A Error Generating Modules

We explain error-generating modules in detail. The order of applying modules, the error category, and the parameters to sample threshold are shown after the module name.

Error category annotation is as follows.

- F: Functional Word Error

- I: Inflection Error

- L: Lexical Choice Error

- O: Word Order Error

- W: Writing System Error

- X: Other

Preposition

The 'preposition' modules apply to words with the Penn Treebank's part-of-speech tag IN, and the dependency tag prep, ROOT, or conj. We categorize them into function word error.

standard_prep (1, F, $\sigma = 0.05$, $\mu = 0.05$)

of, to, in, for, on, with, at $\rightsquigarrow$ ε, *of, to, in, for, on, with, by, at*

prep_from (2, F, $\sigma = 0.05$, $\mu = 0.05$)

from $\rightsquigarrow$ ε, *in, at, of, with, about, since*

prep_into (3, F, $\sigma = 0.05$, $\mu = 0.05$)

into $\rightsquigarrow$ ε (0.2), *in* (0.3), *to* (0.3), *toward* (0.1), *towards* (0.1)

prep_among (4, F, $\sigma = 0.05$, $\mu = 0.05$)

among $\rightsquigarrow$ ε, *in, on, at, about, between, amid*

prep_amongst (5, F, $\sigma = 0.05$, $\mu = 0.05$)

amongst $\rightsquigarrow$ ε, *in, on, at, about, between, amidst*

prep_amid (6, F, $\sigma = 0.05$, $\mu = 0.05$)

amid $\rightsquigarrow$ ε, *in, on, at, about, between, among*

prep_amidst (7, F, $\sigma = 0.05$, $\mu = 0.05$)

amidst $\rightsquigarrow$ ε, *in, on, at, about, between, amongst*

prep_about (8, F, $\sigma = 0.05$, $\mu = 0.05$)

about $\rightsquigarrow$ ε, *in, on, of, to, at*

prep_against (9, F, $\sigma = 0.05$, $\mu = 0.05$)

against $\rightsquigarrow$ *to, for, of, with*

prep_by (10, F, $\sigma = 0.05$, $\mu = 0.05$)

by $\rightsquigarrow$ ε, *in, on, at, for, with, of, from, through, until, till*

prep_since (11, F, $\sigma = 0.05$, $\mu = 0.05$)

since $\rightsquigarrow$ *from*

prep_until (12, F, $\sigma = 0.05$, $\mu = 0.05$)

until $\rightsquigarrow$ *by, for, to, in, up to, when*

prep_till (13, F, $\sigma = 0.05$, $\mu = 0.05$)

till $\rightsquigarrow$ *by, for, to, in, up to, when*

prep_between (14, F, $\sigma = 0.05$, $\mu = 0.05$)

between $\rightsquigarrow$ ε, *in, on, at, about, among, amongst, amid, amidst*

prep_during (15, F, $\sigma = 0.05$, $\mu = 0.05$)

during $\rightsquigarrow$ *in, for, while, when, through, across*

prep_within (16, F, $\sigma = 0.05$, $\mu = 0.05$)

within $\rightsquigarrow$ *with, in, of, on*

prep_after (17, F, $\sigma = 0.05$, $\mu = 0.05$)

after $\rightsquigarrow$ *for, by, over, from, when*

prep_before (18, F, $\sigma = 0.05$, $\mu = 0.05$)

before $\rightsquigarrow$ *on, for, when*

prep_as (19, F, $\sigma = 0.05$, $\mu = 0.05$)

as $\rightsquigarrow$ *by, of, in, on, at, like*

prep_like (20, F, $\sigma = 0.05$, $\mu = 0.05$)

like $\rightsquigarrow$ *as* (0.8), *that* (0.1), *than* (0.1)

prep_than (21, F, $\sigma = 0.05$, $\mu = 0.05$)

than $\rightsquigarrow$ ε (0.2), *to* (0.4), *from* (0.2), *over* (0.1), *beyond* (0.1)

prep_through (22, F, $\sigma = 0.05$, $\mu = 0.05$)

through $\rightsquigarrow$ *in, over, across, into, of, with, by, throughout, thru*

prep_throughout (23, F, $\sigma = 0.05$, $\mu = 0.05$)

throughout $\rightsquigarrow$ *in, over, across, into, of, with, by, through*

prep_above (24, F, $\sigma = 0.05$, $\mu = 0.05$)

above $\rightsquigarrow$ *on, from, to, over*

prep_behind (25, F, $\sigma = 0.05$, $\mu = 0.05$)

behind $\rightsquigarrow$ *after, from, of, out*

prep_below (26, F, $\sigma = 0.05$, $\mu = 0.05$)

below $\rightsquigarrow$ *in, on, by, with, under, through, within*

prep_beyond (27, F, $\sigma = 0.05$, $\mu = 0.05$)

beyond ⤳ *on, over, above, out of, far from*

prep_over (28, F, $\sigma = 0.05$, $\mu = 0.05$)

over ⤳ *on, from, to, above*

prep_across (29, F, $\sigma = 0.05$, $\mu = 0.05$)

across ⤳ *beyond, over, through, throughout, during*

prep_under (30, F, $\sigma = 0.05$, $\mu = 0.05$)

under ⤳ *in, on, by, with, below, through, within*

prep_upon (31, F, $\sigma = 0.05$, $\mu = 0.05$)

upon ⤳ *on, up, up on, over, after, to*

prep_out (32, F, $\sigma = 0.05$, $\mu = 0.05$)

out ⤳ *ε*

prep_toward (33, F, $\sigma = 0.05$, $\mu = 0.05$)

toward, towards ⤳ *to, with, of, for, in, into*

Mark

The 'mark' modules apply to words with the Penn Treebank's part-of-speech tag IN, and the dependency tag mark. We categorize them into function word error.

standard_mark (34, F, $\sigma = 0.05$, $\mu = 0.05$)

as, if, because, so, whether, while, since, although, than, though, once, whereas, whilst, like, except ⤳ *ε*

mark_for (35, F, $\sigma = 0.05$, $\mu = 0.05$)

for ⤳ *ε* (0.2), *to* (0.6), *on* (0.1), *in* (0.1)

Agent

The 'agent' modules apply to words with the Penn Treebank's part-of-speech tag IN, and the dependency tag agent. We categorize them into function word error.

agent_by (36, F, $\sigma = 0.05$, $\mu = 0.05$)

by ⤳ *ε, of, from, with, on*

agent_between (37, F, $\sigma = 0.05$, $\mu = 0.05$)

between ⤳ *ε, by, in, on, at, from, with, about, among, amongst, amid, amidst*

Pcomp

The 'pcomp' modules apply to words with the Penn Treebank's part-of-speech tag IN and the dependency tag pcomp. We categorize them into function word error.

pcomp_of (38, F, $\sigma = 0.05$, $\mu = 0.05$)

of ⤳ *ε*

pcomp_to (39, F, $\sigma = 0.05$, $\mu = 0.05$)

to ⤳ *ε*

pcomp_on (40, F, $\sigma = 0.05$, $\mu = 0.05$)

on ⤳ *at, in, of*

pcomp_for (41, F, $\sigma = 0.05$, $\mu = 0.05$)

for ⤳ *to* (0.4), *at* (0.2), *in* (0.2), *on* (0.2)

pcomp_at (42, F, $\sigma = 0.05$, $\mu = 0.05$)

at ⤳ *on, in, of*

pcomp_by (43, F, $\sigma = 0.05$, $\mu = 0.05$)

by ⤳ *at, in, on, of, from*

Dative

The 'dative' modules apply to words with the Penn Treebank's part-of-speech tag IN, and the dependency tag dative. We categorize them into function word error.

dative_to (44, F, $\sigma = 0.05$, $\mu = 0.05$)

to ⤳ *ε* (0.4), *for* (0.6)

dative_for (45, F, $\sigma = 0.05$, $\mu = 0.05$)

for ⤳ *ε* (0.4), *to* (0.6)

Adverbial Modifier

This module applies to words with the Penn Treebank's part-of-speech tag IN, and the dependency tag advmod. We categorize them into functional word error.

prep_advmod (46, F, $\sigma = 0.05$, $\mu = 0.05$)

at, as ⤳ *ε*

Preposition Insertion

post_vb_prep (47, F, $\sigma = 0.05$, $\mu = 0.05$)

This module inserts *to, in, on, at, by, for, with, of* at the position which satisfies the conditions 1., 2., and 3. below.

1. Penn Treebank's part-of-speech tag of the left word is VB, VBD, VBG, VBN, VBP, VBZ.
2. Dependency tag of the left word is acl, advcl, xcomp, pcomp.
3. 3a or 3b.
3a. Universal part-of-speech tag of the right word is NOUN, DET, ADJ, PROPN.
3b. Universal part-of-speech tag of the right word is SCONJ, and dependency tag of the right word is mark.

Pronoun

These modules apply to word such that lemma is
`-PRON-`.

First Singular

1sng_subj (48, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `nsubj, nsubjpass, conj, appos, attr`.

I $\rightsquigarrow$ ε *i*, *it*, *me*, *we*

1sng_nsubj_obj (49, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `nsubj, nsubjpass, conj, appos, attr`.

me $\rightsquigarrow$ ε, *I*, *my*

1sng_dobj (50, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `dobj`.

me $\rightsquigarrow$ ε (0.1), *I* (0.1), *for me* (0.3), *to me* (0.3), *on me* (0.05), *in me* (0.05), *at me* (0.05), *of me* (0.05)

1sng_pobj (51, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `pobj`.

me $\rightsquigarrow$ ε, *I*, *my*

1sng_dative (52, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `dative`.

me $\rightsquigarrow$ ε (0.1), *I* (0.1), *for me* (0.3), *to me* (0.35), *on me* (0.05), *in me* (0.05), *at me* (0.05)

1sng_poss (53, F, $\sigma = 0.03$, $\mu = 0.03$)

mine $\rightsquigarrow$ *my*, *me*, *ours*, *his*, *hers*

1sng_poss_det (54, F, $\sigma = 0.03$, $\mu = 0.03$)

my $\rightsquigarrow$ *I*, *me*, *mine*, *a*, *an*, *the*, *its*, *his*, *her*, *their*

1sng_reflexive (55, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `dobj, conj, nsubj, nsubjpass, npadvmod, dative, attr`.

myself $\rightsquigarrow$ ε (0.2), *to myself* (0.1), *of myself* (0.1), *for myself* (0.1), *by myself* (0.1), *with myself* (0.1), *myselves* (0.02), *to myselves* (0.02), *of myselves* (0.02), *for myselves* (0.02), *by myselves* (0.02), *with myselves* (0.02), *me* (0.1), *my* (0.02), *mine* (0.02), *yourself* (0.02), *ourselves* (0.02)

1sng_appos_reflexive (56, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `appos`.

myself $\rightsquigarrow$ ε, *me*, *myselves*, *my self*, *for myself*, *by myself*

1sng_pobj_reflexive (57, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `pobj`.

myself $\rightsquigarrow$ *me*, *my*, *my self*, *myselves*, *I*

First Plural

1plu_subj (58, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `nsubj, nsubjpass, conj, appos, nmod, compound, attr`.

we $\rightsquigarrow$ *I*, *us*, *our*, *me*, *they*, *you*

1plu_pobj (59, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `pobj`.

us $\rightsquigarrow$ *we*, *ours*, *our*, *ourselves*

1plu_dobj (60, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `dobj`.

us $\rightsquigarrow$ ε (0.1), *we* (0.1), *for us* (0.3), *to us* (0.3), *on us* (0.05), *in us* (0.05), *at us* (0.05), *of us* (0.05)

1plu_nsubj_obj (61, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `nsubj, nsubjpass, conj, appos, nmod, compound, attr`.

us $\rightsquigarrow$ ε, *we*, *our*

1plu_dative (62, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `dative`.

us $\rightsquigarrow$ ε (0.1), *we* (0.1), *for us* (0.3), *to us* (0.35), *on us* (0.05), *in us* (0.05), *at us* (0.05)

1plu_poss (63, F, $\sigma = 0.03$, $\mu = 0.03$)

ours $\rightsquigarrow$ *our*, *us*, *we*, *mine*, *theirs*, *yours*

1plu_poss_det (64, F, $\sigma = 0.03$, $\mu = 0.03$)

our $\rightsquigarrow$ ε, *we*, *us*, *ours*, *a*, *an*, *the*, *theirs*, *its*, *my*

1plu_reflexive (65, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `dobj, conj, nsubj, nsubjpass, npadvmod, dative, attr`.

ourselves $\rightsquigarrow$ ε (0.2), *to ourselves* (0.1), *of ourselves* (0.1), *for ourselves* (0.1), *by ourselves* (0.1), *with ourselves* (0.1), *ourself* (0.02), *to ourself* (0.02), *of ourself* (0.02), *for ourself* (0.02), *by ourself* (0.02), *with ourself* (0.02), *us* (0.1), *our* (0.02), *ours* (0.02), *yourselves* (0.02), *myself* (0.02)

1plu_appos_reflexive (66, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `appos`.

ourselves $\rightsquigarrow$ ε, *us*, *ourself*, *our selves*, *for ourselves*, *by ourselves*

1plu_pobj_reflexive (67, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `pobj`.

ourselves $\rightsquigarrow$ ε, *us*, *ourself*, *our selves*, *for ourselves*, *by ourselves*

Second
2nd_subj (68, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `nsubj`, `nsubjpass`, `conj`, `appos`, `nmod`, `compound`, `attr`.

you $\rightsquigarrow$ *u*, *yo*, *your*, *yours*, *me*, *they*

2nd_pobj (69, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `pobj`.

you $\rightsquigarrow$ *u*, *yo*, *your*, *yours*, *yourself*, *yourselves*

2nd_dobj (70, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `dobj`.

you $\rightsquigarrow$ ε (0.1), *u* (0.05), *yo* (0.05), *for you* (0.3), *to you* (0.3), *on you* (0.05), *in you* (0.05), *at you* (0.05), *of you* (0.05)

2nd_dative (71, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `dative`.

you $\rightsquigarrow$ ε (0.1), *u* (0.05), *yo* (0.05), *for you* (0.3), *to you* (0.35), *on you* (0.05), *in you* (0.05), *at you* (0.05)

2nd_poss (72, F, $\sigma = 0.03$, $\mu = 0.03$)

yours $\rightsquigarrow$ *you*, *your*, *theirs*, *mine*, *ours*

2nd_poss_det (73, F, $\sigma = 0.03$, $\mu = 0.03$)

your $\rightsquigarrow$ ε, *you*, *yours*, *their*, *a*, *an*, *the*, *its*, *they*

Second Singular
2sng_reflexive (74, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `dobj`, `conj`, `nsubj`, `nsubjpass`, `npadvmod`, `dative`, `attr`.

youself $\rightsquigarrow$ ε (0.2), *to yourself* (0.1), *of yourself* (0.1), *for yourself* (0.1), *by yourself* (0.1), *with yourself* (0.1), *yourselves* (0.02), *to yourselves* (0.02), *of yourselves* (0.02), *for yourselves* (0.02), *by yourselves* (0.02), *with yourselves* (0.02), *you* (0.1), *your* (0.02), *yours* (0.02), *themself* (0.02), *myself* (0.02)

2sng_appos_reflexive (75, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `appos`.

youself $\rightsquigarrow$ ε, *you*, *yourselves*, *your self*, *for yourself*, *by yourself*

2sng_pobj_reflexive (76, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `pobj`.

youself $\rightsquigarrow$ *you*, *your self*, *yourselves*, *your*

Second Plural
2plu_reflexive (77, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `dobj`, `conj`, `nsubj`, `nsubjpass`, `npadvmod`, `dative`, `attr`.

yourselves $\rightsquigarrow$ ε (0.2), *to yourselves* (0.1), *of yourselves* (0.1), *for yourselves* (0.1), *by yourselves* (0.1), *with yourselves* (0.1), *yourself* (0.02), *to yourself* (0.02), *of yourself* (0.02), *for yourself* (0.02), *by yourself* (0.02), *with yourself* (0.02), *you* (0.1), *your* (0.02), *yours* (0.02), *themselves* (0.02), *ourselves* (0.02)

2plu_appos_reflexive (78, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `appos`.

yourselves $\rightsquigarrow$ ε, *you*, *yourself*, *your selves*, *for yourselves*, *by yourselves*

2plu_pobj_reflexive (79, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `pobj`.

yourselves $\rightsquigarrow$ *you*, *your self*, *myselves*, *your*

Third Singular Masculine
3sng_masc_subj (80, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `nsubj`, `nsubjpass`, `conj`, `appos`, `nmod`, `compound`, `attr`.

he $\rightsquigarrow$ *she*, *his*, *him*, *they*, *them*, *their*

3sng_masc_pobj (81, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `pobj`.

him $\rightsquigarrow$ *her*, *he*, *his*, *their*, *them*, *himself*

3sng_masc_dobj (82, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: textttdobj.

him $\rightsquigarrow$ ε (0.1), *her* (0.05), *he* (0.05), *for him* (0.3), *to him* (0.3), *on him* (0.05), *in him* (0.05), *at him* (0.05), *of him* (0.05)

3sng_masc_nsubj_obj (83, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `nsubj`, `nsubjpass`, `conj`, `appos`, `nmod`, `compound`, `attr`.

him $\rightsquigarrow$ ε, *her*, *he*, *his*

3sng_masc_dative (84, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `dative`.

him ⇝ ε (0.1), *her* (0.05), *he* (0.05), *for him* (0.3), *to him* (0.35), *on him* (0.05), *in him* (0.05), *at him* (0.05)

3sng_masc_poss (85, F, $\sigma = 0.03$, $\mu = 0.03$)

his ⇝ ε, *he*, *him*, *her*, *a*, *an*, *the*, *its*, *theirs*

3sng_masc_reflexive (86, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: dobj, conj, nsubj, nsubjpass, npadvmod, dative, attr.

himself ⇝ ε (0.2), *to himself* (0.1), *of himself* (0.1), *for himself* (0.1), *by himself* (0.1), *with himself* (0.1), *himselves* 0.01, *to himselves* (0.01), *of himselves* (0.01), *for himselves* (0.01), *by himselves* (0.01), *with himselves* (0.01), *themself* (0.01), *to themself* (0.01), *of themself* (0.01), *for themself* (0.01), *by themself* (0.01), *with themself* (0.01), *he* (0.045), *his* (0.045), *him* (0.045), *herself* (0.045)

3sng_masc_appos_reflexive (87, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: appos.

himself ⇝ ε, *him*, *themself*, *him self*, *for himself*, *by himself*, *herself*

3sng_masc_pobj_reflexive (88, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: pobj.

himself ⇝ *him*, *him self*, *themself*, *he*, *herself*

Third Singular Feminine

3sng_fem_subj (89, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: nsubj, nsubjpass, conj, appos, nmod, compound, attr.

she ⇝ *her*, *hers*, *they*, *them*, *their*, *he*

3sng_fem_pobj (90, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: pobj.

her ⇝ *she*, *hers*, *their*, *them*, *herself*, *him*

3sng_fem_dobj (91, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: dobj.

her ⇝ ε (0.1), *she* (0.05), *him* (0.05), *for her* (0.3), *to her* (0.3), *on her* (0.05), *in her* (0.05), *at her* (0.05), *of her* (0.05)

3sng_fem_nsubj_obj (92, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: nsubj, nsubjpass, conj, appos, nmod, compound, attr.

her ⇝ ε, *she*, *hers*, *them*, *him*

3sng_fem_dative (93, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: dative.

her ⇝ ε (0.1), *she* (0.05), *him* (0.05), *for her* (0.3), *to her* (0.35), *on her* (0.05), *in her* (0.05), *at her* (0.05)

3sng_fem_poss (94, F, $\sigma = 0.03$, $\mu = 0.03$)

hers ⇝ ε, *she*, *her*, *theirs*, *his*, *they*, *yours*

3sng_fem_poss_det (95, F, $\sigma = 0.03$, $\mu = 0.03$)

her ⇝ ε, *she*, *hers*, *their*, *his*, *a*, *an*, *the*, *its*

3sng_fem_reflexive (96, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: dobj, conj, nsubj, nsubjpass, npadvmod, dative, attr.

herself ⇝ ε (0.2), *to herself* (0.1), *of herself* (0.1), *for herself* (0.1), *by herself* (0.1), *with herself* (0.1), *herselves* (0.01), *to herselves* (0.01), *of herselves* (0.01), *for herselves* (0.01), *by herselves* (0.01), *with herselves* (0.01), *themself* (0.01), *to themself* (0.01), *of themself* (0.01), *for themself* (0.01), *by themself* (0.01), *with themself* (0.01), *she* (0.045), *her* (0.045), *hers* (0.045), *himself* (0.045)

3sng_fem_appos_reflexive (97, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: appos.

herself ⇝ ε, *her*, *themself*, *her self*, *for herself*, *by herself*, *himself*

3sng_fem_pobj_reflexive (98, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: pobj.

herself ⇝ *her*, *her self*, *themself*, *she*, *himeslf*

Third Plural

3plu_subj (99, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: nsubj, nsubjpass, conj, appos, nmod, compound, attr.

they ⇝ *them*, *their*, *he*, *him*, *she*, *her*, *themself*, *themselves*

3plu_pobj (100, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: pobj.

them ⇝ *they*, *their*, *him*, *her*

3plu_dobj (101, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `dobj`.

them ⤳ ε (0.05), *they* (0.05), *her* (0.05), *him* (0.05), *for them* (0.2), *to them* (0.2), *on them* (0.1), *in them* (0.1), *at them* (0.1), *of them* (0.1)

3plu_nsubj_obj (102, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `nsubj, nsubjpass, conj, appos, nmod, compound, attr`.

them ⤳ ε, *they*, *their*, *him*, *her*

3plu_dative (103, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `dative`.

them ⤳ ε (0.05), *they* (0.05), *him* (0.05), *her* (0.05), *for them* (0.3), *to them* (0.35), *on them* (0.05), *in them* (0.05), *at them* (0.05)

3plu_poss (104, F, $\sigma = 0.03$, $\mu = 0.03$)

theirs ⤳ ε, *they*, *them*, *their*, *its*, *his*, *hers*

3plu_poss_det (105, F, $\sigma = 0.03$, $\mu = 0.03$)

their ⤳ ε, *they*, *them*, *theirs*, *a*, *an*, *the*, *its*, *his*, *her*

3plu_reflexive (109, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `dobj, conj, nsubj, nsubjpass, npadvmod, dative, attr`.

themselves ⤳ ε (0.2), *to themselves* (0.1), *of themselves* (0.1), *for themselves* (0.1), *by themselves* (0.1), *with themselves* (0.1), *themself* (0.01), *to themself* (0.01), *of themself* (0.01), *for themself* (0.01), *by themself* (0.01), *with themself* (0.01), *himself* (0.005), *to himself* (0.005), *of himself* (0.005), *for himself* (0.005), *by himself* (0.005), *with himself* (0.005), *herself* (0.005), *to herself* (0.005), *of herself* (0.005), *for herself* (0.005), *by herself* (0.005), *with herself* (0.005), *they* (0.06), *them* (0.06), *their* (0.06)

3plu_appos_reflexive (110, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `appos`.

themselves ⤳ ε, *they*, *them*, *themself*, *them selves*, *for themselves*, *by themselves*, *herself*, *himself*

3plu_pobj_reflexive (111, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `pobj`.

themselves ⤳ *them*, *them selves*, *herself*, *himself*

Third Neuter

3sng_neut_reflexive (106, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `dobj, conj, nsubj, nsubjpass, npadvmod, dative, attr`.

themself ⤳ ε (0.2), *to themself* (0.1), *of themself* (0.1), *for themself* (0.1), *by themself* (0.1), *with themself* (0.1), *themselves* (0.01), *to themselves* (0.01), *of themselves* (0.01), *for themselves* (0.01), *by themselves* (0.01), *with themselves* (0.01), *himself* (0.005), *to himself* (0.005), *of himself* (0.005), *for himself* (0.005), *by himself* (0.005), *with himself* (0.005), *herself* (0.005), *to herself* (0.005), *of herself* (0.005), *for herself* (0.005), *by herself* (0.005), *with herself* (0.005), *she* (0.03), *her* (0.03), *hers* (0.03), *he* (0.03), *him* (0.03), *his* (0.03)

3sng_neut_appos_reflexive (107, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `appos`.

themself ⤳ ε, *they*, *them*, *themselves*, *them self*, *for themself*, *by themself*, *herself*, *himself*

3sng_neut_probj_reflexive (108, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `pobj`.

themself ⤳ *them*, *them self*, *herself*, *himself*

Third Inanimate

3sng_inan_subj (112, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `nsubj, nsubjpass, conj, appos, nmod, compound, attr`.

it ⤳ ε, *its*, *they*, *he*, *she*, *this*, *that*, *which*

3sng_inan_pobj (113, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `pobj`.

it ⤳ *its*, *them*, *him*, *her*, *this*, *that*, *which*

3sng_inan_dobj (114, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `dobj`.

it ⤳ ε, *its*, *them*, *him*, *her*, *this*, *that*, *which*, *for it*, *to it*, *on it*, *in it*, *at it*, *of it*

3sng_inan_dative (115, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `dative`.

it ⤳ ε, *its*, *her*, *him*, *them*, *for it*, *to it*, *on it*, *in it*, *at it*

3sng_inan_poss (116, F, $\sigma = 0.03$, $\mu = 0.03$)

its ⤳ ε, *it*, *theirs*, *their*, *his*, *hers*, *whose*, *a*, *an*, *the*

3sng_inan_reflexive (117, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `dobj`, `conj`, `nsubj`, `nsubjpass`, `npadvmod`, `dative`, `attr`.

itself ⤳ ε (0.2), *to itself* (0.1), *of itself* (0.1), *for itself* (0.1), *by itself* (0.1), *with itself* (0.1), *themself* (0.005), *to themself* (0.005), *of themself* (0.005), *for themself* (0.005), *by themself* (0.005), *with themself* (0.005), *themselves* (0.005), *to themselves* (0.005), *of themselves* (0.005), *for themselves* (0.005), *by themselves* (0.005), *with themselves* (0.005), *itselves* (0.005), *to itselves* (0.005), *of itselves* (0.005), *for itselves* (0.005), *by itselves* (0.005), *with itselves* (0.005), *himself* (0.005), *to himself* (0.005), *of himself* (0.005), *for himself* (0.005), *by himself* (0.005), *with himself* (0.005), *herself* (0.005), *to herself* (0.005), *of herself* (0.005), *for herself* (0.005), *by herself* (0.005), *with herself* (0.005), *she* (0.025), *her* (0.025), *hers* (0.025), *he* (0.025), *him* (0.025), *his* (0.025)

3sng_inan_appos_reflexive (118, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `appos`.

itself ⤳ ε, *it*, *they*, *them*, *themselves*, *it self*, *for itself*, *by itself*, *themself*, *herself*, *himself*

3sng_inan_pobj_reflexive (119, F, $\sigma = 0.03$, $\mu = 0.03$)

Dependency tag: `pobj`.

itself ⤳ *it*, *it self*, *herself*, *himself*

Determiner

These modules apply to word such that Universal part-of-speech tag is `DET`.

art (135, F, $\sigma = 0.05$, $\mu = 0.05$)

a, an, the ⤳ ε (0.2), *a* (0.2), *an* (0.2), *the* (0.3), *this* (0.025), *that* (0.025), *these* (0.025), *those* (0.025)

demonstrative (136, F, $\sigma = 0.05$, $\mu = 0.05$)

Dependency tag: `det`.

this, that, these, those ⤳ ε, *this, that, these, those, a, an, the*

demonstrative_extra (137, F, $\sigma = 0.05$, $\mu = 0.05$)

Dependency tag: `nsubj`, `nsubjpass`, `pobj`, `dobj`, `conj`, `appos`, `attr`, `advmod`, `ROOT`, `quantmod`, `npadmod`

Penn Treebank part-of-speech tag: `WDT`.

this, that, these, those ⤳ ε, *this, that, these, those*

det_no (138, F, $\sigma = 0.03$, $\mu = 0.03$)

no ⤳ *not, non, any*

det_any (139, F, $\sigma = 0.03$, $\mu = 0.03$)

any ⤳ ε (0.4), *some* (0.1), *every* (0.1), *a* (0.1), *an* (0.1), *all* (0.1), *anything* (0.1)

det_some (140, F, $\sigma = 0.03$, $\mu = 0.03$)

some ⤳ ε (0.4), *a* (0.05), *an* (0.05), *the* (0.05), *those* (0.05), *these* (0.05), *few* (0.05), *little* (0.05), *something* (0.05), *somewhere* (0.05), *much* (0.05), *many* (0.05), *so* (0.05)

det_all (141, F, $\sigma = 0.03$, $\mu = 0.03$)

all ⤳ *both, each, every*

det_both (142, F, $\sigma = 0.03$, $\mu = 0.03$)

both ⤳ ε, *all, each, every*

det_each (143, F, $\sigma = 0.03$, $\mu = 0.03$)

each ⤳ ε, *all, both, every*

det_every (144, F, $\sigma = 0.03$, $\mu = 0.03$)

every ⤳ ε, *all, both, each*

ins_det (151, F, $\sigma = 0.05$, $\mu = 0.05$)

This module inserts *a* (0.3), *an* (0.3), *the* (0.3), *this* (0.025), *that* (0.025), *these* (0.025), *those* (0.025) at the position which satisfies the conditions 1. and 2. below.

1. Penn Treebank's part-of-speech tag of the left word is `VB`, `VBD`, `VBG`, `VBN`, `VBP`, `VBZ`, `IN`
2. Penn Treebank's part-of-speech tag of the right word is `NN`, `NNS`, `JJ`, `JJN`, `JJS`

Adverbial Determiner

so (145, F, $\sigma = 0.03$, $\mu = 0.03$)

so ⤳ ε, *such, too*

such (146, F, $\sigma = 0.03$, $\mu = 0.03$)

such ⤳ ε, *so, very*

another (147, F, $\sigma = 0.03$, $\mu = 0.03$)

another ⤳ *other, the other, an other*

other (148, F, $\sigma = 0.03$, $\mu = 0.03$)

other ⤳ *another, others*

there (149, F, $\sigma = 0.03$, $\mu = 0.03$)

there $\rightsquigarrow$ ε (0.5), *here* (0.3), *they* (0.1), *it* (0.1)

here (150, F, $\sigma = 0.03$, $\mu = 0.03$)

here $\rightsquigarrow$ ε (0.5), *there* (0.3), *they* (0.1), *it* (0.1)

Interrogative Word

int_how (120, F, $\sigma = 0.05$, $\mu = 0.05$)

how $\rightsquigarrow$ *what* (0.6), *that* (0.2), *who* (0.1), *which* (0.1)

int_what (121, F, $\sigma = 0.05$, $\mu = 0.05$)

what $\rightsquigarrow$ *how* (0.6), *that* (0.2), *which* (0.1), *who* (0.1)

int_whatever (122, F, $\sigma = 0.02$, $\mu = 0.02$)

whatever $\rightsquigarrow$ *what* (0.4), *however* (0.2), *whichever* (0.2), *whoever* (0.2)

int_who (123, F, $\sigma = 0.03$, $\mu = 0.03$)

who $\rightsquigarrow$ *that* (0.3), *which* (0.3), *what* (0.2), *how* (0.2)

int_whoever (124, F, $\sigma = 0.02$, $\mu = 0.02$)

whoever $\rightsquigarrow$ *whatever*, *whichever*, *however*, *who*

int_which (125, F, $\sigma = 0.05$, $\mu = 0.05$)

which $\rightsquigarrow$ *that* (0.3), *who* (0.3), *what* (0.2), *how* (0.2)

int_whichever (126, F, $\sigma = 0.02$, $\mu = 0.02$)

whichever $\rightsquigarrow$ *whatever*, *whoever*, *however*, *which*

int_that (127, F, $\sigma = 0.05$, $\mu = 0.05$)

Penn Treebank part-of-speech tag: WDT.

that $\rightsquigarrow$ *which*, *who*, *how*, *what*

int_whose (128, F, $\sigma = 0.03$, $\mu = 0.03$)

whose $\rightsquigarrow$ *which*, *who*, *that*, *its*, *his*, *her*, *their*

int_when (129, F, $\sigma = 0.03$, $\mu = 0.03$)

when $\rightsquigarrow$ *where*, *until*, *that*, *what*, *for*, *in*

int_whenever (130, F, $\sigma = 0.02$, $\mu = 0.02$)

whenever $\rightsquigarrow$ *when*, *whatever*, *wherever*, *whichever*, *until*, *that*, *what*, *for*, *in*

int_where (131, F, $\sigma = 0.03$, $\mu = 0.03$)

where $\rightsquigarrow$ *when*, *wherein*, *whereas*, *whereby*

int_wherever (132, F, $\sigma = 0.02$, $\mu = 0.02$)

wherever $\rightsquigarrow$ *where*, *whatever*, *whenever*, *whichever*, *wherein*, *whereas*, *whereby*

int_whither (133, F, $\sigma = 0.03$, $\mu = 0.03$)

whither $\rightsquigarrow$ *when*, *whence*, *that*, *what*

int_whence (134, F, $\sigma = 0.03$, $\mu = 0.03$)

whence $\rightsquigarrow$ *when*, *whither*, *where*, *what*

int_why (F)

Not used in error generation.

why $\rightsquigarrow$ *how*, *when*, *where*, *that*, *what*

int_whether (F)

Not used in error generation.

whether $\rightsquigarrow$ *which*, *what*, *that*, *how*, *whatever*, *whatsoever*, *if*

Punctuation

comma (153, W, $\sigma = 0.1$, $\mu = 0.1$)

, $\rightsquigarrow$ ε (0.9), . (0.05), . . (0.025), ; (0.025)

ins_left_comma (154, W, $\sigma = 0.1$, $\mu = 0.1$)

This module inserts , (0.9), , , (0.025), . (0.025), ; (0.025), : (0.025) at the position satisfying the conditions 1. or 2. below.

  1. 1a. and 1b.
  1a. 1a1. or 1a2.
  1a1. The Penn Treebank's part-of-speech tag of the left word is NN, NNS, NNP, NNPS
  1a2. 1a2a. and 1a2b.
  1a2a. The Penn Treebank's part-of-speech tag of the left word is RB, RBR, RBS
  1a2b. The Universal part-of-sppech tag of the left word is ADV
  1b. The Penn Treebank's part-of-speech tag of the right word is CC, DT, IN, WDT, WP, WP$, WRB
  2. 2a. and 2b.
  2a. The Penn Treebank's part-of-speech tag of the left word is NN, NNS, NNP, NNPS
  2b. 2b1. and 2b2.
  2b1. The Penn Treebank's part-of-speech tag of the right word is RB, RBR, RBS
  2b2. The Universal part-of-speech tag of the right word is ADV

ins_right_comma (155, W, $\sigma = 0.1$, $\mu = 0.1$)

This module inserts , (0.9), , , (0.025), . (0.025), ; (0.025), : (0.025) at the position satisfies the conditions 1. and 2. below.

  1. 1a. or 1b.

1a. The Penn Treebank's part-of-speech tag of the left word is `NN`, `NNS`, `NNP`, `NNPS`
1b. 1b1. and 1b2.
1b1. The Penn Treebank's part-of-speech tag of the left word is `RB`, `RBR`, `RBS`
1b2. The Universal part-of-speech tag of the left word is `ADV`
2. The Penn Treebank's part-of-speech tag of the right word is `VB`, `VBD`, `VBG`, `VBN`, `VBP`, `VBZ`

period (156, `W`, $\sigma = 0.1$, $\mu = 0.1$)

. $\rightsquigarrow \varepsilon$ (0.5), , (0.3), . . (0.05), . . (0.05), : (0.025), ; (0.025), ! (0.025), ? (0.025)

hyphen (157, `W`, $\sigma = 0.05$, $\mu = 0.05$)

– $\rightsquigarrow \varepsilon$ (0.85), -- (0.1), --- (0.025), – – (0.025)

two_hyphen (158, `W`, $\sigma = 0.05$, $\mu = 0.05$)

-- $\rightsquigarrow \varepsilon$ (0.2), – (0.7), --- (0.05), – – (0.05)

ins_hyphen (159, `W`, $\sigma = 0.05$, $\mu = 0.05$)

This module inserts – (0.7), -- (0.25), --- (0.025), – – (0.025), at the position satisfies the conditions 1., 2., 3., or 4. below.

1. 1a. and 1b.
1a. The Penn Treebank's part-of-speech tag of the left word is `JJ`, `JJR`, `JJS`, `NN`, `NNS`, `RB`, `CD`
1b. The Penn Treebank's part-of-speech tag of the right word is `JJ`, `JJR`, `JJS`, `NN`, `NNS`, `VBN`, `VBG`
2. 2a. and 2b.
2a. The Penn Treebank's part-of-speech tag of the left word is `JJ`, `JJR`, `JJS`, `CD`, `NN`, `NNS`
2b. The Penn Treebank's part-of-speech tag of the right word is `RB`.
3. 3a. and 3b.
3a. The Penn Treebank's part-of-speech tag of the left word is `CD`.
3b. The Penn Treebank's part-of-speech tag of the right word is `CD`.
4. 4a. and 4b.
4a. The Penn Treebank's part-of-speech tag of the left word is `VB`, `VBG`, `VBN`, `NN`, `NNS`
4b. The Penn Treebank's part-of-speech tag of the right word is `RP`.

quot (160, `W`, $\sigma = 0.05$, $\mu = 0.05$)

"token with Penn Treebank's part-of-speech tag `` or ''" $\rightsquigarrow \varepsilon$, ', ", '', ``

colon (161, `W`, $\sigma = 0.05$, $\mu = 0.05$)

: $\rightsquigarrow \varepsilon$ (0.4), . (0.2), , (0.2), ; (0.2)

semicolon (162, `W`, $\sigma = 0.05$, $\mu = 0.05$)

; $\rightsquigarrow \varepsilon$ (0.4), . (0.2), , (0.2), : (0.2)

hatena (163, `W`, $\sigma = 0.05$, $\mu = 0.05$)

? $\rightsquigarrow \varepsilon$ (0.1), . (0.75), , (0.1), ? (0.05)

bang (164, `W`, $\sigma = 0.05$, $\mu = 0.05$)

! $\rightsquigarrow \varepsilon$ (0.4), . (0.2), , (0.2), ? (0.2)

Conjunction

conj_and (168, `F`, $\sigma = 0.05$, $\mu = 0.05$)

and $\rightsquigarrow \varepsilon$ (0.95), *but* (0.05)

conj_but (169, `F`, $\sigma = 0.05$, $\mu = 0.05$)

but $\rightsquigarrow \varepsilon$ (0.9), *and* (0.1)

Adjective/Adverb/Noun/Verb

to (152, `I`, $\sigma = 0.05$, $\mu = 0.05$)

Penn Treebank part-of-speech tag: `TO`

to $\rightsquigarrow \varepsilon$ (0.6), *by* (0.2), *for* (0.2)

auxpass (170, `I`, $\sigma = 0.03$, $\mu = 0.03$)

"word with dependency tag `auxpass`" $\rightsquigarrow \varepsilon$

confusion (172, `L`, $\sigma = 0.005$, $\mu = 0.005$)

$x \rightsquigarrow$ "word sampled from confusion set for x"

The confusion set is generated by replacing/deleting/adding suffix.

confusion set = $\{w' | w \xrightarrow[\text{suffix}]{\text{replace}} w' \wedge w' \in$ dictionary$\}$. suffix set = { *ability, able, ably, acy, ade, age, al, an, ance, ancy, ant, ar, arch, archy, ard, arian, ary, aster, ate, ation, ative, bility, ble, bly, ce, cide, cle, cracy, craft, crat, cule, cum, cy, d, dom, drome, e, ed, ee, eer, en, ence, ency, ent, eous, er, erel, ern, ery, es, esce, escent, ese, esque, ess, est, eth, ette, ey, fic, fication, fold, form, free, ful, fy, gamy, gate, gen, gon, gram, graph, graphy, handed, hood, i, ial, ian, ibility, ible, ibly, ic, ical, ically, ice, ician, ics, id, ie, ied, ier, ies, iform, ify, ile, ily, in, ine, ing, ion, ior, iour, isation, ise, ish, ism, ist, ite, itis, itive, itude, ity, ium, ive, ization, ize, le, let, like, ling, log, logical, logist, logue, logy, looking, ly, lysis, man, mancy, mania, manship, men, ment, meter, metry, most, n, nce, ness,*

9

nik, nomy, ock, oid, ology, onomy, or, ory, ose, osis, osity, our, ous, pathy, person, phil, phile, philia, philiac, phobe, phobia, phone, phony, ple, proof, red, rel, ry, s, scape, scope, se, self, selves, ship, some, speak, sphere, ster, t, th, tion, tious, tour, tude, ty, ular, ule, ulous, ure, ward, wards, way, ways, wide, wise, worthy, y, yer}.

synonym (171, `L`, $\sigma = 0.05$, $\mu = 0.05$)

Penn Treebank part-of-speech tag: `JJ`, `JJR`, `JJS`, `NN`, `NNS`, `RB`, `RBR`, `RBS`, `VB`, `VBD`, `VBG`, `VBN`, `VBP`, `VBZ`

$w \rightsquigarrow$ "its synonym sampled from WordNet"

adj_infl (173, `I`, $\sigma = 0.05$, $\mu = 0.05$)

Penn Treebank part-of-speech tag: `JJ`, `JJR`, `JJS`

$w \rightsquigarrow$ "perturb w to other form (e.g. `JJ` → `JJR`) by `lemminflect`"

noun_infl (174, `I`, $\sigma = 0.05$, $\mu = 0.05$)

Penn Treebank part-of-speecht tag: `NN`, `NNS`

$w \rightsquigarrow$ "perturb w to other form (e.g. `NN` → `NNS`) by `lemminflect`"

verb_infl (175, `I`, $\sigma = 0.1$, $\mu = 0.1$)

Penn Treebank part-of-speecht tag: `VB`, `VBD`, `VBG`, `VBN`, `VBP`, `VBZ`

$w \rightsquigarrow$ "perturb w to other form (e.g. `VBG` → `VBN`) by `lemminflect`"

If the Penn Treebank's part-of-speech of the word is `VBG` or `VBN`, extra auxiliary word (have, has, ... or be, ...) is added for 20 % of examples.

Particle

This module applies to words such that its Penn Treebank's part-of-speech tag is `RP`, and dependency tag is `prt`.

part (176, `F`, $\sigma = 0.05$, $\mu = 0.05$)

$x \rightsquigarrow \varepsilon$ (0.65), *out, up, down, about, on, in, off* (0.05)

Contraction

contr (177, `W`, $\sigma = 0.05$, $\mu = 0.05$)

This module removes ' in contraction word (e.g., `'m`).

Deletion

del (178, `X`, $\sigma = 0.01$, $\mu = 0.01$)

This module deletes words in deletion set.

deletion set = { `!`, `"`, `#`, `$`, `%`, `&`, `'`, *'re*, *'s*, `(`, `)`, `*`, `+`, `,`, `-`, `.`, `/`, `:`, `;`, `<`, `=`, `>`, `?`, `@`, `[`, `\`, `]`, `^`, `_`, `` ` ``, *a, about, after, against, am, an, and, are, around, as, at, away, back, back, be, because, been, before, being, but, by, can, could, dare, did, do, does, down, during, for, from, had, has, have, how, if, in, into, is, it, just, like, may, might, of, off, on, or, ought, out, over, shall, should, since, than, that, the, them, there, these, they, this, those, through, to, under, up, was, were, what, when, where, whether, which, while, who, whom, whose, why, will, with, would*, `{`, `}`, `~` }

Writing System

ent (165, `W`, $\sigma = 0.03$, $\mu = 0.03$)

This module applies to named entities detecded by SpaCy's IOB tags. This module conevrts the first letter of the words to lowercase (e.g., Long Island → long island). 80% of the perturbating named entities are fully uncased. 20% of them are partially uncased.

title (166, `W`, $\sigma = 0.01$, $\mu = 0.01$)

This module converts the first letter of the words to uppercase. This module applies to the words such that lowercase, non-digit, and composesd of latin.

uncase (167, `W`, $\sigma = 0.03$, $\mu = 0.03$)

This module converts the first letter of the words to lowercase. This module applies to the words such that non-lowercase, non-digit, and composesd of latin.

del_orth (180, `W`, $\sigma = 0.03$, $\mu = 0.03$)

This module removes white space between 2 words.

ins_orth (181, `W`, $\sigma = 0.03$, $\mu = 0.03$)

This module inserts white space in word and splits the word. This module splits words based on dictionaries, so "football" is more likely to be split into "foot ball" than "foo tball." This rule splits 80% of words so that both separated words are in dictionary. The rest 20% words are separated randomly.

Spell

I classified spelling error into Writing System Error, although it is not Writing System Error. I wonder if there may be better classification.

spell (179, `W`, $\sigma = 0.05$, $\mu = 0.05$)

This module perturbs word spell in the following manner.

1. Sample the number of error generating operation from geometric distribution ($p = 0.9$).
2. Choose operation from delete, swap, insert, or replace with equal probability.
   Delete: delete 1 character.
   Swap: swap consecutive 2 characters.
   Insert: insert character sampled from a learned distribution.
   Replace: replace character to another character sampled from learned distribution.

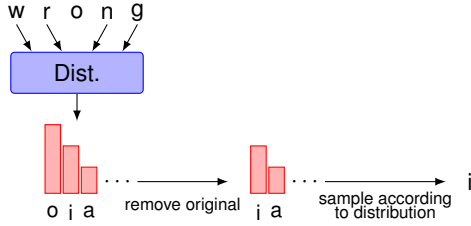The learned distribution is made by a feed-forward neural network.

Figure 1: Spelling Error Generation(e.g., replace)

Word Order

The distance to move word and phrase is sampled from a normal distribution. Therefore, the position after the movement is `float`, not `int`. Finnaly, the word order is determined by `argsort` on the positions represented by `float`.

an_wo (182, `O`, $\sigma = 0.03$, $\mu = 0.03$)

This rule shuffles the order of the consecutive adjectives (word whose Penn Treebank's part-of-speech tag is `JJ`, `JJR`, `JJS`, or `CD`), for example '*big fat*'. This rule can shuffle the order of the concatenation of consecutive adjectives and the following consecutive nouns (word whose Penn Treebank's part-of-speech tag is `NN`, or `NNS`), for example '*big fat cat*'. Whether the rule shuffle the order including nouns is determined by parameter `ratio` (default `ratio` is 0.1).

of_wo (183, `O`, $\sigma = 0.02$, $\mu = 0.02$)

A of B $\rightsquigarrow$ *B of A*. *A*, *B* are both (`ADJ`|`DET`|`NUM`)* `NOUN`+.

pp_wo (184, `O`, $\sigma = 0.02$, $\mu = 0.02$)

This module moves prepositional phrases. We extract phrases which satisfies this rule: (`IN`&`ADP`&`prep`&$\sim$*of*) ((`ADV`|`ADJ`|`DET`|`NUM` |`NOUN`|`PRON`|`PROPN`)&$\sim$`WDT`)+

adv_wo (185, `O`, $\sigma = 0.03$, $\mu = 0.03$)

This module moves adverb (`ADV`). The distance to move words is sampled from a normal distribution ($\mu = 1.0, \sigma = 1.5$).

wh_wo (186, `O`, $\sigma = 0.02$, $\mu = 0.02$)

This module moves *how*, *what*, *who*, *which*, *whose*, *when*, *where*, *whither*, *whence*, *why*, *whether*. The distance to move the words is sampled from a normal distribution ($\mu = 1.5, \sigma = 1.0$).

norm_wo (187, `O`)

This module applies all the words in sentence. This module moves all words in sentence. The distances to move words are sampled from a normal distribution ($\mu = 0.0, \sigma = 0.5 \times x$; x is sampled from a beta distribution for each sentence).

Mask Token Prediction

mask (188, `X`, $\sigma = 0.15$, $\mu = 0.15$)

This module replaces tokens with black mask token (`0x25A8`) and characters with white mask token (`0x25A1`). 80% of the perturbating tokens are entirely masked by black mask token (e.g. *word* $\rightsquigarrow$ `<mask>`). 20% of them are partially masked by white mask token (e.g. *word* $\rightsquigarrow$ *w*`<mask>`*d*). The number of perturbations for a word is sampled from a geometric distribution ($p = 0.8$).