

А. В. Андреев
(Институт лингвистических исследований РАН, Санкт-Петербург)

**ОРФОГРАФИЧЕСКИЕ ПРИНЦИПЫ М. ДАУКШИ:
К ВОПРОСУ ОБ ОТНОСИТЕЛЬНОЙ ХРОНОЛОГИИ
«ПОСТИЛЛЫ» И «КАТЕХИЗИСА»¹**

В статье производится статистический анализ вариативности орфографии «Постиллы» и «Катехизиса» М. Даукши. Показана стабильность основных орфографических принципов в обоих текстах и большее разнообразие графических вариантов в «Катехизисе», что позволяет считать «Катехизис» памятником, более ранним по времени создания, а не только по времени публикации. Обнаружены признаки, подтверждающие предположение о двухэтапной печати текста «Постиллы».

Ключевые слова: история литовского языка, орфография, статистический анализ.

A. V. Andreev
(Institute for Linguistic Studies, RAS, St. Petersburg)

**M. Daukša's spelling principles: on the problem of relative chronology
of *Postilla* and *Cathechism***

In the article, the spelling variance of M. Daukša's *Postilla* and *Cathechism* is statistically analyzed. The stability of basic rules is demonstrated, but a greater diversity in secondary features allows to conjecture "Cathechism" to be written (and not just published) earlier than "Postilla". There is evidence in spelling peculiarities that "Postilla" was indeed published in two stages.

Keywords: history of Lithuanian language, spelling variance, statistical analysis

Хотя «Катехизис» М. Даукши является первым (известным) текстом, изданным на территории Великого Княжества Литовского на литовском языке (1595 г.), а гораздо более обширная «Постилла» была опубликована только через 4 года, давно известно, что подготовка к печати «Постиллы» началась одновременно с «Катехизисом» или даже раньше него; возможно, что к 1595 году текст «Постиллы» был уже частично набран, но

¹ Работа выполнена при финансовой поддержке Российского научного фонда, грант № 17-18-01624

из-за «финансовых затруднений» публикация была отложена (Lebedys 1963: 236). Это было установлено благодаря анализу документов, связанных с историей иезуитской типографии в Вильнюсе (Ivinskis 1953: 47–49). В связи с этим возникает вопрос, возможно ли отыскать в самих этих текстах следы истории их издания и — шире — существуют ли какие-то характеристики текстов, которые могли бы пролить свет на становление орфографических принципов М. Даукши.

В поисках таких характеристик мы обратимся к орфографической вариативности текстов М. Даукши. Логически мыслимы три возможности: 1) вариативность остается более-менее неизменной и достаточно низкой на протяжении текста — это, очевидно, верно для большинства текстов современной печати; 2) вариативность увеличивается за счет накопления ошибок — это было количественно измерено, например, для древнерусских рукописей (Zaliznyak 2008: 134); 3) вариативность будет уменьшаться, потому что по мере работы над текстом орфографические принципы будут стабилизироваться — такие примеры отыскать нелегко, потому что вообще ситуация разработки автором орфографии не слишком часто встречается, но в случае с М. Даукшей мы имеем дело именно с такой ситуацией, так что а priori можно было бы ожидать в рассматриваемых текстах именно такой динамики вариативности.

В Andreev 2013 и Thies 2019: 36–87 выявлены следующие основные элементы орфографической вариативности:

1. ударные / безударные слова
2. использование в качестве знака ударения акута / циркумфлекса / точки
3. использование носового гласного в сочетаниях гласный + носовой согласный
4. удвоенное / одиночное *s* в середине слова
5. передача открытого /e/ через <e> («e caudata») или <ě>
6. передача /i/ через *y/i*

Статус этих вариантов в плане обусловленности, однако, весьма различен. Так, хотя правила распределения ударных и безударных слов остаются по большому счету невыясненными, никто, кажется, не сомневается, что они должны быть как-то связаны с синтаксической и/или интонационной структурой текста. С другой стороны, точка и акут всеми исследователями

признаются просто графическими вариантами одного знака². Остальные варианты образуют своего рода спектр между этими двумя полюсами:

— в Andreev 2013 было показано, что противопоставление акута и циркумфлекса отражает, хотя и не вполне последовательно противопоставление долгих и кратких гласных;

— в Andreev 2021 демонстрируется, что вариативность <y/i> связана в первую очередь с особым выделением глагола-связки *ura* ‘есть (3 л. наст. вр.)’;

— вариативность в написании сочетаний вида VN связана, очевидно, с одной стороны с нейтрализацией фонологического противопоставления носовых/неносовых гласных в позиции перед носовым согласным, а с другой стороны — с влиянием польской орфографии;

— колебания в написании <s/ss> должны объясняться влиянием латинско-немецкой традиции, где противопоставление одиночного и удвоенного написания связано с озвончением интервокального /s/, т. е. с явлением, совершенно несвойственным как литовскому, так и польскому языкам;

— наконец, наличие двух знаков для открытого /e/ труднее всего поддается объяснению, хотя знак *e caudata* характерен для латинских текстов (Bischoff 1990: 122), а употребление уменьшенных надстрочных букв — это скорее черта немецких памятников. Таким образом, здесь мы тоже, скорее всего, имеем дело с взаимодействием разных орфографических традиций; при этом нельзя не отметить, что в латинских текстах *e caudata* используется для передачи дифтонга /ae/, так что М. Даукша здесь творчески переосмысляет правила латинской орфографии³.

Математико-статистический аппарат давно используется для установления истории текстов и их сравнительной хронологии (Faier 2002; Litvinov 2001), однако, как правило, речь идет об исследовании изменений тех или иных измеримых лингвистических характеристик (напр., длины слова) на про-

² До недавнего времени вообще господствовало мнение, что точка — это дефект печати (Zinkevičius 1988: 186), но последние издания позволяют эту точку зрения уверенно опровергнуть.

³ Аналогичное употребление *e caudata* встречается, впрочем, и в древнеанглийских текстах (Cain 2016), хотя М. Даукше оно вряд ли могло быть известно.

тяжении текста методом «скользящего среднего» и т. д. Между тем, существуют математические методы, которые позволяют непосредственно измерять степень вариативности тех или иных элементов текста; мы имеем в виду центральное для теории информации понятия *энтропии* и *относительной энтропии* (Kullback, Leibler 1951), которые находят широкое применение в различных количественных гуманитарных исследованиях, однако филологи их незаслуженно обходят стороной (ср., впрочем, Cisne et al. 2010).

Энтропия последовательности элементов из алфавита A , описываемая формулой
$$H(P) = - \sum_{x \in A} \frac{|P_x|}{|P|} \cdot \log_2 \left(\frac{|P_x|}{|P|} \right)$$
, содержательно обычно определяется как мера «разнообразия» или «избыточности» текста, однако нам представляется более удобным рассматривать ее в структуралистском ключе как степень того, насколько дистрибуция элементов в тексте является дополнительной: если энтропия равна 0, то элементы A являются вариантами одной единицы, а если энтропия максимальна для данного размера алфавита (напр., 1 для алфавита из двух элементов) — то элементы либо функционально противопоставлены, либо находятся в отношении свободного варьирования⁴.

Относительная энтропия текстов P и Q , определяемая формулой
$$D(P||Q) = \sum_{x \in A} \frac{|P_x|}{|P|} \cdot \log_2 \left(\frac{|P_x| \cdot |Q|}{|Q_x| \cdot |P|} \right)$$
, представляет собой меру того, насколько дистрибуция элементов в тексте Q отличается от дистрибуции Q , иными словами, оценку того, насколько правила распределения единиц, выведенные из текста Q применимы к P ⁵.

Материалом для нашего исследования послужили цифровые издания «Катехизиса» и «Постиллы», подготовленные Институтом литовского языка (DK 2006; DP 2006). Ограниченный объем памятников не позволяет проводить надежный статис-

⁴ Следует иметь в виду, что энтропия, как любая статистическая характеристика, позволяет давать только оценку истинного положения дел.

⁵ Относительная энтропия двух текстов в общем случае не выводится из значений энтропии этих текстов по отдельности. Возможна, хотя и нетипична ситуация, когда энтропия двух текстов одинакова, а их относительная энтропия максимальна.

тических анализ варьирования орфографии отдельных форм, поэтому расчет энтропии осуществлялся по отдельности для шести «дифференциальных признаков»: STRESS (ударность / безударность слова), AC (акут / циркумфлекс), AD (акут / точка), VN (неназальный гласный + назальный согласный / назальный гласный / назальный гласный + назальный согласный), SS (одинокое / удвоенное <s>), EA (*e caudata* / <*e^a*>), YI (<y> / <i>).

Измерение энтропии этих признаков обоих текстов в целом и их относительной энтропии дает следующие результаты:

Табл. 1. Значения энтропии и относительной энтропии «Постиллы» и «Катехизиса» для отдельных параметров.

Tabl. 1. Entropy and relative entropy values of Postilla and Cathecism for selected parameters.

	STRESS	AC	AD	VN	SS	EA	YI
$H(DK)$	1,0000	0,8427	0,3501	1,3332	0,8053	0,7478	0,1908
$H(DP)$	0,0000	0,0016	0,0281	0,0144	0,0633	2,1063	0,0001
$D(DK \parallel DP)$	1,0000	0,8121	0,1832	1,3556	0,5721	0,4119	0,1995
$D(DP \parallel DK)$	0,0000	0,0015	0,0214	0,0149	0,0543	1,6609	0,0001

Можно заметить, что:

— энтропия распределения <y/i> наименьшая в обоих текстах, что согласуется с тем, что эти знаки находятся в отношении дополнительной дистрибуции (Andreev 2021);

— энтропия всех признаков, имеющих четкие лингвистические корреляты (STRESS, AC, YI, VN) почти одинакова и в «Катехизисе» и в «Постилле» и относительная энтропия в обе стороны так же невелика, в то время, как у «графических» признаков энтропия существенно отличается;

— в целом энтропия «Катехизиса» выше, чем у «Постиллы»; этот факт заслуживает пристального внимания, поскольку «Постилла» гораздо больше по объему и гораздо разнообразнее по содержанию, так что можно было бы ожидать обратного эффекта;

— бóльшая энтропия признаков VN и YI в «Постилле» по сравнению с «Катехизисом», скорее всего, связана с наличием в «Постилле» значительных фрагментов на польском языке, в котором, разумеется, правила употребления носовых гласных и <y> отличаются от литовского.

Обратимся теперь к анализу динамики энтропии в рамках отдельного текста. Для этого мы будем описывать относитель-

ную энтропию частей этих текстов; тексты разбиваются на две части, примерно равные по объему, посредством процедуры, основанной на так называемых функциях Уолша, позаимствованной нами из биоинформатики (Stamatakis 2019). Ниже приводятся максимальные значения относительной энтропии для каждого признака в обоих текстах для 702 возможных разбиений⁶ (разбиения, для которых достигается наибольшая относительная энтропия обозначены чередующимся белыми и черными кружками):

Табл. 2. Относительная энтропия разбиений текстов «Постиллы» и «Катехизиса» для отдельных параметров.

Tabl. 2. Relative entropy of sections of Postilla and Cathecism for selected parameters.

	DK	DP
STRESS	0,0195 (●○)	0,0238 (●○○○●●)
AC	0,0075 (●○○●)	0,0420 (●○○○●○●●●●○○)
AD	0,0063 (●○○○●●, ○●○○●●○, ○○○●○○●○○●)	0,0134 (○●●●○○)
VN	0,0149 (●○●○○●○○○○○)	0,0074 (●○○○○●○○●●)
SS	0,0480 (●○○○○●●●)	0,0032 (○●●●●○○○○●○○)
EA	0,2364 (●○○●○○○○●●●●)	0,9753 (○●)
YI	0,0286 (○●)	0,0093 (○●○○●○)

Можно видеть, что за вычетом параметра EA оба текста демонстрируют достаточную однородность. Что касается параметра EA, то именно этим в параметре, как видно из предыдущей таблице, сильнее всего проявляется отличие «Постиллы» от «Катехизиса». Более подробный анализ распределения e caudata и $\langle e^a \rangle$ показывает, что доля e caudata резко увеличивается в окрестности DP77. Таким образом, мы можем предположить, что при подготовке издания «Постиллы» в 1595 году были набраны 77 первых страниц, а остальная часть текста набиралась через 4 года.

В остальных параметрах не так просто уловить какие-то четкие закономерности. Однако можно видеть, что в «Катехизисе» неоднородность всех параметров (кроме EA) примерно одинакова, а вот «Постилла» демонстрирует близкую к нулю

⁶ Значение 702 представляет собой количество равновесных разбиений текста на не более чем 12 частей.

неоднородность для всех параметров, не связанных со знаками ударения и ббольшую, чем «Катехизис», неоднородность во всех трех параметрах, связанных с ударением. Это явление требует еще дополнительного изучения, однако поскольку «Постилла», в отличие от «Катехизиса», состоит из большого числа синтаксически и стилистически различающихся фрагментов, то можно высказать предположение о том, что большая неоднородность в распределении знаков ударения свидетельствует в пользу упомянутой в начале гипотезы о том, что расстановка знаков ударения управляется в первую очередь синтаксическими правилами.

Резюмируя вышесказанное, мы можем высказать следующие предположения:

— основные принципы орфографии были выработаны М. Даукшей до создания окончательного текста «Постиллы» и «Катехизиса» и в процессе работы над ними значительно не менялись;

— «Катехизис», тем не менее, следует считать более ранним памятником не только по времени публикации, но и по времени написания, потому что количественные показатели вариативности позволяют охарактеризовать его как своего рода «экспериментальную площадку» для отработки второстепенных принципов орфографии;

— в тексте «Постиллы» можно обнаружить количественно измеримые признаки того, что текст ее набирался в несколько этапов, вероятно, с большим перерывом между этапами;

— получены дополнительные косвенные свидетельства в пользу того, что расстановка знаков ударения у М. Даукши синтаксически, а не фонетически обусловлена.

Литература

- Andreev, A. V. 2013: [Diacritic marks and orthography in M. Daukša's Cathecism of 1595: a quantitative aspect]. *Indoevropėiskoe yazykoznanie i klassicheskaya filologiya* [Indo-European linguistics and classical philology]. 17, 16–25
- Андреев А. В. 2013: Знаки ударения в орфографии Катехизиса 1595 г. М. Даукши: опыт количественного исследования. *Индоевропейское языкознание и классическая филология*, 16–25.
- Andreev, A. V. 2021: [Quantitative corpus-based approach to describing Old Lithuanian orthography] In: *Perevod i yazykovaya adaptaciya v literaturnykh tekstakh srednevekovoj Evropy* [Translation and

- Language Adaptation in the Literary Texts of the Medieval Europe*], SPb. (in print).
- Андреев, А. В. 2021: [Корпусно-статистический подход к описанию старолитовской орфографии] В: *Перевод и языковая адаптация в литературных текстах средневековой Европы*. СПб., 2021 (в печати)
- Bischoff, B. 1990: *Latin Paleography: Antiquity and the Middle Ages*. Cambridge: Cambridge University Press.
- Cain, C. M. 2016: *Some Observations on e-caudata in Old English Texts*. In: *Old English Philology: Studies in Honour of R. D. Fulk*. Ed. By L. Neidorf, R. J. Pascual, T. Shippey. Boydell & Brewer, 233–255.
- Cisne, J. L., Ziomkowski, R. M., Schwager, S. J. 2010: Mathematical Philology: Entropy Information in Refining Classical Texts' Reconstruction, and Early Philologists' Anticipation of Information Theory. *PLoS ONE* 5, 1, 2010 (<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0008661>)
- DK 2006: Mikalojus Daukša. *Katekizmas*, 1595. Prepared by M. Šinkūnas, sponsored by Lithuanian State Science and Studies Foundation (<http://seniejirastai.lki.lt/db.php?source=1>).
- DP 2006: Mikalojus Daukša. *Postilė*, 1599. Prepared by V. Adamonytė, M. Lučinskienė, J. Pajėdienė, M. Šinkūnas, E. Žilinskaitė, O. Aleknavičienė, sponsored by Lithuanian State Science and Studies Foundation (<http://seniejirastai.lki.lt/db.php?source=2>).
- Faier, V. V. 2002: [The integrity of «Ilyad» sub specie statistics]. *Indoevropskoe yazykoznanie i klassicheskaya filologiya* [Indo-European linguistics and classical philology], 6, 202–207.
- Файер, В. В. 2002: Целостность «Илиады» в зеркале статистического анализа. *Индоевропейское языкознание и классическая филология*, 6, 202–207
- Ivinskis, Z. 1953: Die Druckerei der Jesuiten in Vilnius und die ersten litauischen katholischen Bücher. *Commentationes Balticae: Jahrbuch des Baltischen Forschungsinstituts*, 1, 27–68.
- Litvinov, D. A. 2001: [ARATOR as an example of computer-based approach to dating the Latin hexametre]. In: *Slavyanskij stikh: lingvisticheskaya i prikladnaya poëtika* [The Slavic verse: linguistic and applied poetics]. Ed. By M. Gasparov, A. Prokhorov, T. Skulacheva. M: Yazyki slavyanskoj kultury, 404–414.
- Литвинов, Д. А. 2001: Программа ARATOR как пример компьютерных исследований в области датировки латинского гекзаметра. В: *Славянский стих: лингвистическая и прикладная поэтика: материалы межд. конф. 23–27 июня 1998 г.* Под ред. М. Л. Гаспарова, А. В. Прохорова, Т. В. Скулачевой. М: Языки славянской культуры, 404–415.
- Kullback, S., Leibler, R. A. 1951: On information and sufficiency. *Annals of Mathematical Statistics*, 22, 1, 79–86.
- Lebedys, J. 1963: *Mikalojus Daukša: Monografija*. Vilnius.

- Stamatakis, A. 2019: A Review of Approaches for Optimizing Phylogenetic Likelihood Calculations. In: *Bioinformatics and Phylogenetics*. Ed. by T. Warnow. Cham: Springer International Publishing, 2019. 1–19.
- Thies, F. 2019: *Innere Variation bei Daukša: Untersuchungen zu Orthographie und Syntax des altlitauischen Katechismus (1595) und der Postille (1599) von Mikalojus Daukša*. PhD thesis, Universität zu Köln.
- Zaliznyak, A. A. 2008: “*Slovo o polku Igoreve*”: *vzglyad lingvista* [*The Tale of Igor's Campaign from the Viewpoint of a Linguist*]. М.: Рукopисные памятники Древней Руси.
Зализняк, А. А. 2008. «Слово о полку Игореве»: взгляд лингвиста. Издание 3-е, дополненное. М.: Рукописные памятники Древней Руси.
- Zinkevičius, Z. 1988: *Lietuvių kalbos istorija*. T. 3. Senųjų raštų kalba. Vilnius.