

Димитър Попов


Велка Попова

МУЛТИМОДАЛЕН ПОДХОД ЗА
ИЗСЛЕДВАНЕ НА РЕЧТА¹

Dimitar Popov

Velka Popova

A MULTIMODAL APPROACH TO
SPEECH INVESTIGATION



Abstract: The modern concept of multimodality envisages diversity (multiplicity) in the presentation of information channels (auditory and visual) in communication. This fact presents multiple opportunities for a simultaneous and multidimensional interpretation of the variability in an individual's speech in different types of discourse, which could potentially lead to a cross-modal research approach in the future. Multimodal oral discourse (sounding speech) is characterized by a segmental (verbal) component and a multitude of non-segmental (paraverbal, prosodic) parameters which supplement the verbal component in real-time use. The visual channel, on the other hand, contains a series of extra-verbal elements that complement and refine the information conveyed.

Keywords: multimodal linguistics, speech investigation.

„В тона на гласа, в очите и в общия израз се съдържа не по-малко красноречие, отколкото в подбора на думите.“

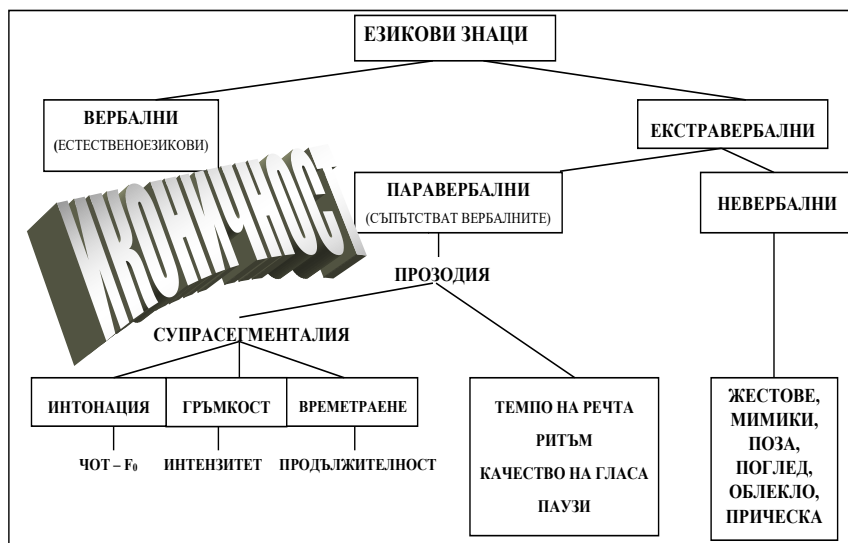
Франсоа дьо Ларошфуко

¹ Това изследване е частично финансирано от Българската национална интердисциплинарна изследователска инфраструктура за ресурси и технологии в полза на българския език и българското културно наследство, част от европейските инфраструктури CLARIN и DARIAH – КЛаДА-БГ, договор ДО1-272/16.12.2019.

Опирайки се на постиженията на съвременните информационни технологии и на създадените мултимедийни програмни пакети, предназначени за постигането на различни цели в отраслите на приложната лингвистика, за съвременния изследовател става възможно илюстрирането и визуализацията на резултатите да бъде представено комплексно и цялостно. За това допринасят и натрупванията в младата *мултимодална лингвистика*, представляваща едно от направленията за едновременно многоизмерно представяне на езиковите факти и явления в рамките на холистичното езикознание. Терминът *мултимодален* се използва от много учени за многоаспектно онагледяване на представата за езика и речта като част от процеса на междуличностната комуникативна интеракция в различни перспективи. Модалността тук не бива да се асоциира с граматическата категория *модалност*. Терминът *мултимодален* в настоящия принос се опира на разбирането за модалност в психологията, неврофизиологията и информатиката, където се тълкува като тип външен стимул, който се възприема с някое от сетивата на човека, на първо място със зрението и със слуха (срв. Кибрик 2010). Оттук произтича и задачата какви ресурси да бъдат подбрани така, че визуалният и слуховият код да бъдат илюстрирани като информационни канали, способстващи за обективното изобразяване на речевите портрети в реално време. При това мултимодалността не означава просто наличие на няколко информационни канала за връзка, тук се има предвид тяхното взаимодействие и интеграцията им. „Мултимодалният подход предполага, че съобщението се разпространява чрез всичките комуникативни модуси. Ако това е така, то всеки модус представлява само частичен носител на глобалното значение на съобщението“ (Kress, van Leeuwen 2001; Kress 2002).

Както беше споменато, понятието „мултимодалност“ се базира на различията в сетивното възприятие и предимно на различията между зрителен и слухов канал. Наред с това, в рамките на всеки от каналите съществуват по-нататък други по-дребни различия, които също се вписват в представата за мултимодалност. Ето защо в мултимодалността на устния дискурс (звучащата реч)

са налични сегментен (вербален) компонент, от една страна, и от друга – множество несегментни (паравербални, прозодични) параметри, които в реално време съпътстват вербалните. Визуалният канал пък включва поредица от екстравербални елементи, като напр. жестикулация, посока на погледа, мимика и други аспекти на „езика на тялото“. Схематичното представяне на знаковете на речта като иконичен модел може да бъде представено таксономично по следния начин:



Фиг. 1. Иконичен модел на езиковите знаци на речта

За целите на изследователския дизайн се използва инструментариумът на информационните технологии и неговото приложение в съвременната мултимодална лингвистика. Информационните технологии се интересуват от модуса (способа, начина, канала) за предаване на информацията.

Модусът (от лат. *modus* – мярка, начин, способ), т.е. каналът за предаване на информация, съдейства, за да бъдат противопоставени различните типове дискурс според характера на тяхната

материална същност и формата им на реализация (срв. Кибрик 2009: 3). Известни са следните типове (модуси) на дискурса:

– устен модус (канал за предаване на информация) – акустичен, с характерна прозодия. Говорещият поражда звуков сигнал с помощта на своя артикулаторен апарат, а адресатът го възприема като акустичен сигнал със слуха си;

– жестов модус, основан на визуалното взаимодействие между говорещия и адресата. Жестовете съпровождат обикновено устната реч;

– писмен модус – визуален, с графични белези. Базиран е на зрителен сигнал в канала за предаване на информация;

– мислителен модус, с помощта на който човек създава вътрешен дискурс, недостъпен за другите (затова е по-малко изследван. Предмет на внимание е предимно на психолози и в частност – в монографията на Л. С. Виготски „Мислене и реч“);

– електронен субмодус. В този предавателен канал на информацията се вписват: телеграмата, радиопредаванията, телевизионните предавания, телефонният разговор, общуването с помощта на пейджъра и телефонния секретар, кореспонденцията по електронната поща във виртуална среда. Типичен пример за съединяване на особеностите на устния и писмения дискурс представлява общуването в режим TALK (или СНАТ) чрез приложенията Viber и Skype, с чиято помощ се реализират аудиално или аудио-визуално разговорите.

Очевидно е, че в тази таксономия устният модус е изходната фундаментална форма за съществуването на езика. Ето защо интересът към неговото изследване дава добри резултати в лингвистичната персонология. Разговорът като типична форма на устната реч представлява особен модус на дискурса. „Разговорният дискурс е важен източник за изследователи, които се интересуват от процеса на конструиране на идентичност, защото в беседата може да се наблюдава както конструирането на идентичност, така и приемането на тази идентичност от другите събеседници. И като добавка – речта е средство за метадискусии

за аспектите на идентичността. Такива метадискусии нерядко се срещат в обичайните спонтанни разговори“ (Кэмерон 2015: 276).

В прагмалингвистичното измерение на дискурса се вписват и неговите тонални характеристики. Комуникативната тоналност на дискурса представя емоционално-стиловия формат на общуването, който се определя от изменящите се нагласи на даден говорещ и от избора на неговите персонални средства за комуникация. Тоналността е съотносима с индивидуалния и уникален тембър на гласа в звучащата реч. М. В. Вербицка например съобщава за пет вида тембрални окраски в английската реч – неутрална, повествователна, лирична, патетична и драматична (Вербицка 2000: 22). Независимо от импресионистичния характер на названията, все пак те отразяват правдоподобно психологическата същност на изразената емоционалност. Възщност емоционалността представлява психологическа окраска на речта, която може да бъде тържествена, неутрална или вулгарна. В този смисъл разглеждана като вид *субективна модалност* обаче (тук конкретно, разбира се, че модалността трябва да се тълкува в нейния граматически смисъл като категория, изразяваща отношението на говорещия към действителността на изказването чрез своите индикатори като наклонение, интонация, частици и др.), наистина отразява отношението на говорещия, а самата дискурсивна тоналност може да има следните разновидности на звучене в речта: официално, сериозно, шеговито, възторжено, дружелюбно, враждебно, саркастично, недоверчиво, безразлично, песимистично, внимателно, замечтано, тъжно, пренебрежително, надменно, агресивно, с отвращение, настойчиво, завистливо, с удивление, развълнувано, хвалебствено, тревожно, раздразнително, с обида, радостно, прикрито, замислено, решително, подбудително, тайнствено, грубо, изплашено, вежливо, ласкаво и пр. От една страна, тези квалификации са израз на емоционално отношение, а от друга – те представляват оформление на речта в определен стилистичен регистър.

Тембърът (фонацията) е инхерентна величина на гласа. В настоящето понятието ГЛАС се използва активно и доста свободно и полисемантично от множество изследователи за целите на анализа на дискурса (срв. Mieroor 2005; Kjær, Palsbro 2008: 279; Slay, Smith 2011). За открояване на звучащите гласове в дискурса е необходимо да бъдат определени действащите лица и да се отчитат особеностите на взаимоотношенията между тях. В широк смисъл и метафорично гласовете се квалифицират като социални агенти на комуникацията, т.е. като личности, които участват явно или скрито в конструирането на идентичности – т.е. те винаги присъстват в дискурса и могат да бъдат представени по различен начин. Гласовете изпълняват регулярно специфични роли в речта и по такъв начин говорещият често конструира своята идентичност (срв. Mieroor 2005: 110–111). Лингвистично това може да се индикира с помощта на дискурсивни маркери за обозначаване на дейксис (от гр. δείξις – показвам, посочвам), а също така – с употреба на модални думи, частици и дори цели фрази, съчетани с определена характерна прозодична огласовка на речта.

Съобщенията в човешката комуникация се предават с думи и звукове, речевите сигнали носят информация за персоналната анатомия, за физиологията, за лингвистичния опит и за менталното състояние на говорещия. Характеристиките на говорещото лице се откриват в неговата реч на всички равнища на описание – от спектралната информация на веригата от звукове до избора на думи и изказвания в определен дискурс. Интерес предизвиква обаче и въпросът за това как речта може да стане носител на информация за характеристиките на говорещия, от една страна, и как – носител на информация за съобщението, от друга. Важно е също да бъдат показани и начините, по които индивидуалността на говорещия може да бъде изразена в речта, и какви са нейните характеристики.

Ако двама говорещи напр. изрекат едно и също изказване, то очевидно ще звучи по различен начин, защото в говорния процес са ангажирани няколко системи – нервна, физиологична, ана-

томична и физическа, а във всяка от тях се откриват специфични за говорещите лица индивидуални особености. Понеже не е възможно две лица да са идентични по физиологията си, характерните различия в техните изговорни системи водят до характеристични разлики в тяхната реч, независимо от факта, че продуцират един и същи израз по смисъл. Предаваните чрез речевия сигнал специфични характеристики, присъщи на дадена говореща личност, могат да информират реципиента и да му дадат възможност да ги опише и да класифицира говорещите по възраст; по произносителен стил; по езика, на който разговарят, както и да разкрие техните емоции, дори здравословното им състояние и пр. Именно по такива налични белези говорещият може да бъде разпознат от слушащия, или казано по друг начин, определени персонални характеристики на едно лице може да се окажат достатъчно физиономични, за да бъде определена и верифицирана неговата идентичност. Тези наблюдения намират необходимото приложение в сферата на сигурността, на съдебната лингвистика, на разследващата журналистика и др. Обединяващият момент тук е да се направи опит за описание на речевите характеристики на дадено говорещо лице, които се изразяват чрез неговия глас.

Немският семиотик Роланд Познер в статията си „Човекът като знак“ пише, че всекидневието на човешкото поведение на различни нива в йерархията на достоверността служи за идентифициране. Всеки човек притежава сума от „лични характеристики на всяко ниво от йерархията; те изграждат неговия специфичен персонален код. Но в повечето ситуации за идентификация на човека се използва съвсем малък обхват от личния код. Обикновено е достатъчен кратък поглед към, да речем, към пуловера на човека или е достатъчно да чуем гласа му. Обикновено се заслушваме или заглеждаме, ако усетим несъответствия (необичайно облекло, дрезгав глас, необичайна походка). В такива случаи обръщаме внимание на постоянните характеристики като физически черти, например“ (Познер 1995: 87). Това наблюдение ангажира вниманието на изследователите да търсят в хода

на анализа характеристикното, физиономичното, уникалното в речевото поведение на човека и още по-конкретно – в неговия глас. Уникалността на индивидуалния човешки глас е неговата невидима „визитна картичка“.

Мултимодалният подход като изследователски дизайн за илюстриране на речевото поведение на човека в реално време дава най-добра, точна и автентична представа за изобразяването и експонирането му като дискурсивен фрейм (или сценарий). И ако сонаграмите, интонограмите и осцилограмите визуализират и обективизират само реализациите на вътрешноприсъщите характеристики на речта (т.е. аудиално, едноканално), то включването и на визуалния канал дава възможност за всеобхватно ситуиране на речевото събитие (дискурса).

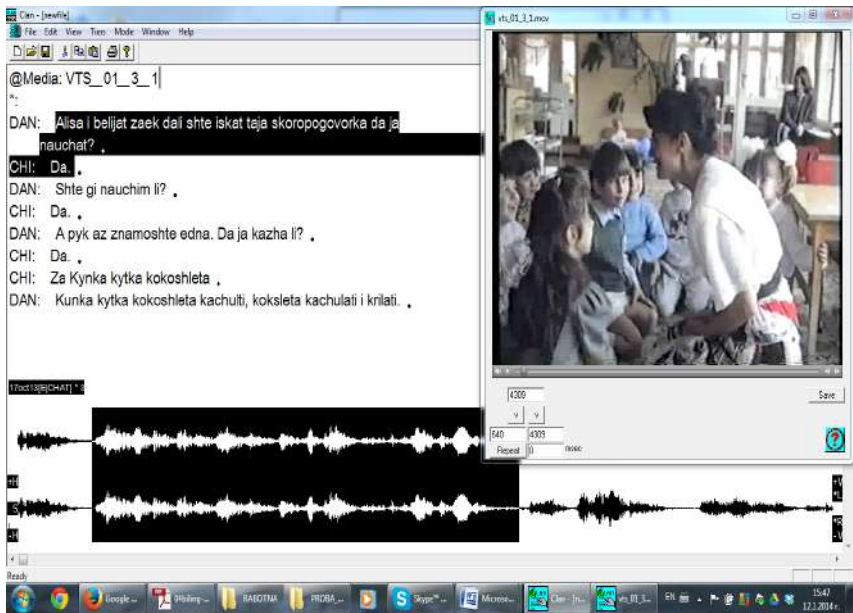
В поредица от екипните приноси на авторите на настоящия доклад, осъществени в Лабораторията по приложна лингвистика на Шуменския университет, е проследено индивидуалното говорно развитие на няколко деца в процеса на усвояването на езика и речта чрез провеждането на лонгитудинални изследвания (срв. Ророва, Роров 2013, 2014, 2015). Като методология за изследването е послужила постановката на Brian MacWhinney (2010), който специално е разработил за целите на анализа на речта и оптималното ѝ многоаспектно визуализиране интерактивна мултимодална система на платформата CHILDES, където с програмата CLAN става възможно интегрираното представяне на речевите данни чрез транскрипти на аудио- и видеозаписи, с помощта на които многостранно да бъде представен конкретен маркиран транскрипт от речевите последователности, който едновременно е линкуван към няколко модуса (канала) на комуникацията:

а) към изпълнимия аудиофайл и неговата чувана (усещана със слуха) в реално време звучаща звукова реализация в SONIC MODE, изобразена като осцилограма в долния край на прозореча на програмата CLAN – Фиг. 2.



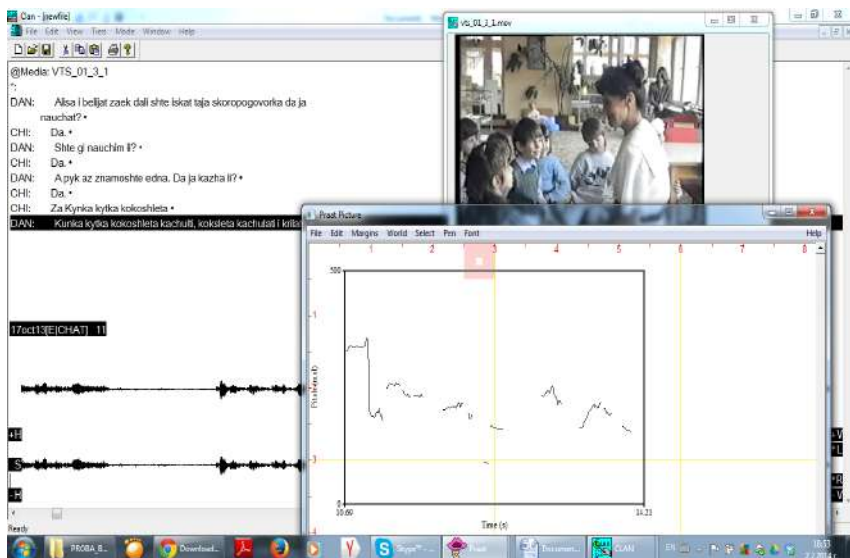
Фиг. 2. Аудиотранскрипция от форма на звуковата вълна.

б) към отделен прозорец на видеофайл, отразяващ в реално време дискурсивния фрейм (отдясно горе) – Фиг. 3.



Фиг. 3. Видеотранскрипция с отворен звуков панел

в) към генерираната в реално време в отделен прозорец (отдясно долу) интонационна реализация на тона (интонограма, визуализирана с програмата Praat) – Фиг. 4.



Фиг. 4. Тоналният контур, експортиран с Praat

Именно такава мултимодално интегрирано многоаспектно представяне на речевото поведение дава оптимални резултати за наблюдението на речевото събитие като процес и от различни зрителни ъгли, така че в хода на анализа да се предоставят добри възможности за правилно разпознаване и идентифициране на целокупния индивидуален слухов образ и адекватното конструиране на речевия портрет на даден говорещ. По този начин върху изследван материал от ексцерпти на българска спонтанна детска реч презентирането на поведението на говорещото дете се осъществява тъкмо в мултимодална перспектива.

Мултимодалната комуникация е вербален процес, неразделно свързан с паравербалните и невербалните компоненти на общуването, който включва в своето единство едновременно съвместно съжителство на различни модалности, които имат за цел да предадат на конкретната речева ситуация някакво социално значение. Към тези модалности се отнасят вербаликата, прозо-

дията (фонацията), кинезиката (мимики, жестове), проксемика (пози, положение на комуникантите един спрямо друг). При това дори и тялото на човека може да бъде и инструмент, а едновременно с това централен ресурс на всички указани равнища на модалността. Многомодалният интерфейс, който обработва паралелно два (аудио, видео) или повече потока от естествените за човека информации, такива като речта, мимиките и жестовете, движението на главата и тялото, постоянно се усъвършенства, за да може да продължи да предоставя все по-оптимални възможности за разпознаването и за разбирането на знаците в човешката комуникация. Модулите, които обезпечават синтез на речта и другите естествени модалности, представляват съществена важна част от многомодалните интерфейси. Това разкрива и нови изследователски хоризонти за бъдещи проучвания в сферата на лингвистичната персонология. Използването на възможностите на съвременния мултимодален интерфейс намира приложение в най-различни отрасли на съвременния живот. Това не само ще улесни общуването в дигиталния и виртуалния свят на човека, но ще помогне за решаването на много практически дилеми в биометричния анализ и по-конкретно във фоноскопските изследвания. С мултимодалното презентизиране на речевото поведение на човека в реално време са свързани впечатленията както от аудиаалния канал, така и от видеоканала (визуалния канал). Ушите чуват звука, докато в същото време очите виждат движенията, лицето, езика и устните. Освен това тактилната информация също се използва в речевата среда. Редица изследователи твърдят, че визуалните сигнали са важни за по-доброто разбиране на произносимата реч. Акцентът в речта може да бъде усилен от един от следните сигнали: от честотата на основния тон, повдигането на веждите, движението на главата, отделния жест или от комбинацията на такива сигнали.

Сред тях разпознаването на качеството на гласа и интонацията имат основно значение за идентифицирането на говорещия, защото те дават индикация за това как да се определят неговата възраст, пол, здравословно състояние, социална принадлеж-

ност, ниво на образование, чужд акцент или диалект на базата на запис с речев материал, съдържащ индивидуалния му начин на говорене и негова акустична визуализация (осцилограма, интонограма, спектрограма) като материално доказателство. Параметрите на гласа, като например честота на основния тон, гръмкост, тембър, темпо, винаги са индивидуални в речта и уникални за съответния речев портрет. Към всичко това допълнително могат да се прибавят и визуалните и тактилните параметри, които тук не са обект на внимание. Важно е обаче да се спомене, че по разчитането на формата и движението на устните биха могли да се определят елементарните единици на визуалната реч (виземите). За гласовия интерфейс обаче са достатъчни само параметрите на слуховата модалност, защото тя е носител на лингвистичната информация за разпознаване и характеризиране на говорещия.

За по-надеждна идентификация на определено лице в последно време в някои страни се използват бимодалните системи, при които биометричните изследвания се правят въз основа на комбинирано разпознаване на поведенческата модалност (гласа) и на физиологическата модалност (ретината на окото, пръстовия отпечатък).

Констатираната множественост на информационните канали дава днес несъмнени възможности за едновременна многоаспектна интерпретация на вариативността в речта на говорещия индивид в реален режим на употреба на езика в различни типове дискурс, поради което в перспектива би могло да се заговори дори и за прилагането на кросмодален подход в научните изследвания, което може да бъде предмет на авторовото внимание в една следваща разработка.

БИБЛИОГРАФИЯ

Вербицка 2000: Вербицкая, М. В. *Теория вторичных текстов (на материале современного английского языка): Автореф. дис д-ра филол. наук.* Москва. // **Verbitska 2000:** Verbitskaya, M. V. *Teoriya vtorichnyh tekstov (na materiale sovremennogo angliyskogo yazyka): Avtoref. dis d-ra filol. nauk.* Moskva.

Виготски 1983: Виготски, Л. С. *Мислене и реч*. София: Наука и изкуство. // **Vigotski 1983:** Vigotski, L. S. *Mislene i rech*. Sofiya: Nauka i izkoustvo.

Кэмерон 2015: Кэмерон, Д. *Разговорный дискурс. Интерпретации и практики*. Харьков: Гуманитарный центр. // **Kjemeron 2015:** Kjemeron, D. *Razgovornyj diskurs. Interpretacii i praktiki*. Har'kov: Gumanitarnyj centr.

Кибрик 2010: Кибрик, А. А. Мультимодальная лингвистика. – В: *Когнитивные исследования – IV*, Москва: ИП РАН, 134–152. // **Kibrik 2010:** Kibrik, A. A. Mul'timodal'naja lingvistika. – V: *Kognitivnye issledovanija – IV*, Moskva: IP RAN, 134–152.

Познер 1995: Познер, Р. Човекът като знак. – В: *Сборник лекции, статии, доклади, есета*. Том I, EFSS, '95. Съставител: проф. М. Н. Попова. София: Пейчински, 78–91. // **Pozner 1995:** Pozner, R. Chovekat kato znak. – V: *Sbornik leksii, statii, dokladi, eseta*. Tom I, EFSS, 1995. Sustavitel: prof. M. N. Popova. Sofia: Psychinski, 78–91.

Kjær, Palsbro 2008: Kjær, A.L., Palsbro, L. National identity and law in the context of European integration: the case of Denmark. – In: *Discourse and Society*. 2008. № 19 (5).

Kress, van Leeuwen 2001: Kress, G., van Leeuwen, T. *Multimodal discourse. The modes and media of contemporary communication*. London: Arnold.

Kress 2002: Kress, G. The multimodal landscape of communication. – In: *Medien Journal* 2002. V. 4, P. 4–19.

MacWhinney 2010: MacWhinney, B. Computational models of child language learning. – In: *Journal of Child Language*, 37, 477–485.

Mieroop 2005: Van de Mieroop, D. An integrated approach of quantitative and qualitative analysis in the study of identity in speeches. – In: *Discourse and Society*, London.

Попова, Попов 2013: Popova, V., D. Popov. Holistic approach to speech interaction in norm and pathology. In: Sovilj M., Subotic M. (Eds.). – In: *Speech and Language 2013. Proceedings of 4th International Conference on Fundamental and Applied Aspects of Speech and Language*, Belgrade, 2013, 40–46.

Попова, Попов 2014: Popova, V., D. Popov. Bulgarian Child Language Corpus – In: Popov D., Skanavis M., Sovilj M. (Eds.). *Parameters, Meaning, Perspectives. Proceedings of 4th International Congress on Ear-*

ly Prevention in Children with Verbal Communication Disorders. Shumen, 2014, 77–85.

Popova, Popov 2015: Popova, V., D. Popov. Multimodal Presentation of Bulgarian Child Language. – In: Ronzhin A., Potapova R., Fakotakis N. (Eds.). *Speech and Computer (SPECOM 2015). 17th International Conference, SPECOM 2015*, Athens, Greece, September 20–24, 2015, Proceedings. Springer International Publishing Switzerland 2015, 293–300.

Slay, Smith 2011: Slay, H., D. Smith. Professional identity construction: Using narrative to understand the negotiation of professional and stigmatized cultural identities. – In: *Human Relations*. Sage. 2011. № 64 (1).