

1. W typowych modelach tłumaczenia jeśli słowa w_1 i w_2 wystąpią w tekście jednokrotnie i przy tym w tym samym zdaniu to nie potrafimy ich odróżnić, tzn. to co można wywnioskować dla w_1 da się też wywnioskować dla w_2 . Oszacuj ile takich par pojawi się w tekście zakładając że mamy 100000 zdań po 20 słów, jest 200000 słów występujących jednokrotnie i że słowa są wybierane do zdań niezależnie. Jak to się zmieni jeśli słów występujących jednokrotnie będzie 100000.

2. Używając program `ns1o1` oblicz ilość słów występujących jednokrotnie w polskim tekście sprawozdań Europarlamentu. Jak to się zmieni jeśli ograniczymy się do słów składających się z małych polskich liter. Jak to się zmieni jeśli z każdego słowa usuniemy ostatnie 4 litery (usunięcie ostatnich liter jest grubym przybliżeniem zastępowania słowa przez formę podstawową).