

≡ 08

## Separando apenas palavras.

Vimos que é necessário criar uma lista apenas com as palavras do nosso **corpus** para “treinar” o corretor ortográfico. Como no processo de tokenização geramos uma lista de sequência de caracteres e não apenas palavras, é preciso gerá-la.

```
def separa_palavras(lista_tokens):  
    lista_palavras = []  
    for token in lista_tokens:  
        if token.isalpha():  
            lista_palavras.append(token)  
    return lista_palavras
```

Considerando o trecho de código acima, assinale a alternativa com a devida entrada `lista_tokens` da função `separa_palavras()` para gerar a seguinte lista de saída:

`['Olá', 'tudo', 'bem']`

*Selezione uma alternativa*

**A** Para gerar `['Olá', 'tudo', 'bem']` a entrada deve ser `['Olá', ',', 'tudo12', 'bem', '?']`

**B** Para gerar `['Olá', 'tudo', 'bem']` a entrada deve ser `['Olá', ',', 'tudo', 'bem', '?']`

**C** Para gerar `['Olá', 'tudo', 'bem']` a entrada deve ser `['Olá,', 'tudo', 'bem?']`