A The Seq2Edits Model

This section contains a short description of the key elements of the Seq2Edits model which we use extensively in this work. For a detailed discussion we refer the reader to the full paper from Stahlberg and Kumar (2020).

Seq2Edits represents sequence transduction as a sequence of edit operations. Each edit operation is applied to a span in the source sentence which is either copied (labelled with `SELF`) or replaced by replacement tokens. Fig. 5 shows that each edit is represented by a 3-tuple consisting of an error tag t_n, the source span end position p_n (the implicit source span start position is p_{n-1}), and the replacement token r_n. The sequence of 3-tuples is predicted auto-regressively by a modified Transformer with multiple target tapes.

Unlike Stahlberg and Kumar (2020), in this work we use Seq2Edits as a corruption model. Thus, each edit corresponds to an artificial error of a certain type (represented by the error tag t_n) rather than a correction. In Sec. 2 we constrain the error tag tape with FSTs (Fig. 1) to force the generation of a specific error type.

B Example Corruptions

The main contribution of this work is using explicit error type tags to control and diversify the output of corruption models since conventional corruption models without tags tend to generate dull and monotonous output. In the examples in Table 9 the untagged corruption model simply deletes a determiner while the tagged corruption model is able to produce a wide variety of human-like writing errors. This observation is supported by Fig. 6 which shows that the untagged corruption model mainly outputs simplistic punctuation errors while the untagged model has a much better coverage of other more complex error types.

C Example Corrections

Table 10 shows example outputs from two systems: with and without using our new $C4_{200M}$ pre-training data set. The system with $C4_{200M}$ tends to be more fluent because it is trained on much more English text from diverse sources. The first example in Table 10 demonstrates that pre-training on $C4_{200M}$ also seems to help learning semantics – the non-$C4_{200M}$ output ("we always stock up on traffic") is grammatically correct, but the $C4_{200M}$ model has learned that "stocking up on traffic" is nonsensical, and that being "stuck in traffic" is probably a better match for the author's intent.

The system with $C4_{200M}$ is able to work well across a wider range of domains by proposing more radical changes such as long-range reorderings to improve the fluency as demonstrated by the movement of "well" in the second example or the major rewrite in the third example of Table 10. However, this can make the $C4_{200M}$ model more prone to subtle changes in semantics, as shown in the last example ("public transport is a cost effective (..) allocation" $\rightarrow$ "public transport is cost effective and ..").

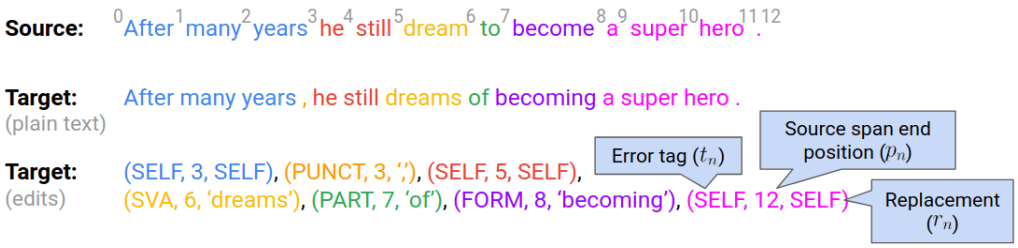

Figure 5: Representing grammatical error correction as a sequence of span-based edit operations. The implicit start position for a source span is the end position of the previous edit operation. SELF indicates spans that are copied over from the source sentence ($\mathbf{x}$). The probability of the first two edits is given by: $P(\text{After many years ,}|\mathbf{x}) = P(t_1 = \text{SELF}|\mathbf{x}) \cdot P(p_1 = 3|\text{SELF}, \mathbf{x}) \cdot P(r_1 = \text{SELF}|\text{SELF}, 3, \mathbf{x}) \cdot P(t_2 = \text{PUNCT}|\text{SELF}, 3, \text{SELF}, \mathbf{x}) \cdot P(p_2 = 3|\text{SELF}, 3, \text{SELF}, \text{PUNCT}, \mathbf{x}) \cdot P(r_2 = ,|\text{SELF}, 3, \text{SELF}, \text{PUNCT}, 3, \mathbf{x})$. Figure and caption are taken from Stahlberg and Kumar (2020).

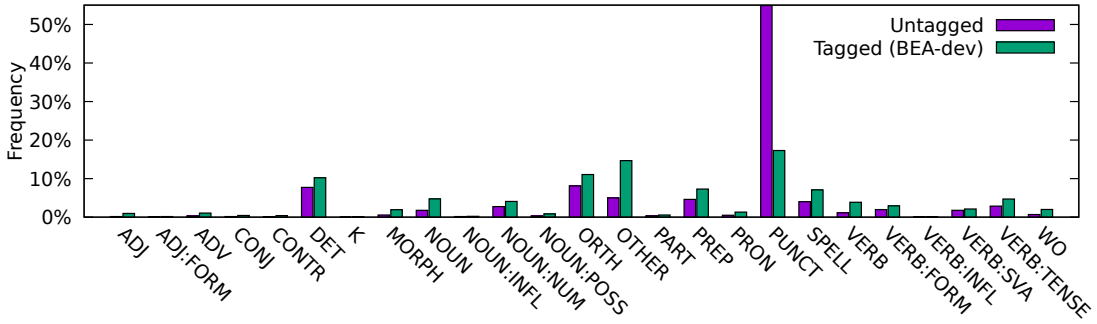

Figure 6: Error tag distributions for untagged and tagged Seq2Edits corruption models. Most of the untagged corruptions are dull punctuation errors (PUNCT) whereas the tagged corruptions are more diverse as they follow the BEA-dev tag distribution in Fig. 3.

Input (clean)	I'm learning a lot and the students are very friendly.
Untagged corruption (1-best)	I'm learning a lot and the students are very friendly.
Untagged corruption (2-best)	I'm learning a lot **and students** are very friendly.
Tagged corruption	
ADJ	I'm learning a lot and the students are very **friendliness**.
ADJ:FORM	I'm learning a lot and the students are very **friendlies**.
ADV	I'm learning a lot and the students are **so** friendly.
CONJ	I'm learning a **lot the** students are very friendly.
CONTR	**I learning** a lot and the students are very friendly.
DET	I'm learning a lot **and students** are very friendly.
K	**I 'm** learning a lot and the students are very friendly.
MORPH	I'm learning a lot and the students are very **friendship**.
NOUN	I'm learning **many things** and the students are very friendly.
NOUN:INFL	I'm learning a lot and the **studentes** are very friendly.
NOUN:NUM	I'm learning a lot and the **student are** very friendly.
NOUN:POSS	I'm learning a lot and the **student's** are very friendly.
ORTH	I'm learning **alot** and the students are very friendly.
OTHER	I'm learning **very much** and the students are very friendly.
PART	I'm learning **up** a lot and the students are very friendly.
PREP	I'm learning **to** a lot and the students are very friendly.
PRON	**Learning** a lot and the students are very friendly.
PUNCT	I'm learning a lot and the students are very **friendly**
SPELL	I'm **lerning** a lot and the students are very friendly.
VERB	I'm learning a lot and the **students very** friendly.
VERB:FORM	I'm **learn** a lot and the students are very friendly.
VERB:INFL	I'm **learnes** a lot and the students are very friendly.
VERB:SVA	I'm learning a lot and the **students is** very friendly.
VERB:TENSE	**I learn** a lot and the students are very friendly.
WO	**I'm a lot learning** and the students are very friendly.
Input (clean)	The British summertime was first introduced in England in 1908.
Untagged corruption (1-best)	The British summertime was first introduced in England in 1908.
Untagged corruption (2-best)	**British** summertime was first introduced in England in 1908.
Tagged corruption	
ADJ	The **English** summertime was first introduced in England in 1908.
ADJ:FORM	The **Britishest** summertime was first introduced in England in 1908.
ADV	The British summertime **was introduced** in England in 1908.
CONJ	**And British** summertime was first introduced in England in 1908.
CONTR	The British **summertime's** first introduced in England in 1908.
DET	**British** summertime was first introduced in England in 1908.
K	British summertime was introduced in **England-in** 1908.
MORPH	The **Britishes** summertime was first introduced in England in 1908.
NOUN	The British **summer** was first introduced in England in 1908.
NOUN:INFL	The British **Summertimes** was first introduced in England in 1908.
NOUN:NUM	The British **Summertimes** was first introduced in England in 1908.
NOUN:POSS	The **British's** summertime was first introduced in England in 1908.
ORTH	The British **summer time** was first introduced in England in 1908.
OTHER	The British summertime was introduced **for first time** in England in 1908.
PART	The British summertime was first introduced **to** in England in 1908.
PREP	The British summertime was first introduced **to** England in 1908.
PRON	**It's** British summertime was first introduced in England in 1908.
PUNCT	The British summer time was first introduced in England in **1908**
SPELL	The **Britishi** summertime was first introduced in England in 1908.
VERB	British summertime was first **invented** in England in 1908.
VERB:FORM	The British summertime was first **introduce** in England in 1908.
VERB:INFL	The British summertime was first **introduceed** in England in 1908.
VERB:SVA	The British summertime **were** first introduced in England in 1908.
VERB:TENSE	The British summertime **is** first introduced in England in 1908.
WO	The British summertime was **introduced first** in England in 1908.

Table 9: Example outputs from tagged and untagged Seq2Edits corruption models.

Source	In my **country, taipei, we always stock in traffic** for about one hour in the morning.
Reference	In my **city, Taipei, we are always stuck in traffic** for about one hour in the morning.
Output (without $C4_{200M}$)	In my **country, Taipei, we always stock up on traffic** for about one hour in the morning.
Output (with $C4_{200M}$)	In my **country, Taipei, we are always stuck in traffic** for about one hour in the morning.
Source	Back to the topic, I don't know **well** any of member of in my family**,** I don't trust them, **honestly**.
Reference	Back to the topic. I don't know any of the members of my family **well.** I don't trust them, **to be honest**.
Output (without $C4_{200M}$)	Back to the topic, I don't know **well** any members of my family**,** I don't trust them, **honestly**.
Output (with $C4_{200M}$)	Back to the topic, I don't know any members of my family **well,** I don't trust them, **honestly**.
Source	What **is the "Family" meaning?**
Reference	What **does "Family" mean?**
Output (without $C4_{200M}$)	What **is the "Family" meaning?**
Output (with $C4_{200M}$)	What **does "Family" mean?**
Source	Basketball has many benefits **not to me but also everyone.**
Reference	Basketball has many benefits**, not just for me but also for everyone.**
Output (without $C4_{200M}$)	Basketball has many benefits **not to me but also everyone.**
Output (with $C4_{200M}$)	Basketball has many benefits **not only for me but also for everyone.**
Source	The manager of **all this** project is Miss June Sid, our science teacher.
Reference	The manager of **this** project is Miss June Sid, our science teacher.
Output (without $C4_{200M}$)	The manager of **all this** project is Miss June Sid, our science teacher.
Output (with $C4_{200M}$)	The manager of **this** project is Miss June Sid, our science teacher.
Source	Volleyball is a sport **play every place, when** I **travel** on the beach I like **plays** with my sister in the sand and **after** we **are going to** the sea.
Reference	Volleyball is a sport **that is played everywhere. When** I **am** on the beach I like **playing** with my sister in the sand and **then** we **go in** the sea.
Output (without $C4_{200M}$)	Volleyball is a sport **played in every place. When** I **travel** on the beach I like **to play** with my sister in the sand and **after** we **go to** the sea.
Output (with $C4_{200M}$)	Volleyball is a sport **played everywhere. When** I **travel** on the beach, I like **playing** with my sister in the sand and **after** we **go to** the sea.
Source	Because public transport is **a cost effective** and better resource allocation in mass transport system.
Reference	Because public transport is **a more cost-effective** and better resource allocation in mass transport system.
Output (without $C4_{200M}$)	Because public transport is **a cost effective** and better resource allocation in **a** mass transport system.
Output (with $C4_{200M}$)	Because public transport is **cost effective** and **there** is better resource allocation in **the** mass transport system.

Table 10: Example outputs on BEA-dev (Bryant et al., 2019) after two stage fine-tuning. The system without $C4_{200M}$ corresponds to the first row in Fig. 4. The outputs with $C4_{200M}$ were generated from the system in Table 6e.