Word Translation Prediction for Morphologically Rich Languages with Bilingual Neural Networks

Ke Tran   Arianna Bisazza   Christof Monz

Informatics Institute, University of Amsterdam

University of Amsterdam

Summary

Choosing the correct surface form requires linguistic features of source and target context:

- in phrase-based SMT, access to source context depends on phrase segmentation
- linguistic features depend on available annotation tools and manual feature engineering

Our approach enables:

- accurate prediction of target translation stem and suffix given fixed amount of context
- automatic learning of relevant features with neural network architecture

This results in:

- significantly higher accuracies than maximum-likelihood baseline
- better ranking of translation options, small but significant BLEU gains in English-to-Russian

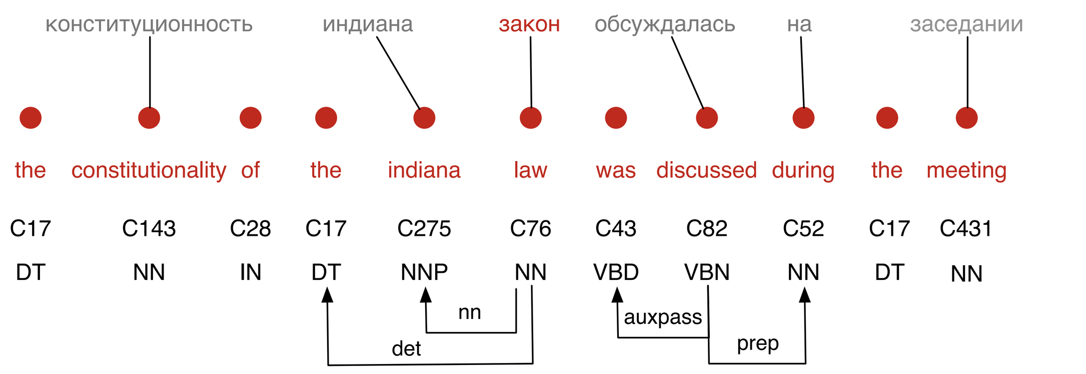
Motivation

[конституционность] [индиана закон] [обсуждалась] [на заседании]

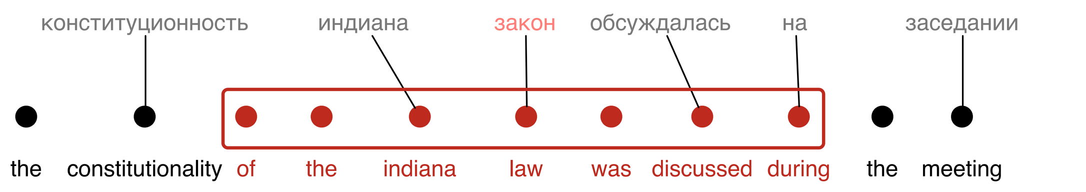
[the constitutionality of the] [indiana law] [was discussed] [during the meeting]

Wrong case translation of "law" due to the rare word "Indiana". The language model does not help in this case → Need a model to improve morphological prediction

Task: Predict target word translation given the source sentence and alignment link



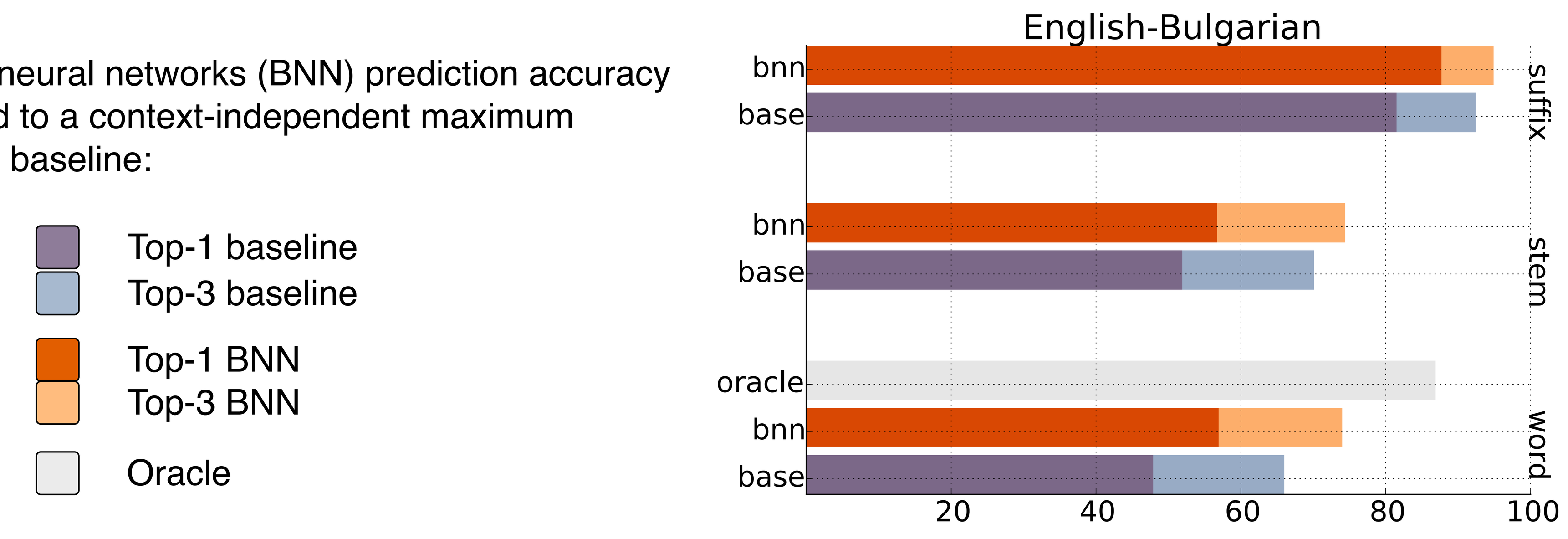
Previous approaches rely on linguistic annotations such as POS, dependency relations,...

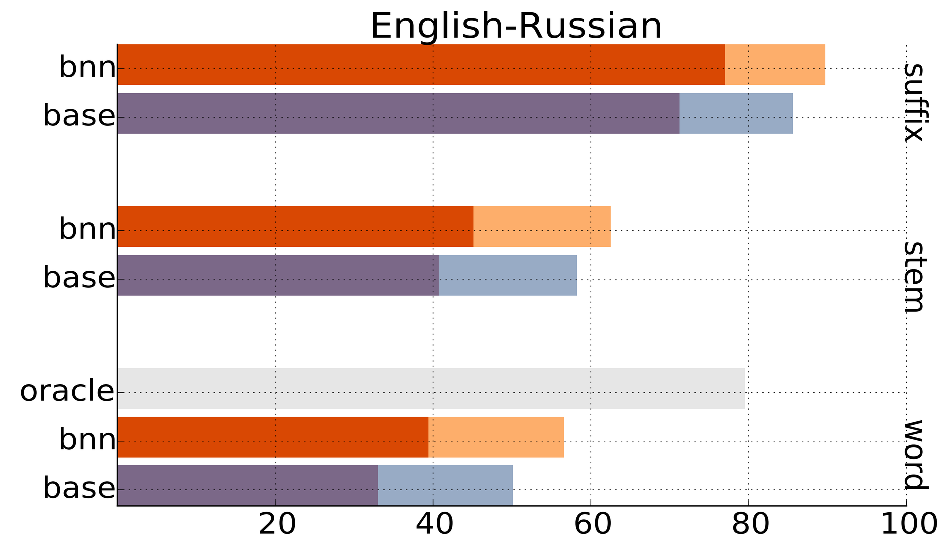


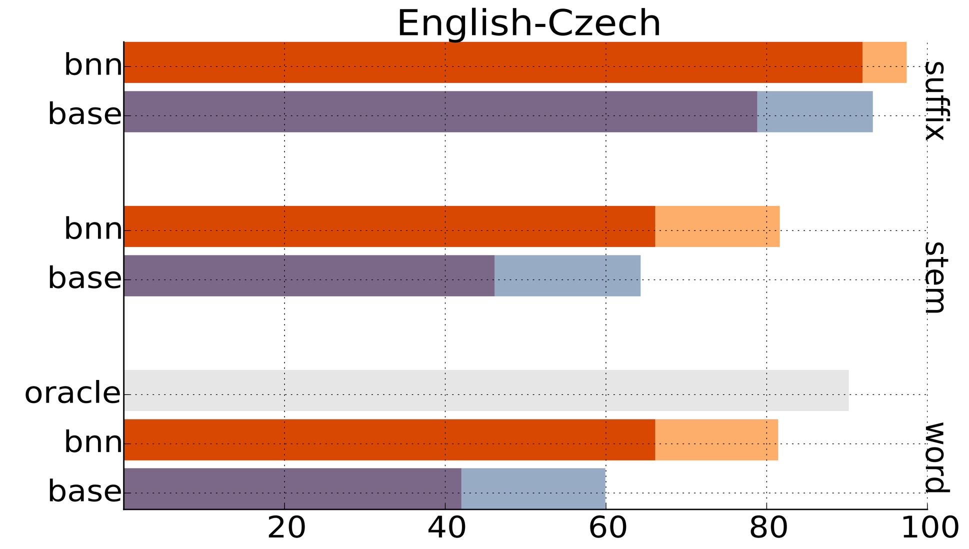
This work: use local context and learn relevant features automatically.

Translation prediction results

Bilingual neural networks (BNN) prediction accuracy compared to a context-independent maximum likelihood baseline:







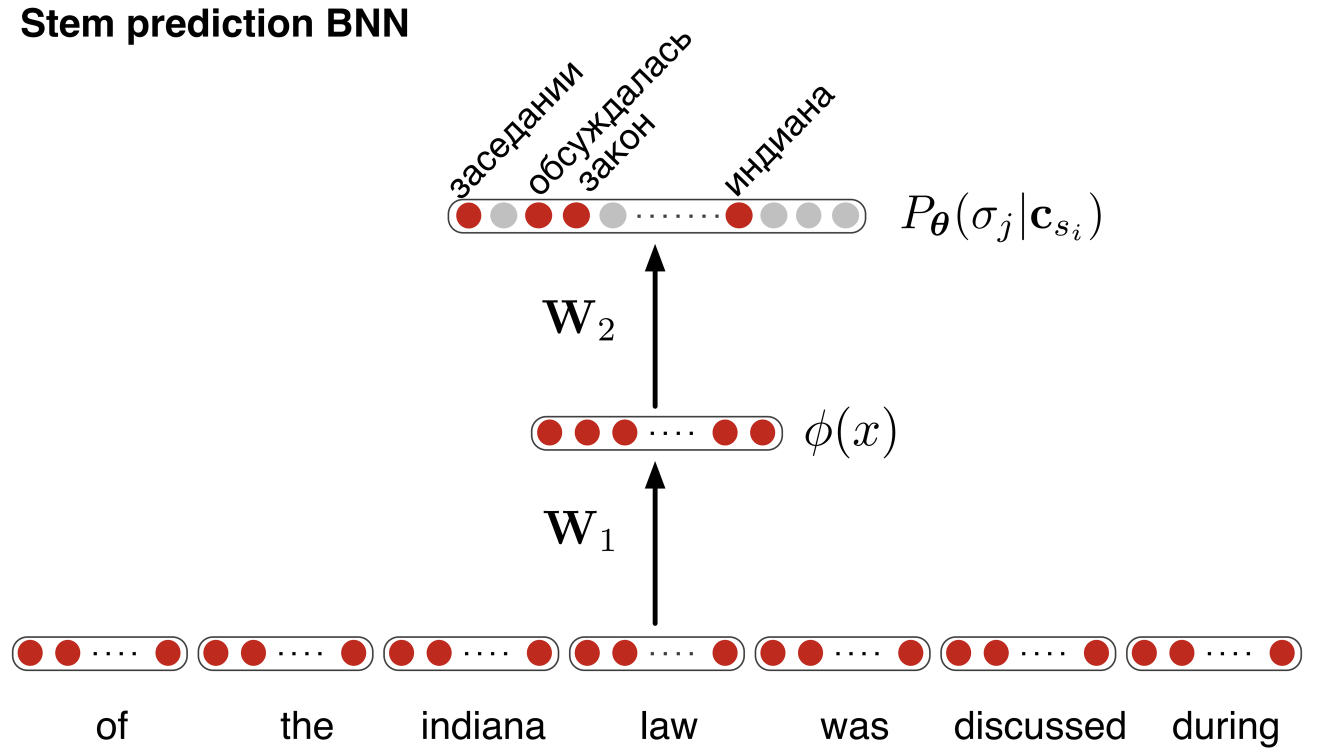
Approach: Bilingual Neural Network (BNN)

Factorize word translation probability into stem and suffix probabilities: $p(t_j|\mathbf{c}_{s_i}) = p(\sigma_j|\mathbf{c}_{s_i})p(\mu_j|\mathbf{c}_{s_i},\sigma_j)$

Conditional probability normalized over the set of translation candidates instead of the whole output vocabulary:

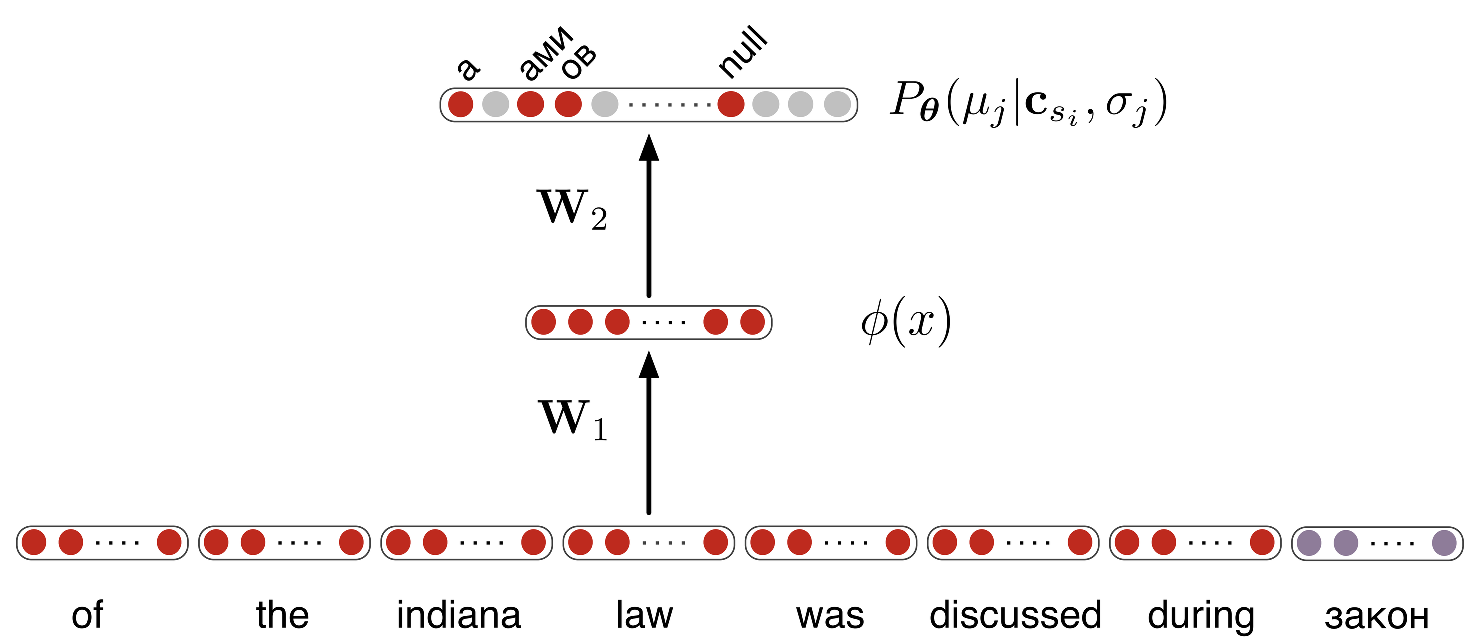
$$p(t_j|\mathbf{c}_{s_i}) = \frac{\exp(\mathbf{W}_2^{[t_j]}\phi(\mathbf{W}_1\mathbf{r}(\mathbf{c}_{s_i})))}{\sum_{t_k\in\mathrm{GEN}(s_i)}\exp(\mathbf{W}_2^{[t_k]}\phi(\mathbf{W}_1\mathbf{r}(\mathbf{c}_{s_i})))}$$

Stem prediction BNN



Input: fixed-size source context window

Suffix prediction BNN



Input: fixed-size source context window + target stem

SMT results

Compute BNN score for each phrase pair, similarly to lexical weighting:

$$P_{\text{BNN-p}}(\tilde{s},\tilde{t},a) = \prod_{i=1}^{|\tilde{s}|}\begin{cases}\frac{1}{|\{a_i\}|}\sum_{j\in\{a_i\}} P_{\text{BNN}}(t_j|\mathbf{c}_{s_i}) & \text{if } |\{a_i\}|>0\\ P_{\text{mle}}(\text{NULL}|s_i) & \text{otherwise}\end{cases}$$

	MLE	BNN	
	p(elf)	stem	suffix
indiana law / индиана закон	0.6	0.6	0.1
indiana law / индиана закона	0.1	0.6	**0.7**
indiana law / индиана законов	0.1	0.6	0.1

Effect of our BNN models on English-to-Russian translation quality (BLEU%) :

SMT system	wmt12 (dev)	wmt13 (test)
Baseline	24.7	18.9
+ stem/suff. BNN	25.1	**19.3***
Base+suff.LM	24.5	19.2
+ word. BNN	24.5	19.2
+ stem/suff. BNN	24.7	**19.6***

Target word coverage analysis of the English-to-Russian SMT system before and after adding the morphological BNN models:

	Base	+BNN
reference/MT-search-space [top-1]	57.6%	59.0%
reference/MT-search-space [top-3]	70.7%	70.9%
reference/MT-search-space [top-30]	86.0%	85.0%
reference/MT-output	50.0%	50.7%

This research was funded in part by the Netherlands Organization for Scientific Research (NWO) under project numbers 639.022.213 and 612.001.218