# Uses TokensRegex library to split further split tokens with -

# Map variable names to annotation keys

tokens = { type: "CLASS", value: "edu.stanford.nlp.ling.CoreAnnotations$TokensAnnotation" }

# Indicates matched expressions should replace the tokens annotation

options.matchedExpressionsAnnotationKey = tokens;

# Indicates that the matched expressions and tokens should be combined

options.extractWithTokens = TRUE;

# Indicates that the matched expressions should be flattened to be just tokens

options.flatten = TRUE;

# Mark units

# Define ruleType to be over tokens

ENV.defaults["ruleType"] = "tokens"

# Case insensitive pattern matching (see java.util.regex.Pattern flags)

ENV.defaultStringPatternFlags = 2

# Indicates that after matching, and computing the result, the result should be placed in the tokens annotation

ENV.defaultResultAnnotationKey = tokens

# Define rule where upon matching tokens with -, the matched token is split

{ pattern: ( /.+-.+/ ), result: Split($0[0], /-/, TRUE) }