A Task-specific Tag Sets

Tables 14 to 16 list the non-trivial tag sets for text normalization, sentence fusion, and grammatical error correction respectively. In addition to the tags listed in the tables, we use the tags SELF and EOS (end of sequence). For sentence splitting and simplification we use the trivial tag set consisting of SELF, NON_SELF, and EOS.

B Example Outputs

Tables 17 to 19 provide example outputs from the *Seq2Edits* model. We use word-level rather than subword- or character-level source positions and collapse multi-word replacements into a single operation in the edit representation examples for clarity.

Example outputs of our sentence fusion system are shown in Table 17. The predicted tags capture the variety of strategies for sentence fusion, such as simple connector particles (SINGLE_S_COORD), cataphoras (SINGLE_CATAPHORA), verb phrases (SINGLE_VP_COORD), relative clauses (SINGLE_RELATIVE), and appositions (SINGLE_APPOSITION). The last example in Table 17 demonstrates that our model is able to produce even major rewrites.

Table 18 compares our model with a full sequence baseline on English text normalization. Correctly predicting non-trivial tags helps our model to choose the right verbalizations. In the first example in Table 18, our model predicts the CARDINAL tag rather than ORDINAL and thus produces the correct verbalization for '93'. In the second example, our model generates a time expression for '1030' and '1230' as it predicted the DATE tag for these

Tag	Description
PLAIN	Ordinary word
PUNCT	Punctuation
TRANS	Transliteration
LETTERS	Letter sequence
CARDINAL	Cardinal number
VERBATIM	Verbatim reading of character sequence
ORDINAL	Ordinal number
DECIMAL	Decimal fraction
ELECTRONIC	Electronic address
DIGIT	Digit sequence
MONEY	Currency amount
FRACTION	Non-decimal fraction
TIME	Time expression
ADDRESS	Street address

Table 14: Semiotic class tags for text normalization copied verbatim from the Table 3 caption of Sproat and Jaitly (2016).

Tag	Description
PAIR_ANAPHORA	Anaphora
PAIR_CONN	Discourse connective
PAIR_CONN_ANAPHORA	Discourse connective + anaphora
PAIR_NONE	None (Control)
SINGLE_APPOSITION	Apposition
SINGLE_CATAPHORA	Cataphora
SINGLE_CONN_INNER	Inner connective
SINGLE_CONN_INNER_ANAPH.	Inner connective + anaphora
SINGLE_CONN_START	Forward connective
SINGLE_RELATIVE	Relative clause
SINGLE_S_COORD	Sentence coordination
SINGLE_S_COORD_ANAPHORA	Sentence coordination + anaphora
SINGLE_VP_COORD	Verb phrase coordination

Table 15: DiscoFuse discourse types. The type descriptions are copied verbatim from Table 7 of Geva et al. (2019). The SINGLE and PAIR prefixes indicate whether the input is a single sentence or two consecutive sentences.

Tag	Description
ADJ	Adjective ("big" → "wide")
ADJ:FORM	Comparative or superlative adjective errors.
ADV	Adverb ("speedily" → "quickly")
CONJ	Conjunction ("and" → "but")
CONTR	Contraction ("n't" → "not")
DET	Determiner ("the" → "a")
MORPH	Morphology
NOUN	Noun ("person" → "people")
NOUN:INFL	Noun inflection
NOUN:NUM	Noun number ("cat" → "cats")
NOUN:POSS	Noun possessive ("friends" → "friend's")
ORTH	Orthography case and/or whitespace errors.
OTHER	Other
PART	Particle ("(look) in" → "(look) at")
PREP	Preposition ("of" → "at")
PRON	Pronoun ("ours" → "ourselves")
PUNCT	Punctuation ("!" → ".")
SPELL	Spelling ("genectic" → "genetic", "color" → "colour")
UNK	Unknown: The annotator detected an error but was unable to correct it.
VERB	Verb ("ambulate" → "walk")
VERB:FORM	Verb form
VERB:INFL	Verb inflection: misapplication of tense morphology.
VERB:SVA	Subject-verb agreement ("(He) have" → "(He) has")
VERB:TENSE	Verb tense (includes inflectional and periphrastic tense, modal verbs and passivization).
WO	Word order ("only can" → "can only")

Table 16: ERRANT tag vocabulary for grammatical error correction copied verbatim from Table 2 of Bryant et al. (2017).

Source	$_0$ Some $_1$ Haemulon $_2$ species $_3$ eat $_4$ plankton $_5$ in $_6$ the $_7$ open $_8$ water $_9$. $_{10}$ Most $_{11}$ seek $_{12}$ small $_{13}$ prey $_{14}$ on $_{15}$ the $_{16}$ seabed $_{17}$. $_{18}$
Reference	Some Haemulon species eat plankton in the open water , but most seek small prey on the seabed .
Edit model	Some Haemulon species eat plankton in the open water (SELF) , but most (SINGLE_S_COORD) seek small prey on the seabed . (SELF)
Edits	(SELF, 9, SELF), (SINGLE_S_COORD, 11, ‘, but most’), (SELF, 18, SELF)
Source	$_0$ It $_1$ is $_2$ a $_3$ fan $_4$ favourite $_5$. $_6$ It $_7$ has $_8$ been $_9$ played $_{10}$ in $_{11}$ almost $_{12}$ every $_{13}$ concert $_{14}$ to $_{15}$ date $_{16}$ since $_{17}$ its $_{18}$ initial $_{19}$ performance $_{20}$. $_{21}$
Reference	Being a fan favourite , it has been played in almost every concert to date since its initial performance .
Edit model	Being (SINGLE_CATAPHORA) a fan favourite (SELF) , it (SINGLE_CATAPHORA) has been played in almost every concert to date since its initial (SELF) performance . (SELF)
Edits	(SINGLE_CATAPHORA, 2, ‘Being’), (SELF, 5, SELF), (SINGLE_CATAPHORA, 7, ‘, it’), (SELF, 21, SELF)
Source	$_0$ Boggs $_1$ was $_2$ decommissioned $_3$ on $_4$ 20 $_5$ March $_6$ 1946 $_7$. $_8$ Boggs $_9$ was $_{10}$ sold $_{11}$ for $_{12}$ scrap $_{13}$ on $_{14}$ 27 $_{15}$ November $_{16}$ 1946 $_{17}$. $_{18}$
Reference	Boggs was decommissioned on 20 March 1946 and sold for scrap on 27 November 1946 .
Edit model	Boggs was decommissioned on 20 March 1946 (SELF) and (SINGLE_VP_COORD) sold for scrap on 27 November 1946 (SELF)
Edits	(SELF, 7, SELF), (SINGLE_VP_COORD, 10, ‘and’), (SELF, 18, SELF)
Source	$_0$ The $_1$ river $_2$ Dulais $_3$ flows $_4$ through $_5$ the $_6$ village $_7$. $_8$ The $_9$ river $_{10}$ Dulais $_{11}$ was $_{12}$ often $_{13}$ referred $_{14}$ to $_{15}$ as $_{16}$ the $_{17}$ Black $_{18}$ River $_{19}$ due $_{20}$ to $_{21}$ pollution $_{22}$ by $_{23}$ coal $_{24}$ dust $_{25}$ from $_{26}$ the $_{27}$ local $_{28}$ mining $_{29}$ industry $_{30}$. $_{31}$
Reference	The river Dulais flows through the village and was often referred to as the Black River due to pollution by coal dust from the local mining industry .
Edit model	The river Dulais (SELF) , which (SINGLE_RELATIVE) was often referred to as the Black River due to pollution by coal dust from the (SELF) local mining industry (SELF) , flows through the village . (SINGLE_RELATIVE)
Edits	(SELF, 3, SELF), (SINGLE_RELATIVE, 11, ‘, which’), (SELF, 30, SELF), (SINGLE_RELATIVE, 31, ‘flows through the village .’)
Source	$_0$ Baldwin $_1$ of $_2$ Bourcq $_3$ married $_4$ Morphia $_5$ and $_6$ Joscelin $_7$ of $_8$ Courtenay $_9$ married $_{10}$ a $_{11}$ daughter $_{12}$ of $_{13}$ Constantine $_{14}$. $_{15}$ Morphia $_{16}$ is $_{17}$ a $_{18}$ daughter $_{19}$ of $_{20}$ Gabriel $_{21}$ of $_{22}$ Melitene $_{23}$. $_{24}$
Reference	Baldwin of Bourcq married Morphia , a daughter of Gabriel of Melitene , and Joscelin of Courtenay married a daughter of Constantine .
Edit model	Baldwin of Bourcq married Morphia (SELF) , a daughter of Gabriel of Melitene (SINGLE_APPOSITION) , and Joscelin of Courtenay married (SELF) a daughter of Constantine . (SELF)
Edits	(SELF, 5, SELF), (SINGLE_APPOSITION, 5, ‘, a daughter of Gabriel of Melitene’), (SELF, 15, SELF), (SINGLE_APPOSITION, 24, DEL)

Table 17: Sentence fusion examples from the DiscoFuse dataset (Geva et al., 2019).

Source	$_0$ Jungle $_1$ Strike $_2$ ” $_3$ Sega $_4$ Force $_5$ July $_6$ 93 $_7$ ($_8$ issue $_9$ 19 $_{10}$) $_{11}$, $_{12}$ pp $_{13}$. $_{14}$
Reference	Jungle Strike sil Sega Force July ninety **three** sil issue nineteen sil sil p p sil
Full seq.	Jungle Strike sil Sega Force july ninety **third** sil issue nineteen sil sil p p sil
Edit model	Jungle Strike (SELF) sil (PUNCT) Sega Force July (SELF) ninety **three** (CARDINAL) sil (PUNCT) issue (SELF) nineteen (CARDINAL) sil (PUNCT) sil (PUNCT) p p (VERBATIM) sil (PUNCT)
Edits	(SELF, 2, SELF), (PUNCT, 3, ‘sil’), (SELF, 6, SELF), (CARDINAL, 7, ‘ninety three’), (PUNCT, 8, ‘sil’), (SELF, 9, SELF), (CARDINAL, 10, ‘nineteen’), (PUNCT, 11, ‘sil’), (PUNCT, 12, ‘sil’), (VERBATIM, 13, ‘p p’), (PUNCT, 14, ‘sil’)
Source	$_0$ Service $_1$ operates $_2$ from $_3$ the $_4$ Courthouse $_5$ at $_6$ 1030 $_7$ am $_8$ and $_9$ 1230 $_{10}$ pm $_{11}$. $_{12}$
Reference	Service operates from the Courthouse at **ten thirty** am and **twelve thirty** p m sil
Full seq.	Service operates from the Courthouse at **one thousand thirty** a m and **one thousand two hundred thirty** p m sil
Edit model	Service operates from the Courthouse at (SELF) **ten thirty** (DATE) am and (SELF) **twelve thirty** (DATE) p m (VERBATIM) sil (PUNCT)
Edits	(SELF, 6, SELF), (DATE, 7, ‘ten thirty’), (SELF, 9, SELF), (DATE, 10, ‘twelve thirty’), (VERBATIM, 11, ‘p m’), (PUNCT, 12, ‘sil’)
Source	$_0$ 168 $_1$ ($_2$ November $_3$ 1991 $_4$) $_5$, $_6$ pp $_7$. $_8$
Reference	**one hundred sixty eight** sil november nineteen ninety one sil sil p p sil
Full seq.	**thousand one hundred sixty eight** sil november nineteen ninety one sil sil p p sil
Edit model	**one hundred sixty eight** (CARDINAL) sil (PUNCT) november nineteen ninety one (DATE) sil (PUNCT) sil (PUNCT) p p (VERBATIM) sil (PUNCT)
Edits	(CARDINAL, 1, ‘one hundred sixty eight’), (PUNCT, 2, ‘sil’), (DATE, 4, ‘november nineteen ninety one’), (PUNCT, 5, ‘sil’), (PUNCT, 6, ‘sil’), (VERBATIM, 7, ‘p p’), (PUNCT, 8, ‘sil’)

Table 18: English text normalization examples from the dataset provided by Sproat and Jaitly (2016).

Source	$_0$ It $_1$ will $_2$ be $_3$ very $_4$ cool $_5$ to $_6$ see $_7$ the $_8$ **las** $_9$ **part** $_{10}$ **mokingjay** $_{11}$! $_{12}$
Reference	It will be very cool to see the **last part of Mokingjay** !
Full seq.	It will be very cool to see the **last mokingjay** !
Edit model	It will be very cool to see the (SELF) **last** (SPELL) **part** (SELF) **of** (PREP) **mokingjay** (SELF) !
Edits	(SELF, 8, SELF), (SPELL, 9, ‘last’), (SELF, 10, SELF), (PART, 10, ‘of’), (SELF, 12, SELF)
Source	$_0$ If $_1$ she $_2$ was $_3$ n’t $_4$ awake $_5$, $_6$ why $_7$ **she** $_8$ **could** $_9$ **n’t** $_{10}$ remember $_{11}$ anything $_{12}$ after $_{13}$ that $_{14}$? $_{15}$
Reference	If she was n’t awake , why **could n’t** she remember anything after that ?
Full seq.	If she was n’t awake , why **she could n’t** remember anything after that ?
Edit model	If she was n’t awake , why (SELF) **could n’t she** (WO) remember anything after that ? (SELF)
Edits	(SELF, 7, SELF), (WO, 10, ‘could n’t she’), (SELF, 15, SELF)
Source	$_0$ **Less** $_1$ channels $_2$ means $_3$ **less** $_4$ choices $_5$. $_6$
Reference	**Fewer** channels means **fewer** choices .
Full seq.	**Less** channels means **fewer** choices .
Edit model	**Fewer** (ADJ) channels means (SELF) **fewer** (ADJ) choices . (SELF)
Edits	(ADJ, 1, ‘Fewer’), (SELF, 3, SELF), (ADJ, 4, ‘fewer’), (SELF, 6, SELF)
Source	$_0$ On $_1$ the $_2$ one $_3$ hand $_4$ travel $_5$ by $_6$ car $_7$ are $_8$ really $_9$ much $_{10}$ more $_{11}$ convenient $_{12}$ as $_{13}$ **give** $_{14}$ the $_{15}$ chance $_{16}$ **to** $_{17}$ **you** $_{18}$ to $_{19}$ be $_{20}$ independent $_{21}$. $_{22}$
Reference	On the one hand **,** travel by car **is** really much more convenient **,** as **it gives you** the chance to be independent .
Full seq.	On the one hand **,** travel by car **is** really much more convenient **,** as **it gives you** the chance to be independent .
Edit model	On the one hand (SELF) **,** (PUNCT) travel by car (SELF) **is** (VERB:SVA) really much more convenient (SELF) **,** (PUNCT) **as give** (SELF) **you** (PRON) the chance (SELF) to be independent . (SELF)
Edits	(SELF, 4, SELF), (PUNCT, 4, ‘,’), (SELF, 7, SELF), (VERB:SVA, 8, ‘is’), (SELF, 12, SELF), (PUNCT, 12, ‘,’), (SELF, 14, SELF), (PRON, 14, ‘you’), (SELF, 16, SELF), (PRON, 18, DEL), (SELF, 22, SELF)

Table 19: Grammatical error correction examples from BEA-dev (Bryant et al., 2019).

spans. The third example demonstrates that the edit model can avoid some of the ‘unrecoverable’ errors (Sproat and Jaitly, 2016) of the full sequence model such as mapping ‘168’ to ‘thousand one hundred sixty eight’.

Finally, the grammatical error correction examples in Table 19 demonstrate the practical advantage of predicting tags along with the edits as they provide useful feedback to the user. The second example in Table 19 shows that our model is able to handle more complex operations such as word reorderings. However, our model fails to inflect “give” correctly in the last example, suggesting that one weakness of our edit model compared to a full

sequence model is a weaker target side language model resulting in less fluent output. This issue can be mitigated by using stronger models e.g. this particular issue is fixed in our ensemble.