

APPLICATION OF THE WUGGY METHOD FOR GENERATION OF PSEUDO-WORDS IN THE BULGARIAN LANGUAGE

Katerina Shtereva

Sofia University "St. Kliment. Ohridski", Bulgaria, katerinashtereva@abv.bg

Boryana Hadzhiyska

Harvard University, United States, hadzhiyska@g.harvard.edu

Tsvetalin Totev

Department of Sensory Neurobiology, Institute of Neurobiology, BAS, Bulgaria, cwetalin@abv.bg

Milena Slavcheva Mihaylova

Department of Sensory Neurobiology, Institute of Neurobiology, BAS, Bulgaria,
milenski_vis@abv.bg

Abstract: Pseudo-words have been largely used in research of reading, short-term working memory as well as in psycholinguistics, neuropsychology, speech therapy. The use of pseudo-words has been widely applied in the diagnosis and treatment of people with developmental disorders such as Autism Spectrum Disorder (ASD), Developmental dyslexia (DD), etc. Pseudo-words resemble real words in a given language, but in fact they carry no meaning and are not part of its lexicon. They are composed of a combination of phonemes that correspond to the phonotactic rules of the language. It is assumed that in the process of reading, pseudo-words are processed largely in a sublexical way, which uses a serial (letter by letter) reading mode, while the process of reading real words is realized via lexical recognition through matching the perceived words with their patterns in memory. Up to the present moment there is not yet fully established and universal method for creating pseudo-words.

The aim of the present work is to adapt the computer program using the so-called "Wuggy" method for generation of pseudo-words for the Bulgarian language. This method requires a rich dictionary in a given language, which must be divided into syllables. Then, each word is divided into its sub-syllable elements: onset, nucleus and coda. In this way, the generated pseudo-words are built entirely on the basis of the existing syllable elements in the specific language, taking into account their position and structure, as well as the allowed consonant clusters. Prior to the present work, this program has been available in Dutch, English, German, French, Spanish, Serbian, Basque and Vietnamese.

The process of adaptation of the program followed several steps. Initially, a dictionary, including the lexical bank of the Bulgarian language, was compiled. The Wuggy program code was adapted for the Bulgarian language and an algorithm was developed for dividing Bulgarian words into syllables in accordance with the basic rules for defining syllables in Bulgarian. In this context, the pseudo-words generated by the program are constructed entirely on the basis of the existing syllabic elements, taking into account their position and structure, as well as the allowed consonant clusters. Subsequently, the Wuggy program, already adapted for the Bulgarian language, was applied for our specific purposes and a selection of linguistic material was made in connection with the peculiarities of the Bulgarian language according to the following characteristics: number of syllables; syllabic structure; parts of speech, etc.

In conclusion, the Wuggy method in its adapted version for the Bulgarian language enables Bulgarian researchers and practitioners to generate pseudo-words for the purposes of research, diagnosis, and therapy. Moreover, the program could be used to select real words that are mutually based on phonological structure, length and frequency, which would contribute to its rapid application in therapeutic and training programs.

Keywords: pseudo-words; sublexical reading way; Wuggy method; children and adolescents with developmental disorders.

The study was funded by grant DN15/6 from 2017 of the National Science Fund of Bulgaria.

The authors are grateful to Emmanuel Keuleers for his help in the process of applying the Wuggy method and program for the Bulgarian language.

ПРИЛАГАНЕ НА МЕТОДА WUGGY ЗА ГЕНЕРИРАНЕ НА ПСЕВДО-ДУМИ В БЪЛГАРСКИЯ ЕЗИК

Катерина Щерева

СУ „Св. Кл. Охридски“, България, katerinashtereva@abv.bg

Боряна Хаджийска

Харвардски университет, САЩ, hadzhiyska@g.harvard.edu

Цветалин Тотев

Направление „Сензорна невробиология“, Институт по невробиология, БАН, България,
cwetalin@abv.bg

Милена Славчева Михайлова

Направление „Сензорна невробиология“, Институт по невробиология, БАН, България,
milenski_vis@abv.bg

Резюме: Използването на псевдо-думи има голямо значение при изследване на четенето, краткосрочната работна памет и области, които са обект на психолингвистиката, невропсихологията, логопедията и др. В диагностиката и терапията на лицата с нарушения в развитието, като разстройство от аутистичния спектър (РАС), дислексия на развитието (ДР) и др., употребата на псевдо-думи е почти ежедневна. Псевдо-думите изглеждат като действителните думи на определен език, докато всъщност нямат значение и не са част от лексикалното му богатство. Те са съставени от комбинация от фонемни, които съответстват на фонотактичните правила на езика. Предполага се, че в процеса на четене псевдо-думите се обработват до голяма степен по сублексикален път, който използва сериен (буква по буква) режим на четене, за разлика от четенето на реални думи, което се реализира чрез лексикалния път на обработка, който действа при холистичното възприемане на думите и включва съвпадение на възприетите думи с техните шаблони в паметта. Въпреки че всички изследователи, които работят в тези области, търсят начини за подбор на качествен изследователски материал, все още няма напълно изграден унифициран метод за създаването на псевдо-думи.

Целта на това изследване е да се представи процесът на адаптиране на компютърната програма и методът Wuggy за генериране на псевдо-думи в българския език и приложението ѝ при изследване на способността за четене на деца и юноши с нарушения в развитието. Този метод изисква да се вземе речникът с всички думи на дадения език, които трябва да са разделени на срички. От своя страна всяка дума се разделя на т.нар. субсричкови елементи: онсет, ядро и кода. По този начин генерираните псевдо-думи се изграждат изцяло на базата на съществуващите сричкови елементи в конкретния език, като се взима предвид тяхната позиция и структура, както и разрешените консонантни клъстери. До този момент тази програма беше достъпна на холандски, английски, немски, френски, испански, сръбски, баски и вьетнамски език.

Реализирането на поставените цели може да се опише в следната последователност. Първоначално се подсигури речник, включващ лексикалната банка на българския език. Осъществи се приспособяването на кода на програмата Wuggy към българския език и се разработи алгоритъм за разделянето на българските думи на срички. Всичко това е съобразено с основните правила за разделяне на българската сричка. В този контекст генерираните от програмата псевдо-думи са конструирани изцяло въз основа на съществуващите силабични елементи, като се отчита тяхното положение и структура, както и разрешените съгласни клъстери. В следствие за целите на нашето изследване се приложи вече адаптираната за българския език програма Wuggy и се направи подбор на лингвистичен материал във връзка с особеностите на българския език по следните характеристики: брой срички; сричкова структура; части на речта и др.

В заключение, методът Wuggy в неговия адаптиран вариант за българския език дава възможност на българските изследователи и практики да генерират псевдо-думи за целите на своята работа по превенция, диагностика и терапия. Не е за подценяване и фактът, че програмата може да се използва и за избор на поредици от реални думи, които са взаимно сходни по фонологична структура, дължина и честота, което би допринесло за бързото му прилагане в терапевтични и тренировъчни програми.

Ключови думи: псевдо-думи; сублексикален път на четене; метод Wuggy; лица с нарушения на развитието.

Това изследване е подкрепено от Фонд научни изследвания на България, договор ДН15/6 от 11.12.2017 г.

1. ВЪВЕДЕНИЕ

Използването на псевдо-думи при изследване на четенето, краткосрочната работна памет и области, които са обект на психолингвистиката, невропсихологията, логопедията и др., е от особено голямо значение през последните години. Псевдо-думите и не-думите изглеждат като действителни думи на определен език, докато всъщност нямат значение и не са част от лексикалното му богатство. Това е един вид нелексикален речник (Rathvon, 2004). От гледна точка на лингвистиката псевдо-думите се характеризират с комбинация от буквени знаци, които съответстват на фонотактичните ограничения на даден език (Trezek, Wang, & Paul, 2010) и не включват звукове или последователности от звукове, които не съществуват на този език. Не-думите за разлика от псевдо-думите са трудни за изговаряне комбинации от звукове и съчетанието им не отговаря на характерното за истинските думи на конкретния език. Вариант на псевдо-думите са така наречените логатоми, които представляват една сричка. Логатомите са важни за изграждането на акустични модели в експерименталната психология във връзка с изследването на процесите учене и памет (Scharenborg, 2007).

Жан Берко е първият изследовател, който използва псевдо-думи в контекста на психолингвистичните експерименти. Затова понякога псевдо-думите се наричат и *wug* [wʌlg], защото тази дума е използвана в първия тест на Берко от 1958 г. Експериментите, използвани от Берко, се прилагат с цел изследване на познавателните възможности на децата по отношение на английската морфология. Изследването установява, че още в много ранна възраст децата са интернализирали редица сложни характеристики на езика, които са извън семантиката и познанието на значението на думите (Welge-Lüssen et al., 2008).

При лица с нарушения на развитието като разстройство от аутистичния спектър (РАС) или дислексия на развитието (ДР), както и такива с придобити нарушения като афазия, употребата на псевдо-думи е почти ежедневна в диагностиката и терапията им. Използването на комбинация от реални и псевдо-думи дава възможност да се сравнят лексикалния и сублексикалния път на четене (Coltheart, Curtis, Atkins, & Haller, 1993; Coltheart et al., 2001). В процеса на четене псевдо-думите се обработват до голяма степен по сублексикален път, който използва сериен (буква по буква) режим на четене. Четенето на реални думи се реализира чрез лексикалния път на обработка, който действа чрез холистичното възприемане на думите и включва сравнение с техните шаблони в паметта. По този начин в процеса на изследване може да се проследи относително независимото използване на двата пътя на четене. По отношение на изследванията, свързани с РАС, трябва да отбележим обстойния обзор на Whalon et al. (2009) на литературата, свързана с инструкциите, прилагани при диагностиката на четенето на децата с РАС. Обзорът обхваща 11 проучвания с 61 участници на възраст от 4 до 17 години. Наборът от инструменти е насочен към изследването на умения, фокусирани върху кода, или фонологично осъзнаване, фониката и плавността, които са необходими за точното и свободно идентифициране на думите в текста като в повечето случаи тестовия материал включва думи и псевдо-думи. Други автори, които изследват морфо-ортографската идентификация отвъд графемно-фонемните способности по отношение на иврит, също използват псевдо-думи като първоначалното им проучване включва 31 деца от шести клас с РАС (Ferman & Var-On, 2017). Интересни данни са констатирани и по отношение на феномена хиперлексия, свързан предимно с лица от аутистичния спектър. Хиперлексията се определя като ранно придобиване на умения за четене без изрично преподаване и наличието на силна ориентация към писмени материали. В този систематичен преглед на отделни случаи (N=82) и групови проучвания (включително 315 участници с хиперлексия) авторите откриват, че хиперлексията е характерна за значителна част от изследваните деца и юноши от аутистичния спектър. Важно е да отбележим, че по-голяма част от инструментите, използвани в тези изследвания, включват набори от псевдо-думи. В други източници, авторите чрез образни изследвания се опитват да изолират невронален корелат на семантичния достъп чрез зрително представени думи и псевдо-думи. Установяват, че псевдо-думите, ако са подобни на истинските думи, могат да предизвикат и семантична обработка (Ostrolenka et al., 2017). Съществуват и противоречиви доказателства за употребата на фонологичния и ортографския път в процеса на четене при децата с РАС, които са резултат от изследване, включващо работа с омофони, спонеризми, четене на думи и псевдо-думи и др. (Saldaña, Carreiras, & Frith, 2009).

При изследвания на лица с дислексия отново с цел установяване на пътя на четене (лексикален или сублексикален) често се използват списъци с кратки и дълги думи и/или псевдо-думи. В допълнение изследователите анализират и моделите на движения на очите по време на четене на думи и псевдо-думи при лица с дислексия и с типично развитие (De Luca et al., 2002; Hutzler, & Wimmer, 2004).

Трябва да отбележим и използването на псевдо-думи в областта на фонологичната терапия, например, чрез модела на “Максималните опозиции” (Bonini, & Keske-Soares, 2018). Интересно е, че тези автори използват експерти за подбора на тестовия материал и по-специално при генерирането на псевдо-думите, което отново подчертава важната роля на използването на псевдо-думите при различни диагностични и терапевтични интервенции и необходимостта от прилагането на бързи и ефективни методи за тяхното създаване.

Въпреки че всички изследователи, които се занимават в тези области на изследвания са отговорни за подбора и качеството на диагностичния материал, няма напълно изграден универсален метод за създаване на псевдо-думи. Има различни методи като най-често се сглобяват срички, като за да се получи една сричка, се взима случаен разрешен онсет, ядро и кода в дадения език (съществуващ в реалните думи), и така се получава новата сричка. От гледна точка на теорията за нелинейната фонологична структура на думата всички срички се делят на онсет “onset” (консонанта или консонантното струпуване, които предхождат гласната) и рима “rime” (гласната и последващите консонантни в сричката). Връх или ядро “nucleus” на сричката в българския език е винаги гласната фонема, а кода (coda) се нарича крайната съгласна или група съгласни, широко се използва и в музикалната терминология и означава ”заклучение”, „край” (Щерева, 2012). Това се повтаря за следващата сричка и т.н. докато не се сглоби дума с произволна дължина. Проблемът на този метод е, че броят на комбинациите нараства експоненциално и не може лесно да се ранкират псевдо-думите според това колко са близки до вече съществуващи думи (виж Ghyselinck, Lewis, & Brysbaert, 2004). Подобни примери могат да се посочат и във връзка с методология за изграждане на псевдо-думи за гръцкия език като авторите създават версия само за съществителни имена на базата на съществуващ корпус Clean Corpus/ILSP, а новата версия наричат NClean (Revithiadou, Ioannou, Chatzinikolaou, & Aivazoglou, 2016). Друга известна и лесна за използване програма за генериране на псевдо-думи е WordGen, която използва лексикалните бази данни CELEX и Lexique за подбор на думи и генериране на псевдо-думи на нидерландски, английски, немски и френски (Duysck, Desmet, Verbeke, & Brysbaert, 2004).

От изложената по-горе информация става ясно, че при различни изследвания на процеса на четене, възниква нужда от подходящ метод за генериране на псевдо-думи.

2. ЦЕЛ

Целта на нашата работа е адаптиране на компютърната програма и методът Wuggy за генериране на псевдо-думи в българския език.

3. МАТЕРИАЛИ И МЕТОДИ

Методът Wuggy дава възможност да се генерират псевдо-думи благодарение на познанията на субсричковата структура, фонотактиката и честотата на възможните им комбинации в съответния език, чрез използването на огромна база лексикални данни. Процедурата включва изчисляване на наличието на даден субсричков елемент, неговото честотно и позиционно място, както и дължината на думата, която го съдържа. До този момент програмата беше достъпна на холандски, английски, немски, френски, испански, сръбски и баски езици (Keuleers, & Brysbaert, 2010).

С цел методът да е ефективен и да се намали броя на думите в програмата, авторите предлагат опростяване чрез съкращаване по дължина и по честота на използваните елементи (Keuleers, & Brysbaert, 2010):

- съкращаване по дължина: създават се само псевдо-думи със същата дължина за всеки субсричков елемент като думата-шаблон на същата позиция в думата (пример: ако шаблонът е стр-а-ст генерираме само псевдо-думи със структура 3-2-2 като стр-е-ст или втр-а-кт);
- опростяване по честота: разглеждаме каква е честотата на всеки вариант за продължаване на субсричковия елемент в дадената дума за езика и махаме всички връзки, които се различават по честота с повече от N. Примерно м-о-ст има биграмите (м), (мо), (ост), (ст) с честота за българския примерно 24, 12, 13, 25; махаме всички връзки за (м) (и после останалите) с честоти по-големи от 24+N и по-малки от 24-N. Например, ако се избере N = 2, но програмата не генерира псевдо-дума се продължава с N = 4; 8; 16 и т.н.

На Таблица 1 се вижда как при наличието на създадена биграмна верига на дадена дума се добавя нова дума, която има някой сходни характеристики, които се отчитат от програмата, но се борави и с новопостъпилата информация.

Таблица 1. Пример за биграмна верига и разширяване на данните с добавяне на нова дума

Биграмна верига:	Добавяне на нова дума:
stamina: ('^', 'st', 'a', '-', 'm', 'i', '-', 'n', 'a', '-', '\$')	statistic: ('^', 'st', 'a', '-', 't', 'i', '-', 's', 't', 'i', 'c', '\$')
(0, '^') : {(1, 'st'): 1}	(0, '^') : {(1, 'st'): 2}
(1, 'st') : {(2, 'a'): 1}	(1, 'st') : {(2, 'a'): 2}
(2, 'a') : {(3, '-'): 1}	(2, 'a') : {(3, '-'): 2}
(3, '-') : {(4, 'm'): 1}	(3, '-') : {(4, 'm'): 1, (4, 't'): 1}
(4, 'm') : {(5, 'i'): 1}	(4, 'm') : {(5, 'i'): 1}
(5, 'i') : {(6, '-'): 1}	(5, 'i') : {(6, '-'): 1}
(6, '-') : {(7, 'n'): 1}	(6, '-') : {(7, 'n'): 1}
(7, 'n') : {(8, 'a'): 1}	(7, 'n') : {(8, 'a'): 1}
(8, 'a') : {(9, '-'): 1}	(8, 'a') : {(9, '-'): 1}
(9, '-') : {(10, '\$'): 1}	(9, '-') : {(10, '\$'): 1}
	(4, 't') : {(5, 'i'): 1}
	(6, '-') : {(7, 'st'): 1}
	(7, 'st') : {(8, 'i'): 1}
	(8, 'i') : {(9, 'c'): 1}
	(9, 'c') : {(10, '\$'): 1}

4. РЕЗУЛТАТИ И ДИСКУСИЯ

Въз основа на характеристиките на програмата, се пристъпи към подсигуряване на лексикална банка на българския език; приспособяване на кода на програмата към българския език и разработване на алгоритъм за разделяне на българските думи на срички, поради липса на такъв речник. Адаптиран и използван е съществуващият код, който е наличен на адрес:

https://github.com/?fbclid=IwAR1TnbAopoU19PJd8zffHeXOR5F_HPqe5Y88qwRuvX-JGk2WUAI_AztPK3E.

Както беше отбелязано чрез обработка от самата програма, всяка дума се разделя на т.нар. субсричкови елементи: онсет, ядро и кода. Генерираните псевдо-думи се изграждат изцяло на базата на съществуващите сричкови елементи в дадения език, вземайки предвид тяхната позиция и структура, както и разрешените консонантни клъстърни. Важно е да отбележим, че думите с еднакъв брой срички в българския език се състоят от еднакъв брой субсричкови елементи, защото липсващата кода или онсет в дадена дума се причислява към общия брой елементи.

Структурата на сричката, съставът и дистрибуцията на клъстърите се подчиняват на универсални закономерности и в същото време отразяват спецификата на дадения език. Сричките биват два вида: затворени и отворени, като могат да съдържат една или повече съгласни. Освен това биват още акцентуирани и неакцентуирани (ударени и неударени). Сричката е най-малката независима произносителна (артикулационна) единица на речта, в която се свързват звуковете при говорене и на която говорещия може да раздели речта си с паузи. Тя няма смислово и граматическо значение. В нейните граници фонемите се групират контрастно в определени структури, които способстват за възприемане и осъзнаване на фонемните противопоставяния. Както беше отбелязано, връх или ядро на сричката в българския език е винаги гласната фонема. Функцията на съгласните в сричката е разграничителна – те определят нейното начало и край: мрежа, яс-но, бук-вар (Бояджиев, Тилков, 1999: 174). В много езици като основен принцип за изграждане на сричката се приема принципът на възходящата звучност. Според този принцип, приложим и за българския език, следващите един след друг вътре в сричката звукове трябва да бъдат контрастни по звучност. Най-важен е основният тонален контраст, следствие от съчетаването на вокалните и консонантните признаци. При групирането на съгласните фонемни спрямо сричкообразуващия център – гласната, качеството на сричковата структура се определя от качествата на предшестващата и следващата фонема. Съобразно това в българския език е невъзможна сричкова структура от сонорна и шумова в началото на сричката, както например в руския: мгла. Тези съчетания в българския език се разделят в две срички: мъг-ла. Характерна особеност на българския език е, че думата не може да завършва с такава сричкова структура, с каквато е възможно да започне. В начални и средишни срички се срещат съчетания, които не се допускат в крайна затворена сричка: слава, но смисъл; траен, но театър, т.е. крайна сонорна съгласна не може да бъде отделена от ядрото на сричката от звук с по-ниска звучност (Бояджиев, Тилков, 1999: 176). Българският език е много богат от гледна точка на морфологичното разнообразие. Думите се променят и се образуват нови словоформи благодарение на поредица от представки, наставки и окончания.

Тук ще представим основните правила за разделяне на българската дума на срички, тъй като те са изключително важни за създаването на алгоритъм, чрез който програмата Wuggy да прави това автоматично

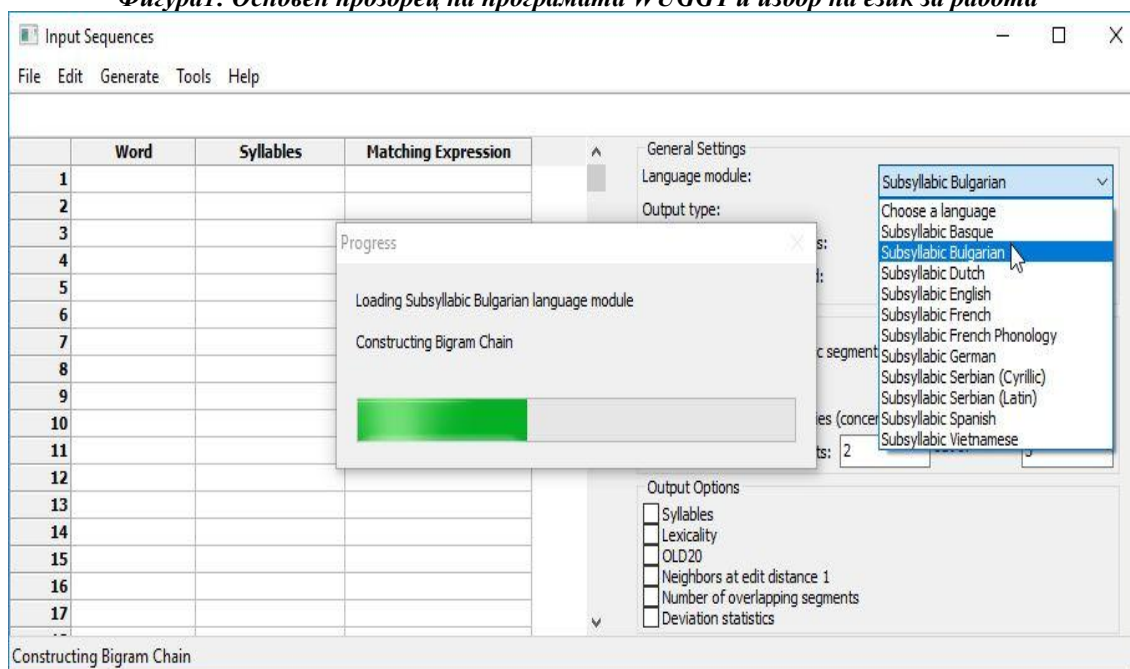
след подаването на съответната дума. Тези правила могат да се систематизират в следната последователност:

- една съгласна между две гласни образува сричка със следващата гласна (о-пе-ра);
- две съгласни между две гласни могат да се разделят от сричковата граница или да останат неделими (тес-то; те-сто);
- двойни еднакви съгласни се разделят между първата и втората сричка (есен-на);
- групи от две и три съгласни подредени на принципа на възходящата звучност обикновено не се разделят от сричковата граница, но може да се правят аналогии и с морфологичното делене (кру – шка; круш-ка);
- поради своята морфологична самостоятелност обикновено се запазват неразчленими и съгласните в представките, наставките и суфиксите (под-будя, без-домен);
- ако обаче следващата сричка започва с гласна представката не запазва своята автономност (безочлив);

Трябва да се отбележат и редица изключения, поради които бяха генерирани много допълнителни условия в създадения алгоритъм за разделянето на българските думи на срички, като например, разпна се разделя на раз-пна или при думи с корени от гръцки произход като например, пневм-, които фигурират обикновено само в началото на думата. От друга страна звукът “в” в българският език има специфично и нетипично поведение на свързване с останалите звукови сегменти, което наложи чрез алгоритъма да се въведе в програмата условието да се направи проверка за наличието на буква “в” и ако тя се намира непосредствено преди втората съгласна, която е “д”, “з”, “к”, “т”, “х”, “ц” “с”, границата на сричката да се поставя преди тази втора съгласна (под-ква-ся). Разбира се, като всеки език и българският има известни редки изключения, с които трябваше да се примирим и приемем известно некоректно сегментиране, като например, думи с дж, тл, пс, и други (кос-тва, оскверня, спретнат, джуд-же, разпна, изгложвам, пневмония и т.н.).

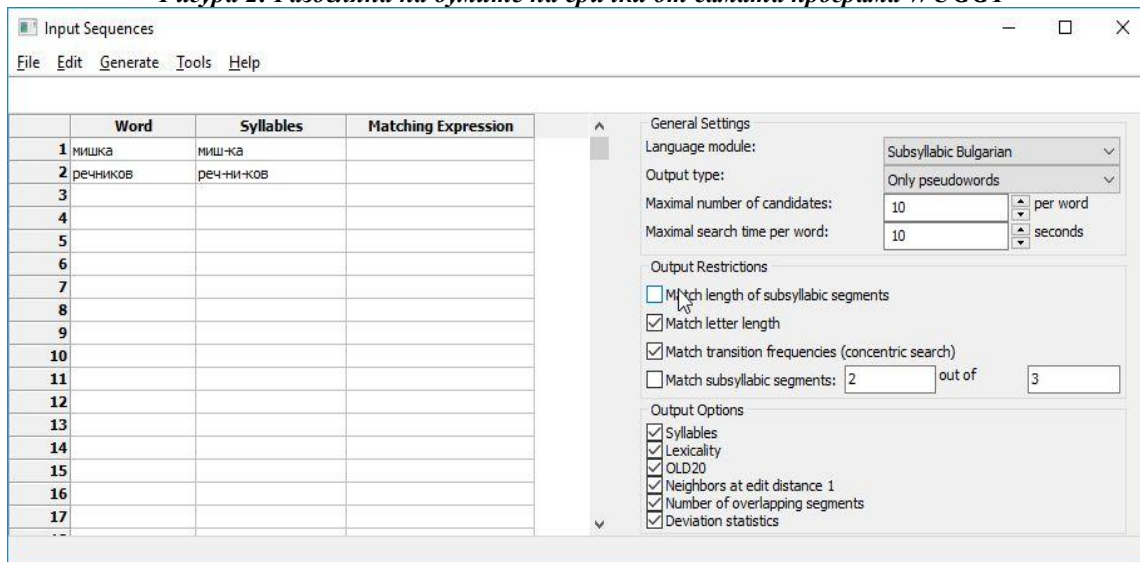
След създаването на алгоритъма и прилагането му в програмата, нашият екип пристъпи към практическото ѝ използване. На Фигура 1 е показан основният прозорец на Wuggy. След стартиране на програмата, трябва да бъде избран езиковия модул от опцията „Общи настройки“. На фигура 1 е „зареден“ модулът с българския език.

Фигура1: Основен прозорец на програмата WUGGY и избор на език за работа



Речниковият фонд, който е наличен в програмата се използва за въвеждането и тестването на лексикалността на генерираните форми. След зареждането на езиковия модул, което може да отнеме няколко минути, могат да се въведат референтните думи в съответната колона. На Фигура 2 се виждат въведените от нас думи „мишка“ и „речников“, след което се избира опцията сричкоразделяне (Syllabify) от падащото меню с инструменти.

Фигура 2: Разделяна на думите на срички от самата програма WUGGY



Фигура 3 показва как от избраната опция за стартиране на подбора (Run) от падащото меню се отваря прозорецът с резултатите, в случая за думите „мишка“ и „речников“ по подразбиране с всички проверени опции и ограничения.

Фигура 3: Изходен прозорец с генерираните псевдо-думи в програмата WUGGY

	Word	Match	Lexicality	Old20	Old20_Diff	Ned1	Ned1_Diff	Overlap_Ratio	Maxdeviation	Summed_Deviation	Maxdeviation_Transition
1	мишка	баг-ка	N	1.8	0.3	4	-4	5/6	0	0	[баг]ка_
2	мишка	хък-ка	N	1.95	0.45	1	-7	5/6	-1	2	[хък]ка_
3	мишка	дяс-ка	N	1.85	0.35	3	-5	5/6	-1	2	[дяс]ка_
4	мишка	цер-ка	N	1.75	0.25	5	-3	5/6	-2	4	[цер]ка_
5	мишка	дум-ка	N	1.7	0.2	6	-2	5/6	0	0	[дум]ка_
6	мишка	пге-ка	N	1.95	0.45	1	-7	5/6	-2	4	[пге]ка_
7	мишка	хаз-ка	N	1.8	0.3	4	-4	5/6	-2	4	[хаз]ка_
8	мишка	гле-ка	N	1.7	0.2	6	-2	5/6	1	2	[гле]ка_
9	мишка	кач-ка	N	1.5	0.0	10	2	5/6	-2	4	[кач]ка_
10	мишка	дог-ка	N	1.8	0.3	4	-4	5/6	0	0	[дог]ка_
11	речников	лав-ни-мът	N	3.75	1.2	0	0	7/9	-2	6	_ лавни[мът]_
12	речников	лав-ни-ков	N	2.85	0.3	0	0	8/9	1	2	[лав]ников_
13	речников	чек-ни-мът	N	3.85	1.3	0	0	7/9	-2	6	_ чекни[мът]_
14	речников	чек-ни-ков	N	2.85	0.3	0	0	8/9	-1	2	[чек]ников_
15	речников	вче-ни-мът	N	4.0	1.45	0	0	7/9	-2	6	_ вчени[мът]_
16	речников	вче-ни-ков	N	3.35	0.8	0	0	8/9	-1	2	[вче]ников_
17	речников	реб-ни-мът	N	3.9	1.35	0	0	7/9	-2	8	_ ребни[мът]_
18	речников	реб-ни-ков	N	2.75	0.2	1	1	8/9	-2	4	[реб]ников_
19	речников	бъз-ни-мът	N	3.95	1.4	0	0	7/9	-2	6	_ бъзни[мът]_
20	речников	бъз-ни-ков	N	3.0	0.45	0	0	8/9	-1	2	[бъз]ников_

Като важен елемент от работата на нашия изследователски екип беше прилагането на програмата Wuggy за подбор на лингвистичен материал за провеждането на планираното изследване. Селектирането на лингвистичния материал за изследването (думи и псевдо-думи) във връзка с особеностите на българския език отговаря на следните характеристики: брой срички - една сричка; две срички; три срички; четири срички; сходна сричкова структура – клъстери (без струпване; две съгласни; три съгласни; четири съгласни); използване на думи, които са различни части на речта – съществително име; глагол; други (прилагателно име, числително име, наречие, местоимение и др.) и генериране на псевдо-думи, които отговарят на съответните характеристики.

5. ЗАКЛЮЧЕНИЕ

В заключение методът Wuggy, приложен за българския език, дава възможност да се генерират български псевдо-думи като се използва познаването на субричковата структура, фонотактиката и честотата на техните възможни комбинации, използвайки огромна лексикална база данни.

Генерираните псевдо-думи са конструирани изцяло въз основа на съществуващите силабични елементи в българския език, като се отчита тяхното положение и структура, както и разрешените съгласни клъстери.

Програмата може да се използва и за избор на реални думи, които са сходни по фонологична структура, дължина и честота, на целевите думи.

Програмата, която до този момент беше достъпна на холандски, английски, немски, френски, испански, сръбски, баски и вьетнамски, вече може да се прилага и за българския език. Програмата може да бъде изтеглена безплатно от <http://autism-vision.bas.bg>.

БЛАГОДАРНОСТИ

Нашият екип изказва благодарност на Emmanuel Keuleers за личното съдействие в процеса на прилагане на метода и програмата Wuggy за българския език.

REFERENCES

- Бояджиев, Т., & Тилков, Д. (1999). Фонетика. Велика Търново
- Щерева, К. (2012). Фонологичното осъзнаване на децата (превенция на дислексия). София. Глосса терапи
- Bonini, J.B., & Keske-Soares, M. (2018). Pseudowords to phonological therapy validated by expert judges. *CoDAS*, 30(5), e20170013. Epub October 22, 2018. <https://doi.org/10.1590/2317-1782/20182017013>
- Coltheart, M., Curtis, B., Atkins, P. & Haller, M. (1993) Models of reading aloud: dual route and parallel-distributed-processing approaches. *Psychological Review*, 100, 589–608. <https://doi.org/10.1037/0033-295X.100.4.589>
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological review*, 108(1), 204–256. <https://doi.org/10.1037/0033-295X.108.1.204>
- De Luca, M., Borrelli, M., Judica, A., Spinelli, D., & Zoccolotti, P. (2002). Reading words and pseudowords: An eye movement study of developmental dyslexia. *Brain and Language*, 80(3), 617–626. <https://doi.org/10.1006/brln.2001.2637>
- Duyck, W., Desmet, T., Verbeke, L. P. C., & Brysbaert, M. (2004). WordGen: A tool for word selection and nonword generation in Dutch, English, German, and French. *Behavior Research Methods, Instruments, & Computers*, 36, 488-499.
- Ghyselinck, M., Lewis, M. B., & Brysbaert, M. (2004). Age of acquisition and the cumulative-frequency hypothesis: A review of the literature and a new multi-task investigation. *Acta Psychologica*, 115, 43-67. Rathvon, Natalie (2004). *Early Reading Assessment: A Practitioner's Handbook*. New York: The Guilford Press. p. 138. ISBN 1572309849.
- Hutzler, F., & Wimmer, H. (2004). Eye movements of dyslexic children when reading in a regular orthography. *Brain and language*, 89(1), 235–242. [https://doi.org/10.1016/S0093-934X\(03\)00401-2](https://doi.org/10.1016/S0093-934X(03)00401-2)
- Keuleers, E., & Brysbaert, M. (2010). Wuggy: a multilingual pseudoword generator. *Behavior research methods*, 42(3), 627–633. <https://doi.org/10.3758/BRM.42.3.627>
- Revithiadou, A., Ioannou, D., Chatzinikolaou, M., & Aivazoglou, K. (2016). Constructing pseudowords for experimental research: Problems and solutions.
- Scharenborg, O. (2007). "Reaching over the gap: A review of efforts to link human and automatic speech recognition research" (PDF). *Speech Communication*. 49 (5): 336–347. doi:10.1016/j.specom.2007.01.009.
- Trezek, B. J., Wang, Y., & Paul, P. V. (2010). *Reading and deafness: Theory, research, and practice*. Clifton Park, NY: Delmar Cengage Learning.
- Welge-Lüßen, A., Hauser, R., Erdmann, J., Schwob, Ch., & Probst, R. (2008). "Sprachaudiometrie mit Logatomen*". *Laryngo-Rhino-Otologie*. 76 (2): 57–64. doi:10.1055/s-2007-997389. PMID 9172631.