

ЕФЕКТЪТ НА ЛЕКСИКАЛНАТА ЧЕСТОТНОСТ И ТИПА МОРФЕМА ВЪРХУ РЕГРЕСИВНАТА АСИМИЛАЦИЯ НА БЪЛГАРСКИТЕ ОБСТРУЕНТИ ПО ПРИЗНАКА ЗВУЧНОСТ

БИСТРА АНДРЕЕВА, МИТКО СЪБЕВ

УНИВЕРСИТЕТ НА СААРЛАНД

andreeva@lst.uni-saarland.de, msabev@lst.uni-saarland.de

Настоящото изследване разглежда връзката между честотността (предсказуемостта) на езиковите единици и регресивната асимилация на обструентите по признака звучност в българския език. Въз основа на данните от речеви корпус и езикови модели на ниво словоформа изчисляваме стойността на изненадата на думи с красловни съгласни /t/, /ʃ/ и /x/ в състава на формообразуващи и словообразуващи морфеме. Чрез смесен линеен модел (LMM) анализираме как предсказуемостта, звучността на следващия обструент и типът морфема влияят върху реализираната звучност на съгласните. Резултатите показват, че езиковата предсказуемост влияе върху прецизността на артикулацията, като модулира степента на регресивната асимилация в зависимост от морфологичния контекст. Установяваме, че обструентите в по-малко предсказуемите думи се произнасят с по-прецизна артикулация, отколкото обструентите в по-предсказуемите думи, особено в състава на словообразуващи морфеме. Това подчертава комплексната динамика между информационна плътност, морфологията и артикулационния процес, което може да даде основа за по-нататъшни изследвания в областта на фонетиката и фонологията.

Ключови думи: лексикална честотност; изненада; регресивна асимилация; звучност

THE EFFECTS OF LEXICAL FREQUENCY AND MORPHEME TYPE ON ANTICIPATORY VOICE ASSIMILATION IN BULGARIAN OBSTRUENTS

BISTRA ANDREEVA, MITKO SABEV

SAARLAND UNIVERSITY

andreeva@lst.uni-saarland.de, msabev@lst.uni-saarland.de

This study investigates the relationship between the frequency (or predictability) of linguistic units and the anticipatory voice assimilation of obstruents in Bulgarian. Using a speech corpus and word-level language models, we calculate the surprisal of word forms ending in the consonants /t/, /ʃ/, and /x/ in inflectional and lexical morphemes. We then employ a linear mixed model (LMM) to analyse how surprisal, the voicing of the following obstruent, and morpheme type affect the voicing realised in the examined consonants.

The findings demonstrate that linguistic predictability affects articulatory precision by modulating the degree of anticipatory assimilation, with different effect sizes for different morpheme types. More specifically, obstruents in less predictable words are articulated with greater precision than those in more predictable words, especially within lexical morphemes. This reveals a complex interaction between information density, morphology and articulation, providing avenues for further research in phonetics and phonology.

Keywords: lexical frequency; surprisal; regressive assimilation; voicing

1. Въведение

Езикът осигурява не само изразителността, необходима за общуването, но и предлага на говорещите множество възможности за избор на начина, по който да кодират своите послания – от времетраенето, спектралните и прозодичните характеристики на звуковете и сричките до избора на думи, структурирането на синтактичните елементи и подреждането на изречението в дискурса. В последните години все по-често се прилагат принципите на теорията на информацията, за да се придобие по-пълна представа за ефекта на информационната плътност (като например честотността и предсказуемостта) на езиковите единици върху вариативността в езика. Резултатите от изследванията показват, че говорещите и слушателите имат достъп до вероятностни разпределения на езиковите единици (Pierrehumbert 2001; Jaeger 2010). Доказано е, че езиковите единици се редуцират или дори заличават, когато са лесно предсказуеми, и че имат по-ясно изразени акустични характеристики, когато са трудно предсказуеми (Aylett, Turk 2004, 2006; Malisz et al. 2018; Ibrahim et al. 2022; Cohen Priva 2015).

Една от количествените мерки за информационната плътност е изненадата (англ. surprisal). Теорията за изненадата (Hale 2001; Levy 2008) е доста успешна в обясняването на трудностите при възприемането, обработката, съхраняването и образуването на речевите единици. Според теорията тази трудност е правопрпорционална на изненадата (или на неочакваността на речевата единица), която обикновено се оценява чрез вероятностни езикови модели. Езиковите модели най-често се обучават на ниво словоформи, но могат да се прилагат и към други типове единици на различни нива на лингвистичното описание.

Установено е, че речеобразуването е чувствително към честотата на употреба на думите (абсолютен брой срещания в корпус) и към вероятността на униграмите (вероятност спрямо други униграми, изчислена на базата на големи корпуси, като всички вероятности се сумират до единица). Например моделът на речеобразуване, предложен от екипа на У. Левълт (Levelt, Wheeldon 1994; Levelt, 1999), предполага взаимодействие между честотността и кодирането на артикулационните процеси. Сричките с висока честотност, които обикновено се срещат в думи с висока честотност, се артикулират по-бързо от сричките с ниска честотност (Carreiras, Perea 2004; Cholin et al. 2006); английските глаголи с висока честотност в минало време се артикулират по-бързо от тези с по-ниска честотност (Losiewicz

1992); в думи и колокации с висока честотност (например „don't“ в „don't know“) се наблюдава по-висока степен на редуция и лениция в сравнение с рядко срещаните (Bybee, Scheibman 1999; Jurafsky et al. 2001). При сречките с висока честотност се наблюдава повече коартикуляция, отколкото при редките сречки (Whiteside, Varley 1998; Croot, Rastle 2004). В експерименти, изследващи коартикулационните ефекти върху лабораторна и актьорска реч, са установени значителни разлики между честите и редките сречки в думи и псевдодуми по отношение на няколко параметъра, определящи гласните в перцептивното пространство, включително траекториите на формантите в рамките на сречката и преходите на формантите на границите на сречките (Benner et al. 2007).

Настоящата статия се фокусира върху връзката между предсказуемостта и регресивната асимилация на обструентите по признака звучност. Хипотезата, залегнала в основата на нашето изследване, е, че говорещите модулират детайлите на фонетичното кодиране, за да поддържат баланса между него и информационната плътност на езиковите елементи, което води до отрицателна корелация между прецизността на артикулационната и акустичната реализация на фонетичните структури, от една страна, и тяхната предсказуемост, от друга. Следователно очакваме, че звуковете в по-трудно предсказуемите (по-нискочестотните) думи ще се артикулират по-прецизно в сравнение със сегментите в по-лесно предсказуемите (по-високочестотните) думи. Също така очакваме, че във формообразуващите морфемии влиянието на изненадата върху регресивната асимилация по звучност ще бъде по-слабо изразено, тъй като тези морфемии сами по себе си са значително по-честотни в езика.

2. Методология на изследването

Анализиран е материал от речевия корпус VulPhonC, версия 3 (Hateva, Mitankin, Mihov 2016), състоящ се от 319 фонетично разнообразни изречения, прочетени от 140 носители на българския книжовен език (81 жени, 59 мъже) със средна възраст от 36.78 години (стандартно отклонение 15.56, медиана 29.50). Записите са осъществени в шумоизолирано ехопотискащо студио с всепосочен микрофон Sennheiser MK 4 и цифрово записващо устройство TASCAM DP32 при честота на дискретизация от 48 kHz и 24-битово квантуване, филтрирани и компресирани до 16 kHz. Корпусът е снабден с орфографска транскрипция и е автоматично сегментиран на ниво фонема (Mitankin, Mihov, Tinchev 2009). За настоящото изследване е ползвана извадка от данните, в която на базата на орфографската транскрипция автоматично са сегментирани и границите на думите. За да проверим нашата хипотеза, анализираме озвучаването на краесловните беззвучни обструенти в личните окончания за сегашно време, 2-ро лице единствено число /s/; сегашно време, 3-то лице множествено число /t/ и минало свършено и минало несвършено време, 1-во лице единствено число /x/ пред 2677 думи, започ-

ващи със звучни, и 3771 думи, започващи с беззвучни обструенти. Сравняваме реализираната звучност на съгласните в личните окончания със звучността на същите фонемни извън състава на формообразуващи морфемни пред 2007 думи, започващи със звучни, и 4364 думи, започващи с беззвучни обструенти.

Както вече беше споменато, изненадата се определя като предсказуемост на дадена единица и може да се използва като мярка за количеството информация, която се предава от тази единица в битове. Тя се изчислява, както е показано в уравнение 1, където „S“ е изненадата (surprisal), „X“ – езиковата единица, а „P“ е вероятността за появата на тази единица в езика или корпуса:

$$(1) S(X_i) = -\log_2 P(X_i)$$

Ако изненадата при дадена единица има ниска стойност, то тя е лесно предсказуема. И обратно, ако изненадата има висока стойност, единицата е по-трудно предсказуема. Изненадата отразява интуицията, че езиковите единици, които са силно предсказуеми, предават по-малко информация от тези, които са неочаквани.

В това изследване използваме езиковите модели GlobalPhone за български език на ниво словоформа (Vu et al. 2010), обучени върху автоматично извлечен интернет корпус с обем 302 милиона думи, които към момента на подаване на статията за печат са налични на адрес <https://www.uni-bremen.de/csl/projekte/aktuelle-projekte/global-phone-language-models>. Въз основа на езиковите модели униграм е изчислена стойността на изненадата на словоформите, в чието краесловие се намират съгласните /t/, /ʃ/ и /x/.

С помощта на скрипт за програмата Praat (Voersma 2024) автоматично са извлечени „профилите на озвучаване“ на съгласните /t/, /ʃ/ и /x/. Скриптът анализира речевия сигнал, като изчислява вероятността за озвучаване във всеки отрязък от съответната съгласна (0 или 1). Получените вероятности се сумират и нормализират спрямо дължината на съгласната, за да се определи общият дял (измерен в проценти) на фонетичната звучност, реализирана в съответната съгласна.

За да проверим нашата хипотеза – а именно, че сегментите в по-трудно предсказуемите (по-нискоочестотните) думи ще се артикулират по-прецизно в сравнение със сегментите в по-лесно предсказуемите (по-високоочестотните) думи – е приложен смесен линеен модел (LMM) с ПРОЦЕНТ ЗВУЧНОСТ, реализирана в съгласните /t/, /ʃ/ и /x/ като зависима променлива, стойността на ИЗНЕДАДАТА на думите, в чието краесловие се намират съгласните /t/, /ʃ/ и /x/, ЗВУЧНОСТ НА СЛЕДВАЩИЯ ОБСТРУЕНТ (звучен/беззвучен) и ТИП МОРФЕМА (формо-/словообразуваща) като независими променливи (фиксиран ефект), както и възможните взаимодействия между тях. Като случаен ефект е зададена променливата РЕСПОНДЕНТ, за да се избегне евентуално случайно въздействие на индивидуалните особености върху резултатите. Линеините смесени модели са изчислени в R (R Core Team 2024) с помощта

на пакетите lme4 (1.1-12) (Bates et al. 2015) и lmerTest (2.0-33) (Kuznetsova et al. 2016). Структурата на модела е представена в уравнение (2):

(2) ПРОЦЕНТ ЗВУЧНОСТ ~ ИЗНЕНАДА * ЗВУЧНОСТ НА СЛЕДВАЩИЯ ОБСТРУЕНТ * ТИП МОРФЕМА + (1|РЕСПОНДЕНТ)

Статистическата значимост на резултата от LMM (значим при $p \leq 0.05$) показва дали изненадата на думите, звучността на следващия обструент и типа морфема влияят върху звучността на анализираната съгласна.

3. Резултати от изследването

Таблица 1 обобщава резултатите от статистическия анализ на ефекта на изненадата, звучността на следващия обструент и типа морфема, както и техните взаимодействия върху комбинаторните фонетични промени при реализацията на краесловните обструенти /t/, /ʃ/ и /x/.

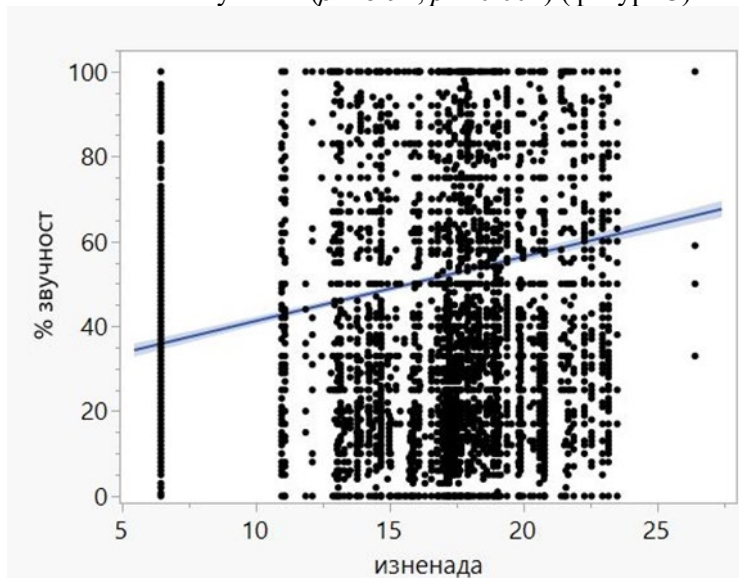
Таблица 1. Дял на звучността в краесловните обструенти: регресионни коефициенти (β), стандартна грешка и статистически резултати от анализа с линеен смесен модел (LMM)

Параметър	B	Стандартна грешка	t	p
Intercept ¹	15.54	1.42	10.96	< 0.001
Изненада	0.52	0.10	5.29	< 0.001
Звучност на следващия обструент [звучен]	61.42	1.19	51.51	< 0.001
Тип морфема [формообразуваща]	5.91	0.90	6.64	< 0.001
Изненада * Звучност на следващия обструент [звучен]	-1.42	0.19	-7.59	< 0.001
Изненада * Тип морфема [формообразуваща]	1.08	0.19	5.79	< 0.001
Звучност на следващия обструент [звучен] *	-0.88	1.53	-0.58	0.563
Тип морфема [формообразуваща]				
Изненада * Звучност на следващия обструент [звучен] * Тип морфема [формообразуваща]	-0.63	0.30	-2.12	0.034

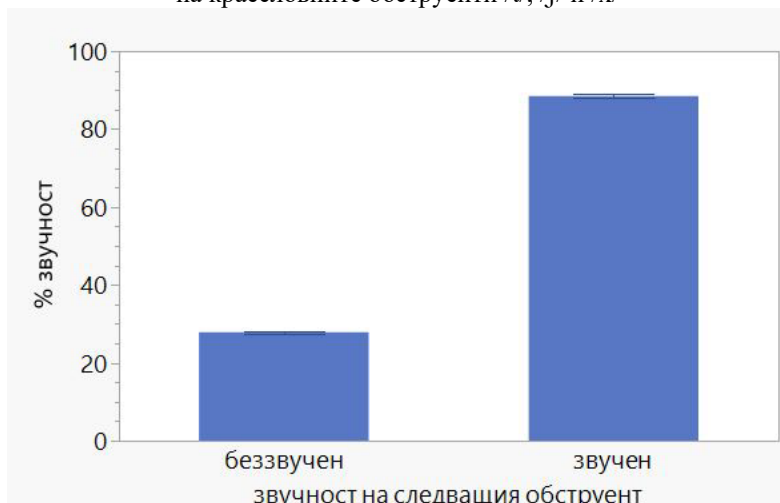
3.1. Основни ефекти

Основният ефект на всеки от трите фактора, т.е. влиянието, което оказва самостоятелно, независимо от другите фактори, върху вариативността при озвучаването на краесловните обструенти, е статистически значим. По-високите стойности на фактора ИЗНЕНАДА, т.е. по-ниската предсказуемост на думите, водят до увеличаване на процента звучност в краесловните обструенти ($\beta = 0.52$, $p < 0.001$) (фигура 1). По отношение на фактора ЗВУЧНОСТ НА СЛЕДВАЩИЯ ОБСТРУЕНТ беше установено, че звучните обструенти, следващи анализираната съгласна, водят до увеличаване на нейната звучност (β

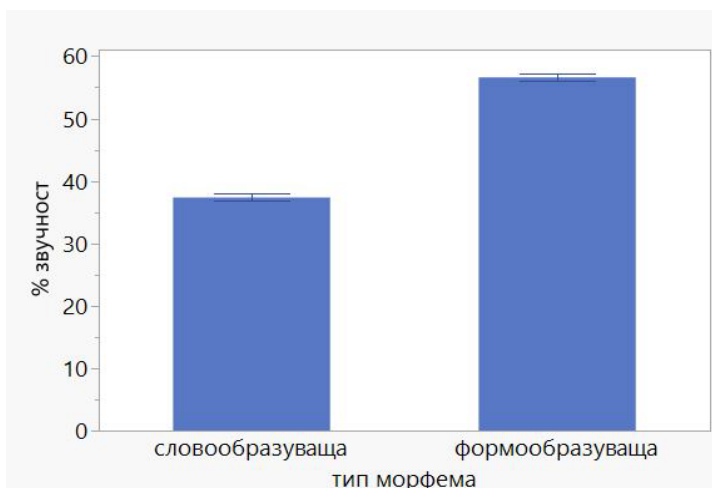
= 61.42, $p < 0.001$), което е очакван ефект, като се има предвид широко застъпената в българския език регресивна асимилация на обструентите по признака звучност. Трябва да се отбележи обаче, че уподобяването по звучност между съседните обструенти не е пълно, въпреки че за анализа са избрани само съседни думи, изговорени бързо, без пауза помежду им (фигура 2). Обструентите във формообразуващите морфемии се характеризират със съществено по-висока звучност ($\beta = 5.91$, $p < 0.001$) (фигура 3).



Фигура 1. Ефектът на фактора ИЗНЕНАДА върху звучността на краесловните обструенти /t/, /ʃ/ и /x/



Фигура 2. Ефектът на фактора ЗВУЧНОСТ НА СЛЕДВАЩИЯ ОБСТРУЕНТ върху звучността на краесловните обструенти /t/, /ʃ/ и /x/



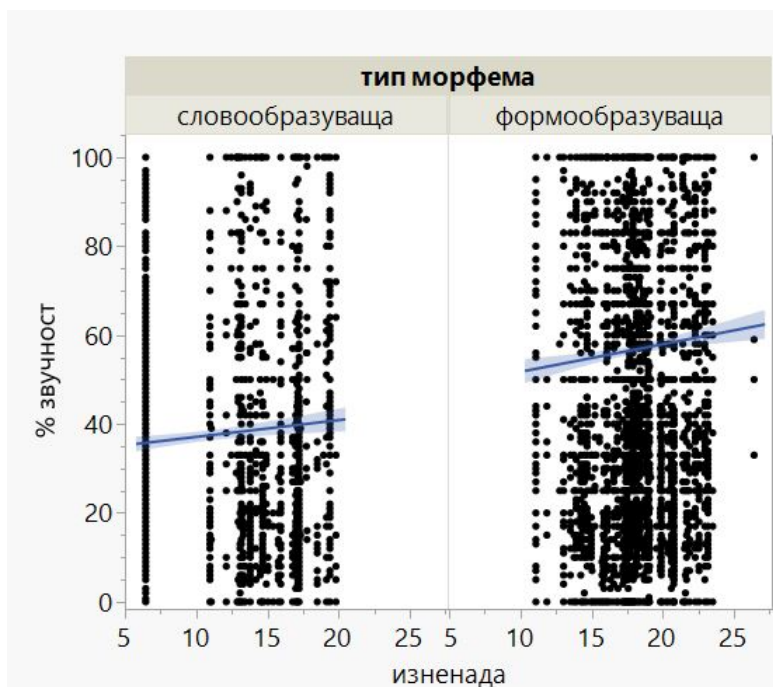
Фигура 3. Ефектът на фактора ТИП МОРФЕМА върху звучността на красловните обструенти /t/, /ʃ/ и /x/

3.2. Взаимодействия

Взаимодействието между факторите ИЗНЕНАДА и ЗВУЧНОСТ НА СЛЕДВАЩИЯ ОБСТРУЕНТ е отрицателно и значимо ($\beta = -1.42, p < 0.001$), което означава, че при следващ звучен обструент обструентите в края на думи с висока степен на изненада се озвучават по-малко (фигура 4). Взаимодействието между факторите ИЗНЕНАДА и ТИП МОРФЕМА е положително и значимо ($\beta = 1.08, p < 0.001$), което означава, че в контекста на формообразуващите морфемии ефектът на изненадата върху звучността се усилва (фигура 5).



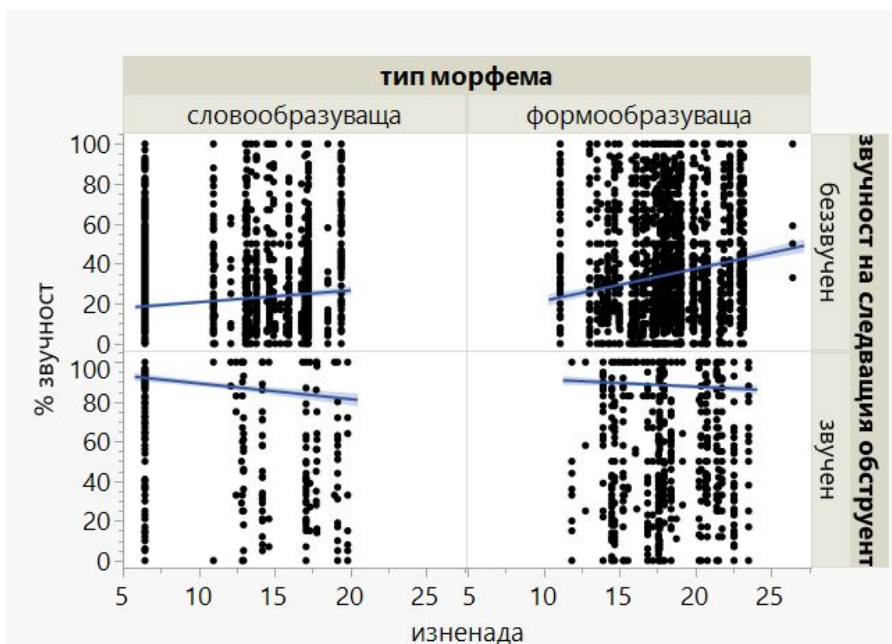
Фигура 4. Ефектът от взаимодействието на факторите ИЗНЕНАДА и ЗВУЧНОСТ НА СЛЕДВАЩИЯ ОБСТРУЕНТ върху звучността на красловните обструенти /t/, /ʃ/ и /x/



Фигура 5. Ефектът от взаимодействието на факторите ИЗЕНАДА и ТИП МОРФЕМА върху звучността на краесловните обструенти /t/, /ʃ/ и /x/

Взаимодействието между трите фактора също е статистически значимо ($\beta = -0.63$, $p < 0.034$), което показва комплексна взаимовръзка между тях при реализирането на звучността на краесловните обструенти. Когато следващата съгласна е беззвучна, ефектът на изненадата при формообразуващите морфемни силно нараства. При тях се наблюдава повече остатъчно озвучаване от предходната гласна при по-висока степен на изненада в сравнение със словообразуващите морфемни. Когато следващата съгласна е звучна, ефектът на изненадата е обратен. Озвучаването пред звучен обструент намалява с увеличаване на изненадата, като този ефект е по-слаб при краесловните обструенти във формообразуващите морфемни (фигура 5).

Този резултат потвърждава, че изненадата (непредсказуемостта) на думата, в която се намира съгласната, и нейните морфологични характеристики оказват значимо влияние върху прецизността в акустичната реализация на обструентите. В частност беззвучните обструенти са по-ясно артикулирани, т.е. се уподобяват в по-малка степен на звучността на следващите обструенти, както в по-трудно предсказуеми думи, така и в словообразуващи морфемни.



Фигура 6. Ефектът от взаимодействието на факторите ИЗНЕНАДА, ЗВУЧНОСТ НА СЛЕДВАЩИЯ ОБСТРУЕНТ и ТИП МОРФЕМА върху звучността на краесловните обструенти /t/, /ʃ/ и /x/

4. Дискусия

Основен фокус на представеното изследване е връзката между честотността (предсказуемостта) на думите и артикулационната прецизност при произношението на краесловните обструенти /t/, /ʃ/ и /x/ пред звучни и беззвучни обструенти. Резултатите предоставят нови данни за връзката между предсказуемостта на словоформите и регресивната асимилация по признака звучност.

Един от най-значимите резултати от изследването е потвърждението на хипотезата, че артикулацията на обструентите в думи с по-ниска предсказуемост е по-прецизна. Този ефект се наблюдава в контекста на асимилацията по признака звучност, като обструентите в думи, които са с по-висока степен на изненада (или по-ниска предсказуемост) в контекста на разговора, се произнасят по-прецизно, за да се намали рискът от неразбиране в процеса на комуникация. В същото време, когато думите са с по-ниска степен на изненада (или по-висока предсказуемост), прецизността в произношението отслабва, тъй като контекстът вече предоставя достатъчно информация за тяхното идентифициране.

Нашите резултати потвърждават изводите от предишни изследвания, които показват, че говорещите адаптират своето произношение според честотността и предсказуемостта на езиковите единици (Aylett, Turk 2004; Jurafsky et al. 2001). Тази динамика между предсказуемост и артикулация може да бъде обяснена с принципа на ефективната комуникация, при която говорещите балансират необходимостта от яснота и артикулационното усилие. Високата информационна плътност, която характеризира непредсказуемите думи, изисква по-прецизна артикулация, за да се предотврати евентуална неяснота или погрешно разбиране. Този резултат подкрепя и теорията на Дж. Хейл (Hale 2001) и Р. Леви (Levy 2008), която посочва, че предсказуемостта е основен фактор при възприемането и обработката на езиковите единици.

Резултатите показват, че регресивната асимилация по звучност се проявява по-силно при по-лесно предсказуеми думи, при които се наблюдава склонност към по-силно озвучаване на беззвучните пред звучни обструенти. Това обстоятелство е свързано с факта, че по-лесно предсказуемите думи предоставят на говорещия възможност да използва по-малко енергия за тяхното артикулиране, което води до озвучаване на иначе беззвучните обструенти.

Морфологичният контекст също оказва значително влияние върху артикулацията на обструентите. Обструентите във формообразуващите морфеми се реализират с по-голяма звучност в сравнение с тези в словообразуващите морфеми. Този ефект се обяснява с честотата на употреба на формообразуващите морфеми, което води до артикулационни промени с цел улесняване на произношението. Обратно, при по-малко използвани или по-сложни словообразуващи морфеми прецизността на артикулацията е по-голяма, за да се избегне евентуална неяснота.

Резултатите от изследването също така показват, че изненадата, типът на морфемата и звучността на следващия обструент не действат изолирано, а взаимно си влияят. Това взаимодействие подчертава сложната природа на фонетичните процеси, при които морфологичният контекст и вероятностните характеристики на езиковите единици играят ключова роля.

Настоящото изследване допълва и разширява предишните трудове в областта на фонетиката и психолингвистиката, като добавя нови данни за взаимодействието между предсказуемост и артикулация. Резултатите са в съответствие с теориите на коартикулацията и артикулационната икономия, но предоставят и данни за по-задълбоченото разбиране за ролята на морфологията и за начина, по който различните фактори се комбинират, за да влияят върху артикулационния процес. Сравнението на различни типове морфеми и тяхното влияние върху звучността на обструентите помага за по-доброто разбиране на морфофонологичните процеси, които се активират в естествения език. В контекста на тези резултати бъдещите изследвания биха се насочили към други морфологични структури и езикови единици, за да

се изследва допълнително взаимодействието между предсказуемостта и артикулацията.

Важно е да отбележим, че документираните честотни ефекти са неочаквани според фонологичната теория, в която асимилацията се описва като споделяне на фонологични признаци между съседни фонемите. Нашите резултати показват, че асимилацията по звучност при групи от беззвучен – звучен обструент е степенувана, а не пълна.

Въпреки че изследването предоставя ценни данни, трябва да се има предвид, че то разглежда конкретни обструенти и морфемите, което ограничава обобщаемостта на резултатите. Бъдещите изследвания биха могли да включват други фонологични и морфологични единици, за да се проучи дали наблюдаваните ефекти се проявяват и в други контексти.

5. Заключение

Изследването доказва значимостта на изненадата (непредсказуемостта) и морфологичните характеристики върху прецизността на артикулацията на обструенти по признака звучност. Резултатите показват, че както езиковата предсказуемост, така и морфологичният контекст влияят върху прецизността на артикулацията, като модулират степента на регресивната асимилация. Високата изненада, т.е. ниската предсказуемост, води до по-ясна артикулация. Това подчертава комплексната динамика между информационната плътност, морфологията и артикулационния процес, което може да бъде използвано като основа за по-нататъшни изследвания в областта на фонетиката, фонологията и морфологията.

Благодарности

Изследването е финансирано от Deutsche Forschungsgemeinschaft (DFG, Project ID 232722074-SFB1102).

БЕЛЕЖКА / NOTE

¹ Intercept (англ.) е свободният член на функцията. Неговата стойност указва пресечната точка с ординатната ос.

ЛИТЕРАТУРА / REFERENCES

- Aylett, Turk 2004: *Aylett, M., A. Turk*. The smooth signal redundancy hypothesis: a functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. – *Language and Speech*, 47, pp. 31–56.
- Aylett, Turk 2006: *Aylett, M., A. Turk*. Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. – *Journal of the Acoustical Society of America*, 119 (1), pp. 30–48.
- Bates et al. 2015: *Bates, D., M. Maechler, B. Bolker, S. Walker, R. H. Bojesen Christensen, N. Singmann, B. Dai*. Package 'lme4': Linear mixed-effects models using

- Eigen and S4, February 20, 2015. [Online]. <http://cran.rproject.org/web/packages/lme4/lme4.pdf>.
- Benner et al 2007: *Benner, U., I. Flechsig, G. Dogil, G., B. Möbius*. Coarticulatory resistance in a mental syllabary. – In: *Proceedings of the International Congress of Phonetic Sciences (Saarbrücken)*, pp. 485–488.
- Boersma 2022: *Boersma, P., D. Weenink*. Praat: doing phonetics by computer, version 6.4.17 (computer program).
- Bybee, Scheibman 1999: *Bybee, J., J. Scheibman, J.* The effect of usage on degrees of constituency: the reduction of don't in English. – *Linguistics*, 37, pp. 575–596. doi: 10.1515/ling.37.4.575.
- Carreiras, Perea 2004: *Carreiras, M., M. Perea*. Naming pseudowords in Spanish: effects of syllable frequency. – *Brain and Language*, 90 (1-3), pp. 393–400. doi: 10.1016/j.bandl.2003.12.003.
- Cholin et al. 2006: *Cholin, J., W. J. Levelt, N. O. Schiller*. Effects of syllable frequency in speech production. – *Cognition*, 99 (2), pp. 205–235. doi: 10.1016/j.cognition.2005.01.009.
- Cohen Priva 2015: *Cohen Priva, U.* Informativity affects consonant duration and deletion rates. – *Laboratory Phonology*, 6 (2), pp. 243–278.
- Croot, Rastle 2004: *Croot, K., K. Rastle, K.* Is there a syllabary containing stored articulatory plans for speech production in English? – In: *Proceedings of the 10th Australian International Conference on Speech Science and Technology (Sydney)*, pp. 376–381.
- Hale 2001: *Hale, J.* A probabilistic Earley parser as a psycholinguistic model. – In: *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics, Stroudsburg, PA*. Association for Computational Linguistics, pp. 159–166. doi: 10.3115/1073336.1073357.
- Hateva, Mitankin, Mihov 2016: *Hateva, N., P. Mitankin, S. Mihov*. BulPhonC. Bulgarian Speech Corpus for the development of ASR technology. – In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation*, pp. 771–774.
- Ibrahim et al. 2022: *Ibrahim, O., I. Yuen, M. van Os, B. Andreeva, B. Möbius*. The combined effects of contextual predictability and noise on the acoustic Realization of German syllables. – *Journal of the Acoustical Society of America*, 152 (2), pp. 911–920.
- Jaeger 2010: *Jaeger, T. F.* Redundancy and reduction: Speakers manage syntactic information density. – *Cognitive Psychology*, 61 (1), pp. 23–62.
- Jurafsky et al. 2001: *Jurafsky, D., A. Bell, M. Gregory, W. D. Raymond*. Probabilistic relations between words: evidence from reduction in lexical production. – In: *Bybee, J., P. Hopper (eds.) Frequency and the Emergence of Linguistic Structure*. Amsterdam, Benjamins, pp. 229–254.
- Kuznetsova et al. 2016: *Kuznetsova, A., P. Bruun Brockhoff, R. Haubo Bojesen Christensen*. lmerTest: Tests in Linear Mixed Effects Models, r package version 2.0-33. [Online]. <<https://CRAN.R-project.org/package=lmerTest>>.
- Levelt 1999: *Levelt, W. J. M.* Producing spoken language: a blueprint of the speaker. – In: *Brown, C. M., P. Hagoort (eds.) The Neurocognition of Language*. Oxford,

- Oxford University Press, pp. 83–122. doi: 10.1093/acprof:oso/9780198507932.003.0004.
- Levelt, Wheeldon 1994: *Levelt, W. J., L. Wheeldon*. Do speakers have access to a mental syllabary? – *Cognition*, 50, pp. 239–269.
- Levy 2008: *Levy, R*. Expectation-based syntactic comprehension. – *Cognition*, 106 (3), pp. 1126–1177. doi: 10.1016/j.cognition.2007.05.006.
- Losiewicz 1992: *Losiewicz, B. L*. The Effect of Frequency on Linguistic Morphology. University of Texas at Austin. PhD thesis.
- Malisz 2018: *Malisz, Z., E. Brandt, B. Mobius, Y. Oh, B. Andreeva*. Dimensions of segmental variability: interaction of prosody and surprisal in six languages. – *Frontiers in Communication*, 3, pp. 1–18.
- Mitankin, Mihov, Tinchev 2009: *Mitankin, P., S. Mihov, T. Tinchev*. Large vocabulary continuous speech recognition for Bulgarian. – In: *Proceedings of Recent Advances in Natural Language Processing (RANLP) 2009*, pp. 246–250.
- Pierrehumbert 2001: *Pierrehumbert, J. B*. Exemplar dynamics: Word frequency, lenition and contrast. – In: Bybee, J., P. Hopper (eds.). *Frequency and the Emergence of Linguistic Structure*. Amsterdam, Benjamins, pp. 137–157.
- R Core Team 2024: *R Core Team*. R: A Language and Environment for Statistical Computing, version 4.4.1. Vienna. <https://www.R-project.org/>.
- Vu et al. 2010: *Vu, N.T., T. Schlippe, F. Kraus, T. Schultz*. Rapid bootstrapping of five Eastern European languages using the rapid language adaptation toolkit. – In: *Proceedings of Interspeech*, 2010, pp. 865–868, doi: 10.21437/Interspeech.2010-292.
- Whiteside, Varley 1998: *Whiteside, S. P., R. A. Varley*. Dual-route phonetic encoding: some acoustic evidence. – In: *Proceedings of the 5th International Conference on Spoken Language Processing (Sydney)*, pp. 3155–3158.

✉ Проф. д-р Бистра Андреева
Институт по езикознание
и езикови технологии
Саарландски университет
66123 Саарбрюкен, Германия

✉ Prof. Dr. Bistra Andreeva
Department of Language Science and
Technology, Saarland University
Building C7.2, Room 5.02
66123 Saarbrücken, Germany

✉ Д-р Митко Събев
Институт по езикознание
и езикови технологии
Саарландски университет
66123 Saarbrücken, Germany

✉ Dr. Mitko Sabeв
Department of Language Science and
Technology, Saarland University
Building C7.2, Room 4.04
66123 Saarbrücken, Germany