

# Fluency Boost Learning and Inference for Neural Grammatical Error Correction: Supplementary Notes

Tao Ge Furu Wei Ming Zhou  
Microsoft Research Asia, Beijing, China  
{tage, fuwei, mingzhou}@microsoft.com

## 1 Tips for implementing self-boost and dual-boost

As discussed in Section 3.2 and 3.3 in our paper submission, the error correction model  $\Theta_{crt}$  and error generation model  $\Theta_{gen}$  are dynamically updated during training. When the training set is not large enough,  $\Theta_{crt}$  and  $\Theta_{gen}$  may not be well learnt (i.e., not fully converged) in the first and the second epoch. In this case, the disfluency candidates generated by  $\Theta_{crt}$  or  $\Theta_{gen}$  are likely to change the original sentence’s meaning, breaking the fluency condition defined in Section 3. For example, the model may generate “Jim likes reading books.” from the sentence “Tom likes reading books.”. Such sentence pairs are obviously undesirable as a training instance for GEC.

Therefore, to avoid the above-mentioned cases, we recommend creating disfluency candidates after training 5-10 epochs if the original training set is not large enough.

## 2 Details for training the RNN language model

The language model we use for helping with decoding and computing a sentence’s fluency score is a 2-layer GRU RNN language model, as introduced in Section 5.2 in our paper submission. The dimensionality of the hidden units and word embedding is 500 and the output vocabulary is 50,000. We train the language model using Adam optimizer with a learning rate of 0.0001. We also apply dropout to non-recurrent connections at a ratio of 0.2.

## 3 Analysis of disfluency candidates

We show the proportion and examples of synthesized errors in Table 1. Note that the stats are from 1,000 sentences randomly sampled from the dis-

Error type	Ratio (%)	Example
Punctuation	11.6	However(.) as mentioned above, waviness causes an appreciable reduction.
Single/Plural	8.8	These theory (theories) map well from the underconnectivity theory of autism.
Verb Form	18.8	But it was <u>be</u> (being) shown in the time of Pausanias in the 2nd century AD.
ArtOrDet	20.5	Anarchists were (the) first to suggest occupying workplaces.
Prep	8.9	<u>In</u> (On) November 9, 2011, Jefferson County declared bankruptcy.
Tense	4.3	Finally, Achilles <u>find</u> (found) his prey.
Confusing word	12.1	Anarchism is often defined as a politic (political) philosophy which holds the states to be undesirable, unnecessary or harmful.
Others	15.0	-
<b>Total</b>	100.0	-

Table 1: Errors synthesized by dual-boost learning. The content in the round brackets are the correct content.

fluency candidates generated by dual-boost learning.

According to Table 1, the most common generated error types are ArtOrDet and Verb Form which are also the most common mistakes that an English learner tends to make. Fluency boost learning enriches the training data with large numbers of generated errors of the common types, strengthening the model’s ability of correcting common errors.