

ДИГИТАЛНИ РЕСУРСИ ЗА ИЗСЛЕДВАНЕ НА ФОНЕТИЧНИТЕ ОСОБЕНОСТИ НА УСТНАТА РЕЧ

Нина Чочева (Пловдив, България)

Обект на изследване е съвременното състояние на съществуващите корпуси на българска разговорна реч, а целта – представяне на новосъздадени електронни ресурси за проучване на фонетичните особености на устната реч.

Съвременните тенденции във фонетичните изследвания се изразяват в описание на националните езици в синхронен аспект и в създаването на корпуси с разговорна реч в електронна форма. Описанието на актуалното състояние на езиците изисква събиране и анализиране на речева продукция. Оказва се, че в сравнение със статичното състояние на писмения език това е нелека задача, защото устната форма се ръководи от различни закони и динамика, които затрудняват нейната обработка и систематизиране. Освен това при съставянето на езиков корпус лингвистът се нуждае, от една страна, от технически умения за работа със звукозаписни средства, а от друга – от базови теоретични знания по физика за характеристиките и разпространението на звука. Голяма част от езиковедите отбелязват, че това е основната причина за все още малкия брой проучвания в тази област.

Корпуси на устна реч на СБКЕ

За разлика от езици като английски, френски, испански, които имат натрупан опит в събирането и съхранението на устната форма в електронен вариант, за книжовния български език този интерес датира едва от края на XX век. Р. Русинов (1995) свидетелства за основания през 90-те години на XX век във Великотърновския университет „Св. св. Кирил и Методий“, на основата на богата фонотека, Архив по българска разговорна реч (БРР). По-амбициозната цел е този архив да прерасне в Център по изучаване на БРР. Затова Русинов определя като първостепенен въпроса за усъвършенстване на методиката за събиране и обработване на речта, както и за изра-

ботване на специфичен понятиен апарат, който да е съобразен с особеностите на тази езикова формация.

Със съжаление се налага да отбележим, че в българското езикознание интересът към процеса на дигитализиране както на устната, така и на писмената форма на родния ни език датира от последното десетилетие и се свързва с членството на България в многоезичния Европейски съюз. Все още не съществува богат референтен корпус на съвременния български книжовен език (СБКЕ), въпреки че в БАН и в Института за български език се наблюдава стремеж да се изготвят такива. Осъзната е нуждата СБКЕ да има своето присъствие в дигиталната информационна среда чрез различните компютърни програми и софтуерни приложения. Все още не са създадени достатъчно големи корпуси на българския книжовен език (БКЕ), които да послужат на специалистите при описание на *всички равнища на езика*. И все пак проекти и инициативи не липсват. Но създадените електронни ресурси са *представителни основно за писмената форма на съвременния български език*. Това са Синтактичен корпус на българския език¹ и Български национален корпус² (БНК). В текстовия архив на БНК само 2,65% от материалите (лекции, парламентарни дебати и субтитри) представят устната форма на СБКЕ. 97,5% от текстовете са ексцерпирани от интернет, а останалата част е предоставена от автори и издатели. Както беше посочено, тези текстове отразяват писмената форма на българския език, която поради статичността си по-лесно подлежи на обработка и анализ. Създаването и обогатяването на БНК несъмнено е важен момент в българската наука, но според нас усилията трябва да са насочени и в посока на съхранение и описание на устната форма на БКЕ. Налага се извоят, че все още интересът към подобни инициативи е незначителен.

Първите корпуси на БРП (Тишева, Джонова 2010), публикувани в интернет, са резултат от работата на чужденец, Хетил Ро Хауге, който пръв събира и публикува материали от проведени изследвания на няколко автори³. По отношение на тези данни Тишева отбелязва,

¹ <http://www.bultreebank.org/>

² http://www.ibl.bas.bg/BGNC_bg.htm

³ Корпус, събран от Красимира Алексова във връзка с изготвянето на дисертацията ѝ; корпус от РР, събран от Цветанка Николова през 70-те години на ХХ век и използван за съставянето на Честотен речник на българската

че „те в голямата си част вече са с исторически характер ... и не отразяват най-актуалното състояние на най-динамичната форма на българския език“ (Тишева 2006: 281).

Към настоящия момент съществува само един по-сериозен корпус на българска разговорна реч (BGSpeech⁴), чиято цел е създаването и поддържането на база от данни, представяща устните форми на СБЕ, както и нейното обогатяването „с данни за конкретни лингвистични проучвания в областта на **морфологията и синтаксиса**“ (Bold мой, Н. Ч.) (Тишева, Джонова 2010⁵). Както става ясно, анализът на фонетичното равнище на устната разновидност на СБКЕ не е приоритет за съставителите на този корпус, който всъщност съдържа транскрибирани текстове на аудио-и видеозаписи на устно общуване „в различни комуникативни ситуации с участието на различни по статус носители на българския език“ (Тишева, Джонова 2011:55). Прави впечатление обаче, че макар и със спазени специфични правила за транскрипция на диалози в спонтанна ситуация (Атанасов 2006), текстовете в рубриките⁶ Транскрибирана разговорна реч (2001–2004) и Транскрибирана устна реч не са придружени със съответния звуков файл. Такъв е наличен единствено в раздела Мултимедиен корпус от устна българска реч⁷.

Налага се изводът, че в разгледаните корпуси на СБКЕ **описанието на фонетичните особености на речта е най-слабо застъпено**. Липсва също така **представителна база от данни с устна реч**, която да послужи за прецизиране, осъвременяване и обогатяване на правоговорната норма на СБКЕ. Съставената от нас база с речта на професионални говорители (радиожурналисти), чието представяне следва, е една от стъпките за запълването на тази липса. Може също така да се приеме, че бъдещо направление на тази **отворена база** от дигитални ресурси е нейното обогатяване чрез словото не само на професионални говорители, а и с това на българи в раз-

разговорна реч; корпус от транскрибирани от Иванка Мавродиева текстове от дебати в Седмото велико народно събрание от 1990 г., достъпни на адрес <http://www.bgspeech.net/bg/links.html>.

⁴<http://bgspeech.net/index.html>

⁵ <http://www.slav.uni-sofia.bg/lilijournal/index.php/bg/issues/summer2010>

⁶<http://www.bgspeech.net/bg/resources.html>

⁷<http://www.bgspeech.net/bg/resources/multimediacorpus.html>

лични комуникативни ситуации с цел проучване на суперсегментните и фоностилистични аспекти например.

Радиожурналистите се приемат за представители на нормативната реч и това е основната причина, поради която се анализират звукозаписи с техния говор. И на второ място, заради отличните технически характеристики на радиозаписите и възможността да се използват многократно. Фонетичното равнище на речта на професионалните говорители не е скорошен обект на проучване в българските изследвания (Холиолчев 1974; Тилков 1974; Русинов 1998; Влахова 2001; Станева 2001; Йотов 2009), но се налага изводът, че сегментните характеристики (стойностите на първи и втори формант) на речта *на говорители от женски пол* са недостатъчно проучена област. Известно е, че първите данни за формантната структура на гласните, произнесени от *българки*, принадлежат на Д. Тилков (Tilkov 1970), т.е. не са актуализирани повече от четири десетилетия. Затова целта на проведения експеримент е, от една страна, създаване на отворена база данни с речта на жени⁸ за проучване на акустичната структура на българските гласни в силна позиция (но не само) и осъвременяване на данните на Д. Тилков, от друга.

Описание на експеримента

Във връзка с проведения експеримент в продължение на една година бяха наблюдавани изговорните стилове на българските радиожурналистки от програма „Хоризонт“ на Българското национално радио (БНР). Избрахме това радио в качеството му на национална медия с установени традиции и професионализъм в представянето на обективни новинарски емисии⁹. Водещите четат предварително подготвен текст с равномерен тон, с умерено темпо, с ясен изговор на думите, без паузи и емоционални окраски, изреченията са кратки с ясно разграничени начало и край. Именно тези характеристики на речта се оказаха подходящи за акустичен анализ на сегментното равнище на речта. Беше анализирана речта на девет български жур-

⁸ Която в бъдеще да се разшири с речта на мъже в различни комуникативни ситуации.

⁹ Изследователски интерес би представлявало проучването на изговорните стилове на журналисти от регионалните клонове на БНР, например от Видин и от Благоевград или от Пловдив и от Шумен.

налистки от програмата „Хоризонт“ на БНР, чиято средна възраст е 47 години.

Дигиталните студийни записи на програмата „Хоризонт“ бяха изтеглени с програмата RadioSure¹⁰ във формат MP3 и конвертирани във формат WAV с програмата Audacity¹¹, с която се осъществи и сегментирането на текста на подходящи фонетични думи. Акустичният анализ (първи и втори формант) на българските ударени гласни се проведе с програмата Speech Analyzer¹².

Емпиричният материал се състои от звукови файлове, в които подбрахме подходящи речеве сегменти, съдържащи българските гласни в силна позиция. Вокалите не са анализирани при следните условия:

– В сложни думи с повече от едно ударение, например *зеленчукопроизводители, европарламента*, тъй като те имат главно и второстепенно ударение, които са с различна сила;

– В акцентно-ритмични единици, които съдържат аббревиатури, например *МВР, ГЕРБ*;

– В позиция пред или след съгласна Й, тъй като при нейното свързване с гласни, се образуват „*съчетания от дифтонгичен тип*“ (ГСБКЕ 1998:127), например *юни, полицай, крайната*.

Звуковите файлове с речта на жени са организирани и управлявани от компютърната програма SARP¹³, която е разширение на Speech Analyzer. SARP има собствена йерархична организация, включваща библиотеки и колекции. Библиотеките посочват езика (BG, FR, UK, US), а колекциите – нормативната реч на естествените му носители. Към настоящия момент библиотеката за български език е съставена само от идиолекта на говорител от мъжки пол. Следователно актуална е необходимостта от обогатяването ѝ с речта на жени, но и на мъже, представители на образцовата устна реч. Вследствие на проведения експеримент библиотеката за български език бе обогатена с произносителните стилове на девет жени.

Ето каква е организацията на колекциите с женски гласове. В библиотеката за български език (BG) на програмата SARP се съз-

¹⁰ <http://www.radiosure.com/>

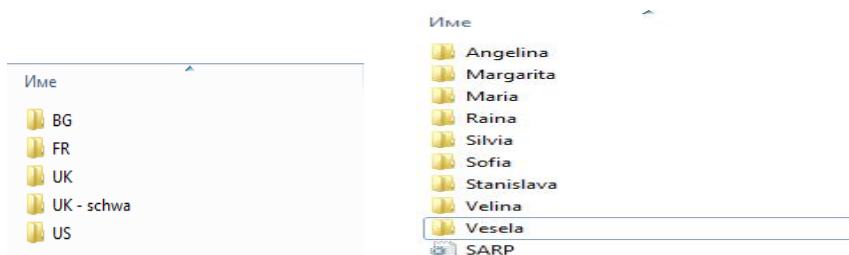
¹¹ <http://audacity.sourceforge.net/?lang=fr>

¹² <http://www-01.sil.org/computing/sa/index.htm>

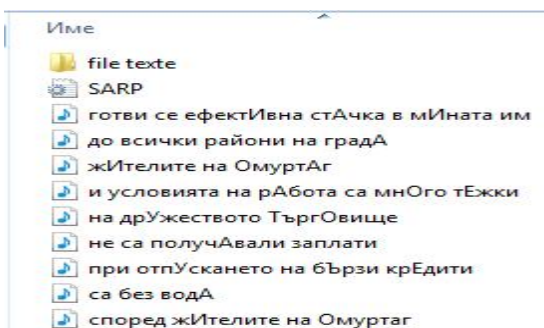
¹³ <http://web.uni-plovdiv.bg/rousni/sarp/>

дава нова папка (колекция) с името на говорителя (фигура 1), която съдържа анализираният звуков файл (фигура 2). Главната буква в наименованието на речевия сегмент посочва кой е изследваният звук в силна позиция (фигура 2).

Фиг. 1. Организация на базата данни на програмата SARP. На първо ниво (вляво) се разполагат библиотеките, на второ ниво (вдясно) – колекциите с имената на говорителите



Фиг. 2. Организация на звуковите файлове в библиотеката за български език – с главна буква е означен изследваният гласен звук



Вследствие на проведените с програмите Speech Analyzer и SARP акустичен анализ се приведоха нови данни за формантните стойности на българските вокали в силна позиция, произнесени от жени, а именно:

	и	е	а	о	у	ъ
F1 (Hz)	421	609	782	641	414	485
F2 (Hz)	2314	1969	1641	1259	1159	1540

В заключение може да се обобщи, че след осъществения експеримент:

1. Се постави началото на отворена база данни с нормативна устна реч на жени от български произход, предназначена за изследване на фонетичните особености на разговорната реч. Тези дигитални ресурси могат да се обогатяват както с данни за гласове на мъже, така и за други езици.

2. Се актуализираха единствените за българския език данни на Д. Тилков (Tilkov 1970) за акустичната структура на българските вокали в ударена позиция, произнесени от жени.

3. Се обогати количествено и качествено колекцията от данни за устна българска реч на програмата SARP.

4. Събраният емпиричен материал може да послужи за сравнителни изследвания в областта на вокализма, както и за проучвания на прозодичните характеристики (честота на основния тон, интензитет, дължина) на устната реч, а също така и за изследване на гласните звукове в неударена позиция по същата методология. Освен това звуковите файлове могат да се използват от чужденци, изучаващи български език.

БИБЛИОГРАФИЯ

Атанасов 2006: Атанасов, А. Проблеми при създаването на езикови корпуси с транскрибирана българска разговорна реч. // Научни трудове Пловдивски университет „Паисий Хилендарски“, том 44, ФИЛОЛОГИЯ, кн. 1, сб. А, с. 289–296.

Влахова 2001: Влахова, Р. Влияние на разговорната реч върху езика на медиите. // Проблеми на българската разговорна реч. София, с. 33–37.

Граматика на съвременния български книжовен език. (Том 1. Фонетика). Велико Търново: Абагар, 1998.

Йотов 2009: Йотов, Ц. Говоренето в условията на нашите съвременни радиопредавания. // Езиковедски изследвания в чест на чл.-кор. проф. д-р Тодор Бояджиев, проф. д-р Венче Попова и проф. Петър Пашов. София, с. 396–402.

Русинов 1995: Русинов, Р. Трета национална конференция по проблеми на разговорната реч. // Проблеми на българската разговорна реч. Велико Търново, с. 5–16.

Русинов 1998: Русинов, Р. Относно факторите и пътищата за по-добро овладяване на книжовната разговорна реч. // Проблеми на българската разговорна реч. Велико Търново, с. 7–1.

Станева 2001: Станева, Х. Книжовно-разговорната реч и медийното пространство. // Проблеми на българската разговорна реч. София, с. 226–232.

Тилков 1974: Тилков, Д. Правоговорната норма и изговорът в телевизионните предавания. // Проблеми на българската книжовна реч. София, с. 38–45.

Тишева 2006: Тишева, Й. Българската разговорна реч в интернет. // Научни трудове на Пловдивски университет „Паисий Хиленадски“, том 44, ФИЛОЛОГИЯ, кн. 1, сб. А, с. 277–288.

Тишева, Джонова 2011: Тишева, Й., М. Джонова. Корпус с устна българска реч – специфика и структура. // Български език, № 3, с. 54–71.

Холиолчев 1974: Холиолчев, Х. За речта на говорителите, редакторите, репортьорите и кореспондентите на радио София. // Проблеми на българската книжовна реч. София, с. 30–37.

Тишева, Джонова 2010: Тишева, Й., М. Джонова. *Електронни ресурси за българската разговорна реч*. Изтеглено 2013 от <http://www.slav.uni-sofia.bg/liljournal/index.php/bg/issues/summer2010>.

Tilkov 1970: Tilkov, D. Le vocalisme bulgare. Les mouvements articulatoires et leur effet acoustique dans la formation des voyelles bulgares. Paris: Klincksieck.

DIGITAL RESOURCES FOR THE STUDY OF PHONETIC FEATURES OF SPOKEN LANGUAGE

Nina Chocheva (Plovdiv, Bulgaria)

The existing corpora of Contemporary Bulgarian Language are intended to analyze the morphological and syntactic features of the language's oral form. The description of the phonetic level is the least studied. The created database of Bulgarian Speech with the program SARP represents an initiative to fill this lack.