Grammar Error Detection and Classification

Task Description

Detect the grammatical errors given the input sentences and classify them into an error type.

Task 1: Error Detection
- Given tokens of a sentence, determine if a grammatical error exists due to that token.

Task 2: Error Classification
- Once the error is detected in a sentence, you must classify this error into one of the 28 classes of error types.
  Please find the description of error types in table 1 of the paper here

Dataset description

You are provided the following files
Data/
- train.txt
- dev.txt
- dev_results.txt
- dev_correction_results.txt

Each line in the data file contains a word, optionally it's correction and the error category.
- Each line in `train.txt` is of the form - Word   [Correction]   [ErrorCategory]
- Each line in `dev.txt` is of the form - Word
- Whereas, each line in `dev_results.txt` is of the form - Word   [Error Category]

Following instances occur in the files
- If the correction replaces the original text at the given location, it should fix the grammatical error.
  - Example
    - there
    - are   is   SVA
    - very
    - limited
    - spaces   space   Nn
    - for
    - us
- But If the correction is `-`, deletion of Word should fix the grammatical error.
  - Example
    - by
    - achieving
    - a
    - better
    - usage
    - of
    - the   -   ArtOrDet
    - land
- If the Word is `-` inserting the Correction in the original text would fix the error.
  - Example
    - this
    - -   has   Vt
    - caused
    - problem   problems   Nn
- In some sentences, corrections happen on a span of tokens. In this case, corrections and its types repeat until the phrase with the error ends.
  - In this example, corrections for on, its, way, to are same. This means that the phrase "on its way to" must be replaced with the correction "its process of" to fix the grammatical error.
    - china
    - has
    - been
    - eagerly
    - speeding
    - up
    - on   its process of   Wci
    - its   its process of   Wci
    - way   its process of   Wci
    - to   its process of   Wci
    - industrialization

Below is a complete list of the error categories in the data files

ERROR TAG   ERROR CATEGORY

Vt   Verb tense
Vm   Verb modal
V0   Missing verb
Vform   Verb form
SVA   Subject-verb-agreement
ArtOrDet   Article or Determiner
Nn   Noun number
Npos   Noun possesive
Pform   Pronoun form
Pref   Pronoun reference
Prep   Preposition
Wci   Wrong collocation/idiom
Wa   Acronyms
Wform   Word form
Wtone   Tone
Srun   Runons, comma splice
Smod   Dangling modifier
Spar   Parallelism
Sfrag   Fragment
Ssub   Subordinate clause
WOinc   Incorrect sentence form
WOadv   Adverb/adjective position
Trans   Link word/phrases
Mec   Punctuation, capitalization, spelling, typos
Rloc-   Local redundancy
Cit   Citation
Others   Other errors
Um   Unclear meaning (cannot be corrected)

What should your program accept ?

Your program must take the testing file which is in `dev.txt` format as input and output the results into a new file in the format of `dev_results.txt`
Correct word / correction is not mandatory.

Bonus

Give the correction along with the error type.
If your programs also outputs a correction for the error, your output file must be of the format `dev_correction_results.txt`
Each line in `dev_correction_results.txt` is of form: Word   [Correction]   [Error type]