

Хетил Ро Хауге

Йовка Тишева

Марина Джонова

BGSPEECH – АКТУАЛИЗИРАНЕ НА ДИЗАЙНА

Kjetil Rå Hauge

Yovka Tisheva

Marina Dzhonova

REDESIGNING BGSPEECH



Abstract: The aim of the article is to describe the implementation of new technologies in the way the data in the Corpus of Spoken Bulgarian are represented. These data are reorganized in order to make explicit the metadata in transcriptions and to use this information to optimize the way researchers can use the data. The complete representation of the data also includes improvement of the advanced search option in publications on the specifics of the Corpus and spoken Bulgarian.

Keywords: Corpus of Spoken Bulgarian, metadata, transcriptions.

Ресурсите за устната реч са предизвикателство за съвременните лингвисти, но и необходимост при проследяването на динамиката на съвременния български език. Поддържането на актуални ресурси за българската устна реч е свързано с постоянното обновяване и допълване на ресурса с нови данни и в същото време с оптимизирането на механизмите за извличане на данни. С тази цел нашият екип се насочи към обновяване на ресурсите, достъпни на сайта за българска устна реч bgspeech.net, с оглед съвременните интернет технологии за визуализация и търсене.

Организирането и класифицирането на данните за устната реч е свързано с подход, съобразен с особеностите на устната

реч и с избраната система за транскрибиране. Както посочват Adolphs и Knight (Adolphs & Knight 2010: 39), за да може да се анализира устната реч, данните за корпуса трябва първо да се съберат, да се транскрибират и да се категоризират така, че изследователите да могат да правят специализирани търсения в него. Устната реч може да се дефинира като система от вербални средства, използвана за устно официално и неофициално, подготвено или спонтанно общуване по различни теми и в различни речеве ситуации. Характеристиките на устната реч са водещи при организирането на ресурсите за нея. Като се придържаме към класификацията на българската устна реч, която се проявява в няколко форми – книжовна разговорна реч (отличаваща се с пълнен произносителен стил), разговорна реч (отличаваща се с различна степен на отдалечаване от пълния произносителен стил), териториални и групови диалекти, екипът се стреми да поддържа данни за устната реч, отличаваща се с непълнен произносителен стил и с редица морфологични, лексикални, синтактични и прагматични маркери, доколкото това е най-динамичната сфера на употреба на устната реч. Макар да се наблюдават някои диалектни особености в речта на част от говорещите, диалектите, както и груповите говори в значителна степен се отличават по сферата си на употреба от разговорната реч и поради това не са специален обект на изследване и отразяване в корпуса.

Sinclair (Sinclair 2005) посочва десет основни принципа, които трябва да се следват при създаването на корпус от устна реч, свързани с подбора и структурирането на данните в корпуса, с неговата представителност и с балансираното представяне на явления в корпуса, както и с отношението между транскрипцията и метаданните към нея. Водещият подход при събирането и структурирането на данните от устната реч е тяхната представителност. Транскрибираните записи, достъпни на сайта, са предимно от *консултативния и ежедневния* регистър, отличаващи се съответно с частично подготвено или спонтанно общуване. Не се включват записи от т.нар. *застинал* регистър, тъй като при него е налице висока степен на официалност и максимална подготвеност на речевото общуване. В ресурса са включени данни от официалното и неофициалното общуване, като са налични данни от академичното общуване, от бизнес общуването, както

и от битовото общуване. Различните речеви ситуации са свързани с неподготвено и частично подготвено общуване с минимум двама участници. Пълнотата на представените речеви ситуации предполага използването на ресурса в различни изследвания – при изследване на изговорните варианти, лексикалните маркери, морфологичните и синтактичните маркери, при изследвания в сферата на речевите актове и прагматиката. Докато при писмената реч може да се говори за различни стилове, при устната реч се говори за регистри (вж. Тишева 2012: 206–207).

При създаването на корпус от устна реч от важно значение е и неговата достъпност и възможността за свързване на търсенето на езикови особености с характеристиките на речевата ситуация, в която е осъществен оригиналният запис. Именно затова на този нов етап от обработването на ресурсите в сайта VgSpeech се насочваме към експлицитното категоризиране на типовете ресурси с възможност за ограничаване на търсенето от страна на изследователите. При транскрибирането на записите метаданните към всяка транскрипция включват информация за самия запис – време и място; типа речево взаимодействие – диалогично, монологично, брой участници; за характеристиките на самите участници – възраст, пол, месторождение, образование; за темата на разговора; за речевата ситуация – официална, неофициална и т.н.

В същото време корпусът съдържа няколко типа ресурси по отношение на съдържанието на ресурса и технологията, с която е обработен и публикуван – транскрипции от периода 2001–2004 г. в текстов формат; транскрипции, направени в xml формат с възможност за автоматична обработка на ресурса и извличане на данни; мултимедиен корпус, в който транскрипциите са в xml формат и са синхронизирани с оригиналния звуков файл. Тук накратко представяме тези ресурси и подхода, който сме избрали при оптимизирането на възможностите за търсене в тях и при избора на единна система за запис на метаданните.

Най-ранните транскрипции, достъпни на сайта, са представени в текстов формат, метаданните за тях също са в текстов формат и няма възможност за автоматично търсене в тях. Основната причина за това е концепцията за транскрибиране на записите с подробно отбелязване на фонетичните особености на речта (вкл.

на нормативните редукции, асимилации и нормативното ударение). Това, разбира се, е важно за изследователите на фонетичните особености на устната реч.

```

# Номер на записа: 0026
# Обратно наименование на записа: интерв
# Транскрибиран: " ... "
# Дата на транскрипцията: 26.02.2004
# Транскрибирана част от целия запис: 00:03:13
# Допълнителни забележки: Участниците са означени с буквите от латинската азбука: A, B,

A: ъ // щелий \този процес преи́мването на \визите / к'а'го спец\ално за \визите подчер\та че то'ва на\ли е едно // (4С) е\дин е\вал
B: да //а// к'ам осмесе\аване на \пълното \право на сво\ободно при\движване на \хора и хали\гали // ъ // на\ли то'ва \право в \шката му п'лно\ка не \блде осмесе\авно то'ваа ко\гато Бя\галия се присъеди\ни към Евро\пейски\иот съ\юз /
B: да
A: к'а'го ос\новното ми възра\жение то'ваа не \блде възможност\а \български специа\листи да се консу\лират за ра\ботни мес\та в Ев\ропа //
B: то'ва чак след присъеди\ването ми то'й по \е \казал че не \стане?
A: присъеди\ването към Евро\пейски\иот съ\юз
B: добре
A: \зачи обясни ми ъ // (4С) на тях спец\ално ми обяс\ни едно \нещо ко'ето \вчера съ го\ворихи / и ко'ето е обяс\нил на гер\манската пос\олница Обригит // ъ / ко'ято го по\питала по к'а'ква \начин на\пример Бя\галия \може да \спре еми\грацията (4С) из\тичането на \мозъци
B: да
A: и то'й \казал два \с \мачините // (4С) \тоест не два \с \мачините но на\ли // \както се \казва им: // не два \мачина \отделил / а два \фактора \могос да с\мажат въз\яание \право \барзо (4С) \барзи \преговори с Евро\пейски\иот съ\юз //
B: да
A: и присъеди\ването // при ко'ето разбира се не \блде при\ет ста\ндарт за\щото по\добно \нещо се наблюдава и в \други стра\ни И\спания / Порту\галия / \начи \маден \пример че Порту\галия пре\ди присъеди\ването към Евро\пейски\иот съ\юз е \имала еми\грацион\на вол\на от три мили\она \души
B: да
A: док'а'то три\найсе мили\она население три мили\она \души / порту\галици се на\дуснали Порту\галия к'а'то съ\гласили \место за \работа \някъде в Ев\ропа
B: да

```

Пример 1.

За да се улесни автоматичното търсене в този ресурс, бяха създадени нормализирани по отношение на правописа версии на транскрипциите. Така изследователите имат достъп до паралелен корпус, в който в зависимост от целта на своето изследване могат да ползват две версии на транскрибираната реч – с подробно отбелязани фонетични особености и с нормализирана фонетика. В оригиналната транскрипция обаче са запазени само фонетичните особености, които са отклонения от книжовния изговор. Тези отклонения са индексирани и отбелязани с червен цвят в текста. При така оформения паралелен корпус е възможно търсенето по два индекса – индекс на отклоненията от книжовния изговор и индекс на думите, употребени в текстовете, като е налице и информация за броя употреби и препратка към всяка от тях.

Индекс/Index	Редактиран/Edited	Оригинал/Original	List of non-standard forms/pronunciation
Click on any wordform after the colon to see it in its context	To return to the index click the word highlighted from the search	Words that are normalised in the column to the left are marked in red	Click to see the form in context in the column to the left
: а: а а а а а а а абе: абе абсолютен: абсолютно абсолютно абсолютно абсолютно аз: те ти аз ти аз аз ти то ти ти ние то те ти аз аз ти ти аз то аз то аз аз ти вие аз ти ти аз аз те то аз ти ти аз то ти тя аз аз ние ти аз той аз аз те ти аз аз той те	Б: и съм съгласен че не става обаче какво общо има вевецето с обектно и принципите за обектно ориентирано програмиране В: не бе мен мен принципите си ме кефят само че не ме кефи как са го измислили във вевецето	Б: [и съм съгласен че ни става / обачикво общо има вевецето с обектно и принципите за обектно ориентирано програмиране , В: не бе / мен мен принципите си ме кеф'ът само че не ме графи как са го измислили във	--] / document framework leko li li li li microsoft microsoft windows а а абсо'лутен абсо'лутен абсо'лутно аз азз ама'егабайта ама'ен б'ах б'езспорно б'еши В'енци влезе вид'ах вий вр'еми вре'то вч'ера gleam glei gol'ama gol'ами говора гол'еми графи графи графи п'еси да

Пример 2.

По отношение на транскрипциите, оформени като xml файлове, беше предприет различен подход. Тези транскрипции включват два типа ресурси – Корпусът от разговорна реч, създаден след 2004 г., и Мултимедийният корпус.

В Корпуса от разговорна реч транскрибираните аудио-записи са представени в писмен вид като поредица от реплики на отделните говорещи. Лингвистичните и паралингвистичните особености на устната реч (напр. елизия, елипса, съкратени форми или фрази, ненормативно ударение, ненормативен изговор, паузи, застъпване на реплики, незавършени изказвания, повторение на фрази, разговорни конструкции, прагматични и дискурсни маркери; жестове, мимики, фонетични паралингвистични средства, като смях) се представят според изискванията на TEI. Метаданните към транскрипциите са анотирани, но не се визуализират при представянето на транскрипцията.

Транскрипция
Участник 1: така [кратка пауза] дай да си говориме как ще се напиеме във събота [средна пауза]
Участник 2: аз на свети Валентин няма намерение...
Участник 1: аз не празнувам свети Валентин аз празнувам Трифон [кратка пауза] отрезан
Участник 2: аз ги празнувам и двете заедно [кратка пауза] не бе [кратка пауза] не [кратка пауза] аз ако се налия...
Участник 1: аз ти знаеш ли аз ако се налия ще ги пребия и двамата [кратка пауза] Огнян и Илиан [кратка пауза] беборанка [средна пауза]
Участник 2: тя как ѝ е името?
Участник 1: ѝлъ
Участник 2: как ѝ е името? [средна пауза]
Участник 1: о са го казах преди малко [средна пауза] със съ беше нещо [средна пауза]
Участник 2: Соня [средна пауза]
Участник 1: не бе [средна пауза]
Участник 2: Силвияааа [средна пауза] Пенка [средна пауза] Петканка Азербейджинова
Участник 1: да: Црепуника Црепунджинова [средна пауза]

Пример 3.

Подобна е и структурата на Мултимедийния корпус на българската устна реч. В този тип ресурс транскрибираният текст на записа е достъпен едновременно с оригиналния звуков файл. Създаването на този ресурс започва от 2009 г. и е свързано с новите възможности за публикуване на данни от устната реч. Корпусът е регистриран и достъпен през Метанет (www.meta-net.eu) от 21.01.2013 г. При създаването му екипът ни се насочи към използването на програмата EXMARaLDA (<http://exmaralda.org>), разработена в Хамбургския университет, тъй като програмата е насочена основно към отразяването на речевите особености и синхронизирането на транскрипцията със съответния аудиозапис, но позволява и отбелязването на паралингвистичните особености, както и на особености, свързани с речевата ситуация. Метаданните са част от самия файл и позволяват класифицирането на записите и търсенето в тях. Според определението на Schmidt et al. (2010: 2–3) разглежданият ресурс е от типа мултимедийни корпуси, чиято основна цел е представянето на особеностите на устната реч, но отразява и паралингвистиката, както и елементи от речевата ситуация. Записите са от автентични диалози. Речта се предава писмено според специално модифицирана система за транскрипция, в която се отчитат особеностите на българската

устна реч и се отбелязват някои от невербалните елементи, като мимики и жестове, но предимство се дава на речевото общуване. В съответствие с целите на BgSpeech създаването на този ресурс не е свързано с изследването на дадено явление в устната реч, а по-скоро цели да представи в цялост актуалното състояние на българската устна реч, като по този начин да обогати съществуващите ресурси за българския език и да даде възможност за по-голяма пълнота в емпиричния материал, използван при изследването на българския език.

Транскрипция

16:58 / 22:58
Ако вашият браузър не поддържа аудиоформата, моля, използвайте Chrome.

Щракнете върху звездичка след номер на елемент от текста, за да стигнете до съответното място в аудиозаписа.

[1]	[1] [1]
Михаил Дюжев [m]	[inhalation] добро утро уважаеми зрител и слушатели все сте със съвременния блок на[ro:]
[2]	[1] [1]
Михаил Дюжев [m]	Двама радио "България" и със "Денят започва" на[ro:] Българската национална телевизия
Михаил Дюжев [m]	железница
[3]	[1] [1]
Михаил Дюжев [m]	<short> аз съм Михаил Дюжев Андрей съм сигурен че вече <short> може би не е представил
Михаил Дюжев [m]	доказващо реч извади

Пример 4.

С цел да се улеснят достъпът и цитирането на данните от ресурсите, при реструктурирането на сайта транскрипциите от Мултимедийния корпус и от Корпуса разговорна реч се представят и в текстов формат.

тема	професионална
------	---------------

Транскрипция

Изчистен преглед

тече реклама аплодисменти

Росен Петров: уважаеми зрители всеки момент при мен ще гостува един певец за когото много може се говори[] наричат го Граф но какво е потеклото му ще разберам след малко вижте за кого ви говоря и така пожелахте да говори Влади Амлов-Графа добре дошъл радвам се да те видя / изглеждаш както винаги добре разбрах че сега и на турне шъ ходиш[] във лятото август месец с &q ;Комиците&q ;

аплодисменти

тече песен и видеоклип

аплодисменти

Влади Амлов: благодаря благодаря и ти също

аплодисменти

Влади Амлов: уникално е искам много да благодаря на &q ;Комиците&q ; за това че[] ни поканиха мен и група &q ;Акага&q ; да се включим / да се включим към това събитие защото[] наистина това което знам от във миналите години е че[] / да тези турнета са много успешни и пак казано за нас е чест че ще бъдиме към този екип и ще си прекарваме освен страхотно ще предложиме и хубави музикални изненади още[] обмислихме програмата но със сигурност едно от парчетата ще е &q ;Невидим&q ; тъй като така това / да най-голямото парче което съм имал ами аз и миналия път когато ти гостувах / да че началото на успеха е едно такова нещо[] нали[] es } / чак такова нещо не очаквахме добре

Пример 5.

Метаданните към транскрипциите бяха обобщени. При подбор на метаданни, които да се визуализират и да могат да се използват при търсенето, беше избран подход, който да е в съответствие с принципите на TEI за публикуване на корпуси от разговорна реч. Отчасти тези принципи съвпадат и с класификацията на данните в писмените корпуси, напр. сферата на употреба, но голяма част от критериите са специфични за данните от устна реч.

На първо място може да се търси само в един от ресурсите в зависимост от това дали изследователят иска да изслуша и оригиналния запис на речта. Вторият критерий е сферата на употреба. Публикуваните транскрибирани записи са от следните сфери: медийна, академична, битово-разговорна, публична/политическа, делова.

Търсене	2013001	2013002	2013003
<input type="text" value="Търсене за ..."/> <input type="button" value="Търси"/>	Мултимедиян корпус	Мултимедиян корпус	Мултимедиян корпус
Тип	2013004	2013005	2013006
<input type="checkbox"/> Паралелен корпус <input type="checkbox"/> Мултимедиян корпус <input type="checkbox"/> Транскрибирана усна реч	Мултимедиян корпус	Мултимедиян корпус	Мултимедиян корпус
сфера	2013007	2013008	2013009
<input type="checkbox"/> медийна <input type="checkbox"/> академична <input type="checkbox"/> битово-разговорна <input type="checkbox"/> публична/политическа <input type="checkbox"/> делова	Мултимедиян корпус	Мултимедиян корпус	Мултимедиян корпус
	2013010	2013011	2013012
	Мултимедиян корпус	Мултимедиян корпус	Мултимедиян корпус
	2013013	2013014	2013015
	Мултимедиян корпус	Мултимедиян корпус	Мултимедиян корпус
	2013016	2013017	2013018
	Мултимедиян корпус	Мултимедиян корпус	Мултимедиян корпус

Пример 6.

Следващите критерии са тип общуване – официално или неофициално, и степен на подготвеност – спонтанно, частично подготвено или подготвено общуване. Първите три критерия позволяват да се правят изводи за появата на определени фонетични и морфологични маркери в определена речева ситуация, напр. в медийната реч при наличие на частично подготвено общуване или в деловата сфера в неофициална ситуация при частично подготвено общуване. Тъй като при транскрибирането на записите е отбелязвана и темата на разговора, обобщихме тази информация, така че тя да е значима при търсенето в целия ресурс. Темата на разговора може да спада към една от следните категории – лична, битова; професионална, обществено значима; професионална. Данните на самия разговор посочват населеното място, в което е направен записът (когато са налице такива данни), формата на общуването – диалогично или монологично, както и колко са участниците в разговора. Третият тип характеристика, отбелязвана в данните за транскрипциите, е свързана с данните за самите участници – тяхната възраст, пол и наличието на познатост между участниците. Възможно е добавянето и на други характеристики, например професия, като те се визуализи-

рат само ако в избрания корпус има данни за професията на участниците. Разбира се, в медийните текстове е от значение отбелязването на професията на гостите, а не на самите журналисти, тъй като това вече е маркирано с типа текст. Така е възможно да се правят обобщения за начина на общуване с други журналисти, със социолози, с артисти, с учени. Във всяка транскрипция се визуализират наличните метаданни.

тип	Мултимедийен корпус
тема	професионална, обществено значима
формат	диалог
сфера	медийна
тип	официално
град	София
подготвеност	частично подготвено
брой участници	3
възраст	
участници	мъже
социална дистанция	общуване между непознати

Пример 7.

Публикуването на ресурсите в два формата – аотиран и изчистен текстов формат, както и експлицитното класифициране на транскрипциите от Корпуса от устна реч и Мултимедийния корпус по наличните метаданни позволява ползването на данните от широк кръг изследователи. Корпусът е предназначен основно за нуждите на лингвистите – за изследване на актуалните тенденции в българската устна реч и за верификация на хипотези в научните изследвания, но може да се използва и като елемент в чуждоезиковото обучение, тъй като аудиозаписът и транскрибираният текст са синхронизирани.

Пълното представяне на българската устна реч е свързано и с представянето на научните изследвания върху нея. Изследователите, преподавателите и студентите могат да се запознаят с пълните данни за транскрипциите, методологията, която е избрана при подбора на записи, които да бъдат представени на сайта,

както и към особеностите, които се отбелязват в транскрипциите. Освен с принципите на създаване и организиране на данните за българската устна реч, изследователите могат да се запознаят и с наблюденията върху особеностите на устната реч, представени от екипа, както и от учени, предоставили своите публикации върху устната реч. Търсенето в този ресурс е улеснено, като публикациите са анотирани по година, автор и ключови думи към публикациите на български и английски. Така например при търсене на думата „частици“ се представят статиите на сайта, свързани с изследването на частиците:

Представените публикации по проблемите на българската разговорна реч и българските диалекти са свързани с цялостното представяне на българската устна реч – като данни от самата устна реч, както и като наблюдения върху нейните отличителни особености. В публикациите може да се търси по ключова дума.

2003

2002

Тишчева, Й. За динамиката в системата на прагматичните частици в Българския език. В: *Lilora Psycholinguistica*. София, Сема РС, 2002. 213-219. [a]

Тишчева, Й., Х. Ро Хаугре. Заемане на прагматични частици. В: *Проблеми на социопсихистиката*, т. 7. Билингвизъм и диплосия – съвременни проблеми. София, Делфи, 2002: 10–17. [a]

2001

Автори

- > Ilijana Krapova (1 публикация)
- > Iskra Likomanova (1 публикация)
- > Kjetil Rå Hauge (1 публикация)
- > Yovka Tisheva (2 публикации)
- > Yovka Tisheva, Ilijana Krapova (1 публикация)
- > Yovka Tisheva, Marina Ozhonova (3 публикации)
- > Атанас Атанасов (2 публикации)
- > Искра Ликоманова (1 публикация)
- > Йовка Тишчева (1 публикация)
- > Йовка Тишчева, Гергана Дичева (1 публикация)
- > Йовка Тишчева, Илияна Кръстова (1 публикация)
- > Йовка Тишчева, Хетил Ро Хаугре (2 публикации)
- > Красимира Алексиева (2 публикации)
- > Мариета Цветкова (1 публикация)
- > Марина Дюнсва (1 публикация)

Пример 8.

На сайта са достъпни данните за публикацията, както и самата публикация или препратка към съответната електронна публикация.

Може да се обобщи, че изследването на устната реч е двустранен процес. Той включва не само емпирични наблюдения и теоретични обобщения, но и поддържането и актуализирането на ресурс с актуални данни от устната реч. Достъпността на ресурсите за българската устна реч от своя страна дава възможност на

изследователите да верифицират своите научни хипотези, както и да наблюдават динамиката на устната реч. Оптимизирането на корпуса от устна реч е направено с оглед представянето на ясната структура на данните с корпуса, подобряване на възможностите за разширяване на ресурса, като се избира модел на представяне на метаданните, приложим при описанието на данни от различни сфери на употреба на устната реч. В съответствие със съвременните принципи за създаване на корпуси от устната реч новите транскрипции в корпуса са достъпни заедно с оригиналния звуков файл. Стремещт на екипа ни е да включва транскрипции, отразяващи речевата ситуация в цялост. Представените в корпуса данни са с независимост на данните от конкретно изследване.

БИБЛИОГРАФИЯ

Тишева, Джонова 2010: Тишева, Й., М. Джонова. Електронни ресурси за българската разговорна реч (инициативата BgSpeech). – В: *Littera et Lingua*, лято 2010. // **Tisheva, Dzhonova 2010:** Tisheva, Y., M. Dzhonova. Elektronni resursi za balgarskata razgovorna rech (initsiativata BgSpeech). – In: *Littera et Lingua*, Summer 2010. <https://naum.slav.uni-sofia.bg/node/1735> [15.01.2021]

Тишева 2012: Тишева, Й. За динамичните модели на диалогичното общуване. – Във: *Време и история в славянските езици. Сборник с доклади от Единадесетите национални славистични четения 19–21 април 2012. Т. I Езикознание*. София: Университетско издателство „Св. Климент Охридски”, 205–212. // **Tisheva 2012:** Tisheva, Y. Za dinamichnite modeli na dialogichното obshtuvane. – In: *Vreme i istoria v slavyanskite ezitsi. Sbornik s dokladi ot Edinadesetite natsionalni slavistichni chetenia 19–21 april 2012. T. I Ezikoznanie*. Sofia: Universitetsko izdatelstvo „Sv. Kliment Ohridski”, 205–212.

Adolphs & Knight 2010: Adolphs, S., D. Knight. Building a spoken corpus: what are the basics? – In: *The Routledge handbook of corpus linguistics*. [eds.] O’Keeffe, A., M. McCarthy. Routledge, 38–52.

Schmidt et al. 2010: Schmidt, T., K. Elenius, P. Trilsbeek. Multimedia Corpora (Media encoding and annotation). Draft submitted to CLARIN WG

5.7. as input to CLARIN deliverable D5.C-3 *Interoperability and Standards*.
http://www.exmaralda.org/files/CLARIN_Standards.pdf [15.01.2021]

Sinclair 2005: Sinclair, J. *Corpus and Text-basic Principles*. – In:
[ed.] Wynne, M. *Developing Linguistic Corpora: A Guide to Good Practice*.
Oxford: Oxbow Books, 1–16.

Text Encoding Initiative (TEI). P5 Guidelines <https://tei-c.org/guidelines/P5/> [15.01.2021]

Сайт за българската устна реч – <http://bgspeech.net>

Система за транскрибиране на устна реч ENMARaLDA – <http://www.exmaralda.org>