



Сравнителният корпус от академична реч GeWiss и българските данни в него

[Адриана Славчева](#)
[Красимира Алексова](#)
[Ласка Ласкова](#)
[Йорданка Велкова](#)

Въведение

В контекста на все по-напредващата интернационализация и мобилност в академичната сфера проучването на културно обусловената специфика на академичния дискурс и езиковите характеристики на академичното общуване придобиват ключово значение. Все още обаче съществуват изключително ограничен брой свободно достъпни корпусни ресурси за контрастивни емпирични изследвания, особено що се отнася до устната академична реч. В настоящата статия се представя един актуален международен проект, който цели да постави основите за такива контрастивни проучвания. В проекта GeWiss научни екипи от Университета в Лайпциг, Германия, Университета „Астон“ в Бирмингам, Великобритания, и Университета във Вроцлав, Полша, работят по създаването на електронно достъпен сравнителен корпус от немска, английска и полска устна академична реч (вж. [Fandrych et al. 2009](#), [Fandrych, Meißner, Slavcheva 2012](#); [Slavcheva, Meißner 2012](#))¹.

От самото начало на работата по създаването на корпуса част от немския екип работеше успоредно по проекта „Контрастивно изследване на българска и немска устна научна реч“, договор № Д002 – 43/14.03.2009, Фонд „Научни изследвания“, МОМН, в рамките на двустранното научно-техническо сътрудничество между Република България и Федерална република Германия („Gesprochene Wissenschaftssprache kontrastiv: Deutsch und Bulgarisch im Vergleich“ – Universität Leipzig, Herder-Institut / Sofia University “St. Kliment Ohridski”). Работата по двата проекта беше синхронизирана, така че базите данни със записи и транскрипции на академична научна реч, реализирани в България от екип с научен ръководител проф. дфн Емилия Денчева и участници проф. дфн Руселина Ницолова, доц. д-р Красимира Алексова, доц. д-р Петя Осенова, гл.ас. д-р Ласка Ласкова, ас. Йорданка Велкова (всички от Катедрата по български език), както и гл.ас. Микаела Петкова-Кесанлис и гл.ас. Деница Димитрова (от Катедрата по германистика и скандинавистика), да бъдат асоциирани към по-големия корпус GeWiss. Записите и транскрипциите, осъществени в България, се поделят на два подкорпуса: на български език и на немски език. Част от данните на български език са събирани още през 2008 г. в изпълнение на проекта „Студентска академична реч“, финансиран от научния фонд на СУ „Св. Климент Охридски“, договор № 164/2008 г. Целта на работата върху всички записи и транскрипции, реализирани в България, бе те да се приведат в съответствие с параметрите на дизайна на GeWiss, за да бъдат интегрирани в него като четвърти субкорпус за контрастивно изследване на българска академична реч (вж. [Алексова, Ласкова, Велкова 2011](#)).

В първата част на статията ще бъдат скицирани дизайнът на корпуса GeWiss, метаданните за неговото описание, както и актуалното състояние по неговото оформяне и достъпа до него. Втората част на статията представя българския субкорпус от студентска научна реч, който предстои да бъде интегриран в GeWiss.

Дизайн на корпуса GeWiss

1.1. Научни цели на проекта GeWiss



GeWiss е замислен като многоезиков сравнителен корпус с централен субкорпус с немска академична реч и паралелни субкорпуса на английски, полски и български като езици за съпоставки, който да предоставя възможност за няколко типа научни изследвания:

- моноезикови проучвания на езиковите характеристики на академичната комуникация в рамките на отделните езици;
- съпоставителни анализи на немска, английска, полска и българска академична реч на носители на езика в страните партньори;
- съпоставителни анализи на академичната реч на студенти и експерти (преподаватели) на техните майчини езици, както и на немски като чужд език;
- проучвания на разликите в академичната комуникация на немски език в немска академична среда в съпоставка с общуването в британска, полска и българска академична ситуация.

За тази цел дизайнът на корпуса е изключително прецизен. *GeWiss* представлява балансиран набор от устни текстове от два основни жанра на академичната комуникация – монологичен (доклади/реферати) и диалогичен (изпитни събеседвания), в сходни научни дисциплини (немска, английска, полска и българска филология) в немския, британския, полския и българския академичен контекст. Корпусът включва, от една страна, устни езикови данни на немски език от носители и неносители на езика, както и данни на английски, полски и български език от носители на езика. Прецизният дизайн на корпуса позволява също така безпроблемното му разширяване в различни аспекти – към корпуса могат да бъдат добавяни нови контрастни езици и академични контексти, други жанрове, както и друг тип чуждоезикови данни, като напр. от училищна комуникация.

1.2. Параметри на дизайна на корпуса

Концепцията, която лежи в основата на *GeWiss*, отчита два основни параметъра: (1) устни академични жанрове и (2) конфигурацията от езици, които говорителите владеят.

В корпуса са включени два устни академични жанра, които могат да се приемат за основни в редица академични общности, независимо от дисциплината, академичната традиция и вида на езиковата общност. Монологичните академични жанрове са представени от жанра *доклад*, който се намира на границата между устните и писмените текстове² и в който е особено важно структурирането на свързания текст. Включването на този жанр в корпуса в две разновидности – *доклад на експерти на научна конференция*, както и *студентски доклад/реферат* от семинарни занятия, дава възможност за сравнение на различните нива на академична компетентност като важен фактор, определящ езиковите характеристики на идиолекта на говорителя в академична среда. Докладите включват също така и дискусиите към тях, които в повечето академични дискурси представляват интегрална част от жанра. Диалогичните жанрове са представени от *изпитното събеседване*, тъй като то има ключово значение за успеха във висшето образование в редица академични контексти.

По отношение на конфигурацията на езиците, които говорителите владеят, в корпуса са включени, от една страна, записи на *носителите на четирите изследвани езика* и от друга, на *неносители на немския език от четирите академични контекста*. По този начин не само може да бъде сравнявана академичната реч на носители и неносители на немския език, но и да бъдат проучени евентуалните разлики в немскоезичния академичен дискурс в различните чуждестранни германистики, както и да се потърси отговор на въпроса до каква степен чуждестранните германистики може да се разглеждат като хибридна форма на националния и на немския научен дискурс.



1.3. Обем на корпуса GeWiss³

В своята първа версия корпусът ще включва общо около 120 часа от транскрибирани и синхронизирани аудио записи – 60 часа от всеки жанр, които съответстват на по 40 часа записи от немския, английския и полския контекст. Таблица 1 представя актуалното състояние на транскрибираните аудиозаписи в GeWiss.

	Немски академичен контекст		Британски академичен контекст		Полски академичен контекст		Общо
	Немски език носители на езика	Немски език неносител и на езика	Английск и език носители на езика	Немски език неносител и на езика	Полски език носители на езика	Немски език неносител и на езика	
Научен доклад	603 мин.	-	447 мин.	321 мин.	295 мин.	295 мин.	32:44 ч.
Студентски доклад	385 мин.	267 мин.	301 мин.	315 мин.	297 мин.	302 мин.	31:07 ч.
Изпитно събеседване	625 мин.	551 мин.	318 мин.	660 мин.	602 мин.	550 мин.	55:06 ч.
Общо	26:53ч.	13:38 ч.	17:46 ч.	21:36 ч.	19:54 ч.	19:07 ч.	118:57 ч.

Таблица 1. Обем на транскрибираните и синхронизирани записи в корпуса GeWiss.

Тъй като транскрипционните дейности по корпуса все още не са приключили напълно, все още не може да бъде дадена точна информация за обема му по отношение на брой думи. По приблизителни изчисления на основата на съществуващи корпуси от академична реч като MICASE⁴ или ELFA⁵ от 1 час записи се получават 7600 – 9000 думи. Ако се приложи по-консервативното съотношение от 1 час : 7600 думи, то приблизителният обем думи на GeWiss се очаква да бъде около 0,9 милиона⁶.

До момента в корпуса са включени около 447 говорители: 128 души, записани в немски, 121 души в британски и 198 души в полски академичен контекст. Макар социално-демографските критерии да не са водещи в дизайна на GeWiss, значителният брой говорители, представени в корпуса, гарантира включването на разнообразни езикови практики, което би могло да позволи генерализиране на наблюдаваните явления за съответния академичен контекст.



Метаданни за описание на корпуса GeWiss

Прецизният дизайн на корпуса прави възможно включването на голямо количество детайлни метаданни, описващи комуникационната ситуация като цяло, записите и съответстващите им транскрипции, както и говорителите. Метаданните на GeWiss са представени в XML формата на EXMARaLDA и се администрират с помощта на EXMARaLDA Corpus Manager (COMA) (вж. Schmidt, Wörner 2009).

Тук накратко ще бъдат описани някои основни параметри от метаданните за комуникативната ситуация и говорителите, които са основни в дизайна на GeWiss и позволяват оформянето на субкорпуси за специфични изследователски цели (вж. Fandrych, Meißner, Slavcheva in print).

1.4. Метаданни за комуникативната ситуация

Метаданните за комуникативната ситуация включват 13 параметъра, описващи ситуацията като цяло, както и информация за нейната локация, за езиците, използвани в нея, както и най-общо за интеракционните условия.

Първият основен параметър от дизайна на корпуса – типът устен текст, е зададен в категорията *Genre*. В съответствие с втория параметър от дизайна – конфигурацията на езиците на говорителите, в категорията *L1 Communication* е посочена информация по отношение на това, дали комуникацията се е провела на майчиния език на централните говорители, или не. Конкретна информация за езиците, използвани в комуникативния акт, може да бъде намерена освен това в секцията *Language(s)*. В GeWiss се прави разлика между основния език на комуникацията, посочен в *Mainlanguageofinteraction*, и други езици, които евентуално са използвани в нея (*Languagealternation*). Освен това в категорията *Degreeoforality* комуникативната ситуация е специфицирана по отношение на степента на подготвеност на текста (напр. свободно говорене, четене или наизустен текст) по преценка на наблюдателя, присъствал на записа, както и на наличните допълнителни материали към комуникативния акт, като манускрипт, презентация и т.н.

1.5. Метаданни за говорителите

Освен някои основни социално-демографски данни, метаданните за говорителите включват информация за тяхното образование, както и за езиците, които те владеят. В съответствие с XML модела на COMA етапите в образователната биография на говорителите са въведени като отделни локации (*Locations*), като в GeWiss се различават следните три типа: училищно образование (*Education*), което се разглежда като ключово за социализацията на говорителя в образователната система на съответната езикова общност и съответно е от основно значение за формирането на базисни академични езикови умения, следване в чужбина (*Studyabroad*) и престой зад граница (*Stayabroad*) – за описание на по-дълги периоди на престой в чужбина с необразователни цели. Освен това в метаданните е посочена и подробна информация за езиковите компетенции на говорителите. Тъй като GeWiss е създаден с идеята да бъде използван за контрастивни изследвания между говорители от различни езикови общности и в частност за сравнително проучване на академичната реч на носители и неносители на немския език, в метаданните е включена информация както за майчиния език на говорителя, така и за чуждите езици, които той владее, дефинирани като всички останали езици след майчиния. Освен това за всички говорители, владеещи немски като чужд език, в категорията *EvaluationoftheLanguagecompetence* е посочено нивото на владеене на немския език. По този начин академичните езикови умения на неносителите на езика могат да бъдат анализирани в съответствие с нивото на общата им езикова компетентност по немски език.



Транскрипциите в корпуса GeWiss

Транскрипциите в GeWiss се изготвят ръчно с помощта на софтуера EXMARaLDAPartitur-Editor (вж. [Schmidt, Wörner 2009](#)) на основата на минималното транскрипционно ниво на конвенцията GAT2 (вж. Selting et al. 2009), разработена за специфичните цели на конверзационния анализ на немска устна реч. В съответствие с тази конвенция транскрипцията представя ортографично звученето на изказванията и тяхната организация по сегменти във вербалния ред на говорителите, без да се съблюдават пунктуационните правила и да се изписват главни букви. Важна особеност с цел четимост на транскрипта за нефилолози е изписването със стандартната ортография на разговорни реализации, следващи книжовната норма. Единствено случаите с големи отклонения от стандартното произношение – диалектни лексеми, жаргонни думи, оказионализми и пр., се изписват според звученето им, напр. *mitnanner* като идиосинкретична форма от *miteinander* в немската устна реч. В хода на транскрипционната работа някои от конвенциите на GAT2 бяха доразвити с цел подобряване на търсенето в корпуса (вж. [Fandrych, Meißner, Slavcheva 2012](#)) и адаптирани от полския, британския и българския научен екип към особеностите на съответните езици (вж. [Lange et al. submitted](#), както и [Алексова, Ласкова, Велкова 2011](#) конкретно за адаптирането на конвенцията към българския език). *Фигура 1* представя изглед от транскрипция на изпитно събеседване от немския субкорпус на GeWiss, която онагледява някои от основните принципи на използваната конвенция.

	251 [04:42.2]	252 [04:43.4]	253 [04:44.254 [04:45.0]	255 [04:46.256 [04:46.1]	257 [04:47.258 [04:49.2]	259 [04:50.260 [04:51.261 [04:52.262 [04:52.0]	263 [04:53.0]	264 [04:54.0]
SB_0311 [v]		hm_hm				hm_hm	ja	
SB_0311 [k]								
LS_0310 [v]	äh (.) sich schon anbie	tet nur	geschichts	(0.3)	darstellungen h° geschichtsschreibung (.) is auch narrativ	°hh	öhmm	((schmatzt)) °hhh was is denn da hh°
LS_0310 [k]								
FL_0264 [v]								
FL_0264 [k]								
BS [v]								
BS [k]								
nn [v]		(0.9)			(0.3)	(0.9)		
nn [k]		Statt						

Фигура 1. Транскрипция от корпуса GeWiss EXMARaLDAPartitur-Editor на основата на транскрипционната конвенция GAT2.

Актуално състояние по оформянето и достъпа до корпуса GeWiss

Към настоящия момент приключва транскрипционната работа по корпуса и започва етапът на неговата електронна обработка. Той включва отстраняване на грешки в синтаксиса, сегментиране на транскрипциите, привеждането им във вид за онлайн публикуване и изработка на интерфейс за достъп през интернет. Първата версия на корпуса се предвижда да бъде публикувана във втората половина 2012 г.

Субкорпусът на български език

Както стана ясно, записите и транскрипциите, реализирани в България, се поделят на две



части в съответствие с езика на говорителите:

- на български език: студентски реферати/презентации и изпити, както и доклади и лекции на преподаватели;
- на немски език: студентски реферати и изпити, както и доклади на преподаватели.

Записите на научна реч на български студенти германисти и на доклади на преподаватели български германисти надвишават 6 часа, като част от тях са транскрибирани с участието и на немските партньори по проекта. Данните от този субкорпус са в основата на няколко публикации – вж. напр. [Petkova-Kessanlis 2012a](#), [2012b](#).

В Таблица 2 представяме данни за записите и транскрипциите на български език, като искаме да подчертаем, че работата по тях не е приключила и че субкорпусът предстои да бъде окончателно проверен и присъединен към големия корпус GeWiss.

	Устен изпит	Реферат/Презентация
Общо времетраене на записите	01:56:10	03:04:11
Средна продължителност	00:06:34	00:23:26
Мин. и максимална дължина	00:01:54 до 00:10:41	00:23:00 до 00:42:08
Транскрибирани записи	01:04:33	03:04:11
Интерактивност	да	ограничена
Спонтанност	относителна	да
Използване на помощни средства	ограничено	да

Таблица 2. Количество и комуникативни особености на субкорпуса на български език.

Докато при транскрибирането на немскоезичните записи, направени от екипа от Катедрата по германистика и скандинавистика, безпроблемно се следваше минималното транскрипционно ниво на конвенцията GAT2, то за нуждите на транскрипциите на български език се наложи тази конвенция да бъде адаптирана. Вариантът, приложен от екипа българисти, отчита особеностите на прилагана в практиката българска конвенция за писмено представяне на устна комуникация, която може да се види на сайта <http://www.bgspeech.net/>. Тя е изработена в Катедрата по български език на СУ „Св. Кл. Охридски“ и е резултат от дългогодишна работа на преподаватели от Катедрата по различни проекти, посветени на българската разговорна реч. Тук само ще посочим, че основни в българска конвенция са два принципа: а) в транскрипцията се отбелязват само отклонения от кодифицираните



правоговорни норми, б) поддържа се такава степен на четивност на транскрипциите, която да позволи използването им както от изследователи лингвисти, така и от по-широк кръг хуманитаристи. Затова тази конвенция се различава съществено от транскрипционните системи, познати от фонетични или диалектоложки трудове. Решението за съвместяване на *GAT минимална транскрипция* и българската транскрипционна конвенция бе взето, тъй като смятаме, че не е необходимо по-детайлно ниво на транскрипция. Изследователите, интересуващи се от специфични фонетични, интонационни и др. особености на устната комуникация, могат да използват за своите нужди звуковия файл, синхронизиран с транскрипцията.

По-важните допълнения към *GAT минимална транскрипция*, отразяващи особености на българския език, предвиждат отбелязване на следните явления в устната комуникация⁷:

- различни типове елизии, напр. *мноо, кво, тва*;
 - депалатализация на съгласната в глаголни и именни форми от типа *вървъ, вървът, мисль, мислът, затварам, отвараме, вечерате, вечерат, зеть, лакътъ, читална, дъщеря*;
- дейотацията в глаголни и местоименни форми, напр. *пеа, пеат, онеа, таа, онаа* и под.;
- смекчаване на съгласните пред гласните **е, и** – предава се писмено с **й**, напр. *ние, тйежко*;
- отмет или премет на ударението, представляващ отклонение от книжовната норма; използва се знакът \ пред ударената гласна, напр. *ход\их, разправ\ял, д\онеси* и др.;
- екане, напр. *голема, хлеб, живехме* и т.н.;
 - свръхякане, напр. *бяли, хлябче, голями, живяли, тяхни, обляни* и т.н.;
 - мекане (окончание -ме вм. -м за 1 л. мн.ч., I и II спр.), напр. *пишеме, четеме, говориме*;
 - аористна гласна **-а**-вм. **-о**- (I спр., I разред), напр. *дадах, донесахме, излязахте, плетаха*;
 - пълната редукция на гласната **е** в безударена позиция, напр. *тиливизори, пипируда*;
- изговор -тъ на членната морфема за ж.р. под ударение, напр. *есънтъ, специалността*;
- т.нар. ъкане, напр. *рък\ътъ, вод\ътъ, глав\ътъ*;
- буквите **я** и **ю** се изписват, въпреки че имат множество звукови стойности, като обаче се следят следните специфични случаи:
 - въпреки наличието на условно писане на **а** (респективно **я**) в транскрипциите се следва правописната норма. Ненормативният изговор от типа [четà] вместо [четъ] в този случай остава неразпознаваем без звуковия файл. В българската конвенция за транскрипция тези случаи се отбелязват с главна буква, напр. *четА* (при изговор [четà] вместо [четъ]).
- Ето и някои допълнителни инструкции:
- числителните бройни се изписват като цели думи (и така, както се чуват): *двайс първи, четринайсе, сто и пеесе*;
- дължина на гласната се отбелязва с две точки, напр. *ху:*;
- абривиатури се изписват по начина, по който са произнесени, напр. *министерство на образованието, българска академия на науките, мевере, бедеже* и др.;
 - според *GAT минимална транскрипция* главни букви не се използват, но българският екип взе решението да се пишат главни букви само при личните имена (собствени



лични, сложни названия и имена на географски и др. обекти) при спазване на правилата за писане на главна буква в *Нов правописен речник на българския език (2002)*, напр. *Българска народна банка, Стара планина, Велико Търново, Петър Атанасов* и т.н.;

- при случаи на вторична номинация се използват кавички (напр. *Софийски университет „Свети Климент Охридски“*);

- сложните думи се изписват според правилата за слято, полуслято и разделно писане на *Нов правописен речник на българския език*;

- диалектни, жаргонни думи или оказионализми в идиолекта на говорещия се изписват по начин, който е най-близо до тяхното звучене.

В транскрипцията се предават всички звукови явления, което означава, че се предава приблизителната звукова стойност на фонационните паралингвистични средства и тяхната функция (семантика) в комуникативната ситуация. Отбелязват се също смехът, покашлянето и др. звукове, които имат отношение към разчитането на звуковата страна и на семантиката на комуникацията, напр.:

- фонационни паралингвистични средства, напр. *а_а_а, ъ_ъ_ъ, м_м_м; тх, хм, аха, ахам, ъхъм, мда* и др.;

- *м_м_м*, докато говорещият мисли или подбира думи, като се уточнява значението – *обмисля отговора, търси подходящи изразни средства* и др.;
- смях от типа *ха-ха, хе-хе, хи-хи, хо-хо*, ако може да се наподобява звуковият състав на кинетичното явление, а ако това е невъзможно – в двойни скоби се изписва *смях*, напр. *не мога да говоря повече по този въпрос ((смях))*;
- едновременното наличие на паралингвистично явление и реч, напр. <<смеене> лято> – говорещият се смее, докато произнесе думата „лято“; <<кашляне> слънце> – говорещият казва „слънце“, кашляйки (вж. Фиг. 2);

27 [02:26.2]	28 [02:27.6*]	29 [02:29.0]
да класифицираща		
значи класифицираща	<<coughing > категория>	освен това? (--)

Фигура 2. Едновременност на вербалното и паралингвистичното явление: говорене и кашляне.

- усмивка без звукова стойност, напр. <<:-) > море>;



- страничен шум от типа отваряне/затваряне на врата, звънене на телефон, музика и др. подобни, които могат да затруднят разбирането на кратки или дълги отрязъци от речта. Всички явления, които пречат на адекватното транскрибиране на речта, се отбелязват в двойни скоби, например: ((звъни телефон)); ((чува се музика)); ((отваря се врата)) и др.

При транскрипциите се отбелязват и паузите, като не се отбелязва какъв е типът ѝ – физиологична (дихателна) или интонационна (лингвистична), а само се уточнява нейната продължителност:

а) много кратка, едва доловима пауза - (.)

б) кратка пауза - (-)

в) средна пауза (0.5 – 0.8 сек) - (--)

г) дълга пауза - (0.8– 0.10 сек) (---)

д) пауза с точно времетраене - (0.5)

Приложената транскрипционна конвенция дава възможност да се отразяват застъпванията между речта на различните участници в комуникацията, тяхното едновременно говорене или застъпване на отделни невербални явления (сочат едновременно нещо, смеят се едновременно и т.н.). Следният модел лежи в основата на отбелязване на застъпванията:

дума [дума]

[дума] дума

Програмният продукт Partitur-Editor позволява застъпващите се вербални или невербални явления да бъдат обособени в отделен интервал (Фиг. 3). В конкретния пример част от репликата на първия говорещ – „съществително“, се застъпва с реплика на втория участник в комуникацията – „добре заповядайте“. Разполагането на застъпващите се части една под друга дава възможност на интересуващите се да ги наблюдават пряко, като освен това могат да извлекат информация за времевата продължителност на застъпването.

	0 [00.0]	1 [02.0]	2 [03.5*]
MK_0006 [v]	първо за съществителното име (.)	съществителното	е самостоятелна:
MK_0006 [nv]			
RN_0007 [v]		добре заповядайте	
RN_0007 [nv]			

Фигура 3. Представяне на едновременното говорене в един времеви интервал.



части от записа, като може да бъде направено предположение за думата, която не се идентифицира напълно:

а) неразбираема сричка – (XXX);

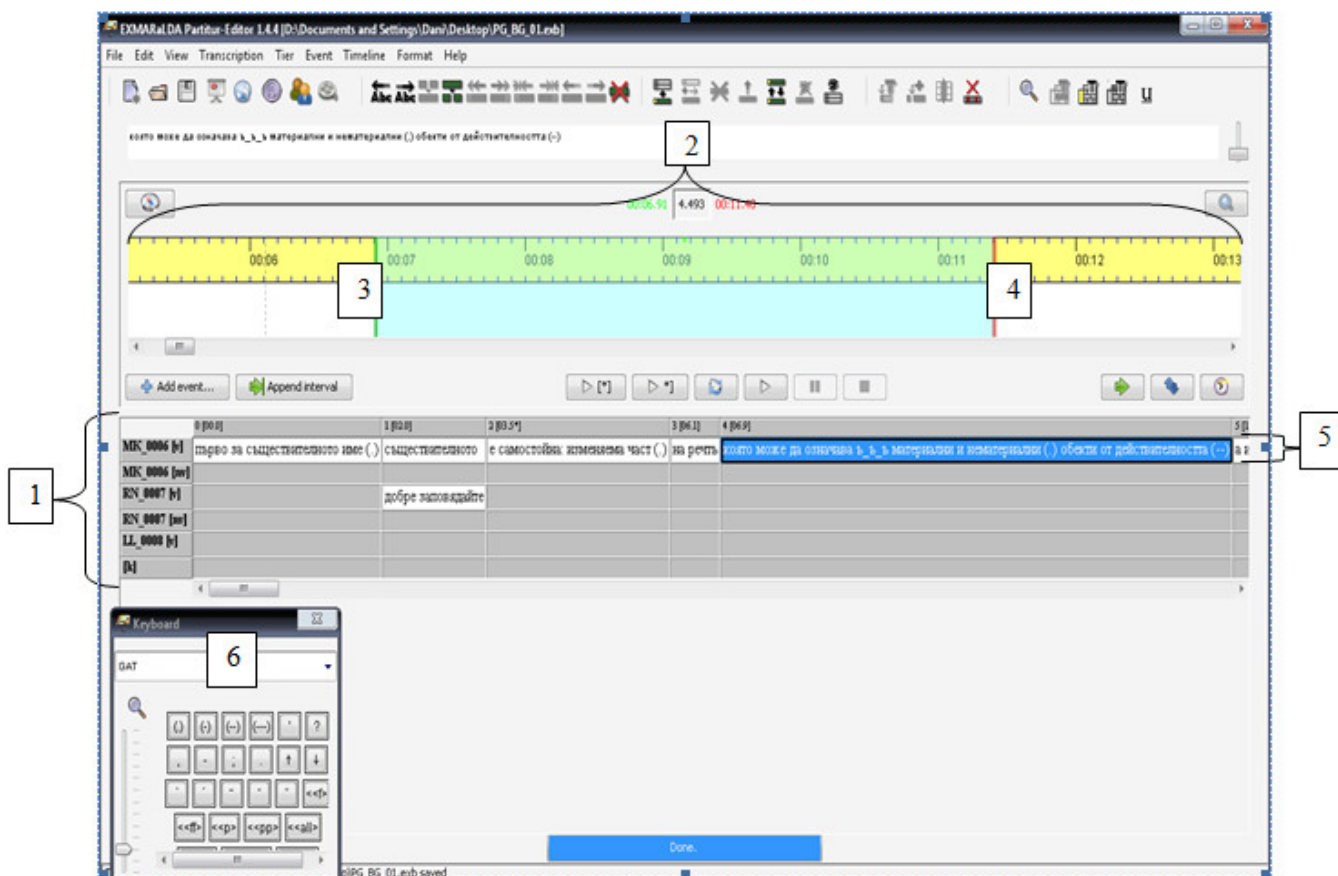
б) неразбираеми срички – (XXX XXX);

в) предположена дума – (ведър);

г) предположение за думи – (ръст/пръст);

д) ако неразбираемият сегмент от записа продължава няколко секунди, в скоби се записва *приблизително* и се уточнява времетраенето: ((неясно, прибл. 3 сек)).

В предходна публикация представихме софтуера за транскрибиране Partitur-Editor, който е част от EXMARaLDA пакета (Алексова, Ласкова, Велкова 2011). Важно е да се отбележи, че при работа с Partitur-Editor могат да се въведат метаданни, свързани с името на проекта, конвенцията за транскрибиране, техническите особености на записа, а освен тях и социално-демографски характеристики на говорещите (вж. и 3.2.). Тук само ще маркираме основни елементи от интерфейса и свързаните с тях функционалности на програмата (Фиг. 4), така че интересуващите се да могат да придобият представа за начина, по който се кодират данните в българския подкорпус от студентска научна реч.



Фигура 4. Общ вид на интерфейса на Partitur-Editor.



Както се вижда, интерфейсът на редактора включва следните основни елементи:

1. Панел за транскрибиране, или още – партитура (нем. Partitur), включващ редове, в които се записва вербалната и невербалната информация.
2. Графика на времевия интервал.
3. Лева граница (начало) на избрания отрязък.
4. Дясна граница (край) на избрания отрязък.
5. Партитурен ред, в чието начало е кодирано името на говорещия. Речта на всеки участник в общуването се записва на отделен ред.
6. Клавиатура – може да се избере клавиатура, която да отговаря на конвенцията за транскрибиране, с която се работи (на екранния кадър се вижда GAT клавиатура).

Партитурните редове са два типа: за транскрипция и за описание (вж. *Фигура 5*). Поради различията в информацията, която партитурните редове могат да съдържат, те се поделят на три категории:

1. Партитурен ред „v“ категория – за вербалното поведение на говорещия.
2. Партитурен ред „nv“ категория – за невербалните явления, напр. кашляне, страничен шум и др., като се уточнява значението на паралингвистичните средства.
3. Партитурен ред „k“ категория – за коментари, отнасящи се до цялата комуникативна ситуация, нейната специфика и особености.

Ако партитурният ред е от типа *транскрипция*, категорията му трябва да бъде „v“; ако е от тип *описание*, редът е задължително „nv“ или „k“, но не може да бъде „v“.

RN_0007 [v]	тивност	(---) да си избират определени (.) форми на думите (.) определени класове думи
RN_0007 [nv]		жестикულიра
LL_0008 [v]		
[k]		чува се шум, в стаята влиза още един участник

Фигура 5. Категории и типове партитурни редове.

Бъдещето на българския субкорпус

Екипът от българисти разполага с много повече записи в сравнение с обработените. Това предполага възможността субкорпусът да се разраства чрез транскрипции и на студентска научна реч, и на образци от научната реч на т.нар. експерти (т.е. преподаватели). Не са обработени все още всички метаданни към файловете, които, както стана ясно, са доста обемни. Част от най-ранните транскрипции трябва да се верифицират отново от експерти, като се провери и анонимизирането на участниците, и синхронизацията на транскрипцията със звуковия файл.

Бележки

1. Научният проект „GeWiss – Gesprochene Wissenschaftssprache kontrastiv: Deutsch im Vergleich



zum Englischen und Polnischen" е финансиран от VolkswagenStiftung в рамките на програмата Deutsch Plus – Wissenschaft ist Mehrsprachig (вж.

<http://www.volkswagenstiftung.de/foerderung/herausforderungen/deutsch-plus-wissenschaft-ist-mehrsprachig.html>). Подробна информация за проекта GeWiss може да бъде намерена на неговата интернет страницата <https://gewiss.uni-leipzig.de/de/>.

2. За описанието на устните и писмени езикови форми в съотношение с комуникативните условия на общуването в континуума близост – дистанция на общуването вж. [Koch, Österreicher 1985](#).

3. В тази част не е включен българският субкорпус, който ще бъде представен подробно във втората част на статията.

4. Корпусът MICASE (MichiganCorpusofAcademicSpokenEnglish) съдържа предимно езикови данни от академична комуникация между носители на английския език. Информация за корпуса може да бъде намерена на <http://micase.elicorpora.info/about-micase>.

5. Корпусът ELFA (EnglishasLinguaFrancaInAcademicSettings) съдържа записи на неносители на английски език във финландски академичен контекст. Информация за него може да бъде намерена на <http://www.helsinki.fi/englanti/elfa/elfacorporus.html>.

6. За съпоставка на GeWiss с други корпуси от академична реч вж. [Fandrych, Meißner, Slavcheva 2012](#).

7. Тази информация и следващите след нея данни са представени и в [Алексова, Ласкова, Велкова 2011](#).

Цитирана литература

- Алексова, Красимира, Ласка Ласкова, Йорданка Велкова. 2011. Корпус от студентска научна реч. *Български език* 3. 72–88.
- Fandrych, Christian, Cordula Meißner, Adriana Slavcheva. 2012. The GeWiss Corpus. Comparing Spoken Academic German, English and Polish. In: Thomas Schmidt, Kay Wörner (Hg.). *Multilingual corpora and multilingual corpus analysis*. Amsterdam: Benjamins. 319–337. (Hamburg Studies in Multilingualism 14).
- Fandrych, Christian, Erwin Tschirner, Cordula Meißner, Stefan Rahn, Adriana Slavcheva. 2009. Gesprochene Wissenschaftssprache kontrastiv: Deutsch im Vergleich zum Englischen und Polnischen. Vorstellung eines gemeinsamen Forschungsvorhabens. – In: Edyta Błachut, Lesław Cirko, Artur Tworek (Hg.). *Studia Linguistica XXVIII*. Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego. 7–30 (Acta Universitatis Wratislaviensis No 3196).
- Koch, Peter, Wulf Oesterreicher. 1985. Sprache der Nähe-Sprache der Distanz: Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. *Romanistisches Jahrbuch* 36.15–43.
- Lange, Daisy, Adriana Slavcheva, Marta Rogozińska, Ralph Morton. (submitted). GAT2 als Transkriptionskonvention für multilinguale Sprachdaten? Zur Adaption des Notationssystems im Rahmen des Projekts GeWiss. In: Christian Fandrych, Cordula Meißner, Adriana Slavcheva (eds.), *Tagungsband der GeWiss-Konferenz vom 27. - 29. 10. 2011*. Heidelberg: Synchronverlag. (Wissenschaftskommunikation).
- Petkova-Kessanlis, Mikaela 2012a (in prep.). Grade sprachlicher Formelhaftigkeit bei der Realisierung der Textsorte "Studentisches Referat" in der Fremdsprache Deutsch. In: Christian Fandrych, Cordula Meißner, Adriana Slavcheva (eds.), *Tagungsband der GeWiss-Konferenz vom 27. - 29.10.2011*. Heidelberg: Synchronverlag. (Wissenschaftskommunikation).
- Petkova-Kessanlis, Mikaela 2012b (in prep.). Varianz bei der Realisierung der Textsorte „Studentisches Referat“ in der Fremdsprache Deutsch und ihre didaktischen Potentiale. *Journal of Linguistics and Language Teaching*.



- Schmidt, Thomas, Kay Wörner. 2009. EXMARaLDA – Creating, analyzing and sharing spoken language corpora for pragmatic research. *Pragmatics* 19. 565–582.
- Selting, Margret et al. 2009: Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). *Gesprächsforschung. Online-Zeitschrift zur verbalen Interaktion* 10. 353–402. <http://www.gespraechsforschung-ozs.de/heft2009/px-gat2.pdf> (23.03.2012).
- Slavcheva, Adriana, Cordula Meißner 2012. GeWiss – a comparable corpus of academic German, English and Polish. In: Proceedings of the LREC-Workshop “Best Practices for Speech Corpora in Linguistic Research” 2012. 7–11 <http://www.lrec-conf.org/proceedings/lrec2012/index.html>

Година: 2012

Книжка: 1

Сподели: [Delicious](#) [Digg](#) [Facebook](#) [Google](#)

[Plus](#) [LinkedIn](#) [Myspace](#) [Orkut](#) [Reddit](#) [StumbleUpon](#) [Twitter](#) [Yahoo](#)

Source URL: <https://naum.slav.uni-sofia.bg/lilijournal/2012/1/slavchevaa>