English-Czech Systems in WMT19: Document-Level Transformer

Martin Popel, Dominik Macháček, Michal Auersperger, Ondřej Bojar and **Pavel Pecina**

Charles University, Faculty of Mathematics and Physics,
Institute of Formal and Applied Linguistics,
Malostranské náměstí 25, 118 00 Prague, Czech Republic
surname@ufal.mff.cuni.cz

Abstract

We describe our NMT systems submitted to the WMT19 shared task in English→Czech news translation. Our systems are based on the Transformer model implemented in either Tensor2Tensor (T2T) or Marian framework.

We aimed at improving the adequacy and coherence of translated documents by enlarging the context of the source and target. Instead of translating each sentence independently, we split the document into possibly overlapping multi-sentence segments. In case of the T2T implementation, this "document-level"-trained system achieves a +0.6 BLEU improvement ($p < 0.05$) relative to the same system applied on isolated sentences. To assess the potential effect document-level models might have on lexical coherence, we performed a semi-automatic analysis, which revealed only a few sentences improved in this aspect. Thus, we cannot draw any conclusions from this week evidence.

1 Introduction

Neural machine translation has reached a point, where the quality of automatic translation measured on isolated sentences is similar on average to the quality of professional human translations. Hassan et al. (2018) report achieving a "human parity" on Chinese→English news translation. Bojar et al. (2018, p. 291) report that our last year's English→Czech system (Popel, 2018) was evaluated as significantly better ($p < 0.05$) than the human reference. However, it has been shown (Läubli et al., 2018; Toral et al., 2018) that evaluating the quality of translation of news articles on isolated sentences without the context of the whole document is not sufficient. It can bias the evaluation results because systems that ignore the context are not penalized in the evaluation for these context-related errors; and vice versa: sys-

tems (or humans) that take the context into account may be unfairly penalized. Läubli et al. (2018) show that while the difference between human and machine translation in adequacy is not significant when evaluated on isolated sentences, it is significant (humans are better) when evaluated on whole documents. This suggests that there are some inter-sentential phenomena where MT applied on isolated sentences is lacking.

Since assessing the performance of document-level systems is one of the goals of WMT19 (Barrault et al., 2019), we decided to build NMT systems trained for translation of longer segments than single sentences. In this paper, we describe our five NMT systems submitted to WMT19 English→Czech news translation task (see Table 1). They are based on the Transformer model (Vaswani et al., 2017) and on our submission from WMT18 (Popel, 2018). Our new contributions are (i) adaptation of the baseline single-sentence models to translate multiple adjacent sentences in a document at once, so the Transformer can attend to inter-sentence relations and achieve better document-level translation quality, as was already showed to be effective by Jean et al. (2017); and (ii) reimplementation of our last year's submission in the Marian framework (Junczys-Dowmunt et al., 2018).

This paper is organized as follows: In Section 2, we describe our training data and its augmentation to overlapping multi-sentence sequences. We describe also the hyper-parameters of our models in the two frameworks. Section 3 follows with a description of the document-level decoding strategies. Section 4 reports and discusses the results of automatic (BLEU) evaluation.

official name	description
CUNI DocTransformer T2T	Document level trained Transformer in T2T.
CUNI DocTransformer Marian	Document level trained Transformer in Marian.
CUNI Transformer T2T 2019	Same model as CUNI DocTransformer T2T, but applied on single sentences (i.e. with no cross-sentence context).
CUNI Transformer T2T 2018	Same model as in the last year (Popel, 2018).
CUNI Transformer Marian	Reimplementation of the last year's model in Marian.

Table 1: Brief descriptions of our WMT19 systems. In the rest of the paper, we omit the CUNI (Charles University) prefix for brevity.

data set	sentence pairs (k)	words (k) EN	CS
CzEng 1.7	57 065	618 424	543 184
Europarl v7	647	15 625	13 000
News Commentary v12	211	4 544	4 057
CommonCrawl	162	3 349	2 927
WikiTitles	361	896	840
EN NewsCrawl 2016–17	47 483	934 981	
CS NewsCrawl 2007–17	65 383		927 348
CS NewsCrawl 2018	12 983		181 004
total	184 295	1 577 819	1 672 360

Table 2: Training data sizes (in thousands).

2 Experimental Setup

2.1 Data sources

Our training data (see Table 2) are constrained to the data allowed in the WMT2019 shared task. "Transformer T2T 2018" and "Transformer Marian" use only the data allowed in WMT2018, which does not include CS NewsCrawl 2018 and WikiTitles. All the data were preprocessed, filtered and backtranslated by the same process as in Popel (2018). We selected the originally English part of newstest2016 for validation, following the idea of CZ/nonCZ tuning in Popel (2018), but excluding the CZ tuning because the WMT2019 test set was announced to contain only original English sentences and no translationese.

2.2 Training Data Context Augmentation

In WMT19, all the training data from Table 2 are available with document boundaries (and unlike in previous years the sentences are not shuffled).[1] We extracted all sequences of consecutive sentences with at most 1000 characters.[2] Our

---

[1] In WikiTitles, each pair of titles is considered a separate document. We decided to upsample this source 23 times, but we have not evaluated the effect of this on the final quality.

[2] The limit of 1000 characters was chosen rather arbitrarily. A 1000-characters long sequence from our training data contains on average about 15 sentences (165 English and 144 Czech words).

context-augmented data consists of pairs of such sequences, where the source sequence has always the same number of sentences as the target sentence. We separate the sentences in each sequence with a special token,[3] so that we can easily extract sentence alignment after decoding. We randomly shuffle the augmented training sequences, but we keep separately the authentic parallel and synthetic (backtranslated) data, so that we can apply *concat backtranslation* (Popel, 2018).

Note that this particular way of context augmentation implicitly upsamples sentences from longer documents relative to sentences from shorter documents. We leave the analysis of this effect and possible alternative samplings for future work.

2.3 Model Hyper-parameters

2.3.1 Tensor2Tensor

Our three systems with "T2T" in the name are implemented in the Tensor2Tensor framework (Vaswani et al., 2018), version 1.6.0. The model and training parameters this year are identical to our last year's (WMT18) submission (Popel, 2018), with just two exceptions: First, we trained on 10 GPUs instead of 8 GPUs, thus using the effective batch size of 29k subwords instead of 23k subwords. Second, we used max_length=200 instead of 150. This means we discard all training sequences longer than 200 subwords. With our 32k joint subword vocabulary, a word contains on average 1.5 subwords. Thus effectively, the sequence-length limit used in T2T training was in most cases lower than 1000 characters – on average it was 785 characters.

2.3.2 Marian

Our two systems with "Marian" in the name use the Marian framework (Junczys-Dowmunt et al., 2018), in the latest stable version 1.7.6. We chose

---

[3] Any token not present in the training data can be used, but it should be included in the subword vocabulary.

Marian for its fast and efficient training and decoding. Due to the good results of "CUNI Transformer" in WMT18 evaluation and lack of time and resources for exhaustive parameter search, we reconstructed all its hyperparameters in Marian wherever possible. Therefore, we trained with the following options:

```
--type transformer --enc-depth 6
--dec-depth 6 --dim-emb 1024
--transformer-dim-ffn 4096
--transformer-heads 16
--transformer-dropout 0.0
--transformer-dropout-attention 0.1
--transformer-dropout-ffn 0.1
--lr-warmup 20000
--lr-decay-inv-sqrt 20000
--optimizer-params 0.9 0.98 1e-09
--clip-norm 5 --label-smoothing 0.1
--learn-rate 0.0002
--exponential-smoothing
```

We used the same learning rate as T2T and estimated the number of warmup training steps so the model consumed approximately the same number of sentences as T2T in warmup. Instead of T2T's default SubwordTextEncoder, we used SentencePiece (Kudo and Richardson, 2018) with its default parameters to obtain a shared vocabulary of 32,000 entries from untokenized training data. We set the maximal sentence length to 150 and decoded with beam size 4.

We could not use Adafactor (Shazeer and Stern, 2018) optimizer as in T2T, because it is not implemented in Marian. We used Adam instead.

We did not set the batch size manually, but used the `--mini-batch-fit` parameter to determine the mini-batch size automatically based on sentence lengths to fit the available memory. We estimated the workspace memory to 13,900 MB as the largest possible on our hardware. We shuffled the training data before training and did not use any advanced reordering to fit more non-padding tokens into a training batch as in T2T.

Another difference is the checkpoint averaging: while our T2T models are (uniform) averages of the last 8 checkpoints from the last 8 hours of training, our Marian models use the exponential moving average regularization method (`--exponential-smoothing`) applied after each update, as suggested by the Marian authors.

2.4 Training

The summary of hardware used for training is in Table 3. First, we trained a non-document models on single sentences, on concatenation of out-domain authentic data and in-domain synthetic

systems	#GPUs	GPU memory	GPU type
T2T 2018	8	11GB	GTX 1080 Ti
T2T 2019	10	11GB	GTX 1080 Ti
Marian	8	16GB	Quadro P5000

Table 3: Hardware used for our systems.

datasets. We trained "Transformer Marian" model for 17 days until the epoch 18. We observed the last improvement in validation BLEU at 15 days and 18 hours of training, in step 1,266M, which we selected as the final model "Transformer Marian". The "DocTransformer T2T" model was trained for 9 days (660k steps).

3 Document-Level Systems

Our document-level models were created by training on the context-augmented data described in Section 2.2. We used different strategies for document-level decoding in Marian and in T2T.

3.1 Decoding in Marian

For "DocTransformer Marian" decoding, we decided to reduce the context to up to three consecutive sentences because decoding of longer contexts was time-consuming and our time was constrained. Each sentence appeared as the first, second or third sentence in a 3-sentence context (1st/3, 2nd/3, 3rd/3) if possible.[4] We experimented also with a 2-sentence context (1st/2, 2nd/2) and no context (1st/1, i.e. the baseline).

We compared dev-set BLEU scores of these six setups and selected the following strategy for the selection of the final translation: For each sentence, if possible and if the translation is "valid", use 2nd/3. If not possible or "valid", use 1st/3, followed by 2nd/2, 1st/2 and 1st/1.

We consider a translation "valid" if it contains the same number of sentences (delimited by a special sentence-boundary character) as the input. We excluded translations containing a given word more than 20 times and translations with a word longer than 49 characters. This rule detected non-meaningful outputs that we observed in validation. We decided to not use 3rd/3 because these translations were the least accurate ones.

Based on the validation BLEU scores, we selected two checkpoints for the final document-level translation. The checkpoint at 2,044M steps

---

[4] For the first sentence in a document only 1st/3 is possible, for the second sentence only 1st/3 or 2nd/3 is possible, etc.

was used for 1st/3, 2nd/3 and 2nd/2. The checkpoint at 1,775M steps was used elsewhere (1st/2 and 1st/1).

3.2 Decoding in T2T

In an initial experiment, we split the test set into non-overlapping sequences of sentences with at most 1000 characters, following the maximum sequence length used in training. We realized that the translation quality is very low, especially close to the end of each translated sequence. Sometimes the number of output sentences (detected based on the special separator character) was different than the number of input sentences. We hypothesized that the reason of low quality is that there are not enough 1000-character sequences in the training data (cf. Section 2.2). With non-overlapping splits, we achieved the best dev-set BLEU, when lowering the limit to about 700 characters.

We further experimented with overlapping splits, where each sequence to be translated consists of

- pre-context: sentences which are ignored in the translation and serve only as a context for better translation of the main content,

- main content: sentences which are used for the final translation,

- post-context: sentences which are ignored, similarly to the pre-context.

Based on a small dev-set BLEU hyperparameter search, we selected the following length limits: pre-context of up to 200 characters (splitting on word boundaries), main content of up to 500 characters (whole sentences only) and post-context of up to 900 characters minus the length of the pre-context and main content (whole sentences only). After the main decoding, we joined together the translations of main contents of all sequences. In rare cases (8 sentences out of 3611), when there were not enough sentences in the translated sequence, we used a single-sentence translation as a backup.

3.3 Post-processing

For T2T systems, we used the same post-processing as last year (Popel, 2018): We deleted the repetitions of phrases of one to four words appearing directly after each other more than two times, and converted the quotation symbols to

system	BLEU uncased	BLEU cased	chrF2 cased
DocTransformer T2T	**31.03**	**29.94**	*0.5628*
Transformer T2T 2018	*30.93*	*29.86*	**0.5630**
Transformer T2T 2019	30.42	29.39	0.5552
DocTransformer Marian	29.17	28.14	0.5466
Transformer Marian	29.20	28.13	0.5474
UEdin	29.00	27.89	0.5516

Table 4: Automatic evaluation on `newstest2019`. Significantly different BLEU scores ($p < 0.05$ bootstrap resampling) are separated by a horizontal line.

„lower and upper". This is considered as standard in Czech formal texts. For Marian, we applied only the conversion of quotation symbols.

4 Results

4.1 Automatic Evaluation

Table 4 reports the automatic metrics of our English→Czech systems submitted to WMT2019, plus the best other system – UEdin (Marian system trained by University of Edinburgh). The automatic metrics are calculated using sacreBLEU 1.3.2 (Post, 2018) and their signatures are:

- BLEU+case.mixed+lang.en-cs+numrefs.1+smooth.exp+tok.13a,

- BLEU+case.lc+lang.en-cs+numrefs.1+smooth.exp+tok.intl and

- chrF2+case.mixed+lang.en-cs+numchars.6+numrefs.1+space.False.

4.2 Explaining the Difference of T2T and Marian

The two comparable systems using the closest possible settings we were able to achieve and identical data, "Transformer Marian" and "Transformer T2T 2018", did not perform equally. The last year's T2T system was around 1.73 BLEU better at the point, where both systems had enough training time to converge. We hypothesize this was caused by the parameters, in which they differ: (i) Marian uses Adam optimizer, T2T Adafactor; (ii) Marian had 8 16GB GPUs and T2T 8 11GB GPUs, it means 128GB vs 88GB in total. We assume Marian is not as effective in memory usage, or we used bigger than optimal memory (and thus batch) size; (iii) Marian uses different batch ordering; (iv) in Marian, we used the exponential moving average, T2T used uniform averaging of the last 8 checkpoints.

4.3 Doc-Level Evaluation

We hypothesized that by providing the translation model with larger attendable context, the resulting translations display larger lexical consistency. We could demonstrate it by finding less examples where an English polysemous word is translated to two or more Czech non-synonymous lemmata within one document.

To evaluate the hypothesis, we word-aligned the source and target sentences using `fast_align` (Dyer et al., 2013).[5] We then lemmatized the aligned words (both English and Czech) using MorphoDiTa (Straková et al., 2014) and considered all instances where a single English lemma was aligned to at least two Czech lemmata in a single document. Since our focus was on evaluating the difference between non-context and document-level models, we selected only the English lemmata with different number of aligned Czech lemmata in the two types of systems. Two pairs of models were compared: "DocTransformer T2T" vs. "Transformer T2T 2019" and "Doc-Transformer Marian" vs. "Transformer Marian". The final pool of examples was evaluated manually.

We found only one and three instances for the Marian and T2T models, respectively, where the document-level variant performed better than the non-context variant. The examples are shown in Table 5. We also found a possible counter-example where the document-level model performed worse than the non-context model, but the evaluation is not clear-cut. The example is shown in Table 6.

Because there are too few examples for any meaningful quantitative analysis, we conclude more data is needed to evaluate the potential benefit a document-level model could have on lexical consistency. By doing manual evaluation, we found the cases where the inter-sentential context is necessary for determining the correct meaning of a polysemous word are rare.

5 Conclusion

We were not able to replicate our last year's T2T system in Marian, but we acknowledge several differences in the setup. We were not able to im-

prove the sentence-level Marian system BLEU by adding a context of up to three sentences. Our document-level trained T2T system achieved an insignificant improvement ($+0.1$ BLEU) over our last year's sentence-level T2T system, but applying this system on sentences led to a significant worsening (-0.6 BLEU).

References

Loïc Barrault, Ondřej Bojar, Marta R. Costa-jussà, Christian Federmann, Mark Fishel, Yvette Graham, Barry Haddow, Matthias Huck, Philipp Koehn, Shervin Malmasi, Christof Monz, Mathias Müller, Santanu Pal, Matt Post, and Marcos Zampieri. 2019. Findings of the 2019 conference on machine translation (wmt19). In *Proceedings of the Fourth Conference on Machine Translation, Volume 2: Shared Task Papers*, Florence, Italy. Association for Computational Linguistics.

Ondřej Bojar, Christian Federmann, Mark Fishel, Yvette Graham, Barry Haddow, Matthias Huck, Philipp Koehn, and Christof Monz. 2018. Findings of the 2018 conference on machine translation (wmt18). In *Proceedings of the Third Conference on Machine Translation, Volume 2: Shared Task Papers*, pages 272–307, Belgium, Brussels. Association for Computational Linguistics.

Chris Dyer, Victor Chahuneau, and Noah A. Smith. 2013. A simple, fast, and effective reparameterization of IBM model 2. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 644–648, Atlanta, Georgia. Association for Computational Linguistics.

Hany Hassan, Anthony Aue, Chang Chen, Vishal Chowdhary, Jonathan Clark, Christian Federmann, Xuedong Huang, Marcin Junczys-Dowmunt, William Lewis, Mu Li, Shujie Liu, Tie-Yan Liu, Renqian Luo, Arul Menezes, Tao Qin, Frank Seide, Xu Tan, Fei Tian, Lijun Wu, Shuangzhi Wu, Yingce Xia, Dongdong Zhang, Zhirui Zhang, and Ming Zhou. 2018. Achieving human parity on automatic chinese to english news translation. *CoRR*, abs/1803.05567.

---

[5] To improve the reliability of automatic word alignments, we trained them on the translations together with the first 500k sentences of CzEng 1.7. Only the intersection of the source-to-target and target-to-source alignments was considered.

Sebastien Jean, Stanislas Lauly, Orhan Firat, and Kyunghyun Cho. 2017. Does neural machine translation benefit from larger context?

Marcin Junczys-Dowmunt, Roman Grundkiewicz, Tomasz Dwojak, Hieu Hoang, Kenneth Heafield, Tom Neckermann, Frank Seide, Ulrich Germann, Alham Fikri Aji, Nikolay Bogoychev, André F. T. Martins, and Alexandra Birch. 2018. Marian: Fast neural machine translation in C++. In *Proceedings of ACL 2018, System Demonstrations*, pages 116–121, Melbourne, Australia. Association for Computational Linguistics.

Taku Kudo and John Richardson. 2018. Sentence-Piece: A simple and language independent subword tokenizer and detokenizer for neural text processing. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 66–71, Brussels, Belgium. Association for Computational Linguistics.

Samuel Läubli, Rico Sennrich, and Martin Volk. 2018. Has machine translation achieved human parity? a case for document-level evaluation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4791–4796, Brussels, Belgium. Association for Computational Linguistics.

Martin Popel. 2018. Cuni transformer neural mt system for wmt18. In *Proceedings of the Third Conference on Machine Translation, Volume 2: Shared Task Papers*, pages 486–491, Belgium, Brussels. Association for Computational Linguistics.

Matt Post. 2018. A Call for Clarity in Reporting BLEU Scores. *CoRR*, arXiv/1804.08771.

Noam Shazeer and Mitchell Stern. 2018. Adafactor: Adaptive Learning Rates with Sublinear Memory Cost. *CoRR*, arXiv/1804.04235.

Jana Straková, Milan Straka, and Jan Hajič. 2014. Open-Source Tools for Morphology, Lemmatization, POS Tagging and Named Entity Recognition. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 13–18, Baltimore, Maryland. Association for Computational Linguistics.

Antonio Toral, Sheila Castilho, Ke Hu, and Andy Way. 2018. Attaining the unattainable? reassessing claims of human parity in neural machine translation. In *Proceedings of the Third Conference on Machine Translation, Volume 1: Research Papers*, pages 113–123, Belgium, Brussels. Association for Computational Linguistics.

Ashish Vaswani, Samy Bengio, Eugene Brevdo, Francois Chollet, Aidan N. Gomez, Stephan Gouws, Llion Jones, Łukasz Kaiser, Nal Kalchbrenner, Niki Parmar, Ryan Sepassi, Noam Shazeer, and Jakob Uszkoreit. 2018. Tensor2tensor for neural machine translation. *CoRR*, abs/1803.07416.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems 30*, pages 6000–6010. Curran Associates, Inc.

6 Appendix

source	[...] to meet Craig Halkett's header across goal. The hosts were content to let Rangers play in front of them, knowing they could trouble the visitors at set pieces. And that was the manner in which the crucial **goal** came. Rangers conceded a free-kick [...]
T2T	A to byl způsob, jakým přišel rozhodující **cíl** (*aim*).
T2T-doc	A to byl způsob, jakým přišel rozhodující **gól** (*goal*).
source	Elizabeth Warren Will Take "Hard Look" At Running For President in 2020, Massachusetts Senator Says Massachusetts Senator Elizabeth Warren said on Saturday she would take a "hard look" at running for president following the midterm elections. During a town hall in Holyoke, Massachusetts, Warren confirmed she'd consider **running**. "It's time for women to go to Washington and fix our broken government and that includes a woman at the top," she said, according to The Hill. [...]
T2T	Na radnici v Holyoke v Massachusetts Warrenová potvrdila, že uvažuje o **útěku** (*escape*).
T2T-doc	Na radnici v Holyoke ve státě Massachusetts Warrenová potvrdila, že o **kandidatuře** (*candidacy*) uvažuje.
source	At 6am, just as Gegard Mousasi and Rory MacDonald were preparing to face each other, viewers in the UK were left stunned when the coverage changed to Peppa Pig. Some were unimpressed after they had stayed awake until the early hours especially for the **fight**. [...]
T2T	Na některé to neudělalo žádný dojem, když zůstali vzhůru až do časných ranních hodin, zvláště kvůli **rvačce** (*crawl*).
T2T-doc	Na některé to neudělalo žádný dojem, když zůstali vzhůru až do ranních hodin, zejména kvůli **zápasu** (*match*).
source	[...] she felt "terrified of retaliation" and was worried about "being publicly humiliated." The 34-year-old says she is now seeking to overturn the **settlement** as she continues to be traumatized by the alleged incident. [...]
Marian	Čtyřiatřicetiletá žena tvrdí, že se nyní snaží o zrušení **osady** (*village*), protože je nadále traumatizována údajným incidentem.
Marian-doc	34letá žena tvrdí, že nyní usiluje o zrušení **vyrovnání** (*compensation*), protože je nadále traumatizována údajným incidentem.

Table 5: Examples of non-context model errors corrected by the document-level models.

source	New cancer vaccine can teach the immune system to 'see' rogue cells New cancer vaccine can teach the immune system to 'see' rogue cells and kill them Vaccine teaches immune system to recognise rogue cells as part of treatment Method involves extracting immune cells from a **patient**, altering them in lab They can then 'see' a protein common to many cancers and then reinjected A trial vaccine is showing promising results in **patients** with a range of cancers. One woman treated with the vaccine, which teaches the immune system to recognise rogue cells, saw her ovarian cancer disappear for more than 18 months. The method involves extracting immune cells from a **patient**, altering them in the laboratory so they can "see" a protein common to many cancers called HER2, and then reinjecting the cells.
T2T	Nová protinádorová vakcína může naučit imunitní systém „vidět" zlovolné buňky Nová protinádorová vakcína může naučit imunitní systém „vidět" zlovolné buňky a zabít je. Vakcína učí imunitní systém rozpoznávat zlovolné buňky jako součást léčby Metoda zahrnuje odebrání imunitních buněk z **pacienta** a jejich změnu v laboratoři. Pak mohou vidět protein, který je společný mnoha nádorům, a znovu ho vstříknout. Zkušební vakcína vykazuje slibné výsledky u **pacientů** s řadou nádorových onemocnění. Jedna žena léčená vakcínou, která učí imunitní systém rozeznávat zlovolné buňky, byla svědkem vymizení rakoviny vaječníků na více než 18 měsíců. Metoda spočívá v odebrání imunitních buněk z **pacienta**, jejich přeměně v laboratoři, aby mohli „vidět" protein, který je společný mnoha nádorům nazývaným HER2, a poté reinjekci buněk.
T2T-doc	Nová protinádorová vakcína může naučit imunitní systém „vidět" zlovolné buňky Nová protinádorová vakcína může naučit imunitní systém „vidět" zlovolné buňky a zabít je Vakcína učí imunitní systém rozpoznávat zlovolné buňky jako součást léčby Metoda zahrnuje extrakci imunitních buněk z **pacienta**, jejich změnu v laboratoři Poté mohou „vidět" bílkovinu společnou mnoha nádorovým onemocněním a poté ji znovu nasadit Zkušební vakcína vykazuje slibné výsledky u **pacientů** s řadou nádorových onemocnění. Jedna žena léčená touto vakcínou, která učí imunitní systém rozpoznávat zlovolné buňky, byla svědkem vymizení rakoviny vaječníků na více než 18 měsíců. Tato metoda zahrnuje odebrání imunitních buněk od **pacientky** (*female patient*), jejich změnu v laboratoři, aby mohly „vidět" bílkovinu, která je společná mnoha nádorům nazývaným HER2, a poté reinjekci buněk.

Table 6: The example of an error introduced by a document-level model.