

А. С. Выренкова, Б. В. Орехов, Т. И. Резникова

Типологическая база данных глаголов,
обозначающих звуки животных¹

Животный мир представлен в лексической системе языка достаточно разнообразно. Разнообразнее, чем кажется на первый взгляд. Неспециалист прежде всего вспомнит о названиях животных. Действительно, их много, у каждого своё происхождение и судьба в языке: какие-то стали частотными, вошли в состав фразеологизмов, а какие-то чрезвычайно редки или вовсе забыты носителями. Однако имена животных не существуют в языке изолированно, они сочетаются с определёнными глаголами и прилагательными, причём делают это не хаотически, а в соответствии с последовательной системой, не всегда очевидной. Скажем, про зайца можно сказать, что он «прыткий» («Уже лихорадка начинала подкрадываться к нему; уже священник, «мастер против порчи», два раза в виде очень прыткого зайца с бородой и косичкой перебежал ему дорогу»), И.С. Тургенев «Дым»), а про ворону по-русски так, скорее всего, не говорят. Лингвист вправе задаться вопросом о границах такой сочетаемости и о том, как и почему она складывается. Более того, такая система для разных языков может отличаться, и вполне возможно, что в каком-то языке и про ворону, и про зайца можно будет сказать «прыткий», но для того, чтобы это выяснить, необходимо провести специальные лингвистические изыскания — в обычном словаре такую информацию не найти.

Один из интереснейших классов глаголов, связанных специфическими отношениями с животными, это класс глаголов звука. Наш повседневный опыт подсказывает нам, что животные издают звуки, часто хорошо узнаваемые, то есть лай собаки трудно спутать с мычанием коровы или кудахтаньем курицы. Но отражено ли это разнообразие в языке? С детства нам знакомы звукоподражательные слова: *кар-кар, гав-гав, кря-кря*, которые закреплены за каким-либо животным. Однако этот набор ограничен, а животных гораздо больше. Звукоподражания известны для одомашненных существ и тех животных, которые обитают вблизи поселений человека. Но уже для звуков, издаваемых дикими животными (вроде, например, клекота птиц), звукоподражаний мы, скорее всего, не найдем. Существенно полнее мир звериных звуков представлен в глагольной лексике, ср. *мычать, блеять, чирикать, ворковать* и под. При этом глаголы с лингвистической точки зрения интереснее: дело в том, что их семантика богаче, чем у звукоподражательных слов. Этот факт обусловлен тем, что глаголы наряду с исходным значением животного звука нередко развивают переносные значения, и иногда, как мы увидим далее, используются даже в «незвуковых» контекстах.

Кроме того, у глаголов есть еще одна примечательная особенность: часто один и тот же глагол употребляется для обозначения звуков сразу нескольких животных. Важно при этом, что в разных языках такие «объединения» необязательно совпадают. Например, мыши и птенцы «пищат» и в башкирском (*тупелдәү*), и в коми-зырянском (*ĉipкуну*). Но в хантыйском мышь ещё и «визжит» (*s'äliti*), как свинья, собака или птица. В русском языке мы вряд ли сможем найти такой глагол, который одинаково естественно звучит в сочетании и с мышью, и со свиньей, и с собакой, и с птицей.

Изучением лексических систем языков в сопоставлении занимается лексическая типология, дисциплина, сравнительно недавно возникшая по аналогии с типологией грамматики. Хотя само по себе сравнение слов в различных языках имеет давнюю традицию, лексическая типология как научная область состоялась только в тот момент, когда у неё появился свой методологический аппарат: «исследования в этой области могут стать продуктивными только тогда, когда описания разных языков будут следовать единому теоретическому подходу, единой методике и единому плану описания <...> именно в этом случае можно будет получить главное: сопоставимые результаты по разным языкам»². Одним из возможных теоретических подходов к лексической типологии является «фреймовый»: он предполагает в качестве общей основы для сопоставления языков набор элементарных ситуаций («фреймов»), которые покрываются лексемами исследуемого семантического поля. Соответственно, разные языки сравниваются на основании того, как они распределяют эти ситуации между своими лексическими единицами. Такая работа не просто выявляет сходства и различия между отдельными языками, но и показывает, какие явления окружающей нас действительности чаще всего противопоставляются в естественных языках лексическими средствами. «Так, в плавании — это активное плавание, плавание неодушевленного предмета по течению, статичное положение на поверхности воды и движение судов и на судах»³.

Применительно к глаголам, обозначающим звуки животных, набор элементарных ситуаций, служащих основой для сопоставления, задается самим списком животных. И языки, как и в случае других лексических полей, различаются стратегиями распределения этих ситуаций между лексемами. Действительно, как отмечалось выше, в разных языках комбинации животных, объединяемых в одном звуковом глаголе, не совпадают, и выявление частотных и, напротив, редких моделей в этой зоне — одна из задач типологического исследования. Наиболее радикальная стратегия, которую можно себе представить в этой связи, — это объединение всех животных звуков в одной лексеме. Такая модель представлена, например, в японском, где имеется универсальный глагол *naku*, обозначающий крики всех животных вообще.

Однако для межъязыкового сравнения существенны не только особенности лексического комбинирования звуков различных животных, но и переносные значения соответствующих глаголов. Дело в том, что многие лексемы, которые первоначально употреблялись только по отношению к животным, со временем стали приобретать метафорические значения и выступать применительно к людям, предметам и явлениям природы⁴. Исследование этого аспекта поведения лексики представляет особый интерес в типологической перспективе. Если в русском языке

воркуют голуби и влюблённые, то наблюдается ли такое же метафорическое значение у глагола *ворковать* в других языках? Ответить на этот вопрос можно только собрав и классифицировав по фреймам (ситуациям) значительный материал разных по структуре, происхождению и географической локализации языков.

Однако ручная классификация значительного объёма данных, который необходим для обоснованных выводов, затруднительна. Поэтому собранный материал удобнее всего поместить в базу данных.

На сегодняшний день уже существует несколько баз данных, результирующих лексико-типологические исследования. Среди них база данных признаковой лексики⁵, предикатов боли⁶, глаголов плавания. В этих базах собраны данные, полученные с помощью опроса информантов — основного метода получения лингвистической информации в лексической типологии.

При этом упомянутые базы данных отличаются своеобразным методологическим дуализмом. С одной стороны, как уже говорилось, такого рода электронные инструменты нужны для облегчения работы исследователя по обработке и систематизации материала. С другой стороны, в самой архитектуре базы обычно бывают заложены принципы, сформулировать которые возможно только уже проведя концептуальную обработку данных. То есть базы как инструмент анализа своей структурой во многом уже отражают результаты анализа. В случае с глаголами, обозначающими звуки животных, речь идёт о классификации вторичных значений по типам метафорического переноса. Такая классификация может появиться только как результат внимательного рассмотрения собранного языкового материала. В то же время база будет малофункциональна без введения соответствующего классификационного поля, которое позволило бы извлекать примеры сходных значений (скажем, случаи, в которых глагол, обозначающий звук, издаваемый каким-либо животным, в применении к человеку начинает обозначать плач или смех).

Информацию, содержащуюся в базе данных глаголов звуков животных⁷, можно условно разделить на две категории. В первой находятся сведения, касающиеся глаголов и их употребления в прямом значении, то есть в таких контекстах, которые подразумевают участие в описываемой ситуации собственно животного: *корова мычит, собака лает, ворона каркает*. Именно эта информация может дать представление о том, как именно происходит распределение глаголов по животным в смысле их сочетаемости. Например, в грузинском языке одна лексема служит для обозначения звуков, издаваемых утками и лягушками. Для носителя русского языка, в котором присутствуют и *крякать*, и *квакать*, этот языковой факт наверняка будет воспринят как необычный. В то же время во многих языках собакоподобные животные (собственно собака, волк, иногда лисица) «объединяются» глаголом со значением «рычать», и в этом ни носитель русского, ни носители большинства других известных нам языков не найдут ничего необычного.

Однако те же самые глаголы часто используются в речи в непрямом значении. У лексем, первоначально означавших специально звуки животных, как уже говорилось выше, можно наблюдать метафорический перенос. Релевантная для этого — переносного — значения информация составляет вторую группу полей базы данных.

Рассмотрим эти поля предметно. Каждая запись в базе данных со-

держит информацию о языке, для которого эта запись сделана. Количество таких языков для лексико-типологического исследования обычно, во-первых, меньше, чем для аналогичных проектов в грамматической типологии, а, во-вторых, включает и близкородственные пары языков (опять-таки в отличие от традиционного построения выборки в грамматической типологии). Тому есть ряд системных причин, среди которых и трудности качественного сбора материала, требующего плотной работы с носителем языка, и зачастую полная невозможность обращения к имеющимся справочникам, которые попросту не содержат нужной информации, в то время как типологи, изучающие грамматические явления, могут обойтись соответствующими справочниками, фиксирующими всё необходимое. Всё это затрудняет привлечение материала большого количества языков. Кроме того, для исследований лексики (например, лексической типологии) оказывается не столь важна степень родства между языками⁸. В нашей базе данных около 30 языков (и база продолжает пополняться), и среди них есть такие, которые относятся к одной группе по генетической классификации языков (например, русский и сербский).

Еще одно поле содержит информацию о релевантном для данной записи животном. Иными словами, название какого животного сочетается с интересующим нас глаголом. С исследовательской точки зрения это поле базы также проблематизировано. Например, не любой язык способен дать лингвистически специфичный материал для каждого животного. С одной стороны, по отношению к ряду животных употребляются глаголы звучания, изначально использующиеся для других, не звериных, источников звука. Так, в привычном для русского языка словосочетании *слон трубит* глагол *трубить* по отношению к животному *слон* очевидным образом вторичен. С другой стороны, не все животные издают звуки. А из тех, что издают их, не все имеют для этого процесса специфическое с точки зрения языка описание. Скажем, в современном русском языке нет специального (и даже регулярного) глагола, обозначающего звук, издаваемый лисой. Хотя, возможно, такой глагол существовал в древнерусском языке, насколько можно судить по тексту «Слова о полку Игореве»: «Уже бо бѣды его пасеть птиць подобію, вльци грозу въсрожать по яругамъ, орли клетомъ на кости звѣри зовуть, лисици *бреиуть* на чръленыя щиты».

Наконец, списки самих названий животных для разных языков могут быть неодинаковыми. Скажем, некоторые языки могут не различать членов пар «корова — бык», «ворон — ворона» и т. д.

С точки зрения иерархии представления информации, центральным в этой группе полей, естественно, является поле, называемое глагол в данном языке. Оно является своего рода «входом» для всей записи и единицей классификации.

К глаголу прикреплен его переводной эквивалент, то есть простой перевод глагольной лексемы на русский язык. Такой перевод может быть не вполне точен и по значению либо шире, либо уже соответствующего ему русского глагола. Если мы наблюдаем несоответствие семантики, то должны сделать пояснения на этот счёт и занести их в базу. Однако диапазон расхождения для прямого значения глаголов, обозначающих звуки животных, не слишком велик и поэтому пояснения необходимы далеко не всегда.

Каждая запись также содержит поле для фиксации примера употребления глагола в прямом значении.

Более сложную информацию представляет вторая группа полей, описывающих метафорическое употребление глаголов.

Во-первых, каждый глагол в своем производном значении должен иметь другой переводной эквивалент, очевидно, не совпадающий с переводом первичного значения. Так, звук *murista*, издаваемый целым рядом животных в финском языке (кошка, лев, морской лев, медведь, кабан, волк, собака), можно обозначить русским глаголом *рычать*, а его метафорическое употребление лежит в области физиологических звуков, издаваемых человеком, и должно иметь перевод *урчать в животе*.

И для прямого значения, и для метафорического в базе заложены поля, описывающие грамматические конструкции, в которых употребляются интересующие нас глаголы. Это сделано для того, чтобы зафиксировать возможную разницу между употреблениями глаголов в разных значениях. Так, русский глагол *мычать* может быть употреблен в переносном значении в составе конструкции «мычать, что...», но такая конструкция была бы невозможна при употреблении глагола в прямом значении.

Центральным в группе полей, описывающих метафорические переносы глаголов, является поле, фиксирующее тип результата при метафорическом переносе (фрейм) — своего рода теги, по которым в будущем можно будет легко находить в базе случаи одинаковых метафорических переносов. Заполнение этого поля представляет собой наибольшие трудности, отчасти затрагивающие проблему методологического дуализма, о котором говорилось ранее. Заранее заданный для участников проекта список таких типов оказался слишком приблизительным и не учитывал всего реального многообразия языкового материала. В результате нередки были случаи, которые заставляли экспертов по конкретным языкам «притягивать» имеющиеся теги к не слишком подходящим переносам. Иными словами, в какой-то части априорного списка типов результата при метафорическом употреблении явно не хватало необходимых пунктов. Одновременно с этим в процессе внимательного рассмотрения собранного в базе данных материала обнаружились и такие теги, необходимость которых либо не подтверждалась реальными языковыми данными, либо могла быть поставлена под сомнение из-за их неудачной формулировки. Так, к примеру, в ходе доработки списка тегов было принято решение убрать из него пункт, описывающий «веселый, открытый смех», так как материал, появившийся в результате работы с информантами, обнаруживал очевидное дублирование в данном случае других разновидностей смеха, не отличимых друг от друга по языковым данным.

Предварительный список тегов и процесс его улучшения требуют специального описания. Пока ограничимся тем, что отметим главное: типы результата при метафорическом переносе глаголов звуков животных могут быть приблизительно поделены на разнообразные звуки, издаваемые человеком, природными объектами и артефактами, а также не звуки (например, глаголы, используемые для описания ощущения и состояния, как русское *куковать*).

При работе с этой второй группой полей, описывающих метафорическое употребление глаголов, крайне важным является поле, фиксирующее пример такого употребления и его перевод. Только это поле способно дать представление о реально стоящем за глаголом значении. Некачественный пример в данном случае может выглядеть следующим образом: *кукует как кукушка (о человеке)*. Такое высказывание не дает

понимания лингвистической природы употребления глагола. Действительно ли глагол *куковать* регулярно употребляется в интересующем нас языке с метафорическим переносом на человека или это всего лишь окказиональный случай, порожденный в идиостиле информанта? Каковы границы семантики такого употребления? На все эти вопросы с помощью приведенного примера ответить невозможно. Напротив, метафорическое употребление корейского глагола *swik-swik-keli-ta* (“шипеть”, в прямом значении употребляется в сочетании со змеей) довольно прозрачно поступает при рассмотрении таких примеров: *Pwungsen-eyse palam-i* [X] *swik.swik-keli-mye* [V] *say-e nao-ass-ta* “Из воздушного шарика, свистя, выходил воздух”; *Yele kay-uy sinhothan-i* [X] *swik-swik-keli-mye* [V] *sos-a-ol-lass-ta* “Несколько сигнальных пуль взлетело, свистя”.

В первоначально разработанной архитектуре базы данных были и другие поля, дальнейшая работа с которыми породила сложности такого уровня, что приходится поставить их будущее в общей структуре базы под сомнение. Среди них поле оценки. Замысел, подразумевавший наличие этого поля, был в том, чтобы фиксировать оценочные компоненты значения глагола для разных языков. Скажем, глагол *ржать* в русском языке в метафорическом употреблении всегда подразумевает негативную оценку говорящим. В то же время на практике при работе с информантами и при дальнейшей обработке материала чрезвычайно трудно разграничить объект приложения такой оценки: оценивается таким образом действие, его субъект, производимый звук или что-то еще. Вероятнее всего, это поле будет удалено из окончательного списка полей базы.

Такая архитектура базы данных, как мы надеемся, должна охватывать большинство потребностей исследователей, которые захотели бы воспользоваться ею для своих целей. Как кажется разработчикам, эти цели могут лежать, прежде всего, в двух плоскостях. Первую можно условно обозначить как лексикографическую. База данных могла бы стать основой для нового типа словаря, фиксирующего гораздо больше нюансов значений, чем это обычно происходит. Такой словарь мог бы быть полезен и изучающим новый для себя язык, и опытным переводчикам. Вторую плоскость можно назвать исследовательской, она связана с чисто академическим интересом возможных пользователей, которые захотят осуществить на этом материале свои исследовательские проекты.

База будет представлена в Интернете в свободном доступе.

¹ Данное научное исследование (№ 14-05-0060) выполнено при поддержке Программы «Научный фонд НИУ ВШЭ» в 2014 г.

² Рахилина Е.В., Резникова Т.И. Фреймовый подход к лексической типологии // Вопросы языкознания. 2013. № 2. С. 3-31.

³ Там же.

⁴ См. Рахилина Е. В. Звуки Му // Проблемы грамматики и типологии : сборник статей памяти В. П. Неद्याкова : сб. науч. тр. М. : Знак, 2010. – С. 283–302.

⁵ Кюсева М. В., Резникова Т. И., Рыжова Д. А. Типологическая база данных адъективной лексики // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 29 мая - 2 июня 2013 г.). В 2-х т. Т. 1: Основная программа конференции. Вып. 12

(19). М. : РГГУ, 2013. P. 407-418.

⁶ Kostyrkin A. V., Panina A. S., Reznikova T. I., Bonch-Osmolovskaia A. A. Constructing a lexico-typological database (for a study of pain predicates) // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 30 мая - 3 июня 2012г.). В 2-х т. Т. 1: Основная программа конференции. Вып. 11 (18). М. : РГГУ, 2012. P. 288-295.

⁷ Интерфейс базы доступен в Интернете по адресу: <http://nevmenandr.net/cgi-bin/zvukimu.pl>

⁸ Reznikova T., Rakhilina E., Bonch-Osmolovskaya A. Towards a typology of pain predicates // Linguistics. Volume 50, Issue 3 (May 2012). P. 421-465.