

## ТРИНАДЕСЕТА МЕЖДУНАРОДНА ОЛИМПИАДА ПО ЛИНГВИСТИКА

**Иван Держански**

*Институт по математика и информатика – БАН*

**Резюме.** Статията е посветена на проведената през 2015 г. Международна олимпиада по лингвистика. Представени са резултатите на българските участници, част от задачите и техни решения.

*Keywords:* linguistics, Olympiad, problem solution.

От 20 до 24 юли 2015 г. в Благоевград беше проведена **Тринадесетата международна олимпиада по лингвистика**. В Олимпиадата взеха участие 166 ученици, съставлящи 43 отбора от 29 страни. За първи път свои състезатели изпратиха Бангладеш, Казахстан и Франция. Бяха представени също Австралия, България, Великобритания, Естония, Индия, Ирландия, Испания, Канада, Китай, Латвия, о-в Ман, Нидерландия, Пакистан, Полша, Румъния, Русия, САЩ, Сингапур, Словения, Тайван, Украйна, Унгария, Чехия, Швеция, Южна Корея и Япония. От Австралия, България, Великобритания, Китай, Индия, Казахстан, Полша, Русия, САЩ, Украйна, Унгария, Швеция, Южна Корея и Япония имаше по два отбора.

Българските участници в олимпиадата бяха избрани въз основа на резултатите от състезанието по лингвистика в рамките на Зимните математически състезания, националния кръг на Олимпиадата по математическа лингвистика и две контролни работи, проведени по време на подготовката на разширения състав на отбора (Слънчев бряг, 29 май – 7 юни). Те бяха:

– I отбор: **Тина Владимирова** (СМГ), **Надежда Димитрова** (128. СОУ – София), **Борислав Георгиев** (164. ГПИЕ — София), **Валентин Димов** (91. НЕГ – София);

– II отбор: **Данаил Пенев** (1. АЕГ – Варна), **Радина Добрева** (МГ – Варна), **Здравко Иванов** (СМГ), **Ралица Дарджонова** (СМГ).

Ръководители на отборите бяха **Веселин Златилов** (София) и **Виолета Витлянова** (Варна), а **Александър Велинов** беше председател на местния организационен комитет.

Както винаги, Олимпиадата включваше индивидуално и отборно състезание. В работата на задачната комисия, съставила темите за двете състезания,

взеха участие **Божидар Божанов** (председател), **Иван Держански**, **Милена Венева**, **Елица Миланова**, **Павел Софрониев** и **Тодор Червенков**. Първите трима работиха и в международното жури на Олимпиадата.

В индивидуалното състезание осемте български ученици спечелиха един *златен медал* (**Данаил Пенев**), три *сребърни* (**Валентин Димов**, **Ралица Дарджонова** и **Тина Владимирова**) и три *бронзови* (**Здравко Иванов**, **Надежда Димитрова** и **Радина Добрева**). Отборът **България II** се оказа на второ място (между двата отбора на САЩ) по среден резултат в индивидуалното състезание.

\*\*\*

Ето три от задачите от индивидуалното състезание.

**Задача №3 (Александър Пиперски).** Soundex е алгоритъм за кодиране на имена. Разработен е в САЩ в периода 1918 – 1922 г. от Робърт Ръсел и Маргарет Кинг Одел, за да се улесни търсенето на подобно звучащи фамилни имена. В средата на XX век Soundex е широко използван в САЩ за анализ на резултатите от преброяванията на населението в периода 1890 – 1920 г.

По-долу е дадена примерна карта с данни от преброяването на население през 1910 г. Можете да видите Soundex-кода за *Wilson*, който е W425:

| LOUISIANA               |               |            |            |
|-------------------------|---------------|------------|------------|
| HEAD OF FAMILY          |               | E.D.       | SHEET      |
| W 425                   | Wilson, Alice | 118        | 17         |
| COLOR                   | AGE           | BIRTHPLACE |            |
| B                       | 42            |            |            |
| COUNTY                  | St. Landry    |            | CITY       |
| OTHER MEMBERS OF FAMILY |               |            |            |
| NAME                    | RELATIONSHIP  | AGE        | BIRTHPLACE |
| Eugene                  | W             | 46         |            |
| Regina                  | D             | 15         |            |
| Walter                  | S             | 13         |            |
| Louisa                  | D             | 12         |            |
| Camila                  | D             | 7          |            |
| Correll                 | S             | 7          |            |
| Ludson                  | S             | 4          |            |

FORM 10-636 (4-20-01)  
1910 CENSUS INDEX - FAMILY

U.S. DEPARTMENT OF COMMERCE  
BUREAU OF THE CENSUS

(Източник: <https://familysearch.org/learn/wiki/en/Soundex>)

Даден е списък от фамилни имена и съответстващите им Soundex-кодове в случаен ред. Някои символи са изпуснати:

|                                                                                                                                                                       |                                                                                                      |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------|
| Allaway, Anderson, Ashcombe,<br>Buckingham, Chapman, Colquhoun, Evans,<br>Fairwright, Kingscott, Lewis, Littlejohns,<br>Stanmore, Stubbs, Tocher, Tonks,<br>Whytehead | S312, T_6_,_ 5_3, C42_, T520, L_42,<br>A536, C155, _623, S356, _252, _152,<br>_330, A251, A400, L2_0 |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------|

(а) Опишете постъпково как се генерира Soundex-код.

(б) Определете съответствията между фамилните имена и Soundex-кодовете и възстановете изпуснатите символи.

(в) Генерирайте Soundex-кодове за следните фамилни имена:

*Ferguson, Fitzgerald, Hamnett, Keefe, Maxwell, Razey, Shaw, Upfield.*

**Решение.** Най-важното в тази задача е да се проумее, че алгоритъмът Soundex не може да знае как се произнасят имената, а просто обработва последователности от букви, но е направен така, че подобно звучащите имена с голяма вероятност да имат еднакъв код.

(а) 1. Първата буква остава както е.

2. *h* и *w* се изтриват.

3. Всички съгласни букви се заменят с цифри (букви, чиито най-често срещани произношения са подобни, се групират):

|         |             |    |   |    |   |
|---------|-------------|----|---|----|---|
| bpv (f) | cgjkqs (xz) | dt | l | mn | r |
| 1       | 2           | 3  | 4 | 5  | 6 |

4. Две или повече еднакви цифри поред се съкращават до една.

5. Всички гласни (*a, e, i, o, u, y*) се изтриват.

6. Оставят се само първите три цифри или се добавят нули в края, за да станат една буква и три цифри в кода.

(б) *Allaway*: A400, *Anderson*: A536, *Ashcombe*: A251, *Buckingham*: B252, *Chapman*: C155, *Colquhoun*: C425, *Evans*: E152, *Fairwright*: F623, *Kingscott*: K523, *Lewis*: L200, *Littlejohns*: L342, *Stanmore*: S356, *Stubbs*: S312, *Tocher*: T260, *Tonks*: T520, *Whytehead*: W330.

(в) *Ferguson*: F622, *Fitzgerald*: F326, *Hamnett*: H530, *Keefe*: K100, *Maxwell*: M240, *Razey*: R200, *Shaw*: S000, *Upfield*: U143.

**Задача №4 (Ксения Гилярова).** Дадени са изречения на езика уамбая и българските им преводи:

1. *Ngarlu nya*. Ти танцува.

2. *Gulugba ngu*. аз ще спя.

3. *Bardbi ga bungmanya*. Старицата избяга.

4. *Ngunybulugi barrawunguji*. Лекарят има къща.

5. *Ngajbi nginya*. аз те видях.

6. *Dagumajba nguuni ngunybulugi*. Тя ще удари лекаря.

7. *Alanga gu ngarlujba janjingunya*. Момичето ще танцува с кучето.
8. *Ngajba gununyu*. Той ще те види.
9. *Ngirra nga gijilulu*. Аз откраднах парите.
10. *Yagujba nyu barrawuwaji*. Ти ще го оставиш без къща.
11. *Daguma ngiyinga bungmanyani*. Старицата ме удари.
12. *Bardbi ga darranggunguji*. Той избяга с пръчката.
13. *Alayulu ngiya juwa*. Тя намери мъжа.
14. *Juwani gina ngajbi gijilungunya*. Мъжът я видя с парите.
15. *Yagu gininya*. Той ме остави.
16. *Bungmaji gina ngajbi balamurruwaji*. Той видя стареца без копие.
17. *Dirragba gu bungmaji*. Старецът ще скача.

(а) Преведете на български:

1. *Alayulujba nguynyu bungmanyani*.
2. *Yagu gininya*.
3. *Janji darranguwaji*.
4. *Ngirra nya alanga*.
5. *Daguma nyinga*.
6. *Dirragbi ga balamurrungunya*.

(б) Преведете на уамбая:

1. Ти ще ме оставиш.
2. Лекарят спа.
3. Мъжът ще избяга с парите
4. Той ще открадне кучето.
5. Момичето те видя

Забележка: Езикът уамбая е от западнобарклийското семейство. Говори се от около 50 души в Северна територия, Австралия.

**Решение.** Правила:

– Словоред: V P (S/O); S/O P V Poss, V P Poss; S Poss.

– V = глагол (формата за бъдеще време може да се получи от формата за минало време, като в края заменим **-bi** с **-ba**, ако го има, а в противен случай – като прибавим **jba**).

– S = подлог (съществително). Подлогът на преходен глагол получава окончанието **ni**.

– O = допълнение (съществително).

– P = местоимения (подлог + допълнение) + време:

– подлог:

– 1-во л. **ngV-**,

– 2-ро л. **nyV-**,

– 3-то л.: непреходен глагол — **gV-**, преходен глагол — мъжки род **gVnV-**, женски род **ngVyV-**;

– допълнение: 1-во л. **-ngV**, 2-ро л. **-nyV**, 3-то л. **-Ø**;

– V са гласни звукове (в мин. вр. всички са **i**, само последната е **a**; в бъд. вр. всички са **u**).

– Poss = притежаемо: показател **-ngu** или **-wa** в зависимост от това, дали става дума за присъствието, или за отсъствието на нещо, следван от **-ji** за мъжки или **-nya** за женски род.

(а) 1. *Alayulujba nguyunyu bungmanyani*. Старицата ще те намери.

2. *Yagu gininya*. Той те остави.

3. *Janji darrangguwaji*. Кучето няма пръчка.

4. *Ngirra nya alanga*. Ти открадна момичето.

5. *Daguma nyinga*. Ти ме удари.

6. *Dirragbi ga balamurrungunya*. Тя скочи с копие.

(б) 7. Ти ще ме оставиш. *Yagujba nyungu*.

8. Лекарят спа. *Gulugbi ga ngunybulugi*.

9. Мъжът ще избяга с парите. *Juwa gu bardba gijilulunguji*.

10. Той ще открадне кучето. *Ngirrajba gunu janji*.

11. Момичето те видя. *Ngajbi ngiyinya alangani*.

**Задача №5 (Александър Пиперски).** Дадени са 25 полустиишия на сомалийски език, написани в размера масафо:

1. ogaadeen ha ii dirin

2. duul haad amxaaraa

3. kaa dooni maayee

4. amba waa ku daba geli

5. dakanka iyo qaantee

6. anaa been dabaadee

7. galbeed uga dareershaan

8. dalkaad adigu joogtiyo

9. dar alliyoo heshiis iyo

10. mase waa dayoobeen

11. dacalkaaga kuma shuban

12. miyaan duudsiyaayaa

13. doodaye maxaad oran

14. daliilkii ku siiyaye

15. miyaad iigu duurxuli

16. dorraad adigu kama dhigin

17. ma deldelin raggoodii

18. deelqaadkan aad tiri

19. diigaanyo ciidana

20. wax ma kala dillaallaa

21. duunyada ka qaadoo

22. diinkiyo dugaaggiyo

23. dildillaaca waaberi

24. dibnahaaga kama qiran

25. hobyoo wixii ka soo degey

За да разберете по-добре устройството на масафо, ето 10 полустиишия, които са получени от автентични полустиишия чрез случайно разместване на думите в рамките на едно полустиишие. Може някои от тях да удовлетворяват правилата на стихосложението, но при повечето не е така:

1. u anigaa lehe diin

2. waad nimankaad ma diidi

3. qoran daftarkaaga kuma

4. fuushaan kaama dusha

5. helo dabacayuun kulaan

6. kuu miyuu tari wax dafir

7. kuu daalasaayee nin

8. shareecada dikrigiyo

9. dumarkii furayaan ma

10. ogaadee diyaar kuu

- (а) Опишете структурата на полустушието в размер масафо.  
(б) Дадени са още 10 полустушия в размер масафо. Пет от тях са автентични, а другите пет са получени чрез случайни размествания. Кое какво е?

1. war ismaaciil daarood
2. dir miyaad wadaagtaan
3. labadaad ka duudiye
4. ka jannadaad daahiye
5. adiga iyo deriskaa

6. digaxaarka mariyoo
7. ciid iyo doolo diraac
8. nooma keeneen darka
9. kala deyaayaa miyaan
10. wuxuun kaa danqaabaan

### Решение

(а) Всеки ред съдържа по 9 или 10 гласни букви, които могат да се срещат поединично или по двойки (ако това са дълги гласни, значи една дълга се брои за две кратки). Не навсякъде обаче може да има двойна гласна. Ако броим гласните букви от края на реда (така че 1 е последната, 2 е предпоследната и т.н.), можем да видим, че двойни гласни има само на позиции 1 – 2, 3 – 4, 6 – 7 и 8 – 9.

Можем да представим размера схематично така:

$$(o) \frac{oo}{\infty} \frac{oo}{\infty} \circ \frac{oo}{\infty} \frac{oo}{\infty}, \quad \left| \begin{array}{l} \circ = V (a, e, i, o, u) \\ \infty = VV (aa, ee, ii, oo, uu) \end{array} \right.$$

(б) На изложените правила не отговарят петте реда 36, 39, 42, 43 и 44, в които има двойни гласни на позиции, различни от изброените по-горе. Знаем, че има пет реда с отклонения, следователно останалите редове са автентични.

\*\*\*

Вече върви подготовката за Четиринадесетата международна олимпиада по лингвистика, която ще се проведе в град Майсур (Индия) през последната седмица на юли 2016 г.

## THIRTEENTH INTERNATIONAL LINGUISTICS OLYMPIAD

**Abstract.** The paper is dedicated to the International Linguistics Olympiad in 2015. What are presented are the results of the Bulgarian participants, some of the tasks and their solutions.

✉ **Dr. Ivan Derzhanski, Assoc. Professor**  
Institute of Mathematics and Informatics  
Acad. G. Bonchev Street, block 8  
1113 Sofia  
E-mail: iad58g@gmail.com