Transfer from a single Distant Language:
Data Hallucination helps a lot more than cross-lingual transfer for distant languages. In a few cases, cross-lingual transfer hurts performance over training only on hallucinated data! (see examples to the right)

Transfer from multiple related languages: Beneficial in many cases where such languages are available. (See examples for Karelian and Tatar below)

Standing Issues:
1. Handling of different writing scripts is still an issue for character-level cross-lingual transfer. Maybe use IPA? (g2p systems lacking, though)
2. The benefit of the adversarial language discriminator is inconsistent.
3. Can we train a single system for all languages?

Novel Two-Step Attention for Disentangled Inputs
First, encode the tag sequence and the lemma:
For each decoding step,
  a) get context from tag attention
  b) obtain a tag-informed decoder state
  c) attend over lemma
  d) produce output character

Additional Biases
1. encourage monotonic attention: use an additional copying task (see training regime below)
2. encourage attention coverage of the two sources:
3. Language discriminator over the encoder outputs (with gradient reversal):

Data Hallucination
1. Find a "stem"-like region based on character alignment that remains unchanged
2. Randomly replace the inside characters
(examples from Greek)

Cross-Lingual Training Regime
1. Train only on copying task over all languages large batch size and learning rate
2. Train on both inflection (80%) and copying (20%) tasks for all languages upsample the low-resource language learning rate decay and restart the optimizer
3. Train only on the test language inflection task small batch size scheduled sampling

Analysis
Transfer from a single Distant Language:
Data Hallucination helps a lot more than cross-lingual transfer for distant languages. In a few cases, cross-lingual transfer hurts performance over training only on hallucinated data! (see examples to the right)

Transfer from multiple related languages: Beneficial in many cases where such languages are available. (See examples for Karelian and Tatar below)

Disentangled Attentions lead to interpretable attention matrices
See examples from Kazakh and Modern Greek below, where the appropriate morphological tags are attended when the corresponding affixes are generated.

Average Results over 100 language pairs:

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wu and Cotterell (2019)</td>
<td>48.5</td>
</tr>
<tr>
<td>this work:</td>
<td></td>
</tr>
<tr>
<td>two-step attention and</td>
<td>48.8</td>
</tr>
<tr>
<td>single language transfer</td>
<td></td>
</tr>
<tr>
<td>+ data hallucination</td>
<td>60.1</td>
</tr>
<tr>
<td>+language adversarial loss</td>
<td>60.8</td>
</tr>
<tr>
<td>+multi-language transfer</td>
<td>63.8</td>
</tr>
<tr>
<td>oracle</td>
<td>68.2</td>
</tr>
</tbody>
</table>

Highlights
Morphological Inflection is the task where, given a lemma, e.g. aguar and a set of morphological tags, e.g. V; PRS; 2; PL; IND; one has to generate the correctly inflected form, e.g. aguà

In low-resource settings this task is still very challenging. We combine several techniques:
1. a novel two-step attention for the decoder
2. data hallucination
3. multi-tasking with a simple copying task
4. cross-lingual transfer from multiple languages

and achieve state-of-the-art results over 44 test languages (from the SIGMORPHON 2019 challenge), with a gain of more than 15 points over the baseline.

Average Results over 100 language pairs:

Highlights
Video (with quick 1 min highlights!)
Code: https://github.com/antonisa/inflection
Skype: antonis.anastasopoulos
Email us for any questions!