

рода выделяется чистой показателем пола; но здесь оказывается подчеркнуто обозначенным. И это мешает равнозначному употреблению таких слов женского рода наряду со словами мужского рода в современном языке, хотя со времени революции именно эта особенность значения способствовала распространению подобных имен женского рода.

Некоторые другие словообразовательные типы также претерпевают изменения, нередко связанные с рядом различных сложных причин. Это требует большой осторожности в использовании материалов, представленных обратным словарем, и, безусловно, для каких-либо выводов относительно словообразования необходимы дополнительные сведения о времени существования слова, его стилистической и диалектной принадлежности, которые могут быть достаточно достоверно установлены лишь на основании толковых словарей. Пока что необходимость этой трудоемкой работы, кажется, не принимается во внимание составителями обратных словарей современных языков, лишенных ссылок на источники.

Необходимость обратных словарей очевидна. Можно ожидать, что в ближайшее время у нас появятся такие словари различных типов¹.

После выхода в свет словаря под ред. Бильфельдта и словаря Греше и Креше они могут быть составлены достаточно быстро и уже на более высоком научном уровне, так как указанные словари дают возможность определить круг вопросов, требующих предварительного решения при составлении обратного словаря.

Р. В. Бахмурина

¹ Работа по составлению обратного словаря в настоящее время ведется в Ленинграде И. К. Зборовским с группой сотрудников словарного сектора Института русского языка. Вызывает некоторое беспокойство намерение И. К. Зборовского включить в обратный словарь материалы картотеки полностью из данного еще Большого академического словаря русского языка, что затруднит использование обратного словаря для изучения словообразования даже по сравнению с упомянутыми словарями.

СЛОВАРЬ ПОЛЬСКОГО ЯЗЫКА XVI В.¹

(Важный вклад в развитие лингвистической статистики)

Словарь польского языка XVI в. является продолжением словаря древнепольского языка («Słownik staropolski»), который давно уже был подготовлен и с 1953 г. выходит отдельными выпусками (к настоящему времени издание доведено до буквы G). Работа над словарем ведется пятью лабораториями под руководством профессоров С. Бонка, С. Храбца, В. Курашкевича, С. Роспонда и В. Ташницкого. Общую координацию осуществляет отдел истории литературного языка и теории литературы Польской Академии наук под руководством проф. М. Р. Майеновой. В феврале 1954 г. план и проблематика словаря были обсуждены и одобрены на заседании языковедческого комитета. О работе над словарем доложил на IV Международном съезде славистов В. Курашкевич.

Словарь польского языка XVI в., как его задумали авторы, — это прежде всего словарь исторический. Ему чужда всякая идея нормативности, отбора собранного материала, он должен быть прежде всего полным. Заметим, что такая постановка вопроса стала возможной только благодаря тому, что авторы поставили себе тесные, хотя и довольно условные хронологические рамки (1501—1600 гг.). Авторы ограничиваются также в основном обследованном печатных текстов, не привлекая пока рукописных материалов.

¹ «Słownik polszczyzny XVI wieku. Zeszyt próbny», Wrocław, 1956 (PAN, Instytut badań literackich).

Как указывается во введении к проблемному выпуску, удача подобного словаря целиком зависит от того, насколько полон материал, подвергающийся обработке, насколько он показателен для центральных, существеннейших процессов, происходивших в языке XVI века (стр. XII). Надо сказать, что список отобранных источников достаточно внушителен. Он включает в себя: произведения по математике, астрономии и астрологии (4), по медицине и лекарственному делу (4), по педагогике (2), по хозяйству и праву (6), по орфографии и словарю (4), по географии (включая описания путешествий) (4), хроники и повести (10), апокрифы и жития святых, биографии философов (7), проповеди (7), религиозную полемическую прозу (13) (из них в форме диалогов 3), молитвенники и псалтыри (6), религиозную морализующую прозу (8), светскую политическую литературу (11) (из них в форме диалогов 2), литературу в стихах (разных жанров) (82) (из них диалоги и драмы 14), переводы библии (6). Общим объемом обследованного материала составляет 7 млн. слов.

Перейдем к анализу отдельной словарной статьи. Для примера возьмем слово *głowa* «голова». Словарное слово (*hasto*) дается в древнепольской транскрипции (которая в данном случае совпадает с современной), в скобках указана частота употребления слова в обследованных текстах (1018 раз). Приводятся отдельные формы, например родительный единственного числа *głowy* (163 раза) и *głowej* (2 раза), причем

отмечены авторы, у которых эти формы употреблялись. Далее указывается первое значение (обычно это исходное значение; наряду с этим историческим критерием используется критерий частотности). Как правило, для объяснения подбирается соответствующее по значению слово современного польского языка. В качестве вспомогательного средства определения значения используются латинские эквиваленты данного слова (в данной статье: *caput*). Указывается в скобках, если обратиться к рассматриваемой статье, что в этом первом значении слово встретилось 660 раз. Далее приведено более 50 примеров, иллюстрирующих как синтаксическое употребление слова, так и его смысловые связи. Интересно, что примеры из словарей даны вместе с латинскими оригиналами; ср. *diffunditur morbus a capite; Idzie od głowy do drugich czlonkow niemoc* (J. Maczynski, *Lexicon Latino-polonicum*, 1564, 139d).

Наконец, указываются фразеологические единицы, в которых слово выступает в данном значении. Фразеология трактуется авторами очень широко, основным здесь является критерий совместного употребления слов. Как указывается во введении, частое совместное употребление может отражать тенденцию к дальнейшей лексикализации или же стилистические тенденции того времени, а иногда являться результатом поисков эквивалентов для некоторых латинских слов. Так, парное сочетание *glowa a rozum* являлось переводом латинского *iudicio*. Авторы словаря при анализе фразеологии (стр. XXXV) выделяют следующие группы: 1) фраза (*gazes*), в которой наличествует подлежащее и сказуемое; 2) неглагольное сочетание (*wyrażenie*), в котором определяемое не есть глагол; 3) глагольное сочетание (*zwrot*), в котором определяемое — глагол; 4) парное сочетание (*szereg*), состоящее из однородных членов.

В нашей статье фразеологическая часть открывается рядом неглагольных сочетаний, например *psia glowa* «собачья голова», причем указано, что это сочетание употреблено 3 раза, и приведены все три контекста. Затем идут глагольные сочетания, например: *glową chwiać, kinać glową, glową kinać* «кивать головой» (объяснено словами: «дать знак одобрения или неодобрения»). Опять указана частотность (10 раз) и приведены все контексты. По этому же принципу строится словарная статья и для других значений слова.

Наиболее важным в данном словаре является то, что впервые в истории польской лексикографии словарь польского языка XVI в. строится, как мы видели, на статистической основе.

Как указывают авторы, статистика должна прежде всего помочь ответить на вопрос о языковых формах XVI в., показав господствующие, возникающие и умирающие тенденции развития лексики и морфологии. Если составители современного словаря могут в какой-то мере опираться на свое языковое чутье при снабжении того или иного слова определенными

пометами, то историк языка более чем кто-либо нуждается в строго объективных критериях для определения стилистической характеристики слова.

М. Р. Майенова¹ считает, что статистические данные, приведенные в словаре, будут полезны как для тех, кто интересуется нормами и различными стилями общенародного языка, так и для исследователей стиля определенного литературного направления, жанра, школы, эпохи, а также для исследователей индивидуально-художественного стиля автора.

Майенова следующим образом представляет себе работу стилиста со словарем. При чтении текста стилист формулирует определенную гипотезу о стилистической значимости (*wartość*) данного слова или выражения. Обращение к словарю позволяет подтвердить эту гипотезу или отбросить ее и сформулировать новую. Пусть исследователь встретил в «Псалтыри» Кохановского слово *sąd* в значении «посуда» (в современном языке «суд»). Ему кажется, что для второй половины XVI в. употребление слова в данном значении есть архаизм. Поскольку словарь фиксирует все употребления данного слова у всех авторов, гипотезу эту нетрудно проверить.

С другой стороны, простой просмотр словарных статей дает стилисту возможность поставить ряд интересных вопросов, при решении которых опять-таки может использоваться словарный материал. Например, форма *abo* «или» встретилась 1353 раза, в то время как *aboc* только 5 раз — один раз у старшего из взятых писателей (Бернат) и 4 раза у одного и того же автора (Рей). Возникает вопрос, идет ли здесь речь о форме архаичной или, может быть, разговорной, или же употребление такой формы — особенность стиля писателя. Сопоставление данной формы с другими дает возможность получить на эти вопросы ответ.

Таковы, по мнению авторов, некоторые из возможных употреблений словаря. К этому хотелось бы добавить, что богатый стилистический материал, собранный авторами, а также анализ применявшихся методов могут быть весьма полезными для развития самой лингвистической стилистики. В самом деле, в настоящее время объем статистически обследованного материала очень незначителен. Имеющиеся статистические словари для ряда языков (немецкого, английского, французского, испанского, русского и некоторых других²), как пра-

¹ M. R. Ma y e n o w a, *Słownik polszczyzny XVI-wiecznej z punktu widzenia potrzeb stylistyki, «Z polskich studiów slawistycznych. Prace język. i etnogen...»*, Warszawa, 1958.

² F. W. K a e d i n g, *Häufigkeitswörterbuch der deutschen Sprache, Steglitz, 1897—1898*; G. E. V a n d e r B e k e, *French word book, New York, 1929*; V. G. H o z, *Vocabulario usual, vocabulario común y vocabulario fundamental, Madrid, 1953*; H a r r y H. J o s s e l s o n, *The Russian word count and frequency analysis of grammatical categories of standard literary Russian, Detroit, 1953*.

вило, не отражают каких-либо определенных лингвистических задач, поставленных их авторами (в особенности это относится к частотному словарю Йосселсона). К тому же работа составителей этих словарей чрезвычайно затруднилась тем обстоятельством, что при изучении современного языка очень трудно остановить свой выбор на тех или иных произведениях, статистика которых должна отражать распределение вероятностей отдельных слов во всем современном языке. Ясно отграничив период, подлежащий обследованию, и обследовав достаточно богатый материал, авторы польского словаря собрали материал, чрезвычайно важный для общей методологии и проблематики лингвистической статистики.

В частности, было бы интересно проверить на этом материале известную формулу лингвистической статистики о соотношении между длиной слова (т. е. числом его фонем) и его частотностью. Согласно этой формуле, если обозначить количество фонем в слове через k , а через r его ранг, т. е. номер в списке по убывающей частотности, то имеет место следующее соотношение:

$$\frac{k}{\log r} \approx \text{const}$$

(т. е. чем чаще употребляется слово, тем оно короче, и наоборот)¹. Установлена также еще одна интересная закономерность, а именно — связь между числом значений слова (s) и его частотой (f), которая выражается следующей формулой:

$$\frac{s}{\sqrt{f}} \approx \text{const}$$

(т. е. чем чаще употребляется слово, тем больше у него значений, и наоборот)².

Эти закономерности чрезвычайно интересны с точки зрения применения к анализу языка методов теории информации. Материал словаря особенно важен в этом отношении: соотношение между частотой и длиной какого-то слова, а также количеством его значений в языке XVI в. можно сравнить с частотой, длиной и количеством значений этого же слова в современном языке. Тем самым можно было бы установить, какое значение имеют для развития языка основные принципы теории кодов. В частности, можно было бы проверить гипотезу о том, что в процессе развития языка «форма языка, сохраняющаяся благодаря самому факту ее употребления и сохранения, обязательно принимает форму, очень близко напоминающую оптимальную форму распределения»³.

Материал по фразеологии, собранный в словаре, может помочь при ответе на

следующий весьма важный вопрос: в какой мере степень свободы или связанности отдельных единиц в словосочетании (ср., например, известную классификацию акад. В. В. Виноградова) может быть выражена в статистических терминах. Можно предложить следующий подход: зафиксировать первое слово сочетания и выяснить, какова вероятность того, что следующее слово будет таким-то (условная вероятность появления какого-то слова после данного слова). Ясно, что если условная вероятность велика, то слова фразеологически сильно связаны, если же она мала, то речь идет о сочетании более свободного характера. Конкретный анализ материала поможет проанализировать зависимость между семантической и статистической связанностью.

Некоторые важные обобщения уже сделаны на основе анализа материалов словаря.

Одним из актуальнейших вопросов лингвистической статистики является нахождение статистических критериев для оценки «богатства» или, наоборот, «бедности» лексики произведения. Оценка эта играет большую роль при решении проблем атрибуции. Формулы, предложенные П. Гиро и Дж. Херданом⁴, являются настолько громоздкими, что вряд ли могут практически использоваться. В. Курашкевич⁵ показал возможность оценки этой величины средствами, доступными каждому стилисту. Курашкевич исследовал графики зависимости появления новых слов от длины текста, причем длина текста в тысячах печатных знаков откладывалась по оси абсцисс X , а количество новых слов — по оси ординат Y . Если заносить ряд значений X и Y на чертеж, то получается ломаная линия, очень близкая к параболе, описываемой функцией $Y = K \sqrt{X}$, где K — некий коэффициент. Курашкевич использует этот коэффициент как меру «богатства» словаря. Этот метод оказался весьма эффективным и позволил сделать много наблюдений, касающихся языка и стиля польских писателей XVI в. За неимением места мы на них не останавливаемся. Отметим лишь некоторые общие выводы, полученные Курашкевичем при анализе текстов: 1) при прочих равных условиях текст стихотворный оказывается более «богатым», чем текст прозаический, причем коэффициент K зависит и от размера стиха; 2) первые произведения писателя имеют коэффициент K гораздо более низкий, чем дальнейшие его произведения. Иными словами — чем более зрело творчество писателя, тем выше коэффициент K ; 3) оригинальные произведения дают более высокий коэффициент K , чем переводы, выполненные тем же писателем.

Курашкевич подчеркивает, что полученная им оценка «богатства» словаря весьма ус-

¹ См., например, P. Guiraud, Les caractères statistiques du vocabulaire, Paris, 1954.

² См. G. K. Zipf, The meaning-frequency relationship of words, «Journal of general psychology», vol. 33, half 2, 1945.

³ Н. Винер, Кибернетика и общество, М., 1958, стр. 100 (перевод с англ.).

⁴ G. Herdan, Language as choice and chance, Groningen — Noordhoff, 1956.

⁵ W. Kuraszkiewicz, Statystyczne badanie słownictwa polskich tekstów XVI wieku, «Z polskich studiów slawistycznych...».

ловна, однако это вовсе не снижает ее ценности. Во-первых, статистическую оценку богатства можно усложнить, введя в рассмотрение деление на служебные и подлинно значимые слова, ушитывая количество значений отдельных слов и т. п. Во-вторых, — и это главное — даже при помощи такой приближительной оценки, как коэффициент К, удалось получить весьма ценные результаты. И, наконец, эти результаты, добытые совершенно объективным способом, не противоречат субъективному восприятию читателя, но уточняют его, делая возможным точное сравнение отдельных произведений. Как правильно указывает Курашкевич, было бы очень интересно исследовать этим методом произведения авторов XVIII — XX вв.

Разумеется, при той большой и ответственной работе, которую проделали авторы словаря, не обошлось и без ошибок. Но сами эти ошибки весьма поучительны для развития лингвистической статистики.

В. Курашкевич еще в 1951 г. показал, что при анализе текста больших размеров (больше 300 тыс. печ. зн.) достаточно брать материал с каждой пятой страницы, чтобы получить 50% всех слов¹. Этот метод, который, очевидно, дает большую экономию, вначале и был положен в основу статистической работы. В дальнейшем, как указывает Храбец², от этого метода пришлось отойти, так как он, во-первых, не дает полноты лексического материала и, во-вторых, не дает статистической картины индивидуального стиля писателя. По этому поводу хотелось бы отметить следующее: метод выборочного анализа, вообще говоря, вовсе не плох, все зависит от цели исследования. Например, при отборе определенного словаря-минимума, т. е. какого-то количества наиболее часто встречающихся слов, этот метод несомненно дает хорошие результаты. Если же нужно выявить самое распределение частот (вероятностей) между всеми словами (последний подход особенно важен при применении к языку методов теории информации³), то в этом случае необходим полный анализ текста, и лучше ограничить количество материала, чем проводить выборочное обследование.

В связи с анализом рецензируемой работы, как и любой другой значительной работы по лингвистической статистике, возникает следующий важный методический вопрос. Как указал в свое время замечательный русский статистик проф. А. А. Чу-

пров, особенности проявления закона больших чисел при анализе языкового материала требуют обследования весьма обширного языкового материала⁴. В связи с этим любая статистическая работа является весьма трудоемкой. В частности, подготовительная работа по словарю польского языка XVI в. велась коллективом, состоящим из 30 питавших сотрудников, в течение 9 лет (работа начата в 1949 г.). Группа, состоящая из 5 человек, могла обработать текст длиной в 10 тыс. слов в течение месяца.

Целесообразно ли в дальнейшем вести подобную статистическую работу вручную или, может быть, уже сегодня имеет смысл поставить вопрос о применении для нужд современной лингвистической статистики электронно-вычислительных машин? Как показывают опыты итальянца Р. Буза (Roberto Busa) и американца Эллисона, машины, применяемые для статистики, могут: а) разбивать текст на отдельные элементы (предложения, слова, синтагмы, основы, окончания); б) изготовлять для каждого слова отдельные карточки с указанием источника, номера предложения и с требуемым контекстом; в) производить классификацию карточек по заданным критериям⁵.

В частности, Эллисон при помощи электронно-вычислительной машины UNIVAC получил за 1000 рабочих часов машины словарь, фиксирующий все употребления слов в библии, на основе анализа 750 тыс. слов — работа, которую обычным способом можно было бы выполнить за 23 года.

Ясно, что для решения тех многообразных задач, которые поставили себе составители польского словаря XVI в., машина могла бы выполнить лишь некоторую предварительную работу (заполнение карточек, фиксация всех вариантов написания слова, отождествление вариантов в тех случаях, когда для него могут быть даны формальные правила, фиксация контекста заданной длины, указание источника, расположение по алфавиту, предварительные подсчеты). В случае необходимости машина могла бы «видеть» кривые Курашкевича. Машина не могла бы, однако, выделять отдельные значения слова (если не исходить из чисто синтаксического критерия), типы фразеологических единиц, в которых слово выступает (если эти типы не определены статистически). Машина не могла бы произвести отбора наиболее характерного иллюстративного материала. В целом же работа словарного коллектива могла бы быть значительно облегчена.

Дело состоит, однако, в том, что прежде чем использовать машины для серьезной статистической работы, необходимо точно определить программу работы, наши требования, ввести точные определения для

¹ W. Kuraszkiiewicz i J. Łukasiewicz, Pośe różnych wyrazów w zależności od długości tekstu, «Pamiętnik literacki», roczn. XLII, zes. 1, 1951, стр. 168—182.

² St. Hrabec, Teoretyczne założenia słownika polszczyzny XVI wieku i ich praktyczna realizacja, «Z polskich studiów slawistycznych...».

³ Ср. с этой точки зрения резко отрицательную рецензию известного специалиста по теории информации Б. Манделброта на частотный словарь Йоселсона («Word», vol. 12, № 1, 1956, стр. 157).

⁴ См. А. А. Чупров, Очерки по теории статистики, 2-е изд., СПб., 1910, стр. 280—281, примеч. 2.

⁵ См. E. Agricola, Elektronische Analyse und Synthese in der Sprachwissenschaft, «Forschungen und Fortschritte», Jg. 32, Hf. 3, 1958.

основных единиц¹, подвергаемых счету. С этой точки зрения анализ работы над словарем может дать, как мне представляется, весьма много поучительного.

В целом, проведенная польскими лингвистами работа над словарем польского языка XVI в. несомненно представляет собой важный вклад не только в польскую лексикографию, но и в общую теорию лингвистической статистики.

И. И. Ревзин

¹ См. И. И. Ревзин, О соотношении структурных и статистических методов в современной лингвистике, сб. «Вопросы статистики речи (материалы совещания)», Л., 1958.

J. de Vries. Altnordisches etymologisches Wörterbuch. — Leiden, Lf. 1—2, 1957; Lf. 3—4, 1958.

Исландский язык с его чрезвычайно своеобразными путями развития, его архаичным строем, его лексикой, хранящей многие элементы, восходящие еще к общегерманской эпохе, наконец, замечательная, весьма богатая и самобытная литература на этом языке — это все привлекало и продолжает привлекать внимание скандинавистов, германистов и индоевропейцев к данному языку. В этом свете становится понятным, хотя и достойно удивления, что за последние десять лет появилось три этимологических словаря исландского языка¹.

Рецензируемый словарь де Фриза², выходящий с 1957 г. (к настоящему времени вышло уже 4 выпуска — до слова *hríp*), намного превосходит вышеуказанные словари. Наряду с готским словарем Фейста, немецким словарем Клуге-Митцка, шведским словарем Хелльквиста он несомненно будет использован не только всеми германистами, но и индоевропейцами. Такому высокому качеству словаря весьма способствовало то, что де Фриз начал работу над словарем, будучи признанным специалистом в области германской мифологии, религии и скандинавской литературы. Отсюда — обилие и точность филологических указаний, проникновенная разработка всех вопросов, связанных с историей и культурой древних скандинавов. Принципы отбора слов не отличаются существенно от других словарей. Включены слова как исландские, так и норвежские, включены также имена личные, имена богов, многие геогра-

фические названия, даются также названия пародов.

Слова расположены в алфавитном порядке, омонимы даются разными словарными статьями. Каждая статья открывается грамматической пометой (часть речи, род существительных и т. д.), затем следуют соответствия в рунических надписях (с точным указанием памятников), соответствия в новоскандинавских языках (исландском, фарерском, норвежском, датском, шведском). Весьма примечательно, что тут же дается написание и значение слова, если оно отличается от исландского в нескандинавских языках (немецком, французском, ирландском, английском и его диалектах — иотландском, оркнейском и т. д.), будучи заимствовано этими языками. Особо даются заимствования из германских языков в финские (причем не только в финский, но и саамский, ливский, эстонский, удмуртский и т. д.). Далее идут соответствия в других германских языках, а иногда и восстанавливаемая общегерманская форма (особенно в случаях, когда фонетическое развитие каким-либо образом затемнено); затем — бесспорные соответствия в других индоевропейских языках. Наконец — ссылки на однокорневые слова. Особо выделяется обзор литературы по спорным этимологиям.

Слова, принадлежащие к одному и тому же индоевропейскому корню, помимо того, что они анализируются в соответствующих словарных статьях, сводятся также в одну группу, рассматриваемую в конце одной из статей. Две иллюстрации:

1. Часть словарной статьи *hark* «шум»: «Широко разветвленное гнездо, которое можно обзорно представить в связи со следующими и.-е. корнями:

(s) *kerg*, ср. *hark* «шум», *harka* «тащить со скрежетом», *harki* «мусор», *herkir* «опустошитель, огонь», *horkull* «шум, стыл», *hráki* «слюна», *hrókr* «ворона», *hrækja* «плавать», *skark* «шум, суматоха», *skerkir* «крикун», *skrækr* «крик», *skrök* «ложь»

kerk, *krenk*, ср. *hringja* 3 «звонить»
(s) *kerd* » *skrati* «тролль»
(s) *kerb* » *hróp* «брань», *skrafa* 2 «болтать»
(s) *kerp* » *hrafn* «ворон», *skrafa* «болтовня»
(s) *kerm* » *harmr* 2 «ястреб»
(s) *krei* » *hrina* «кричать», *skríkja* «сойка»
(s) *kreu* » *hrjóta* «ворчать»

2. Часть словарной статьи *gjá* «пропасть, овраг»: «Широко представленный индоевропейский корень **gnei* «звать», ср.

детерминатив n, ср. *gína* «звать»
» m » *geimi* «море»,
» *gína* «отверстие»
» l » *geil* «долина»,
» *gil* «ущелье»
» gh » *geiga* «сбиться с пути»
» p » *gífr* «чудовище»
» b » *geipa* «болтать вздор».

¹ См.: F. Holthausen, Vergleichendes und etymologisches Wörterbuch des Altwestnordischen, Göttingen, 1948; A. Jönnesson, Isländisches etymologisches Wörterbuch, Bern, 1951—1956.

² Авторам настоящей статьи приняты во внимание следующие рецензии: F. de Tollenaere, «Leuvense Bijdragen», Jg. XLVII, afl. 1—2, Bijblad, 1958, стр. 58—61; S. Gutenberg, «Deutsche Literaturzeitung», Jg. 79, Hf. 9, 1958, стр. 771—774; K. G. Ljunggren, «Arkiv för nordisk filologi», bd. 73, hf. 1—2, 1958, стр. 98—99.