

Задача 2. Разгледайте долните два текста.

Задание 1: Пресметнете *средното разстояние* и *отклонението* между думите *закупуване* и *земите* в Текст 1

и между

обезмаслено и *мляко* в Текст 2

и ги напишете. Следете за появяване на думите (пример в експеримента) в показаната последователност, в рамките на едно изречение. Следвайте разсъжденията на стр. 6-7 от лекцията за колокации. За всеки текст ще се произведе таблица 5 (с 1 ред)

Начало Текст 1 -----

Удължават срока за закупуване на земите към стопанските дворове

Комисията по земеделие и гори към парламента ще разгледа възможността за удължаване на срока за закупуване и ползване на земите, които са принадлежащи площи към сгради в стопанските дворове на ликвидирани селскостопански организации. Писмено предложение за промяна в нормативния акт беше направено още през м. октомври м. г., когато изтече срокът, в който собствениците на сгради, построени върху земеделски земи, получиха възможността за закупуване на терените, без да заплащат дължимото обезщетение за ползване на земите. Големия брой собственици, желаещи закупуване при този преференциален режим, е сред мотивите за удължаване на срока за придобиване собственост върху земите. Отнемащите много време допълнителни процедури при закупуване, като изработване на парцелирани планове на земите в стопанските дворове и издаването на актове за собственост, са също част от мотивите, наложили промени в нормативната база. Целта на въведения преференциален режим на закупуване е бързото и окончателно приключване на отношенията с държавата и регламентиране правата на собственост на притежателите на земите, принадлежащи към стопанските дворове. ----- *Край Текст 1*

Начало Текст 2 -----

Здравословно хранене с обезмаслено мляко

Поемането на по-малко мазнини с храната става възможно благодарение на разнообразните видове обезмаслено прясно и кисело мляко, които се появиха на българския пазар през последните 4-5 години. Обезмаслено прясно краве мляко се предлага от “Верей” и “Пършевица” с масленост 0,1% или 2% в опаковки от картон или пластмасове бутилки от 1 л. Ако пиете кафето си с обезмаслено прясно мляко всеки ден, ще поемате 100 калории по-малко. На основата на пълноценно българско кисело мляко се произвежда и богат асортимент от обезмаслено кисело мляко. “Култура” на фирмата “Бор-Чвор” е обезмаслено краве мляко с ниска масленост, добавка на калций и специална закваска, което е получило златен медал и диплом на Международния пролетен панаир в Пловдив през 2007 г. Това обезмаслено мляко с масленост 0,1% се продава в опаковки по 400 г. “Верей Виталия” е обезмаслено краве мляко с добавка на парченца плодове и житен микс, създадено на основата на първокачествено свежо краве кисело мляко. “Виталия” замества сладкишите на дамите, които се стремят към прием на обезмаслено мляко. Поради традицията да се консумира краве масло, в България не е развито примитивното производство на обезмаслено овче и биволско мляко. ----- *Край Текст 2*

Указания към задание 1. Не разглеждайте “колокационните прозорци”, споменати в лекциите, а работете директно с намерените биграми в оригиналните изречения. Отклонението s (квадратен корен от дисперсията) се смята по формулата

$$s = \text{SQRT} \left(\sum_{i=1, \dots, n} (d_i - d)^2 / n \right)$$

където:

- n е броят на примерите, при които двете разглеждани думи се срещат като колокации в нашия експеримент,
- d_i е отдалечеността за i -тия пример и
- d е средното аритметично на отдалечеността на разглежданите думи в намерените примери.

Задание 2. Отговорете какво е вашето мнение – дали думите *закупуване* и *земите* в Текст 1 и *обезмаслено* и *мляко* в Текст 2

са устойчиви двуместни словосъчетания, или са често повторение на думи с композиционна семантика в конкретните текстове? Може да не отговорите точно, но се ангажирайте с разсъждения по въпроса и изкажете предположение.

Решение.

За текст 1, имаме 6 примера на срещане на двойката *закупуване* и *земите* в Текст 1. Нанесли сме ги по-долу в различни цветове:

Удължават срока за **закупуване** на **земите** към стопанските дворове

Комисията по земеделие и гори към парламента ще разгледа възможността за удължаване на срока за **закупуване** и ползване на **земите**, които са принадлежащи площи към сгради в стопанските дворове на ликвидирани селскостопански организации. Писмено предложение за промяна в нормативния акт беше направено още през м. октомври м. г., когато изтече срокът, в който собствениците на сгради, построени върху земеделски земи, получиха възможността за **закупуване** на терените, без да заплащат дължимото обезщетение за ползване на **земите**. Големият брой собственици, желаещи **закупуване** при този преференциален режим, е сред мотивите за удължаване на срока за придобиване собственост върху **земите**. Отнемащите много време допълнителни процедури при **закупуване**, като изработване на парцелирани планове на **земите** в стопанските дворове и издаването на актове за собственост, са също част от мотивите, наложили промени в нормативната база. Целта на въведения преференциален режим на **закупуване** е бързото окончателно приключване на отношенията с държавата и регламентиране правата на собственост на притежателите на **земите**, принадлежащи към стопанските дворове.

Разстоянията между думите са както следва /ако не съм сгрешила при броенето/:

Пример №	1	2	3	4	5	6
Разстояние	2	4	12	17	8	17

Средно разстояние $d = (2+4+12+17+8+17)/6 = 60/6 = 10$

Отклонението

$$s = \text{SQRT} \left(\left((2-10)^2 + (4-10)^2 + (12-10)^2 + (17-10)^2 + (8-10)^2 + (17-10)^2 \right) / 6 \right)$$

$$\text{SQRT} \left((64+36+4+49+4+49)/6 \right) = \text{SQRT} (206/6) = \text{SQRT} (34,33)$$

Понеже $5*5=25$, а $6*6=36$, отклонението s е число между 5 и 6, по-близо до 6. И така, според 2-рия параграф на текста на лекцията на стр. 7, голямото отклонение показва, че тези две думи не са устойчиво словосъчетание. Тук сме приели вашата умна забележка, че запетаята като разделител на фразите в изречението следва да се брой за отделна дума, понеже тя е още едно индиректно доказателство, че двете целеви думи могат да бъдат разделени на произволна дистанция.

За текст 2, имаме 10 примера на срещане на двойката *обезмаслено* и *мляко* в Текст 2. Нанесли сме ги по-долу в различни цветове:

Здравословно хранене с **обезмаслено мляко**

Поемането на по-малко мазнини с храната става възможно благодарение на разнообразните видове **обезмаслено** пряно и кисело **мляко**, които се появиха на българския пазар през последните 4-5 години. **Обезмаслено** пряно краве **мляко** се предлага от “Веря” и “Пършевица” с масленост 0,1% или 2% в опаковки от картон или пластмасове бутилки от 1 л. Ако пиете кафето си с **обезмаслено** пряно **мляко** всеки ден, ще поемате 100 калории по-малко. На основата на пълноценно българско кисело мляко се произвежда и богат асортимент от **обезмаслено** кисело **мляко**. “Култура” на фирмата “Бор-Чвор” е **обезмаслено** краве **мляко** с ниска масленост, добавка на калций и специална закваска, което е получило златен медал и диплом на Международния пролетен панаир в Пловдив през 2007 г. Това **обезмаслено мляко** с масленост 0,1% се продава в опаковки по 400 г. “Веря Виталия” е **обезмаслено** краве **мляко** с добавка на парченца плодове и житен микс, създадено на основата на първокачествено свежо краве кисело мляко. “Виталия” замества сладкишите на дамите, които се стремят към прием на **обезмаслено мляко**. Поради традицията да се консумира краве масло, в България не е развито промишленото производство на **обезмаслено** овче и биволско **мляко**.

Съответно имаме разстоянията:

Пример №	1	2	3	4	5	6	7	8	9	10
Разстояние	1	4	3	2	2	2	1	2	1	4

$$d = (1+4+3+2+2+2+1+2+1+4)/10 = 22/10 = 2,2$$

$$s = \text{SQRT} \left((3*(1-2,2)^2 + 4*(2-2,2)^2 + (3-2,2)^2 + 2*(4-2,2)^2) / 10 \right) =$$

$$\text{SQRT} \left((3*(-1,2)^2 + 4*(-0,2)^2 + (0,8)^2 + 2*(1,8)^2) / 10 \right) =$$

$$\text{SQRT} \left((3*1,44 + 4*0,04 + 0,64 + 2*3,24) / 10 \right) = \text{SQRT} (11,6/10) = \text{SQRT} (1,16)$$

Това s е около 1,08. Отклонението е относително малко – думите винаги се срещат близо една до друга, когато се появяват заедно (и това не е чудно, понеже формират едно NP за разлика от примера с текст 1). Но средното е сравнително голямо – 2,2.

Не можем да се ангажираме с предположение дали това е колокация. Трябва да изследваме по-голям корпус, като вземем под внимание и срещанията на думите в “компанията” на други думи.