

### 3. házi feladat

## Megoldás

Ferenczi Zsanett

### 1. A feladat

Számold ki az "i want chinese food" mondat valószínűségét! Két számítást végezz:

- egyet a simítatlan, 6. oldalon található táblázat alapján
- egyet az add-one simított táblázat alapján

Ez utóbbi esetben  $P(i|<s>) = 0.19$  és  $P(</s>|food) = 0.40$  legyen.

Ezt követően vizsgáld meg, hogy a simítatlan vagy a simított valószínűség nagyobb-e. Indokold meg, miért.

### 2. A megoldás

#### 2.1. Simítatlan valószínűség

$P(<s> \text{ i want chinese food } </s>) =$

$P(i|<s>) \cdot P(\text{want}|i) \cdot P(\text{chinese}|\text{want}) \cdot P(\text{food}|\text{chinese}) \cdot P(</s>|\text{food}) =$

$0,25 \cdot 0,33 \cdot 0,0065 \cdot 0,52 \cdot 0,68 =$

$18,96 \cdot 10^{-5}$

## 2.2. Add-one simított valószínűség

$$\begin{aligned} P_{Laplace}(<s> \text{ i want chinese food } </s>) &= \\ P(i|<s>) \cdot P(\text{want}|i) \cdot P(\text{chinese}|\text{want}) \cdot P(\text{food}|\text{chinese}) \cdot P(</s>|\text{food}) &= \\ 0,19 \cdot 0,21 \cdot 0,0029 \cdot 0,052 \cdot 0,4 &= \\ 24,06 \cdot 10^{-7} \end{aligned}$$

## 2.3. Megfigyelés, indoklás

A simítatlan valószínűség nagyobb, mint az add-one simított, mivel a simítás hatására lecsökkennek az eredetileg nem-nulla darabszámú bigramok valószínűségei. Mindez azért történik, hogy az egyszer sem előforduló bigramok számát növelni tudjuk eggyel, és a valószínűségek összege is 1 maradjon. Így azon esetekben, amikor nem 0 az együttes előfordulás (jelen esetben minden egyes bigram esetén), csökken a valószínűség, így az egész mondat valószínűsége is kisebb lesz az így kapott bigram-valószínűségeket összeszorozva.