Tokens (Lexical Analysis)

The **tokens' regular expressions** are as follows. The capitalized names are the actual token types to be returned to the parser.

Words:	returned token type

(vowel | vowel n | consonant vowel | consonant vowel n |
consonant-pair vowel | consonant-pair vowel n)^+

if it ends in a vowel lower case	**WORD1**
if it ends in I | E	**WORD2**

Where consonants are:
  b **d** g h **j** k m **n** p r s t **w y z** (note that d,j,w,y,z do not start consonant pairs)
consonant pairs are:
  by
  gy
  hy
  ky
  my
  ny
  py
  ry
  ch (note that c is not listed as a consonant)
  sh
  ts
and vowels are:
  a, i, u, e, o (but I and E uppercase are allowed)

Punctuation:	returned token type
.	**PERIOD** (for end of a sentence)

Sample words are:		showing the parts:
aoi	WORD1 i ending	(a-o-i)
okashii	WORD1 i ending	(o-ka-shi-i)
agE	WORD2 E ending	(a-gE)
tatakI	WORD2 I ending	(ta-ta-kI)
nyuuryoku	WORD1 u ending	(nyu-u-ryo-ku)
panda	WORD1 a ending	(pan-da)
hon	WORD1 n ending	(hon)

Reserved words:	returned token type

  (English translation are found below the words for your convenience)
  These are a subset of Words.

Verb Markers:

masu	**VERB**
masen	**VERBNEG**
mashita	**VERBPAST**
masendeshita	**VERBPASTNEG**
desu	**IS**
deshita	**WAS**

Particles:

o	**OBJECT**
wa	**SUBJECT**
ni	**DESTINATION**

watashi, anata, kare, kanojo, sore (I/me) (you) (he/him) (she/her) (it)	**PRONOUN**
mata (Also)	**CONNECTOR**
soshite (Then)	**CONNECTOR**
shikashi (However)	**CONNECTOR**
dakara (Therefore)	**CONNECTOR**

eofm	**EOFM** (end of file marker)

Assumptions:

You may assume that each token is preceded and followed by blanks.
e.g.   watashi wa gakkou ni ikI mashita .

You may assume that all words are in **lower case** except for I and E in some cases.

You may also assume that there is an **eofm marker** at the end of the input file.