

A Supplemental Material

In this Appendix, we report the detailed algorithm for unsupervised PBSMT, a detailed ablation study using NMT and conclude with some example translations.

Algorithm 2: Unsupervised PBSMT

- 1 Learn bilingual dictionary using [Conneau et al. \(2018\)](#);
 - 2 Populate phrase tables using Eq. 3 and learn a language model to build $P_{s \rightarrow t}^{(0)}$;
 - 3 Use $P_{s \rightarrow t}^{(0)}$ to translate the source monolingual dataset, yielding $\mathcal{D}_t^{(0)}$;
 - 4 **for** $i=1$ **to** N **do**
 - 5 Train model $P_{t \rightarrow s}^{(i)}$ using $\mathcal{D}_t^{(i-1)}$;
 - 6 Use $P_{t \rightarrow s}^{(i)}$ to translate the target monolingual dataset, yielding $\mathcal{D}_s^{(i)}$;
 - 7 Train model $P_{s \rightarrow t}^{(i)}$ using $\mathcal{D}_s^{(i)}$;
 - 8 Use $P_{s \rightarrow t}^{(i)}$ to translate the source monolingual dataset, yielding $\mathcal{D}_t^{(i)}$;
 - 9 **end**
-

A.1 NMT Ablation study

In Table 4 we report results from an ablation study we performed for NMT using the Transformer architecture. All results are for the *en* \rightarrow *fr* task.

First, we analyze the effect of different initialization methods for the embedding matrices. If we switch from BPE tokens to words, BLEU drops by 4 points. If we used BPE but train embeddings in each language independently and then map them via MUSE ([Conneau et al., 2018](#)), BLEU drops by 3 points. Finally, compared to the word aligned procedure used by [Lample et al. \(2018\)](#), based on words and MUSE, the gain is about 7 points. To stress the importance of initialization, we also report performance using random initialization of BPE embeddings. In this case, convergence is much slower and to a much lower accuracy, achieving a BLEU score of 10.5.

The table also demonstrates the critical importance of the auto-encoding and back-translation terms in the loss, and the robustness of our approach to choice of architectures.

A.2 Qualitative study

Table 5 shows example translations from the French-English *newstest* 2014 dataset at different

	en \rightarrow fr	fr \rightarrow en
<i>Embedding Initialization</i>		
Concat + fastText (BPE) [default]	25.1	24.2
Concat + fastText (Words)	21.0	20.9
fastText + Align (BPE)	22.0	21.3
fastText + Align (Words)	18.5	18.4
Random initialization	10.5	10.5
<i>Loss function</i>		
without \mathcal{L}^{lm} of Eq. 1	0.0	0.0
without \mathcal{L}^{back} of Eq. 2	0.0	0.0
<i>Architecture</i>		
without sharing decoder	24.6	23.7
LSTM instead of Transformer	24.5	23.7

Table 4: **Ablation study of unsupervised NMT.** BLEU scores are computed over *newstest* 2014.

iterations of the learning algorithm for both NMT and PBSMT models. Prior to the first iteration of back-translation, using only the unsupervised phrase table, the PBSMT translations are similar to word-by-word translations and do not respect the syntax of the target language, yet still contain most of the semantics of the original sentences. As we increase the number of epochs in NMT and as we iterate for PBSMT, we observe a continuous improvement in the quality of the unsupervised translations. Interestingly, in the second example, both the PBSMT and NMT models fail to adapt to the polysemy of the French word “langue”, which can be translated as “tongue” or “language” in English. These translations were both present in the unsupervised phrase table, but the conditional probability of “language” to be the correct translation of “langue” was very high compared to the one of “tongue”: $P(\text{language}|\text{langue}) = 0.92$, while $P(\text{tongue}|\text{langue}) = 0.0005$. As a comparison, the phrase table of a Moses model trained in a supervised way contains $P(\text{language}|\text{langue}) = 0.633$, $P(\text{tongue}|\text{langue}) = 0.0076$, giving a higher probability for “langue” to be properly translated. This underlines the importance of the initial unsupervised phrase alignment procedure, as it was shown in Figure 1.

Finally, in Table 6 we report a random subset of test sentences translated from Russian to English, showing that the model mostly retains the semantics, while making some mistakes in the grammar as well as in the choice of words and entities. In Table 7, we show examples of translations from German to English with PBSMT, NMT, and PBSMT+NMT to show how the combination of these models performs better than them individually.

Source	Je rêve constamment d’eux, peut-être pas toutes les nuits mais plusieurs fois par semaine c’est certain.
NMT Epoch 1	I constantly dream, but not all nights but by several times it is certain.
NMT Epoch 3	I continually dream them, perhaps not all but several times per week is certain.
NMT Epoch 45	I constantly dream of them, perhaps not all nights but several times a week it ’s certain.
PBSMT Iter. 0	I dream of, but they constantly have all those nights but several times a week is too much. ”
PBSMT Iter. 1	I had dreams constantly of them, probably not all nights but several times a week it is large.
PBSMT Iter. 4	I dream constantly of them, probably not all nights but several times a week it is certain.
Reference	I constantly dream of them, perhaps not every night, but several times a week for sure.

Source	La protéine que nous utilisons dans la glace réagit avec la langue à pH neutre.
NMT Epoch 1	The protein that we use in the ice with the language to pH.
NMT Epoch 8	The protein we use into the ice responds with language to pH neutral.
NMT Epoch 45	The protein we use in ice responds with the language from pH to neutral.
PBSMT Iter. 0	The protein that used in the ice responds with the language and pH neutral.
PBSMT Iter. 1	The protein that we use in the ice responds with the language to pH neutral.
PBSMT Iter. 4	The protein that we use in the ice reacts with the language to a neutral pH.
Reference	The protein we are using in the ice cream reacts with your tongue at neutral pH.

Source	Selon Google, les déguisements les plus recherchés sont les zombies, Batman, les pirates et les sorcières.
NMT Epoch 1	According to Google, there are more than zombies, Batman, and the pirates.
NMT Epoch 8	Google’s most wanted outfits are the zombies, Batman, the pirates and the evil.
NMT Epoch 45	Google said the most wanted outfits are the zombies, Batman, the pirates and the witch.
PBSMT Iter. 0	According to Google, fancy dress and most wanted fugitives are the bad guys, Wolverine, the pirates and their minions.
PBSMT Iter. 1	According to Google, the outfits are the most wanted fugitives are zombies, Batman, pirates and witches.
PBSMT Iter. 4	According to Google, the outfits, the most wanted list are zombies, Batman, pirates and witches.
Reference	According to Google, the highest searched costumes are zombies, Batman, pirates and witches.

Table 5: **Unsupervised translations.** Examples of translations on the French-English pair of *newstest* 2014 at different iterations of training. For PBSMT, we show translations at iterations 0, 1 and 4, where the model obtains BLEU scores of 15.4, 23.7 and 24.7 respectively. For NMT, we show examples of translations after epochs 1, 8 and 42, where the model obtains BLEU scores of 12.3, 17.5 and 24.2 respectively. Iteration 0 refers to the PBSMT model obtained using the unsupervised phrase table, and an epoch corresponds to training the NMT model on 500k monolingual sentences. At the end of training, both models generate very good translations.

<i>Russian</i> → <i>English</i>	
Source	Изменения предусматривают сохранение льготы на проезд в общественном пассажирском транспорте.
Hypothesis	The changes involve keeping the benefits of parking in a public passenger transportation.
Reference	These changes make allowances for the preservation of discounted travel on public transportation.
Source	Шесть из 10 республиканцев говорят, что они согласны с Трампом по поводу иммиграции.
Hypothesis	Six in 10 Republicans say that they agree with Trump regarding immigration.
Reference	Six in 10 Republicans say they agree with Trump on immigration.
Source	Metcash пытается защитить свои магазины IGA от натиска Aldi в Южной Австралии и Западной Австралии .
Hypothesis	Metcash is trying to protect their shops IGA from the onslaught of Aldi in South Australia and Western Australia.
Reference	Metcash is trying to protect its IGA stores from an Aldi onslaught in South Australia and Western Australia.
Source	В них сегодня работают четыре сотни студентов из столичных колледжей и вузов.
Hypothesis	Others today employs four hundreds of students from elite colleges and universities.
Reference	Four hundred students from colleges and universities in the capital are working inside of it today.

Table 6: **Unsupervised translations: Russian-English.** Examples of translations on the Russian-English pair of *newstest* 2016 using the PBSMT model.

<i>German → English</i>	
Source	Flüchtlinge brauchen Unterkünfte : Studie warnt vor Wohnungslücke
PBSMT	Refugees need accommodation : Study warns Wohnungslücke
NMT	Refugees need forestry study to warn housing gap
PBSMT+NMT	Refugees need accommodation : Study warns of housing gap
Reference	Refugees need accommodation : study warns of housing gap
Source	Konflikte : Mehrheit unterstützt die Anti-IS-Ausbildungsmission
PBSMT	Conflict : Majority supports Anti-IS-Ausbildungsmission
NMT	Tensions support majority anti-IS-recruiting mission
PBSMT+NMT	Tensions : Majority supports the anti-IS-recruitment mission
Reference	Conflicts : the majority support anti ISIS training mission
Source	Roboterautos : Regierung will Vorreiterrolle für Deutschland
PBSMT	Roboterautos : Government will step forward for Germany
NMT	Robotic cars will pre-reiterate government for Germany
PBSMT+NMT	Robotic cars : government wants pre-orders for Germany
Reference	Robot cars : Government wants Germany to take a leading role
Source	Pfund steigt durch beschleunigtes Lohnwachstum im Vereinigten Königreich
PBSMT	Pound rises as UK beschleunigtes Lohnwachstum .
NMT	£ rises through rapid wage growth in the U.S.
PBSMT+NMT	Pound rises by accelerating wage growth in the United Kingdom
Reference	Pound rises as UK wage growth accelerates
Source	46 Prozent sagten , dass sie die Tür offen lassen , um zu anderen Kandidaten zu wechseln .
PBSMT	52 per cent said that they left the door open to them to switch to other candidates .
NMT	46 percent said that they would let the door open to switch to other candidates .
PBSMT+NMT	46 percent said that they left the door open to switch to other candidates .
Reference	46 percent said they are leaving the door open to switching candidates .
Source	Selbst wenn die Republikaner sich um einen anderen Kandidaten sammelten , schlägt Trump noch fast jeden .
PBSMT	Even if the Republicans a way to other candidates collected , beats Trump , yet almost everyone .
NMT	Even if Republicans are to donate to a different candidate , Trump takes to almost every place .
PBSMT+NMT	Even if Republicans gather to nominate another candidate , Trump still beats nearly everyone .
Reference	Even if Republicans rallied around another candidate , Trump still beats almost everyone .
Source	Ich glaube sicher , dass es nicht genügend Beweise gibt , um ein Todesurteil zu rechtfertigen .
PBSMT	I think for sure that there was not enough evidence to justify a death sentence .
NMT	I believe it 's not sure there 's enough evidence to justify a executions .
PBSMT+NMT	I sure believe there is not enough evidence to justify a death sentence .
Reference	I certainly believe there was not enough evidence to justify a death sentence .
Source	Auch wenn der Laden gut besucht ist , ist es nicht schwer , einen Parkplatz zu finden .
PBSMT	Even if the store visited it is , it is not hard to find a parking lot .
NMT	To be sure , the shop is well visited , but it 's not a troubled driveway .
PBSMT+NMT	Even if the shop is well visited , it is not hard to find a parking lot .
Reference	While the store can get busy , parking is usually not hard to find .
Source	Die Suite , in dem der kanadische Sänger wohnt , kostet am Tag genau so viel , wie ihre Mama Ewa im halben Jahr verdient .
PBSMT	The suite in which the Canadian singer grew up , costs on the day so much as their mum Vera in half year earned .
NMT	The Suite , in which the Canadian singer lived , costs day care exactly as much as her mom Ewa earned during the decade .
PBSMT+NMT	The suite , in which the Canadian singer lived , costs a day precisely as much as her mom Ewa earned in half .
Reference	The suite where the Canadian singer is staying costs as much for one night as her mother Ewa earns in six months .
Source	Der durchschnittliche BMI unter denen , die sich dieser Operation unterzogen , sank von 31 auf 24,5 bis Ende des fünften Jahres in dieser Studie .
PBSMT	The average BMI among those who undergoing this operation decreased by 22 to 325 by the end of fifth year in this study .
NMT	The average BMI among those who undergo surgery , sank from 31 to 24,500 by the end of the fifth year in that study .
PBSMT+NMT	The average BMI among those who undergo this surgery fell from 31 to 24.5 by the end of the fifth year in this study .
Reference	The average BMI among those who had surgery fell from 31 to 24.5 by the end of their fifth year in the study .
Source	Die 300 Plakate sind von Künstlern , die ihre Arbeit im Museum für grausame Designs in Banksys Dismaland ausgestellt haben .
PBSMT	The 250 posters are by artists and their work in the museum for gruesome designs in Dismaland Banksys have displayed .
NMT	The 300 posters are posters of artists who have their work Museum for real-life Designs in Banksys and Dismalausgestellt .
PBSMT+NMT	The 300 posters are from artists who have displayed their work at the Museum of cruel design in Banksys Dismaland .
Reference	The 300 posters are by artists who exhibited work at the Museum of Cruel Designs in Banksy 's Dismaland .
Source	Bis zum Ende des Tages gab es einen weiteren Tod : Lamm nahm sich das Leben , als die Polizei ihn einkesselte .
PBSMT	At the end of the day it 's a further death : Lamb took out life as police him einkesselte .
NMT	By the end of the day there was another death : Lamm emerged this year as police were trying to looseless him .
PBSMT+NMT	By the end of the day , there 's another death : Lamm took out life as police arrived to him .
Reference	By the end of the day , there would be one more death : Lamb took his own life as police closed in on him .
Source	Chaos folgte an der Grenze , als Hunderte von Migranten sich in einem Niemandland ansammelten und serbische Beamte mit Empörung reagierten .
PBSMT	Chaos followed at the border , as hundreds of migrants in a frontier ansammelten and Serb officers reacted with outrage .
NMT	Chaos followed an avalanche as hundreds of thousands of immigrants fled to a mansdom in answerage and Serbian officers responded with outrage .
PBSMT+NMT	Chaos followed at the border as hundreds of immigrants gathered in a bush town and Serb officers reacted with outrage .
Reference	Chaos ensued at the border , as hundreds of migrants piled up in a no man 's land , and Serbian officials reacted with outrage .
Source	'' Zu unserer Reise gehörten viele dunkle Bahn- und Busfahrten , ebenso Hunger , Durst , Kälte und Angst '' , schrieb sie .
PBSMT	'' To our trip included many of the dark rail and bus rides , such as hunger , thirst , cold and fear , '' she wrote .
NMT	'' During our trip , many included dark bus and bus trips , especially hunger , Durst , and cold fear , '' she wrote .
PBSMT+NMT	'' Our trip included many dark rail and bus journeys , as well as hunger , thirst , cold and fear , '' she wrote .
Reference	'' Our journey involved many dark train and bus rides , as well as hunger , thirst , cold and fear , '' she wrote .

Table 7: **Unsupervised translations: German-English.** Examples of translations on the German-English pair of *newstest* 2016 using the PBSMT, NMT, and PBSMT+NMT.