**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 translation quality (BLEU%)

**Approach: Bilingual Neural Network (BNN)**

Factorize word translation probability into stem and suffix probabilities:

\[ p(t_j | c_s) = p(\sigma_j | c_s) p(\mu_j | c_s, \sigma_j) \]

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

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

**Motivation**

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 translation quality (BLEU%)

**Translation prediction results**

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

**SMT results**

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

\[ P_{\text{BNN}}(\delta, t, a) = \prod_{i=1}^{i=m} \left( \frac{1}{1 + \sum_{j \neq \delta} P_{\text{BNN}}(t_j | c_s)} \right) \]

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

<table>
<thead>
<tr>
<th>SMT system</th>
<th>wmt12 (dev)</th>
<th>wmt13 (test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline + BNN</td>
<td>25.1</td>
<td>19.3*</td>
</tr>
<tr>
<td>+ stem/suff. BNN</td>
<td>24.5</td>
<td>19.2</td>
</tr>
<tr>
<td>+ word. BNN</td>
<td>24.5</td>
<td>19.2</td>
</tr>
<tr>
<td>+ stem/suff. BNN</td>
<td>24.7</td>
<td>19.6*</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Target word coverage analysis</th>
<th>Base + BNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>reference/MT-search-space (top-1)</td>
<td>57.6%</td>
</tr>
<tr>
<td>reference/MT-search-space (top-3)</td>
<td>70.7%</td>
</tr>
<tr>
<td>reference/MT-search-space (top-10)</td>
<td>66.0%</td>
</tr>
<tr>
<td>reference/MT-output</td>
<td>50.0%</td>
</tr>
</tbody>
</table>

**This work:** use local context and lean relevant features automatically.