



М. ЯНАКИЕВ, Н. В. КОТОВА

О НЕКОТОРЫХ ПРИНЦИПАХ ЭДИЦИОННОЙ ПАЛЕОСЛАВИСТИКИ

Новые веяния в лингвистике, связанные с идеями Ф. Соссюра и И. А. Бодуэна де Куртенэ, совсем не ощущаются или ощущаются очень слабо в филологических исследованиях фактов исчезнувших, ныне мертвых языков и более ранних этапов развития современных языков. Особенно подчеркнуто воздерживаются от использования новых теоретико-лингвистических трактовок филологи, занимающиеся публикацией старых текстов. Издания старых памятников письменности, осуществленные во второй половине XIX в. наиболее компетентными, наиболее эрудированными филологами того времени, поражают современного лингвиста своей обдуманностью и компактностью, легко дают ему элементарную и наиболее часто нужную информацию, так что он привыкает считать эти издания совершенными. Таково положение в палеофилологии вообще, таково оно и в палеослави- стике.

Авторитет И. В. Ягича и С. Н. Северьянова как эдиторов настолько стабилен, что в последние несколько лет часть их изданий старославянских памятников была факсимилирована и отпечатана без изменений. Но и тща- тельно подготовленные в последнее время публикации старых памятников славянского происхождения следуют без изменений эдиционным традициям И. В. Ягича, В. Н. Щепкина, С. Н. Северьянова.

Однако с каким бы почтением современный лингвист ни относился к грандиозной эдиторской работе, проделанной в XIX в. и начале XX в., он не может не испытывать трудностей, пытаясь извлечь из имеющихся публи- каций старых текстов более «современную» лингвистическую информацию: 1) о количественных и качественных характеристиках тех типов фактов, которые имеют соответствия в аллоэмической терминологической системе (графемы, фонемы, морфемы, их варианты — аллографы, аллофоны, алломорфы — и их сочетания, синтагмы); 2) об отношениях между гра- фемами, рассматриваемыми как компонент письменной практики, и фоне- мами — соответствиями графем в устной речевой практике; 3) о структуре языка (*langue*), породившего изданный текст, о степени ее «чистоты», т. е. о ее стилистических «подструктурах».

Настоящая статья имеет целью привлечь внимание лингвистов к основным, принципиальным проблемам, возникающим при попытке пред- ставить издание старого славянского текста, соответствующее современ- ному состоянию теоретической лингвистики, при максимальном исполь- зовании возможностей современной техники.

Г л о т т о м е т р и ч е с к и е п р о б л е м ы

Прежде всего возникают проблемы, связанные с количественными характеристиками текста, — проблемы, которые мы будем называть глоттометрическими¹. Первая из них — в какой степени издатель текста должен быть и его исследователем. Если издатель поставит своей целью дать очень подробное исследование издаваемого им текста, издание может задержаться на десятилетия, и работающие в данной области не смогут использовать ценную информацию²; наряду с этим, чем детальнее исследование, тем менее полезным оно делает само издание текста, который с увеличением объема исследования превращается в не особенно нужное приложение. Вот почему разумным решением проблемы было бы дать читателю возможно более компактную информацию о содержащемся в тексте, а не толкования этой информации.

Такое решение логически ведет к решению другой проблемы: только ли издаваемым текстом должен заниматься издатель или он должен заниматься и «памятником», т. е. «внешней историей» текста³, отношением его к другим подобным текстам? Изучение связей между издаваемым текстом и другими подобными текстами является частью именно подробного исследования текста, которое, если мы не хотим превратить его в некоторую совокупность весьма общих и потому не особенно убедительных утверждений, может задержать издание⁴.

Какие еще глоттометрические проблемы можно решить после такого решения двух указанных выше проблем?

Тексты, с которыми мы имеем дело в палеославиистике, состоят из двух типов фактов — частых (схнотипы) и редких (спаннотипы)⁵. Частые как правило более простые или элементарно простые, т. е. настолько простые, что мы рассматриваем их репрезентанты как атомарные (в первоначальном смысле этого слова, неразложимые на более простые) факты. Редкие обычно сложны, т. е. составлены из более простых и более частых фактов. Для пользующегося изданием было бы целесообразно индексирование составляющих текст фактов, т. е. классификация по различным признакам с указаниями, где эти факты находятся в тексте (сигнатуры, адреса). Составление хороших индексов к изданию текстов — работа, требующая высокой лингвистической квалификации, работа, при которой издатель должен принимать во внимание не только требования лингвистики, но и психологию ищущего информацию человека.

Индекс должен быть построен так, чтобы можно было максимально быстро находить извлеченную из текста информацию об интересующих нас

¹ О термине «глоттометрический» см.: Н. В. Котова, М. Янакиев. Составление некоторых количественных характеристик славянских языков. В сб. «Славянская филология», вып. 7. М., 1968, стр. 58.

² Так, например, от издания снимков Ассеманиева евангелия (1929 г.) до издания кириллической транслитерации текста (1955 г.) Й. Курцем прошла четверть века, причем к транслитерации не приложен ожидаемый лингвистический анализ памятника. Но практика публикации исследования текста отдельно от издания самого текста, ставшая традицией в палеославиистике, неудобна для лингвистики. См., например, монографию А. Маггулиэс. *Der altkirchenslavische Codex Suprasliensis*, Heidelberg, 1927, которая не является частью издания Супрасльского сборника.

³ Ср. по этому вопросу: Л. П. Жуковская. Лингвистические данные в текстологических исследованиях. В сб. «Изучение русского языка и источниковедение». М., 1969.

⁴ Примером того, насколько «поиски памятника в тексте» могут увеличить объем работы, является I том издания Тихонравовского дамаскина (см.: Е. И. Демин. Тихонравовский дамаскин. Болгарский памятник XVII в. Часть I. Филологическое введение в изучение болгарских дамаскинов. София, 1968). Здесь на 235 страницах большого формата определяется место Тихонравовского дамаскина среди других болгарских дамаскинов, а публикация самого текста будет осуществлена в отдельном томе.

⁵ См.: Н. В. Котова, М. Янакиев. Там же, стр. 59.

типах фактов без опасности утонуть в море сигнатур, но и без многократного перелистывания, без необходимости делать подсчеты, если мы интересуемся «удельным весом», частотой, долей количества фактов определенного типа, но и без опасности пропуска редких типов фактов, даже единичных (*ἄπαξ λεγόμενα*).

При индексировании сихнотипов (наиболее частых типов фактов) текста перечисление адресов их репрезентантов теряет смысл. Если, например, сихноморфема « \therefore » (четверсточие) в паримейнике Григоровича реализуется около восьми тысяч раз, т. е. в среднем по сорок раз на каждой странице, давать в перечне графем сигнатуры всех четверсточий явно бессмысленно. Однако и следовать за И. В. Ягичем, который в отличном издании «Codex Marianus», например, для чрезвычайно частого же указывает только *saepissime occurit*, не соответствовало бы современному состоянию лингвистики.

Современная наука располагает проверенной методикой для описания больших массивов фактов — математической статистикой. Согласно математической статистике массово появляющийся тип фактов характеризуется совокупностью чисел, параметров некоторой хорошо аппроксимирующей распределение его появлений в тексте математико-статистической модели. Количественные лингвистические исследования, проведенные до настоящего времени, показывают, что почти всегда хорошо аппроксимирует распределения частых типов фактов в текстах модель, называемая обычно «нормальным распределением». А нормальное распределение определяется двумя параметрами: 1) средней частотой типа фактов в одинаковых по размеру отрезках текста (\bar{x}); 2) дисперсией частот этих появлений в равных отрезках текста около средней частоты (σ^2) или ее квадратным корнем, средним квадратическим отклонением (σ).

Вот почему для удовлетворительного описания сихнотипов фактов в издаваемом тексте необходимо каждый из них характеризовать двумя числами (\bar{x} и σ^2 или σ). А для того, чтобы \bar{x} и σ получали быстрое и легкое толкование и могли быть сразу использованы в сопоставительных исследованиях, единица измерения (моно- X -а ⁶, обычно «монофона» ⁷) размера текстовых сегментов, на основе которых вычисляются \bar{x} и σ , должна быть фиксированной и всегда соблюдаться; размером же текстовых сегментов (проб) удобнее всего брать размер, равный числу единиц измерения, кратному десяти (лучше всего дека- X -а, гекто- X -а, кило- X -а) или по крайней мере пяти. Пробы, путем математико-статистической (глоттометрической) обработки которых вычисляются \bar{x} и σ , образуют «репрезентативную выборку». Для вычислений наиболее удобно, если количество проб будет кратным десяти (например, 10 гекто- X). Минимальное количество проб определяется правилами, установленными математической статистикой.

У филологов существует неверное представление о точности количественного лингвистического описания текста. Описание считается «точным», если в нем указывается с точностью до единицы количество репрезентантов всех типов фактов. Распределение репрезентантов типов фактов в тексте, согласно математической статистике, характеризуется удовлетворительно по крайней мере двумя параметрами, один из которых (σ) можно вычислить, лишь разделив текст на части (пробы). А сообщение количества фактов с «точностью» «до единицы» дает представление только о (средней) частоте типа и лишь при условии, что известен размер всего описываемого текста. Так, например, в индексе лексем, представляющем часть во многих отношениях образцового издания «Изборника 1076 г.», наряду с сигнату-

⁶ См.: М. Я н а к и е в. Основы теории орфографии. «Вопросы языковедения», 1963, № 5, стр. 49—50.

⁷ См.: Н. В. К о т о в а, М. Я н а к и е в. Там же, стр. 65—67.

рами всех репрезентантов каждой лексемы, реализованной в тексте, указывается и количество этих репрезентантов. Однако нетрудно на конкретных примерах увидеть, что «точное» до единицы число с математико-статистической (глоттометрической) точки зрения недостаточно показательно. Так, из индекса лексем, включенного в издание «Изборника 1076 г.», видно, что лексемы **бывати** и **въпрось** имеют одинаковое количество реализаций в тексте — 32. А дисперсия появлений лексемы **бывати**, которую вычисляем, разделив текст на 5 равных частей ($\sigma^2 = 2,3$) более чем в 28 раз меньше дисперсии появлений **въпрось** для тех же пяти равных частей ($\sigma^2 = 65,3$), т. е. математик-статистик нисколько не усомнился бы в существенном различии между распределениями этих двух лексем, несмотря на одинаковое количество их реализаций в тексте⁸.

Точности в описании сихнотипов можно достичь, охарактеризовав их поведение в тексте не только оценкой \bar{x} , но и оценкой σ . Лишь для редко реализующихся типов фактов (спаниотипов) достаточно исчерпывающе указать в индексе сигнатуры (адреса) их реализаций. Естественно возникает вопрос: где граница между сихно- и спаниотипами фактов? Ее следует искать в минимизации усилий, которые должен затратить читатель, чтобы получить необходимую ему информацию от изданного текста. Практика показывает, что лингвист убеждается в наличии некоторого класса фактов после того, как ему приведут десять—двадцать примеров. А если для какого-либо типа фактов в индексе указано более двадцати сигнатур, они или не будут все до конца просмотрены читателем, или читатель решит, что в данном типе следовало бы искать подтипы. Вот почему было бы разумным раздроблять, где это возможно, статьи индекса, пока в каждой статье не окажется самое большее по двадцати сигнированных репрезентантов. При количестве репрезентантов более двадцати мы можем говорить о сихнотипе, который, с одной стороны, подлежит «раздроблению», а с другой — статистической обработке.

В лексикальных индексах этот принцип автоматически почти соблюдается, так как для каждой аллолексы адреса указываются отдельно, хотя и объединены в одну статью; и только небольшое количество статей, главным образом для неизменяемых слов, перегружено более чем двадцатью адресами. Так, в индексе «Изборника 1076 г.» в «раздроблении» нуждаются статьи только около сорока лексем, и только пять из них — статьи флективных лексем («изменяемых слов»): **себе**, **быти**, анафорическое местоимение **и**, **ты** и **богъ**. Только у этих пяти лексем есть аллолексы, реализованные в тексте значительно более двадцати раз.

С индексированием грамматических, фонематических и графематических типов фактов положение иное. О его специфике мы будем говорить ниже, здесь ограничимся лишь одним общим замечанием: современный лингвист рассматривает все типы фактов аналогичным образом, и поэтому индексированию в принципе подлежат не только лексические, но и все прочие типы лингвистических объектов — грамматические морфемы, фонемы, графемы. Кажущиеся различия в их описании объясняются различной частотой их реализации. Указание адресов реализаций очень частых фактов бессмысленно, но это не означает, что их редкие варианты не заслуживают специального внимания и адресования. Принцип очень просто формулируется следующим образом: реализации всех спаниотипов фактов (типов фактов, имеющих менее двадцати реализаций в тексте) даются с их адресами в соответствующем индексе, а обо всех сихнотипах фактов (т. е.

⁸ Оценка сделана на основе так называемого F-критерия Фишера. См.: В. Ю. Урбан. Биометрические методы. М., 1964, стр. 163—164.

типах с более чем двадцатью реализациями в тексте) дается глоттометрическая (обработанная математико-статистически) информация.

При адресовании и глоттометрической обработке описываемых типов фактов возникает еще одна проблема — проблема выбора способа связи адресов с измерением объема текста. Существующая традиция адресовать факты указанием номеров листов и строк неудобна для глоттометрии, так как она не дает возможности для глоттометрических сопоставлений: страницы одних памятников значительно больше страниц других, величина букв также различна.

По причинам, которые мы здесь не обсуждаем⁹, наиболее целесообразно адресовать лингвистические факты указанием номеров их начальной монофонии или, если эти факты не имеют фонетической стоимости, указанием номеров монофон, между которыми они стоят. Т. е., говоря не вполне точно, номеруются звуки текста, и по их номерам находится место всех других фактов (реализаций графем, фонем, морфем). Конечно, нумеровка фонем — процесс трудоемкий, но скрупулезность хорошего эдителя известна, а введение предлагаемой здесь практики обещает столько удобств в будущем в связи с машинной обработкой заключенной в тексте информации, что «монофонное» адресование должны были бы встретить с радостью издатели старых текстов.

Графематика и фонематика

Вторая группа проблем, возникающих при современном издании старых текстов, связана с выявлением отношений между письменной и устной реализациями языка. Подробное обсуждение этой проблемы выходит за рамки настоящей статьи. Здесь только необходимо подчеркнуть, что при анализе письменной специфики текста (его графематики) мы можем абстрагироваться от ее фонематических коррелятов, так как непосредственно нам дана именно графематика. Но при этом нельзя абстрагироваться от каких бы то ни было зрительно воспринимаемых результатов человеческого труда, вложенного в текст. В сущности, подобное требование стремились удовлетворить палеографические разделы хороших изданий старых текстов и во времена И. В. Ягича и С. Н. Северьянова, но в настоящее время мы в состоянии достичь гораздо больших результатов.

С одной стороны, современная фотографическая и полиграфическая техника, если их использовать в эдиционной и текстологической работе хотя бы в той степени, в какой их использует криминалистика, в состоянии обеспечить публикацию, которая даже при простом просмотре даст больше информации о тексте, чем просмотр самого рукописного текста, и притом сделает возможным его одновременный просмотр большим количеством людей. С другой стороны, легкость факсимильного издания старых текстов побуждает издавать только или почти только клише страниц памятников. Такая практика с основанием не одобряется лингвистической общественностью¹⁰. Факсимильное издание требует больших средств на всякий отпечатанный экземпляр, чем средства, необходимые для обычных черно-белых и даже цветных фотокопий страниц памятника, а качество отпечатков в нем не выше качества фотокопий.

Это не означает, что типографское воспроизведение факсимиле, особенно цветных, совсем устарело. Как часть издания текста оно все еще остается лучшим способом предохранения оригинала от изнашивания при палеографических справках, требующих зрительного ознакомления с содержащейся в тексте информацией. Конечно, наряду с воспроизведением факсимиле было

⁹ См. об этом: Н. В. Котова, М. Янакиев. Там же, стр. 65—67.

¹⁰ См.: Р. М. Цейтлин. *Юсип Врана. Вуканово еванђеље. «Советское славяноведение»*, 1969, № 2, стр. 116.

бы уместно включить в комплект издания цветной диафильм для демонстрации страниц текста, позволяющей их рассматривать в значительно увеличенном формате и большему числу людей (при преподавании и исследовании текста).

Если в тексте есть трудно читаемые места и особенно если мы имеем дело с палимпсестом, издатели памятника должны постараться использовать все достижения современной фотографии, чтобы сделать ясно видимым все, что было написано или нарисовано на пергамене или бумаге. При известном старании даже без специальной аппаратуры можно получить факсимиле водяных знаков на бумаге, которые и до сих пор перерисовываются или описываются. Не следует забывать, что палеославистика все еще только «знает» о двух глаголических памятниках — смытых текстах, по которым написаны части Боянского и Зографского евангелий. Отчасти палимпсест, но со смытым греческим текстом, представляет собой и паримейник Григоровича, не говоря о более поздних памятниках.

Но обработка графематики добросовестно подготовленного к изданию текста не сводится к фототехническим операциям по выявлению нечитаемого. Традиционная же палеография нуждается в значительном усовершенствовании, чтобы удовлетворить современного лингвиста-теоретика. В настоящей статье, естественно, могут быть предложены для обсуждения только самые основные проблемы такого усовершенствования.

Теория анализа почерков сейчас интенсивно развивается как база для операций, помогающих криминалистам. Поставленная в соответствие с современными лингвистическими теориями описания языка, она могла бы кое-чему научить и эдителя старых текстов. Можно и нужно упорядочить понятийную систему, которую применяют палеографы при описании многообразия графем. Этого можно достичь, пользуясь аппаратом аллограф и графических дифференциальных элементов¹¹.

Что касается фонематики текста, то при ее описании следует помнить, что эдитор ее извлекает из графематики с помощью трансформационных процедур, характер и обоснование которых необходимо показывать читателю в возможно более доступной форме. При этом в современном издании нельзя по традиции пропускать фонетическую информацию, содержащуюся в пунктуационных графемах. Хорошее описание употребления пунктуационных графем может натолкнуть на серьезные размышления многих современных синтаксистов.

При описании графем, графемных дифференциальных элементов, фонем и фонемных дифференциальных элементов также необходима информация о количественном (глоттометрическом) распределении их появлений в тексте. И здесь применимы принципы описания частых (более 20 появлений) фактов с помощью \bar{x} и σ , а редких — исчерпывающим адресованием. Таким образом можно преодолеть скудость графематических и фонематических разделов описаний, составляющих часть изданий старых текстов.

М о р ф е м а т и к а и л е к с и к а

Сказанное выше применимо и к морфемам несловам, реализуемым часто или редко в тексте. Если эдитор признает, что морфема, а не слово, является основным, элементарным смысловым компонентом языковой системы (это теоретически сейчас признают все лингвисты), он постарается информировать читателя о распределении в тексте не только лексем (в большинстве случаев редко появляющихся морфем, спаниоморфем), но и о распреде-

¹¹ Интересную попытку в этом направлении сделала в своей кандидатской диссертации Е. М. Сморгунова. См. ее автореферат: Графика смоленской скорописи XVII в. М., 1967.

лении более часто реализуемых грамматических морфем (сихноморфем). В традиционной эдиторской практике информация о сихноморфемах так же лапидарна, как и информация о графемах и фонемах.

Однако описание морфем неслов, как и морфем слов (лексем), ставит и новые проблемы вследствие того, что эти языковые факты являются носителями семантики, о которой издатель также должен дать информацию. Очень многие памятники славянской древности, как известно, представляют собой переводы с греческого (на востоке и юге) и латинского (на западе), причем переводы, очень близко следующие оригиналу. Вот почему опасно исследовать старые славянские тексты без сопоставления с оригиналами, с которых они переведены. Это обязывает эдителя палеослависта приводить в издании и оригинал (греческий или латинский) рядом со славянским текстом, тем более, что оригинал часто опубликован в трудно доступных старых книгах.

Иноязычный оригинал облегчает работу эдителя при описании семантики реализуемых в издаваемом тексте морфем (именно морфем, а не лексем). Однако недостаточно приводить в индексе иноязычные морфемы только после их славянских соответствий. К сожалению, даже от этого воздержались издатели «Изборника 1076 г.»; к тому же они помещают греческий текст не рядом со славянским, а отдельно, в конце издания, что затрудняет справки. Необходим также отдельный индекс, в котором иноязычные морфемы были бы даны в алфавитном порядке, а рядом с ними указывались бы их славянские соответствия. Аналогичное решение разумно и для современных переводов старых славянских морфем. Отдельный алфавитный индекс современных переводов старых славянских морфем был бы важным источником для семантических наблюдений над историей славянских языков.

Ст р у к т у р а я з ы к а т е к с т а

Все рассмотренные до сих пор проблемы, относящиеся к изданию текста, как, вероятно, уже ясно, затрагивают способы подачи более компактной, но и более исчерпывающей информации о тексте как готовом продукте языковой деятельности человеческого коллектива определенной эпохи и определенной национальности. Поскольку, однако, современный лингвист стремится в тексте (*parole*) найти структуру, устройство породившего текст механизма языка (*langue*), эдитор лучше всех, очевидно, знающий текст, должен ориентировать читателя и в этом отношении.

Так как этот механизм, называемый еще грамматикой, представляет совокупность разрешений и запретов определенных типов сочетаний морфем, эдитору следовало бы написать «грамматику», соответствующую издаваемому тексту, одновременно исчерпывающе указав и отклонения от этой «грамматики» в тексте, т. е. дав сведения и о стилистических особенностях языковой структуры, породившей текст. Однако в данном случае не идет речь о новой проблематике. Что в хороших традиционных изданиях проблемы структуры языка, породившего текст, ставились и отчасти решались, можно доказать, напомнив рассуждения Ягича о национальности переписчика Мариинского четвероевангелия — рассуждения, которые могут быть поучительными и для современного специалиста по структурной лингвистике, и для современного специалиста по стилистике.