

Москва, 14–17 июня 2023 г.

Corpus of Accentuated Byzantine Written Monuments and Methods of Its Markup

Evdokimova A. A.

Institute of Linguistics, Russian Academy of Sciences /
1 bld. 1 Bolshoy Kislovsky Lane, 125009 Moscow
arochka@gmail.com

Abstract

This article presents a corpus of Byzantine accentuated texts (BGAT) created since 2008. It currently includes 1010 Byzantine inscriptions, 950 papyri from various collections from the 1st to the 9th centuries, 132 seals from the collection of Dumbarton Oaks, and a selection of 100 Athos manuscripts from the 8th to the 15th centuries. Based on the collected data, we developed a method for markup such texts, which later makes it possible to create a database of accentuated texts from them and, based on the entire corpus, to train neural networks for classifying texts according to accentuation systems and recognizing them in images. As a result of marking texts, in addition to the previously known Alexandrian, Byzantine, and Dorian systems of accentuation, new accentuation systems were identified, including logical or semantic, with a shift of an accent mark to the right, with a displacement of an accent mark to the left, and mixed. For each group of monuments, their variants of using accentuation systems, especially the Alexandrian one, were identified, which show different aspects of the accentuation of the Byzantine Greek language. When creating a glossary of accentuated word forms based on the corpus, we determined that there were lexemes that retained their accentuation regardless of the influence of the dialect, meter, or traditions characteristic of the masters. However, a comparison of identical texts, even found in the same region of the Byzantine Empire, showed that the accentuation was not replicated when quoting.

Keywords: corpus BGAT, markup methods, accentuation systems, Byzantine Greek language, diacritical marks, Greek epigraphy, Byzantine sigillography, Byzantine manuscripts, papyri.

DOI: 10.28995/2075-7182-2023-22-1071-1081

Корпус акцентуированных византийских письменных памятников и методы его разметки

Евдокимова А. А.

Институт языкознания РАН /
125009, Москва, Большой Кисловский пер. 1 стр. 1
arochka@gmail.com

Аннотация

В данной статье презентуется корпус византийских акцентуированных текстов (BGAT), создаваемый с 2008 г. В его состав на сегодняшний день входят 1010 византийских надписей, 950 папирусов из разных коллекций с I по IX вв., 132 печати из собрания Dumbarton Oaks и выборка из 100 афонских рукописей от VIII по XV вв. На основе собранных данных была разработана методика разметки такого рода текстов, позволяющая в дальнейшем сделать из них базу акцентуированных текстов и на основе всего корпуса обучать нейронные сети классификации текстов по системам акцентуации и распознаванию их на изображениях. В результате разметки текстов кроме известных ранее александрийской, византийской и дорийской систем акцентуации были выявлены новые системы, среди которых логическая или смысловая, со сдвигом ударения вправо, со сдвигом ударения влево и смешанная. Для каждой группы памятников были выявлены свои варианты использования систем акцентуации, в частности александрийской, которые показывают разные аспекты акцентуации византийского греческого языка. При создании глоссария акцентуированных словоформ по материалам корпуса было определено, что существовали лексемы, сохраняющие свою акцентуацию независимо от влияния диалекта, стихотворного размера или традиций, свойственной мастерам. Однако, сопоставление одинаковых текстов даже в одном регионе показало, что акцентуация не заимствовалась при цитировании.

Ключевые слова: корпус BGAT, методика разметки, система акцентуации, Византийский греческий язык, диакритические знаки, греческая эпиграфика, византийская сфрагистика, византийские рукописи, папирусы.

[1] Введение

Акцентуированные греческие надписи византийского периода (IV-XV вв.), особенно граффити, показали, что, несмотря на использование в рукописях с X в. последовательной акцентуации, названной византийской, параллельно с ней существовали другие системы акцентуации. Среди них самая старая александрийская [3, 7, 9, 13, 15, 30], характерная для папирусов I-III вв., которая встречается и в надписях XIII-XV вв. из Северной Греции, Каппадокии, Грузии [18, 22].

С распространением технологий музеи и архивы занялись фотофиксацией своих коллекций с последующим представлением их в виде баз данных или электронных корпусов. Так в научный оборот стали поступать ранее не учтенные источники, хранившиеся в запасниках, и памятники, известные по изданиям, в которых отсутствовали фотографии. Как следствие, одной из актуальных задач для византистики стало описание всех коллекций в онлайн-формате и создание корпусов, разделенных по жанровым, географическим и иным принципам [ср. 11, 12, <http://papyri.info/>]. Частично это было реализовано при создании базы данных РНИ-7, где в виде текстов представлены папирусы, надписи и печати. Анализ этой базы с точки зрения акцентуации показал, что исследователи могут быть непоследовательны и отражать только те случаи акцентуации, которые показались им значимыми или унифицировать их, игнорируя данные критического аппарата в издании. К орфографии памятников применялся другой принцип. Она сохранялась, как в изданиях, откуда собирались надписи, что позволяет посчитать статистику для типов замен одних графем на другие. Поскольку существовала традиция при публикации византийских памятников письменности унифицировать знаки акцентуации в пользу позднейшей византийской системы или вовсе их пропускать, то сейчас в результате цифровизации источников появились новые данные по бытовавшим в Византии системам акцентуации и стало возможно их более детальное изучение. Однако, до сих пор не существует электронной базы акцентуированных греческих надписей и папирусов, включающей в себя фотографии, прорисовки, расшифровки, переводы, географические карты, лингвистические, палеографические и исторические комментарии, а также изданного корпуса подобных текстов или представленного онлайн. Все это делает затруднительным решение вопроса, какие системы акцентуации бытовали в византийский период, как они между собой коррелировали и как они влияли на произошедшие фонетические изменения. Однако, решение данного вопроса необходимо не только для истории систем акцентуации в греческом языке, но и для истории греческого языка в целом, а также как дополнительный источник для датировки памятника. Что привело нас к необходимости собрать корпус акцентуированных памятников в виде изображений и разметить их. В процессе сбора корпуса, начатого в 2008 г., и разметки данных проводились исследования, посвященные акцентуированным греческим памятникам разных регионов Византийской империи [18, 19, 20, 22, 23, 24, 25].

[2] Материал и метод

Собранный корпус акцентуированных памятников содержит:

- 1010 надписей из Каппадокии, Константинополя, Киева, Грузии, Болгарии, Северной Греции, Афин, Италии, большая часть которых датируется X-XV веками. Для каждого региона производилось ранжирование по типу надписи (лапидарная, граффито [22], дипинти etc) и по типу материала: на металле, на камне, на стеатите, на дереве, на фреске [25].
- 950 папирусов из разных коллекций с III по IX вв. Примерное соотношение акцентуированных папирусов и неакцентуированных иллюстрируют коллекции Кельнского университета - из 420 папирусов 60 акцентуированы [19].
- выборку из рукописей, хранящихся на Афоне (ок. 100 экземпляров от VIII по XV вв.) [24].
- 142 акцентуированные печати, отобранные из 11984 экземпляров коллекции Dumbarton Oaks <https://www.doaks.org/resources/seals>. Распределение акцентуированных печатей по векам показало, что на XI-XIII вв. приходится рост акцентуированных сфрагистических памятников [23].

Все памятники в корпусе при разметке были расклассифицированы в отношении 8 систем акцентуации, существовавших в Византии: александрийской¹ [1, 4, 9, 15], дорийской [3], византийской² [16, 18], смысловой или логической [24, 25], со сдвигом вправо [23], со сдвигом влево, соединения систем со сдвигами в одну систему, обусловленную поэтическим метром [20, 23], и смешанной, соединяющей черты любых из них.

Для классификации по системам учитывались диакритические знаки:

- акут,
- гравис,
- циркумфлекс,
- придыхание (легкое и густое),
- трема.

[3] Технические вопросы и подготовка к разметке

3.1 Связанные с текстовым представлением акцентуации

Традиционно для древнегреческого языка шрифты построены на последовательном проставлении диакритических знаков над ограниченным набором гласных. Так возникает проблема представления в уникоде различных знаков акцентуации и диакритики, если они отличаются от традиционно принятых, например, циркумфлекс над *ο* или трема над *ε*. Кроме передачи знаков в случае систем со сдвигами ударения, когда они оказываются над согласными или между двумя буквами, важной оказывается возможность поставить знак так, как он представлен в надписи, для возможности дальнейшего поиска по корпусу всех этих особенностей. Например, *λε'ω* выглядит, достаточно наглядным, однако шрифта, где ударение можно было бы сместить еще правее и написать его над любым согласным не существует. Другой вариант решения кроме разработки шрифта, содержащего все необходимые символы, это представить все слова в виде буквенных формул на латинице и задавать в них через определенные правила местоположения знаков, как и сами типы знаков в более общем виде. Один из способов представления таких формул приводится нами в разделе 4.3.

3.2 Обработка картинок и сопутствующих иллюстративных материалов

Каждый акцентуированный памятник представлен в корпусе в виде фотографии или прорисовки. В случае, если авторы изданий опубликовали и то, и другое, то оба материала учитываются, так как в них бывают разночтения. Чаще прорисовка отражает ту интерпретацию, которой придерживается издатель. При беглом анализе фотографий из одного региона видно различие в материале, на котором выполнена надпись (рис.1 и рис.2), и в использовании разного типа письма, маюскульного (рис. 2) или минускульного (рис. 3).



Picture 1: Италия, надпись на мраморе



Picture 2: Италия, надпись на металле

1 два и более ударений в слове, маркировка безударных слогов, ударение дальше, чем третий слог от конца, ударение на первый элемент дифтонга, ударения на сонорные согласные, гравис как маркер безударного слога, сдвиг ударения влево.

2 ударение точно над гласной, не далее третьего слога от конца, четкие правила употребления разных знаков акцентуации.



Picture 3: Италия, надпись на металле минускулом

Все фотографии фиксируются в учетной документации, которая помимо экстралингвистических данных (дата, регион, палеографические характеристики etc) указывает и некоторые формальные параметры для дальнейшего анализа систем акцентуации:

- наличие акцентных знаков над всеми словами,
- смещение ударения, с указанием типа смещения,
- наличие придыхания внутри слова,
- ударение над сокращениями,
- ударение над первым элементом дифтонга,
- ударения над безударными слогами,
- ударение дальше, чем третий слог от конца.

Каждая надпись или папирус имеет справочную карточку в общем учетном файле вида (рис. 4)³:

	<p>ВСН1590 – 1455-1512 гг., Мистра (bch23.121.10.1) Палеография: маюскульная надпись, буквы с имитацией объема, сокращение "К," для <u>καί</u> внутри слова. Акцентуация: ударения стоят почти над всеми словами и над сокращениями. Густое придыхание внутри слова над второй из следующих подряд гласных, ударение в дифтонге со смещением влево на первый гласный.</p>
--	---

Рис. 4: Пример сопроводительной карточки к надписи.

Каждая из акцентуированных печатей или рукописей представлена в виде фотографий с лингвистическими комментариями и гиперссылкой на электронную базу:



Феофилакт (XI в.) BZS.1958.106.1725
 +ΟΜΩ
 ΝΥΜΟΝ
 ΣΟΝΟΙΚΕΤ
 ΜΑΚΑΡ
 ΣΚΕΠ
 ὁμόνυμόν σον οἰκέτ(ην), μάκαρ, σκέπ(οις)
 Придыхание ровно над гласной, акут над 3 и 4 словом, в 4 слове сдвинут вправо.

Рис. 5 BZS.1958.106.1725 Феофилакт (XI в.)

³ В скобках ссылка на файл с разбором всех акцентуированных слов

[4] Результаты и дискуссия

4.1 Теги и их варианты при разметке в зависимости от категории

В результате анализа данных, вошедших в корпус на сегодняшний день, были выявлены 8 систем акцентуации. Для удобства пользования каждой из систем приписывался тег (alex(andrine), dor(ic), byz(antine), mix(ed), log(ic), l-or (left-oriented), r-or (right-oriented), l-r-or (left-right-oriented)), и все акцентуированные слова маркировались этими тегами. В случае, если акцентуация в слове могла быть интерпретирована с точки зрения разных систем, то приписывались теги всех систем. За норму отсчета принималась позиция ударения по словарю, т.е. как принято в византийской системе акцентуации. Поскольку акцентуация литературных папирусов, стихотворных легенд печатей и текстов надписей часто связана с метрическими законами, использовалась отдельная система тегов для поэтических размеров (hex(ametr), 12 (=dodekasyllabos) etc.) и маркировки типов текстов (pr(osaic), po(etic), doc(umental) etc). Для надписей указывался их жанр: f(unerary) погребальные, bu(ilding) строительные, d(edication) посвячительные, inv(ocation) молитвы etc.

Независимо от системы акцентуации, все акцентуированные позиции маркировались по критериям:

- соответствует ударению по словарю или правилам грамматики: tg (true grammatic = соответствует форме в словаре или акцентному типу согласно морфологической парадигме) или нет — ntg;
- произошло ли смещение ударения (shift), если нет, то тег nsh;
- куда произошло смещение ударения влево — sh_l (shift_left) или вправо — sh_r (shift_right) и на сколько;
- в случае дифтонга: первый или второй компонент оказался под ударением, d1 и d2 соответственно. Если знаков несколько и они оказались на разных частях дифтонга, то для каждого знака указывается его вариант через /;
- соответствует типу ударения, ожидаемому в этой позиции tt (true type) или нет — ntt.

Такие маркеры помогают определить при анализе, к какой из систем будет отнесена та или иная словоформа. Что в дальнейшем позволит не только выстроить удобную систему поиска, но и выявить, какие слова сохраняют собственную систему акцентуации независимо от общей системы акцентуации памятников, в которых они встречаются.

4.2 Соотношение орфографии и акцентуации

В византийских греческих надписях нет унификации орфографии и понятия нормы [ср. 14, 27, 28, 29], поэтому нельзя однозначно сказать, меняется ли функция у ударения в случае монофтонгизации дифтонгов или сокращения долготы гласной или ее формальной замены при написании на другой более краткий или долгий вариант. Поскольку часто в разбираемых источниках слова пишутся не так, как в словарях, это приводит к возникновению нетипичных сочетаний ударений. Например, так появляется циркумфлекс над о или над ε или два и больше ударения в слове. Эти особенности также маркируются отдельно для удобства поиска следующими тегами: orph (для отличной от словаря орфографии) и n1ac (no one accent, для нескольких ударений в слове). Так достигается основная задача разметки — учесть особенности написания каждого слова и при этом в сводном глоссарии представить эти особенности в зависимости от системы акцентуации. Сам глоссарий строится как дерево, в котором от унифицированной по словарю словоформы идут ветви по выбранной системе акцентуации. Для удобства анализа, создания системы поиска и дальнейшего использования при создании базы данных или обучения нейронной сети распознаванию и распределению фрагментов текстов, каждая словоформа представлена в виде картинка, вырезанной из конкретного памятника. Такая картинка сопровождается комментариями: 1) лингвистическим; 2) палеографическим; 3) акцентологическим со всеми необходимыми для дальнейшей работы тегами. Примеры слов из глоссария, вырезанных из афонских рукописей, представлены на рисунках 6-9:

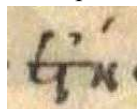


Рис. 6: ἐτή

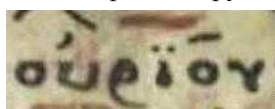


Рис. 7: οὐρεῖον

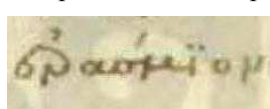


Рис. 8: ἐράσιον

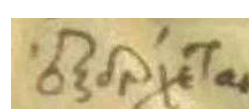


Рис. 9: ἐξέρχεται

Такая разметка показывает, что в некоторых случаях акцентуация копируется вместе с орфографией. Можно сказать, что ряд слов имели устойчивое написание по этим двум показателям (акцентуация и орфография) в определенные периоды или в конкретных регионах. В случае регионов мы можем

отнести эти особенности к письменной фиксации диалектного произношения. Встречаются также примеры акцентуации, которые объясняются фонологически, как произошедшие под влиянием произношения. Среди них появление густого придыхания внутри слова на стыке двух гласных, после дифтонга или гласной над следующей гласной, чтобы избежать зияния, например, *δέησιν* или *παλαιόλογος*. Что, в свою очередь, помечается тегом *pron* (=pronunciation).

4.3 Подготовка к использованию алгоритма для полуавтоматической разметки и принципы кодирования текста

Анализ результатов распознавания надписей и фрагментов рукописей разными онлайн программами OCR для древнегреческого языка показал, что знаки акцентуации были распознаны неточно и часто унифицировались согласно принятым в византийской системе акцентуации правилам. Поэтому использование результатов этого распознавания в дальнейшей работе на данном этапе невозможно. Это привело к необходимости после выполненной вручную разметки ряда надписей создать алгоритм полуавтоматической разметки и потом проверить его работу на практике.

На данном этапе работы все тексты корпуса фиксируются в общем файле, разделенном по разделам: рукописи, папирусы, лапидарные надписи, фресковые надписи, надписи на предметах прикладного искусства, граффити, печати. К каждой надписи составляется отдельный учетный лист (рис. 10), где представлены все акцентуированные слова по строкам, и к каждому слову указаны соответствующие теги.

расшифровка	прав. написание	перевод	формула	место находки	дата	знаки	граммат.	сдвиг	дифтонг	без гема	соответ. типа	орф. тип	акцентуации
αἰσθησῖν	αἰσθησῖν	чувство	d1sb/d2aCCVLCVtC	Феры		1104 sb, a, t	ntg	sbsb_l, nash	d1sb/d2a	nfr	tt		alex
αὐτᾶ	αὐτᾶ	эти	d1sbCVg	Феры		1104 sb, g	ntg	sbsb_l, ngsh	d1		tt		alex
δοστ' οτην	δοσπότην	господин	CVSCabVS(CVLC)	Феры		1104 a	ntg	ash_lb			tt		alex, l-or
ἐμπικραίνων	ἐμπικραίνων	рассердившись	sbVSCCVtCCabCVLC	Феры		1104 sb, t, a	ntg	ash_lb	db	nfr	tt		alex, l-or
ἐυβαλές	εὐβαλές	пышно цветущий	d1sbCVCVSaC	Феры		1104 sb, a	ntg	rbsh_l, ash_rb	d1		tt		byz
θησαυρίσας	θησαυρίσας	сохраненным	CVLCdCtaCdC	Феры		1104 ta	tg	nsh		nfr	tt		byz
καί	καί	и	Cd1g	Феры		1104 g	ntg	gsh_lb	d1b		tt		alex, l-or
καταφύγευσαις	καταφυτεύσαις	обсаженным	CVVCVCd2CdC	Феры		1104 a	tg	nsh	d2		tt		byz
πάλιν	πάλιν	обратно	CVcir2CVtC	Феры		1104 cir1, t	ntg	cirsh_r		nfr	ntt		byz, r-or
τοῦ	τοῦ	артиклъ В.п.	CVSCg	Феры		1104 g	ntg	ash_r			tt		r-or
τούτου	τούτου	вот этого	Cd1aCVSC	Феры		1104 a	ntg	ash_l	d1		tt		r-or
τρως	τρως	роскошь	CVCVLCirC	Феры		1104 cir1	tg	nsh			ntt		byz

Рис. 10: Надпись из Адрианополя 976-1064 гг.

Из таких учетных листов, собранных вместе и ранжированных по алфавиту, собирается общий глоссарий. Приведем фрагмент, сделанный на основе надписей разного типа для лексемы «έτος» (рис. 11):

расшифровка	прав. написание	перевод	формула	место находки	дата	материал	знаки	граммат.	сдвиг	дифтонг	бсг	ор	соответ. типа	тип акцентуации	жанр	тип текста	в формуле	название файла	публикация
	ἐτ' ἔς	ἔτους	года	VsrbCa(d)C	Адрианополь	976-1064	мрамор	rb1, a2	ntg	ash_r_c	nd		rbnnt, att	r-or	ded	pr	да	DSC02233adrianopol	Asdracha 108
	ἐτ	ἔτ(ουc)	года	aCVS	Афины	1062	граффито	a/rb	ntg	a/rbsh_l	nd		att/rbnnt	l-or	f	pr	да	DSC00413	А. Капустин 11, №17
	ετ'	ἔτ(ουc)	года	C/VSa	Италия	1083-1084	мрамор	a	ntg	ash_r	nd		att	byz+r-or	bu	pr	да	DSC0248ital	Guillou 131
	ἐτ	ἔτ(ουc)	года	sbaVS(d)C	Комопины	4 февраля 1111	мрамор	arb	ntg	nsh	nd		rbnnt, att	byz+l-or	f	pr	да	DSC08791	Asdracha 63
	ἔτος	ἔτος	года	rbVSaCVSC	Италия	1131-1132	мрамор	arb	ntg	rbsh_l, nash	nd		rbnnt, att	byz+l-or	f	pr	да	DSC03221ital	Guillou 112
	ετ'	ἔτ(ουc)	года	C/VSa	Афины	1153	граффито	a	ntg	nsh	nd		tt	byz	f	pr	да	DSC02388athen	Orlandos 39
	ετ'	ἔτ(ουc)	года	C/Vsa	Афины	1160	граффито	a	ntg	ash_r	nd		tt	byz	f	pr	да	DSC02389athen	Orlandos 40
	ἐτ	ἔτ(ουc)	года	rbCVS	Италия	7 августа 1186	мрамор	asb	ntg	sbsb_l+ash_l	nd		rbnnt, att	byz+l-or	f	po	да	DSC03212ital	Guillou 104
	ἐτόc	ἔτους	года	sbVS/C(d)C	Накос	1309	фреска	sb	ntg	sbsb_l	nd		tt	byz+l-or	bu/f	pr	да	DSC08594	Μουτσοπούλου 20.1
	ε'τ'(ουc)	ἔτ(ουc)	года	Vsa(CdC)	Афины	1314	граффито	a	ntg	ash_r	nd		tt	r-or	f	pr	да	DSC02385athen	Orlandos 26
	ἔτ(ουc)	ἔτ(ουc)	года	sbaVS/C(dC)	Трикалаон, Метеора	1548	фреска	arb	tg	nsh	nd		rbnnt, att	byz	bu	pr	да	DSC08594	Μουτσοπούλου 22.3
	ἔται	ἔται	года	VsrbCa(d)	Трикалаон, Каламба	1766	фреска	asb	tg	sbsb_l, nash	nd		tt	byz+l-or	bu/f	pr	да	DSC08595	Μουτσοπούλου 24.2

Рис. 11: Фрагмент подготовительного файла для глоссария с лексемой ἔτους

При этом к каждой лексеме делается сопроводительный формульный лист (табл. 1), составленный согласно византийской системе акцентуации для удобства анализа орфографических вариаций:

нач. форма / В.п. ед.ч.	перевод	формула	Р.п. ед.ч	формула	Д.п. ед.ч	формула	Им.п. мн.ч./ В.п. мн.ч.	формула	Р.п. мн.ч.	формула	Д.п. мн.ч.	формула
ἔτος	год	VSsbaCVSC	ἔτεος	VSsbaCVSVSC	ἔτεϊ	VSsbaCVSt	ἔτη	VSsbaCVL	ἔτων	VSsbCVLcirC	ἔτεσι	VSsbaCVSCV
			ἔτους	VSsbaCdC								

Табл. 1: Формульный лист для ἔτος

Для обработки нового материала используется следующий порядок действий:

Сначала ставится тег по названию знака, который использован: a=acute (´), g=gravis (˘), cir = circumflex (ˆ), t=trema (¨), rb = rough breathing (ˆ), sb = smooth breathing (˘). Для определения знака в дальнейшем с использованием нейросети для каждого из них собраны графические варианты с учетом всех палеографических особенностей. То же сделано и для случаев соединения знаков, которые прописываются отдельно, например, rba, sba, ta, tcig и т.д. В случае, если над одной лексемой присутствуют два знака на разных слогах, то проставляются теги по числу вхождений и, если они одинаковые, то для определения их порядка в слове им присваивается номер от 1 до 5. Если циркумфлекс оказывается над о или ε, то ставится дополнительный тег ntg. Такой же тег добавляется в случаях, когда придыхание зафиксировано не на первой гласной в слове, если только это не дифтонг. На втором шаге определяется местоположение знака. Если ровно над гласной, то ставится тег nsh (no shift), если нет, то маркируется сдвиг вправо sh_r или влево sh_l. При сдвиге цифра рядом с тегом показывает на сколько элементов сдвинуто: 1 — на предыдущую или последующую букву, 2 — на пробел дальше и 3 — на букву через одну. Если знак попадает между буквами, то добавляется тег b (=between). В случае, если знак длинный и попадает на две буквы, это маркируется отдельно l (=long), а если точкой, то p (=point). Для обучения нейронной сети автоматическому определению, находится ли знак над гласной или нет, собран из размеченного вручную материала отдельный корпус примеров для каждой гласной и каждого типа гласных, иллюстрирующий, когда знак ровно над гласной, а когда нет. Все случаи со сдвигами маркируются также тегом ntg.

На следующем этапе предлагается все слова приводить в базе к схематичному виду, например, для ἀνέκενϑοῦ — sbVCVsaCVSVLCCVt, где VS - краткий гласный (ε, ο), VL — долгий гласный (η, ω), V — гласный, который может быть и кратким, и долгим (α, ι, υ), C — согласный. Все буквы заданы списками с вариантами их написания по палеографическим схемам (пример таковых, [21]). Маленькими буквами обозначены акцентные знаки, зафиксированные после того знака, к которому

относятся. Если придыхание оказывается до гласного, то тег пишется до гласного. Если знак сдвинут и попадает между буквами, то в схеме после него ставится тег *b*. В случае горизонтальных лигатур (рис. 12) буквы, в них включенные, пишутся в строчном виде *Cv*, где первый знак с большой буквы, а второй с маленькой и в случае вертикальных лигатур (рис. 13) *C/C*, где первой пишется верхняя буква. Ряд классических лигатур для дифтонгов, как δ — *ov*, сразу записываются в виде *(d)*. В фигурных скобках пишутся принятые сокращения с их расшифровкой, например, вместо $\overset{\circ}{\circ} = \delta v$ {*VgC*}.

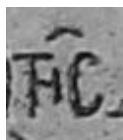


Рис. 12 пример горизонтальной лигатуры

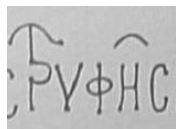


Рис. 13: Пример вертикальной лигатуры

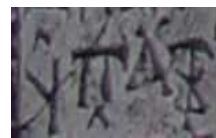


Рис. 14: Пример для кодирования с дифтонгом и лигатурой

Если в слове две последовательные гласные, то определяется дифтонги они или нет. Для этого проверяется, является ли вторая гласная *v* или *i* и написана ли она без тремы. В случае дифтонга под ударением маркируется на какой элемент дифтонга падает ударение и ставятся соответствующие теги *d1* и *d2*. Если ударение оказывается между, то пишется тег *db*. Что приводит к корректировке схемы, если она изначально была взята не из ручной разметки, где эти теги проставляются сразу, например, для слова $\acute{\upsilon}\lambda\acute{\alpha}\tau\omicron\upsilon$ (см. рис. 14) *VrbCVaC/(VSV)* переходит в *VrbCVaC/(d)*.

После анализа состава слова по описанным выше схемам определяется корректно ли согласно грамматике и словарям место ударения, соответствует ли слово своему акцентному типу и акцентной парадигме. Для удобства определения можно использовать правила, описанные Тронским И.М. [30, 49-50] и представленные через введенные выше принципы кодирования. Например, для акута — окситоны (ударение на последнем слоге): *VCVCVsaC*, *CVCCVSCVLaC*, парокситоны (ударение на предпоследнем слоге): *d/VCd/VLaCVLC*, *VCVsaCd*, *CVSCVaCVS*, пропарокситоны (ударение на 3 слоге от конца слова): *VaCCVLCVSC*, *CVsaCVSCVSC*, *CVSCVLaCVCV*. для циркумфлекса — периспомены (ударение на последнем слоге): *VCVSCVLCirC*, пропериспомены (ударение на предпоследнем слоге): *VCVLCdcirVSC*. для грависа — *CVSCVsg*, *VCVCVsgC*, *CVCCVSCVlgC*. Дополнить возможные схемы можно, используя файл, сделанный из словаря древнегреческого языка [8] путем кодировки по нашей системе, где один столбец слово, другой его код. При необходимости в третий столбец можно добавить морфологические показатели по классификатору. Тогда этот файл можно будет использовать как проверочный для идентификации возникших в слове орфографических вариантов.

Если после слова стоит энклитика и с нее переносится ударение, то это маркируется отдельно тегом *en* после всех тегов этого ударения. Это необходимо, чтобы разграничить случаи, когда в слове появлялось два и более ударения на безударных слогах, как было принято в александрийской системе акцентуации.

В случае наличия тремы маркируется тегом *tr*, если ее положение соответствует ее функциям, а если оно фонетически не обусловлено, то тегом *ntr*.

В случае наличия традиционных сокращений слов и ударения на них, ставится тег *abr*.

Для простановки тегов, фиксирующих типы систем акцентуации, используются следующие принципы:

alex — в случаях, когда в итоговой схеме гравис оказывается не на последнем слоге тег *gntg* или сочетание тегов *g*, *ntg*, *gsh_1l*. А также фиксируется ударение на первом элементе дифтонга *d1*, или тег *sh_1*, или слово содержит несколько ударений и тег *n1ac*, но за ним нет энклитики и как следствие тега *en*.

byz ставится, если есть теги *nsh*, *d2*, *tt*, *tg*.

g-or — если в графе со сдвигом находится тег, содержащий в своем составе сочетания *sh_g*.

l-or — если в графе со сдвигом находится тег, содержащий в своем составе сочетания *sh_1l*.

l-g-or — если в графе со сдвигом находится тег, содержащий в своем составе сочетания *sh_1l* и *sh_g*.

log — при маркировке ударением сокращений и наличия тега *abr*.

4.4 Одинаковые тексты и их акцентуация

Греческие надписи, особенно, на фресках часто содержат фрагменты одинакового текста. Это могут быть цитаты из литургических текстов или формульные словосочетания или фразы. Отдельно при разметке выявляются одинаковые тексты и их фрагменты в разных регионах, поскольку анализ в

рамках византийских надписей Грузии показал, что даже в одном регионе одинаковый текст может акцентуироваться по-разному. Рассмотрим на примере трех надписей (рис. 15, 16 и 17) из Грузии [5,6] и Касторьи [10]:



Рис. 15: Грузия XIVв.

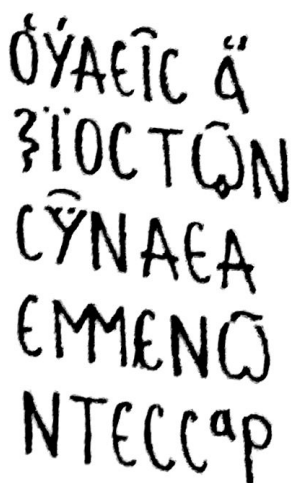


Рис. 16: Грузия XIVв.

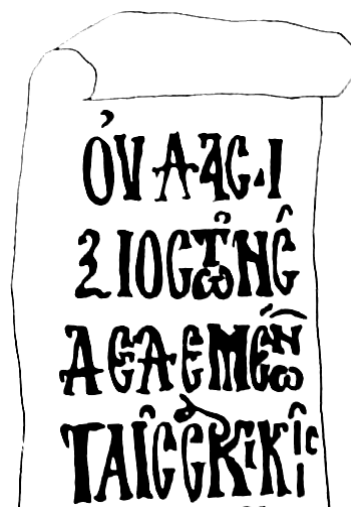


Рис. 17: Касторья 1432 г.

Близкие по времени три варианта акцентуации цитаты из херувимской песни: «Οὐδεὶς ἄξιός τῶν συνδεδεμένων ταῖς σαρκικαῖς» (d2sbCd2gC VsbaCVVSC CVLcirC CVCCVSCVSCV SaCVLC Cd2cirC CVCCVCd2cirC) показывают, что при разнице орфографии везде использованы элементы александрийской системы, но в разных местах. В таблице с формулами для каждой надписи подчеркнуты места, которые можно интерпретировать, как александрийскую систему (табл. 2):

Грузия XIVв.	Грузия XIVв.	Касторья 1432 г.
<u>d1rbb</u> Cd <u>Cg</u> VrbaCVLVSC CVScirbC C(VL <u>Cg</u>)CVSCVS()	<u>d1rb</u> / <u>d2a</u> Cd2cirC VrbaCVtVSC CVLcirC <u>CVcirt</u> CCVSVS(CC)VSC <u>VLcir</u> C CVSC CVC()	<u>d1sb</u> C(d)C VCVVSC C/VLcirC C()CVSCVSCV SaC/VL(C) Cd2cirbC V/C(CC)VC(<u>d1</u>) <u>cir</u> C

Табл. 2: Кодировка надписей, цитирующих одинаковый текст.

Выявление таких формульных обозначений для характерных элементов разных систем акцентуации позволит их в дальнейшем выделять автоматически.

4.5 Синонимия знаков акцентуации или проблема наложения значения

По данным разных памятников в процессе разметки было выявлено, что некоторые знаки акцентуации использовались в нескольких значениях одновременно, независимо от ведущей системы акцентуации [26]. Этот факт позволяет нам говорить применительно к знакам акцентуации о таких явлениях как синонимия, омонимия и других применимых традиционно к лексике понятиях. В таких случаях при разметке теги приводятся через «/», что потом позволит учитывать оба варианта и определить самые частотные сочетания. Например, волнистая линия могла быть циркумфлексом рис. 16 во второй строке или знаком сокращения, как на рис. 17 в третьей строке, а также знаком переноса (тег hy (=hyphen) [9, 26]), как на том же рис. 16 в 4 строке. При этом последний можно интерпретировать и как циркумфлекс на безударном слоге, что можно закодировать так: CVcirtCCVSVS(CC)VSCVLcir/hyC.

[5] Выводы

На данном этапе исследования для разработки системы тегирования и кодирования были вручную размечены надписи из Каппадокии, как граффити, так и фресковые, надписи из Грузии, надписи на металле из Италии, граффити из Киева, Константинополя, Афин, Кипра, Крита, надписи на фресках из Кипра, Северной Греции и ряд надписей из других регионов для сравнения. В результате было выявлено, что использование систем акцентуации в разных регионах обусловлено не только

материалом и типом надписи, но господствующим диалектом. В Каппадокии такой характеристикой стала особая традиция использования тремы, частые ударения на сокращениях, как элемент смысловой системы. Анализ подборки надписей на металле из Италии показал, что в них часто используется александрийская система акцентуации и различные вариации сочетаний систем со сдвигом вправо и влево, как в сфрагистических памятниках. В ряде случаев наблюдается ошибочный выбор знака или их путаница по палеографическим причинам, когда знак по начертанию похож на точку, что нами также было зафиксировано и на материале печатей. Грузинские надписи, цитирующие литургические тексты, предпочтительно используют элементы александрийской системы акцентуации. При этом не наблюдается связи между сохранением правильной орфографии и использованием правильной, с точки зрения византийской системы, акцентуации. Однако, некоторые частотные слова сохраняли свою акцентуацию независимо от других факторов.

Применение предложенного типа кодирования для остального корпуса позволит проверить, насколько бытовавшие в разное время варианты написания слов отклоняются от описанных ранее в исследованиях акцентных парадигм. Как влияет на эти отклонения выбранная автором система акцентуации и можно ли говорить о влиянии орфографических вариаций, в целом, на изменения в акцентуации. Последовательное кодирование памятников позволит автоматически тегировать источники по системам акцентуации и создать систему корректировки акцентуации в процессе распознавания такого рода текстов.

Предполагается в дальнейшем написать на основе сформулированных принципов тегирования программу, которая бы осуществляла частичную автоматическую разметку подобных текстов. Результаты этой разметки после верификации также можно будет использовать для обучения нейронных сетей классификации текстов по их системам акцентуации и распределению по акцентным парадигмам. Последнее применимо не только для византийского греческого, но и для других языков и письменных памятников, содержащих акцентные знаки. Что было нами выборочно проверено на материале славянских языков, представленного в книге Дыбо В.А. [16].

Acknowledgements

Выражаю огромную благодарность своим коллегам: Дыбо В.А., Казанскому Н.Н., Афиногенову Д.Е., Степановой Е.В., Степаненко В.П., Алексеенко Н.А., Б.Л. Фонкичу, Д.И. Коломацкому, которые поддерживали меня на разных стадиях работы над корпусом.

References

- [1] Biondi A. Gli accenti nei papyri greci biblici. — Roma: Papyrologica Castrociquitana, 1983.
- [2] Hinge G. Der dorische Akzent in den hellenistischen Korpora der archaischen Meliker. — 2001–2006 — Access mode: <http://alkman.glossa.dk/akzent.html>.
- [3] Fournet Jean-Luc. Hellénisme dans l'Égypte du Vie siècle. La bibliothèque et l'oeuvre de Dioscore d'Aphrodité MIFAO 115/1 Caire: 1999. P. 22.
- [4] Fournet Jean-Luc. L'influence des usages littéraires sur l'écriture des documents : perspectives, Proceedings of the 20th International Congress of Papyrologists. Copenhagen: 1994. P. 418-422.
- [5] Kauhchischvili T. Korpus der griechischen Inschriften in Georgien — Tbilisi: 1999 — Vol.1 (on Georgian).
- [6] Kauhchischvili T. Korpus der griechischen Inschriften in Georgien — Tbilisi: 2002 — Vol. 2-3 (on Georgian).
- [7] Laum B. Das Alexandrinische Akzentuationssystem unter Zugrundelegung der theoretischen Lehren der Grammatiker und mit Heranziehung der praktischen Verwendung in den Papyri. — Paderborn: Druck und Verlag von Ferdinand Schoeningh, 1928.
- [8] Liddell H.G., Scott R. A Greek-English Lexicon. Oxford: Clarendon Press, 1940. (https://www.perseus.tufts.edu/hopper/text?doc=Perseus%3Atext%3A1999.04.0057%3AAalphabetic+letter%3D*a%3Aentry+group%3D1%3Aentry%3D*a)
- [9] Murphy D. (1995) Hyphens in Greek Manuscripts, Greek, Roman, and Byzantine Studies, 36, pp. 293-314.
- [10] Mutsopulu N. K. Contribution a la morphologie de l'écriture hellénique: album d'inscriptions byzantines et post-byzantines. — Thessaloniki: 1977.
- [11] Reggiani N. Digital Papyrology I. Methods, Tools and Trends. — Berlin/Boston: 2017.
- [12] Reggiani N. Digital Papyrology II. Case Studies on the Digital Edition of Ancient Greek Papyri. — Berlin/Boston: 2018.
- [13] Seider R. Paläographie der griechischen Papyri. — Stuttgart: Anton Hiersmann, 1967. — Bd. 1.
- [14] Ševčenko I. (1981) Levels of style in Byzantine prose, Jahrbuch der Österreichischen Byzantinistik 31.1. P. 289–312.
- [15] Turner E. G. Greek manuscripts of ancient world. — Oxford: Clarendon Press, 1971.

- [16] Дыбо В.А., Замятина Г.И., Николаев С.Л. Основы славянской акцентологии. — Москва: Наука, 1990.
- [17] Евдокимова А.А. Языковые особенности греческих граффити Софии Киевской. — Москва: 2008 (диссертация).
- [18] Евдокимова А.А. Александрийская и византийская системы акцентуации в греческих акцентуированных надписях провинциального происхождения. // Индоевропейское языкознание и классическая филология 13. — Санкт-Петербург: 2009, с. 164–171.
- [19] Евдокимова А.А. Акцентуированные папирусы из коллекции Кельнского университета (предварительные замечания). // Индоевропейское языкознание и классическая филология 14. — Санкт-Петербург: 2010, с. 277–295.
- [20] Евдокимова А. А. Системы акцентуации греческих папирусов и Парфений Алкмана. // Аристей. — Москва: 2010, с. 41–52.
- [21] Евдокимова А.А. Палеография греческих граффити Софии Константинопольской, Софии Киевской и Софии Новгородской: сравнительный анализ. // Аристей VII. — Москва: 2013, с. 132–180
- [22] Евдокимова А.А. Диалог византийской и александрийской систем акцентуации в греческих граффити из разных балканских памятников. // Балканские чтения 15. — Москва: 2019, с. 31–38.
- [23] Евдокимова А.А. Системы акцентуации в легендах византийских моливдовулов XI–XII вв. Из коллекции Dumbarton Oaks. // В XII международной византийский семинар ΧΕΡΣΟΝΟΣ ΘΕΜΑΤΑ: Империя и полис. — Симферополь: 2021, с. 111–126.
- [24] Евдокимова А.А. Системы акцентуации в греческих рукописях из коллекции Афона. Ч. I. // Индоевропейское языкознание и классическая филология 25.1. — Санкт-Петербург: 2021, с. 341–368.
- [25] Евдокимова А.А. Системы акцентуации в византийских надписях на фресках богородичных церквей Троодоса (Кипр) // XIII международный византийский семинар ΧΕΡΣΟΝΟΣ ΘΕΜΑΤΑ: Империя и полис. — Симферополь: 2022, 83-92.
- [26] Евдокимова А.А. Синонимия акцентных знаков или знак переноса (ἐνωτικόν) в византийских надписях // Индоевропейское языкознание и классическая филология 26.1 — Санкт-Петербург: 2022, с. 387-401.
- [27] Казанский Н.Н. Между письменным и устным текстом: древнегреческое «наивное» письмо. // Исследования по языкознанию: к 70-летию члена-корреспондента РАН А.В. Бондарко. — Санкт-Петербург: 2001, с. 246–256.
- [28] Казанский Н.Н. Вариативность диалектной нормы и история древнегреческого языка. // Язык, литература, эпос: к 100-летию со дня рождения академика В. М. Жирмунского. — Санкт-Петербург: 2001, с. 82–99.
- [29] Кисилиер М.Л. Анализ средневековых текстов и социолингвистические факторы (на материале «Луга духовного» Иоанна Мосха) // Юрьевские чтения: Материалы междисциплинарной конференции молодых филологов. Выпуск II. — Санкт-Петербург: 2002, с. 41–48.
- [30] Тронский И. М. Древнегреческое ударение — Москва, Санкт-Петербург: 1962.