min	0.32	0.27
BLEU	max	0.81	0.78
	min	0.13	0.15

lation and a set of candidate ones; clearly, the more each candidate resembles the reference, the better the score. Our test has been built by identifying 1000 segments from the TM such that every correspondent T_i contains at least two identical literal translation from different authors ($\bar{t}_i^h = \bar{t}_i^k$ with $h \neq k$, and $a_i^h \neq a_i^k$). We considered each of those translations as the reference translation t_i^{ref} of the related source segment s_i^{test}, with i ranging from 1 to 1000. We removed $\{s_i^{test}\}$ set from the TM, considering the remaining set of translations as the candidates set. Given the large number of existing evaluation metrics, we focused on those we retain the two more representative ones in the literature: Word Error Rate (WER), and BLEU, described in Papineni et al. (2002). WER, one of the first automatic evaluation metric used in machine translation, is computed at word level by using the Levenshtein distance between the candidate and the reference, divided by the number of words in the reference[5]. BLEU, instead, measures n-gram precision, i.e., the proportion of word n-grams of the candidate that are also found in the reference. We produced our own implementations of WER and BLEU, by normalizing WER values between 0 and 1. For each s_i^{test}, our similarity function retrieved the best match (if available), and each of its related Italian translations has been compared with t_i^{ref}, saving both the best and worst scores for each measure. Finally, all the best scores and the worst ones, have been averaged and reported in Table 2 as max and min values, respectively. This process has been performed on both literal translations and translations with contextual information. Concerning max values, regardless the presence of contextual information, the BLEU score is greater than WER. WER counts word errors at the surface level. It does not consider the contextual and syntactic roles of a word. This result is encouraging since BLEU, instead, gives us a measure of the translations' quality, as reported in White et al. (1993) considering two different aspects: lower-order n-grams tend to account for adequacy, while higher-order n-grams tend to account for fluency (we used n-grams with n ranging from 1 to 4).

Instead, when we examine min values, the results are completely reversed. Both measures indicate that values are very low: in particular BLEU shows that the quality of the provided translations is very poor (15%). This situation brought us to repeat the BLEU measurement taking into account only the exact matches of the test set: in this

[5] Several variants exist, most notably TER (Translation Edit or Error Rate), in which local swaps of sequences of words are allowed. However, WER and TER are known to behave very similarly, as described in Cer et al. (2010).

case, the minimal values arise to the 40%, still showing a bad quality of some translations. Based on these results, we believed necessary to integrate a revision activity for the translators' work, as it was originally scheduled within the project.

Another information we can deduce from Table 2 is that, for both measures, the translations accounting for contextual information achieve better results than those achieved by literal ones. The reason for this phenomenon could depend on the fact that, as stated in Firth (1957), similar words or similar linguistic structures, tend to occur in similar contexts. Therefore, we can assume that similar literal translation parts present similar contextual information parts. Indeed, in the previous example, we can see that the word "witness" is a contextual information that appears in all the provided suggestions (see "testimone" in Figure 1).

5 Conclusions and Future Works

In this paper, we introduced some components of the CAT system we developed to support the collaborative translation of the Babylonian Talmud in Italian. Differently from the functionalities a CAT system typically offers to ease and speed up the translation of a text, the translation (and revision) of an ancient text as the BT requires specific features. In general, the environment must create the conditions to allow a user, in the double role of translator and scholar, to produce correct translations, which often require, especially for texts such as the BT, several textual additions. We opted for a TM instead of a Machine Translation approach for a number of reasons, both related to the intrinsic nature of the BT (high number of quotations, recurrence of fixed expressions and dialectical narration) and to the lack of NLP tools for ancient Semitic languages. We focused here on the description of the Translation Memory and on some aspects regarding the relative translation suggestion component and its evaluation. Among the most interesting features we introduced in the system, there are: i) the manual segmentation, necessary to compensate to the lack of punctuation marks, required for automatic sentence splitting, ii) translation suggestions, which take into account both the author of the translation and the context (very important also for the revision process), iii) the possibility to distinguish between literal translations and explicative additions, iv) the possibility to add references to quotations, v) the organization of the translations into hierarchical subdivisions. Concerning future works, we plan to automatically populate the TM with biblical quotations, taking into account the work described in HaCohen-Kerner (2010), and to integrate a Language Identification tool both to improve the performance of the suggestion component and to allow advanced searches on the translated corpus on a language basis. Furthermore, we are working on the development of a language independent, open source and web-based CAT system, since we believe most of the solutions adopted for the translation of the BT could be generalized to assist in the translation of the majority of ancient texts.

6 Acknowledgements

This work has been conducted in the context of the research project TALMUD and the scientific partnership between S.c.a r.l. "Progetto Traduzione del Talmud Babilonese"

and ILC-CNR and on the basis of the regulations stated in the "Protocollo d'Intesa" (memorandum of understanding) between the Italian Presidency of the Council of Ministers, the Italian Ministry of Education, Universities and Research, the Union of Italian Jewish Communities, the Italian Rabbinical College and the Italian National Research Council (21/01/2011).

References

Gordon, I. (1996) Letting the CAT out of the bag - or was it MT?. In: *Proceedings of the* 8^{th} *International Conference on Translating and the Computer*, Aslib, London, 1996.

Bar-Haim, R., Sima'an, K. and Winter, Y. (2005) Choosing an Optimal Architecture for Segmentation and POS-Tagging of Modern Hebrew. In: *Proceedings of the Association for Computational Linguistics Workshop on Computational Approaches to Semitic Languages*, 2005, pp.39-46.

Itai, A. (2006) Knowledge Center for Processing Hebrew In: *Proceedings of the* 5^{th} *International Conference on Language Resources and Evaluation - Workshop "Towards a Research Infrastructure for Language Resources"*, Genoa, Italy, 22 May, 2005.

Covington, M.A., and McFall, J.D. (2010) Cutting the Gordian Knot: The Moving-Average Type-Token Ratio (MATTR). *Journal of Quantitative Linguistics*, 17(2).

Navarro, G. (2001) *A guided Tour to Approximate sentence Matching*, Technical report, Department of Computer Science, University of Chile.

Mandreoli, F., Federica M., Martoglia, R., and Tiberio, P. (2002) Searching Similar (Sub)Sentences for Example-Based Machine Translation. In: *Atti del Decimo Convegno Nazionale su Sistemi Evoluti per Basi di Dati*, Portoferraio (Isola d'Elba), Italy, 19-21 June, 2002.

Wu, C. F. J. (1986) Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis. *The Annals of Statistics* 14(4):1261-1295.

Papineni, K., Roukos, S., Ward, T., and Zhu, W.J. BLEU: A Method for Automatic Evaluation of Machine Translation. In: *Proceedings of the* 40^{th} *Annual Meeting on Association for Computational Linguistics*, Stroudsburg, PA, USA, 2002, pp. 311-318.

Cer, D., Manning, C. D., and Jurafsky, D. (2010) The Best Lexical Metric for Phrase-Based Statistical MT System Optimization. In: *Human Language Technologies, the* 11^{th} *Conference of the North American Chapter of the Association for Computational Linguistics*, Los Angeles, USA, June 2-4, 2010, pp. 555-563.

White, J. S., O'Connell, T. A., Carlson, L. M. (1993) Evaluation of Machine Translation. In: *Proceedings of the workshop on Human Language Technology Association for Computational Linguistics*, Stroudsburg, PA, USA, 1993, pp 206-210.

Firth, J. R. (ed) *Papers in Linguistics 1934*. Oxford University Press, London.

HaCohen-Kerner, Y., Schweitzer, N., and Shoham, Y. (2010) Automatic Identification of Biblical Quotations in Hebrew-Aramaic Documents. In: *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval*, Valencia, Spain, October 25-28.