A Supplemental Material

A.1 Morphological tagger

We adapt the parser's encoder architecture for our morphological tagger. Following notation in Section 2, each word w_i is represented by its context-sensitive encoding, $\mathbf{h}_i$ (Eq. 2). The encodings are then fed into a feed-forward neural network with two hidden layers—each has a ReLU non-linearity—and an output layer mapping the to the morphological tags, followed by a softmax. We set the size of the hidden layer to 100 and use dropout probability 0.2. We use Adam optimizer with initial learning rate 0.001 and clip gradients to 5. We train each model for 20 epochs with early stopping. The model is trained to minimized the cross-entropy loss.

Since we do not have additional data with the same annotations, we use the same UD dataset to train our tagger. To prevent overfitting, we only use the first 75% of training data for training[8]. After training the taggers, we predict the case for the training, development, and test sets and use them for dependency parsing.

A.2 Results on morphological tagging

Table 9 and 10 present morphological tagging results for German and Russian. We found that German and Russian have similar pattern to Czech (Table 5), where morphological case seems to be preserved in the encoder because they are useful for dependency parsing. In these three fusional languages, contextual information helps character-level model to predict the correct case. However, its performance still behind the oracle.

We observe a slightly different pattern on Finnish results (Table 11). The character embeddings achieves almost similar performance as the oracle embeddings. This results highlights the differences in morphological process between Finnish and the other fusional languages. We observe that performance of the encoder representations are slightly worse than the embeddings.

Feature	baseline	embedding		encoder	
		char	oracle	char	oracle
Case	35.2	35.7	**100**	80.8	**99.7**
Gender	56.8	63.6	**100**	75.7	**78**
Number	59.1	67.1	**100**	78.3	**93.9**
All	34	34.3	**100**	63.6	**78.5**

Table 9: Morphological tagging results for German.

Feature	baseline	embedding		encoder	
		char	oracle	char	oracle
Case	71.6	80.5	**100**	90.4	**98.5**
Gender	87.7	97.4	**100**	**69.9**	57.3
Number	83.7	94.5	**100**	**85.7**	83.8
All	71.3	77.2	**99.9**	**56.9**	47.2

Table 10: Morphological tagging results for Russian.

Feature	baseline	embedding		encoder	
		char	oracle	char	oracle
Case	56	96.7	**100**	88.9	**91.4**
Number	56.4	97.4	**100**	81.9	**89.5**
All	55.8	**95**	91.6	74	**82.7**

Table 11: Morphological tagging results for Finnish.

---

[8]We tried other settings, i.e. 25%, 50%, 100%, but in general we achieve best result when we use 75% of the training data.