Integration of Translation Memory with Neural Machine Translation Based on Formal and Semantic Coverage of the Input Sentence

Yaling Wang, *Waseda University, yaling.wang@moegi.waseda.jp,*
Yves Lepage, *Waseda University, yves.lepage@waseda.jp*

***Abstract*— Neural Machine Translation (NMT) has made progress in recent years. But translation Memories (TM) have long been used by translators as a tool to suggest similar translations. Integrating TM into NMT has been proved to be efficient for improving translation quality. We propose to use a simple method to integrate them without altering the NMT model architecture. We retrieve similar sentences covering the sentence to translate and use them as annotations on the input of an NMT system. Our results show that our method can outperform a baseline model in some cases. The improvements are mainly for the translation of sentences with a length ranging from 10 to 20 words. This shows that similar sentences offer more contextual information useful for translation, in comparison with a baseline model.**

***Index Terms*—machine translation, retrieval, similar sentences in form and meaning.**

I. Introduction

MACHINE Translation (MT) is the use of computers to perform translation. It offers better results with the emergence of Neural Machine Translation (NMT), which led to significant increase in *translation accuracy*, measured by comparison to reference translation. However, as with many other neural network techniques, the *interpretability* of NMT is poor: errors are difficult to interpret, i.e., to trace back to the training data. Translation Memories (TM) are tools used by human translators. They are databases containing parallel sentences of high quality. They allow translators to retrieve sentences similar to an input sentence, and they return the corresponding translations as suggestions for translation. A main advantage of translation memories is that they ensure consistency and *interpretability* across translations because common or similar parts in sentences can easily be identified. TM suggests high-quality translations when there are highly similar sentences and they provide references for translations. The combination of TM and MT is found Computer-Assisted Translation (CAT) environments [1], where an MT system is often used as back off when highly similar sentences from the TM cannot be retrieved. TMs can also be integrated within MT systems, like phrase-based statistical MT (PBSMT) and NMT systems. We introduce this in detail in Section II.

We propose a method to leverage a TM in machine translation without altering the model architecture. Our method consists in retrieving a subset of sentence pairs from a parallel data set that covers the sentence to translate and performing translation by means of concatenation. Figure 1 shows the process of translation in our proposal.

Suppose that we translate from English to German. In detail, for a sentence to be translated, we firstly retrieve sentences from the parallel data set and obtain some similar sentences which cover the input sentence. For example, the sentence to translate '*I want to go to school.*' is covered by a series of similar sentences '***I want to go to** hospital.*' and '*This is a beautiful **school.***'. The corresponding translation of similar sentences is obtained from the parallel data set. '*Ich will ins Krankenhaus.* ' and '*Das ist eine schöne Schule.* ' That is,

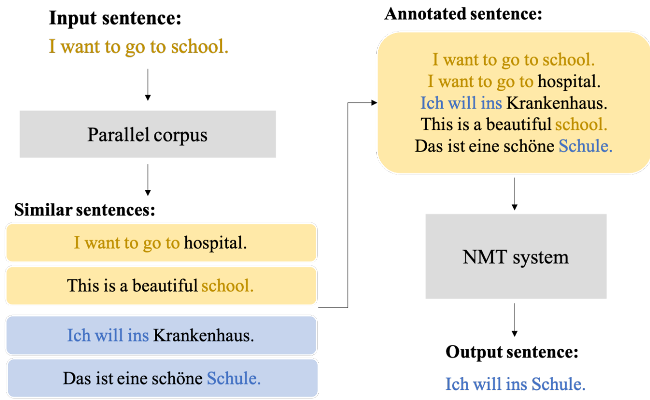

Fig. 1. Overview of translation based on retrieval

by retrieval, we acquire similar sentence pairs. We consider these translation equivalents as annotations and concatenate the input sentence with similar sentence pairs into a new piece of data. By using this kind of piece of data to train a neural machine translation model, we expect that similar sentences and their translations can offer reference during translation and boost translation quality.

This paper is structured as follows. Section II documents related work. In Section III, we introduce the retrieval method used in this paper. Section IV shows the experimental setup and Section V presents the results.

II. Related Work

Many works propose to integrate TM with different MT systems. In the beginning, TM-MT integration made use of example-based MT [2]. They proposed a generalized TM system that broke the limitation of the entire sentence during retrieval and matched with sub-sequences in the input sentence.

PB-SMT was also integrated with TM. In [3], they first use TM to retrieve matches and then replace the mismatched parts with an SMT system as the final translation result. In [4], unlike the previous method, the proposed model makes use of

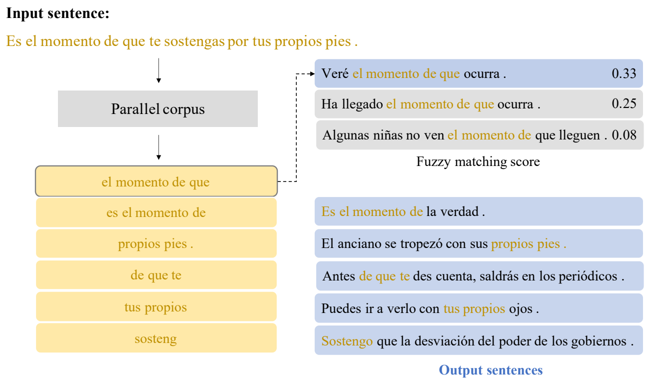

Fig. 2. Overview of the formal coverage retrieval

the TM to assist in the translation of each phrase during SMT decoding.

With the rise of NMT, more research attempted to incorporate TM with NMT. In [5], they firstly retrieve a subset of sentence pairs, and then both the source sentence and these sentence pairs are used to perform the translation in their new model – search engine guided NMT (SEG-NMT). In [6], they encode TM matches in to NMT using an extra decoder. In [7], they propose a simple but powerful data augmentation method for TM-NMT integration by means of the concatenation of source data and retrieved fuzzy targets. All these approaches led to better performance. In [8], they propose a retrieval method to cover a sentence in form and meaning with fewer retrieved sentences.

Our method aims to retrieve a subset of sentence pairs from parallel data set covering the sentence to translate and perform translation by means of concatenation without altering the NMT model architecture.

III. Retrieval of Sentences

In this section, we explain the retrieval process of similar sentences to cover an input sentence. There are two types of coverage: formal coverage and semantic coverage, respectively explained in Subsections III-A and III-B. The different possibilities of concatenating the sentences, retrieved by use of both types of coverage, with the input sentence is explained in Subsection III-C.

A. Formal Coverage

Formal coverage means that the sentence to translate and retrieved sentences have some common (sub)words or n-grams. Formal coverage retrieval is performed in 3 steps. Figure 2 shows the overview of formal coverage retrieval.

1) Matching N-grams

Firstly, we iterate each n-gram of the given sentence, and match the longest common n-grams at each start position to ensure the maximal formal coverage and minimum number of common n-grams. For the n-gram without any match in the corpus, we derive subword tokens to match.

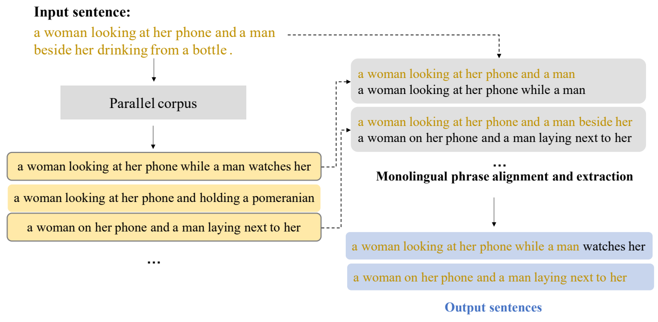

Fig. 3. Overview of the semantic coverage retrieval

2) Fuzzy Matching Selection

After matching the covering n-grams, we retrieve all the sentences that contain each n-gram and subword from the corpus. For each n-gram, there are a certain number of candidate sentences containing the same n-gram. To reduce the number of candidates, we use the fuzzy matching score to rank candidates and select the sentence with the highest score.

The fuzzy matching score between two sentences is based on the edit distance between sentences in terms of tokens. The fuzzy match score is defined as:

$$\mathrm{FM}(s_i, s_j) = 1 - \frac{\mathrm{EditDistance}(s_i, s_j)}{\max(|s_i|, |s_j|)} \tag{1}$$

where $\mathrm{EditDistance}(s_i, s_j)$ computes the edit distance between sentences, and $|s|$ denotes the length of sentence s. We compute the fuzzy matching score using Levenshtein distance [9].

3) Trimming redundancies

As sentence matching is separated from n-gram matching, the retrieved sentences tend to over-cover the query. We trim these redundant sentences in the last phase.

B. Semantic Coverage

Semantic coverage consists in retrieving sentences which have a similar meaning as the input sentence, but do not necessarily have words in common. To retrieve sentences in semantic coverage, there are 3 steps. Figure 3 shows the overview of semantic coverage retrieval.

1) Retrieving similar sentences

We search for the top k similar sentences to the input sentence using the distributed method based on sentence embeddings. Sentences can be retrieved by measuring the cosine similarity of sentence embeddings. The contextual similarity score is defined as:

$$\mathrm{EM}(s_i, s_j) = \cos(\mathbf{s}_i, \mathbf{s}_j) = \frac{\mathbf{s}_i \cdot \mathbf{s}_j}{\|\mathbf{s}_i\| \times \|\mathbf{s}_j\|} \tag{2}$$

where $\|\mathbf{s}\|$ denotes the norm of vector $\mathbf{s}$.

Here we use the pre-trained sentence-BERT model [10] to represent sentences in vectors. Efficient semantic search of a sentence vector space is facilitated by the Faiss library [11].

2) Similar phrase match extraction

Similar to formal coverage, we also expect to know which parts of retrieved sentences have the same meaning with input sentence. However, retrieved sentences in semantic do not necessarily have common words with input sentence.

We attempt to extract phrase matches using the monolingual phrase alignment approach based on pre-trained word embeddings proposed in [12]. This method delivers word alignments based on a matrix of cosine similarity between pre-trained word embeddings. Each pair of phrase matches corresponds to an alignment score rated by a phrase extractor.

3) Screening candidate phrase matches

Phrase match candidates are sorted by the rank of the contextual similarity score of sentences containing these phrases. For these phrase match candidates, we select those with an alignment score larger than a threshold and which contributes to the increase of coverage.

C. Concatenation

By retrieval, we get some similar sentence pairs both in form and in semantic. Before using NMT system, we need another step – concatenation.

We concatenate the input sentence, i.e., the sentence to translate, and its similar sentence pairs into a piece of data. There are different formats for concatenation:

- source only
- target only
- all source sentences followed by all target sentences
- source followed by target for each sentence pair

Figure 4 shows an example of 4 types of concatenation.

We can combine the input sentence and similar sentences in source language ('*source only*'). For example, we combine the input sentence '*I want to go to school.*' and similar sentences in source language '*I want to go to hospital.*', '*This is a beautiful school.*' as a whole.

We can also combine the input sentence and similar sentences in target language ('*target only*'). Or we can combine input sentence and similar sentences in both source and target language but by different orders. In the first case, all source sentences are followed by all target sentences. For example, '*I want to go to hospital.*' and '*This is a beautiful school.*' are followed by '*Ich will ins Krankenhaus.*' ' and '*Das ist eine schöne Schule.*' While in the second case, for each sentence pair, the source sentence are followed by the target sentence. For example, there are two pairs of similar sentences. we combine similar sentences by this order: '*I want to go to hospital.*' followed by its translation '*Ich will ins Krankenhaus.* Then, '*This is a beautiful school.* ' followed by its translation '*Das ist eine schöne Schule.*'

After concatenation, we regard these data as an annotated input sentence, and we use this annotated sentence as the input of the machine translation.

IV. Experimental setup

A. Data

We use the parallel corpus Multi30k [13]. The Multi30k corpus contains multilingual image descriptions for multilingual and multimodal research. The languages used are English, German and French. Some statistics for the corpus are given in Table I.

Fig. 4. Example of 4 types of concatenation

TABLE I
STATISTICS OF THE CORPUS (MULTI30K)

Language	#sentences	Avg. Length	Vocab. Size
en	30,014	13.02	10,214
de	30,014	12.44	18,722
fr	30,014	13.62	11,794

All sentences are tokenized. We randomly divide the data set into 3 parts: training set (80%), validation set (10%) and test set (10%).

B. Evaluation

The automated evaluation metrics used is BLEU (BiLingual Evaluation Understudy) [14]. The BLEU score is the geometric mean of the probability of n-grams in the hypothesis to be present in the references, with a brevity penalty (BP).

$$\text{BLEU} = \text{BP} \sqrt[N]{\prod_{n=1}^{N} p_n} \tag{3}$$

There is one reference translation per test sentence. All evaluations are carried out on tokenized data. We use the implementation of SacreBLEU [15].

C. Baseline system

We compare our proposal to a baseline. Our baseline model is trained using the same NMT model but with different input data. The input data for the baseline system is the sentence to translate without anything else.

Our NMT model follows the Seq2seq architecture [16] implemented in the OpenNMT-py toolkit [17]. Further configuration details are given in Appendix A.

TABLE II
RESULTS OF DIFFERENT FORMATS FOR ANNOTATION

Annotation	BLEU score (en ! de)
source only	25.9
target only	26.4
all source sentences followed by all target sentences	**27.1**
source followed by target for each sentence pair	26.8

TABLE III
STATISTICS FOR RETRIEVAL RESULTS (IN NUMBER OF RETRIEVED SENTENCES PER SENTENCE)

Language	Formal coverage				Semantic coverage			
	mean	stdev.	median	mode	mean	stdev.	median	mode
de	5.03	2.22	5	4	2.81	1.42	3	2
en	5.50	2.22	5	5	2.22	1.14	2	2
fr	5.40	2.32	5	4	2.69	1.33	2	2

V. RESULTS

A. *Annotation Selection*

In Section III-C, we mentioned that there are 4 formats for concatenation or annotation. We use these different annotated sentences for machine translation and select the format with the best performance. We use English as the source language and German as the target language.

Table II shows the results for the different formats of annotation. From the results, the format '*target only*' performs better than '*source only*'. This suggests that direct translations of similar sentences in the target language improve the translation quality by copy mechanism.

In addition, the similar sentence pairs perform better than source only and target only. It shows that the corresponding relationship between similar sentences in source side and target side provides the model with more information.

The annotated sentences with format '*all source sentences followed by all target sentences*' perform best among all the formats. For each similar sentence pairs, they offer common n-grams in the sentence to translate and similar sentences in the source language and corresponding translations in the target language for translation. The model better captures translation information on this distance between similar sentences and its translations. We select this format in later experiments.

B. *Retrieval of Sentences*

We use English, German and French sentences as query sentences to retrieve similar sentences separately. For the retrieval results, we focus on the number of similar sentences (retrieved sentences) per sentence. This is because our retrieval method aims of giving the maximal coverage of input sentence with the least number of retrieved sentences. When the number of similar sentences is less, it means that the common n-grams between the input sentence and the retrieved sentences are longer and more complete. We expect this kind of situation.

Table III gives some statistics for the retrieval results. Figure 5 shows the frequency distribution of the number of retrieved sentences per input sentence. The number of retrieved sentences in the 3 languages is similar. For formal coverage, the average number of retrieved sentences is about 5, which means that 5 n-grams in the retrieved sentences cover the input sentence. The value of the standard deviation is also relatively small. The most frequent number of retrieved sentences is 4, 5, or 6. There are only a few cases where the number of retrieved sentences is greater than 10. This means that, in general, our retrieval method can cover the sentence to translate with a small number of similar sentences. Compared to formal retrieval, the number of similar sentences in semantic retrieved is less. This is because we select the sentences which are top k similar and contribute to increasing the coverage of the input sentence. We consider these similar sentences as a supplement.

Figures 8 and 9 show some examples for retrieval results.

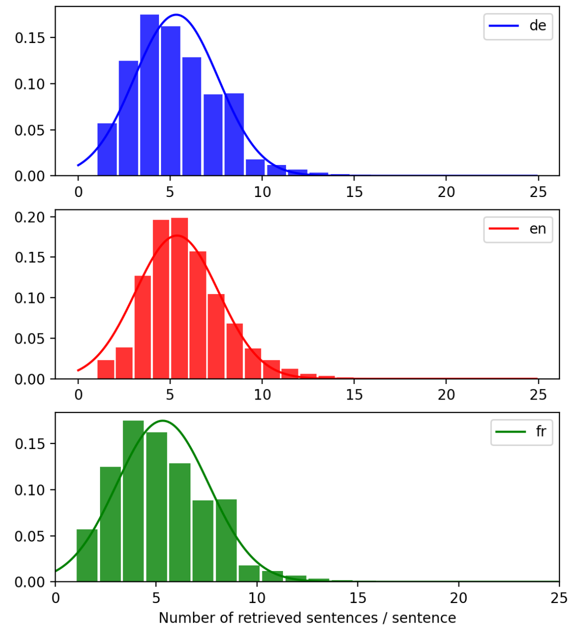

Fig. 5. Frequency distribution of number of retrieved sentences per sentence

TABLE IV
EVALUATION RESULTS OF TRANSLATION (IN BLEU)

Translation task	Baseline		Ours			
			Formal coverage		Semantic coverage	
de ! en	29.6	0.8	**30.5**	0.9	**30.6**	0.8
de ! fr	30.6	0.8	**31.8**	0.8	29.7	0.8
en ! de	27.4	0.8	27.1	0.8	26.1	1.0
en ! fr	42.2	1.2	41.8	1.2	**47.2**	1.0
fr ! de	24.3	0.8	24.1	0.8	23.6	0.8
fr ! en	38.8	0.9	**39.6**	0.9	**42.5**	1.2

C. *Translation Based on Retrieval*

We use annotated data that contain the sentence to translate and similar sentences obtained from retrieval and perform machine translation experiments in the following language pairs: de–en, de–fr and en–fr, each in both directions.

We train a baseline model that only uses the sentence to translate without retrieval. We train our proposed model using the input sentence to translate along with the retrieved similar

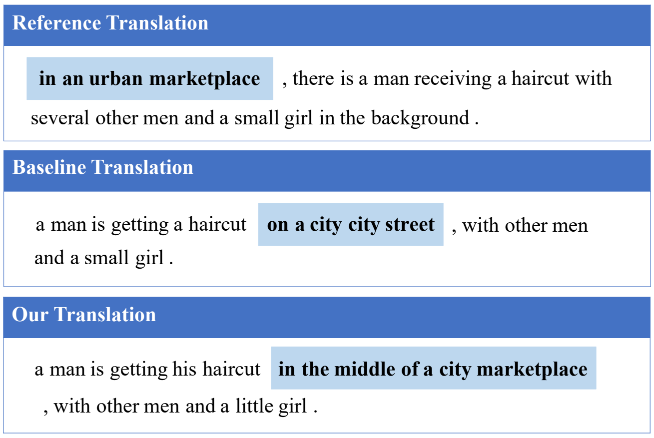

Fig. 6. Example for comparison of translations (de ! en)

sentences. Evaluation is done by computing the BLEU score on the test set.

Table IV summarizes the evaluation results. When using formal coverage retrieval results, our models outperform the baseline model in three translation tasks: de ! en, de ! fr and fr ! en. In the other cases, although our models do not exceed the baseline system, confidence intervals, as shown in Table IV, indicate that the baseline model and our models perform similarly. For instance, for the direction en ! de, confidence intervals of 0.8 do not allow to say that a baseline of 27.4 is really better than our model with 27.1. As the main difference is the language of query sentences, i.e, the source language, we might think that the differences in BLEU observed by the difference in morphology of the source and target languages explains the results. In general, the result shows that the formal coverage retrieval method contributes to improving the translation quality or performs similarly compared to the baseline system.

When using semantic coverage retrieval, our models outperform the baseline model in three translation tasks: de ! en, en ! fr and fr ! en. This is the same number as for formal coverage, but one language direction is different: en ! fr instead of de ! fr. A large improvement is obtained in the direction: fr ! en. In this translation task, the model using semantic coverage retrieval outperforms the baseline model by 3.7 BLEU points, which is largely more than the model using formal coverage retrieval. Our method leads to an even larger improvement in the translation task en ! fr using semantic coverage retrieval. The BLEU score increases by 5.0 points over the baseline model, whereas the model using formal coverage retrieval does not exceed the baseline system. We conclude that our proposed method with semantic coverage is especially efficient for the language pair en–fr, in both directions.

Figure 6 shows an example for comparison of translations output by the baseline model and our proposed models in the direction de ! en. In this example, the n-gram '*in an urban marketplace*' in the reference translation is not accurately translated by the baseline model. But our model finds the same meaning in a reference translation thanks to a similar sentence in the annotated data. Figure 7 shows more examples of translation results.

TABLE V
TRANSLATION RESULTS FOR DIFFERENT SENTENCE LENGTHS (IN BLEU, DE ! EN)

Length of sentence	# sentence	BLEU score Baseline	BLEU score Ours
<10	448	31.3	31.0
10-20	2,207	30.1	**31.1**
>20	247	25.9	25.5

Based on this result, we found that for shorter sentences (length less than 10 words), our model delivers similar performance as the baseline model. However, for sentences with a length between 10 and 20 words, our model offers better translations due to the information found in similar sentence pairs. In order to confirm the impression left by this observation, we split the test set into 3 parts by the length of the sentence to translate and we compare the performance on these 3 separate subsets.

Table V shows the results for the 3 separate subsets containing sentences with different lengths. The sentences of a length between 10 and 20 words account for the most part of the test set. Our model outperforms the baseline model on this subset by 1.0 BLEU point. However, for sentences of length more than 20, both models cannot perform well.

VI. Conclusion

We integrated TM with NMT by using results of retrieval of similar sentences. We annotated the input of the NMT system with such retrieved sentences by different means of concatenation, and we found the format with the best performance. The results of translation based on retrieval show that, for some translation tasks, our system can perform better than a standard NMT system without retrieval. The effect is shown to be larger for sentences with a length between 10 and 20 words.

For future work, firstly, we want to address data sets in which the lengths of the sentences are different. Secondly, we will integrate TM with other NMT methods such as Transformers. At last, we will use more metrics to evaluate the quality of translation from different points of view and explore more possibilities offered by the integration of TM with NMT.

Appendix A
Configuration for the NMT model

The training time for each step is about several seconds. Training an entire model requires about 500 seconds, i.e., less than 10 minutes. The configurations are shown in the Table VI.

Appendix B
Retrieval Times

Our retrieval method requires to build and store an index file before the first access to allow faster retrieval later. Tables VII, VIII and IX show the time for different sizes of datasets. The unit of data in the table cell is the second (s). Experiments were run on a single GPU (GeForce RTX 2080Ti).

Input sentence	Translation	Reference
one lady in a plaid coat eating cotton candy .	une femme en manteau à carreaux mange de la barbe .	une femme en veste écossaise mangeant de la barbe à papa .
two men and a woman are inspecting the front tire of a bicycle .	deux hommes et une femme untersuchen le vorderrad d' un vélo .	deux hommes et une femme inspectent le pneu avant d' un vélo .
a young lady is performing yoga along with the rest of her class .	une jeune femme en train de faire du yoga avec le rest ihres .	une jeune femme fait du yoga avec le reste de son groupe .
un petit chien avec un ruban rouge sur sa tête marche dans l' herbe .	ein kleiner hund mit einer roten ruban auf seinem kopf .	ein kleiner hund mit einem roten band auf dem kopf läuft durch das gras .
trois femmes en rouge de l' équipe de basket russe suivant le ballon .	drei frauen in roter équipe suivant suivant .	drei frauen in roten trikots aus der russischen basketballmannschaft laufen dem basketball hinterher .
une photo de cyclistes lors d' une course avec l' arrière-plan rendu flou par leur vitesse .	ein photo bei einem rennen mit dem hintergrund des vitesse .	eine aufnahme von radfahrern bei einem rennen , wobei der hintergrund aufgrund ihrer geschwindigkeit verschwommen ist .
ein thaiboxer übt zum aufwärmen vor dem kampf einen beinhochtritt .	a thaiboxer band is practicing for the aufwärmen in front of the net .	this thai boxer is practicing a high leg kick as a warm up before his fight .
ein mann mit einem rucksack springt von einem pier .	a man with a backpack jumps off a pier .	a man wearing a backpack is jumping off a pier .
a woman sleeping alone in a bed .	eine frau schläft allein in einem bett .	eine frau schläft allein in einem bett .
two people walking across a street .	zwei personen gehen über eine straße .	zwei menschen überqueren eine straße .

Fig. 7. Random examples for translation results

Sentence to translate	Similar sentences	Reference translation
zwei männer stehen am herd und bereiten essen zu .	**zwei männer stehen am** strand und schauen in die ferne . **two men are on** a beach looking in the distance . zwei bäckereimitarbeiter mit roten schürzen **bereiten essen zu .** two bakery employees wearing red aprons **are preparing food .** 4 personen stehen neben einem feuer **und bereiten** eine mahlzeit zu . 4 people standing next to a fire cooking a meal zwei männer und eine frau stehen vor einem **herd und** kochen . two men and one women are standing in front of a **stove** cooking . zwei männer , mit dem rücken zur kamera gewandt , kochen **am herd** . the backside of two men cooking at a **stove** .	two men are at the stove preparing food .
eine gruppe von personen , die im park grillen .	**eine gruppe von personen , die** mit biergläsern an einem tisch sitzen und lächeln . **a group of people** at a table with glasses of beer smiling . zwei männer , die am strand **grillen .** two men **barbecuing** at a beach **.** eine gruppe von kindern **, die im** gras rennen . a group of children running in the grass . eine gruppe von menschen trinkt bier **im park** . a group of people are enjoying beers **in the park** .	a group of people having a barbecue at a park .

Fig. 8. Random examples for retrieval results in language pair de ! en

Sentence to translate	Similar sentences	Reference translation
une équipe de construction travaillant sur plusieurs endroits endroits de la route .	les ouvriers **de construction** travaillent sur une route dans la nuit . construction workers work on a road into the night . une femme orientale **travaillant sur** une chaîne de production . an oriental woman **working on** an industry line . un couple parlant sur le bas-côté **de la route .** a couple on the side **of the road** talking **. une équipe de** cinq hommes sur scène . **a** five man drill team on stage . un homme âgé travaillant avec **plusieurs** lignes de corde . an elderly man working with **several** lines of string .	a construction crew working at several spots on the road .
un garçon balance une batte de base-ball , et un receveur se tient derrière lui .	un jeune garçon brandit **une batte vers une grosse balle de** baseball . a young boy swings **a bat at a large** baseball . **un petit garçon** en tenue de baseball brandissant **une batte . a small boy** in baseball attire swinging **a bat .** deux équipes de garçons jouant au football **, et un** garçon est en l' air avec le ballon derrière lui . two teams of boys playing soccer **and** one boy is up in the air with the ball behind him .	a boy swings a baseball bat , and a catcher stands behind him .

Fig. 9. Random examples for retrieval results in language pair fr ! en

TABLE VI
CONFIGURATION FOR THE NMT MODEL

Encoder	
Type	LSTM
Embedding Dimension	500
Number of layers	2
Size of hidden layer	500
Decoder	
Type	StackedLSTM
Embedding Dimension	500
Number of layers	2
Size of hidden layer	500
Number of parameters	18,368,003
Optimizer	SGD
Learning rate	1.0

TABLE VII
BUILDING TIMES (IN SECONDS)

Function	Number of sentences		
	29k	290k	2,900k
Training suffix array	0.14	4.15	57.47
Encoding sentences	14.57	126.29	1290.97
Training Faiss	0.05	0.51	5.20
Training (total)	36.06	153.75	1409.99

TABLE VIII
LOADING TIMES (IN SECONDS)

Function	Number of sentences		
	29k	290k	2,900k
Loading suffix array	0.06	0.57	5.58
Loading Faiss	0.02	0.42	4.16
Loading (Total)	21.10	21.82	35.81

TABLE IX
RETRIEVAL TIMES (IN SECONDS)

Function	Number of sentences		
	29k	290k	2,900k
Retrieval in form	0.13	1.63	16.85
Retrieval in meaning	1.12	2.87	5.13
Retrieval (both)	1.25	4.50	21.98

REFERENCES

[1] M. Federico, A. Cattelan, and M. Trombetti, “Measuring user productivity in machine translation enhanced computer assisted translation,” in *Proceedings of the 10th Conference of the Association for Machine Translation in the Americas: Research Papers*, 2012.

[2] M. Simard and P. Langlais, “Sub-sentential exploitation of translation memories,” in *Machine Translation Summit*, vol. 8, 2001, pp. 335–339.

[3] P. Koehn and J. Senellart, “Convergence of translation memory and statistical machine translation,” in *Proceedings of AMTA Workshop on MT Research and the Translation Industry*, 2010, pp. 21–31.

[4] K. Wang, C. Zong, and K.-Y. Su, “Integrating translation memory into phrase-based machine translation during decoding,” in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Sofia, Bulgaria: Association for Computational Linguistics, Aug. 2013, pp. 11–21. [Online]. Available: https://aclanthology.org/P13-1002

[5] J. Gu, Y. Wang, K. Cho, and V. O. Li, “Search engine guided neural machine translation,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.

[6] Q. Cao and D. Xiong, “Encoding gated translation memory into neural machine translation,” in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Brussels, Belgium: Association for Computational Linguistics, Oct.-Nov. 2018, pp. 3042–3047. [Online]. Available: https://aclanthology.org/D18-1340

[7] B. Bulté and A. Tezcan, “Neural fuzzy repair: Integrating fuzzy matches into neural machine translation,” in *57th Annual Meeting of the Association-for-Computational-Linguistics (ACL)*, 2019, pp. 1800–1809.

[8] Y. Liu and Y. Lepage, “Covering a sentence in form and meaning with fewer retrieved sentences,” in *Proceedings of the 35th Pacific Asia Conference on Language, Information and Computation (PACLIC 35)*, 2021, pp. 1–10.

[9] V. I. Levenshtein, “Binary codes capable of correcting deletions, insertions and reversals.” *Soviet Physics Doklady*, vol. 10, no. 8, pp. 707–710, feb 1966, doklady Akademii Nauk SSSR, V163 No4 845-848 1965.

[10] N. Reimers and I. Gurevych, “Sentence-bert: Sentence embeddings using