

ОТЗЫВ официального оппонента

о диссертации И. Д. Семеновой «Параметрический анализ лексики карачаево-балкарского языка на тюркском фоне» (Воронеж, 2017 – 339 стр.), представленной на соискание ученой степени кандидата филологических наук по специальности 10.02.02 – Языки народов Российской Федерации (карачаево-балкарский язык)

Работа Индиры Джотдаевны Семеновой выполнена в рамках квантитативного подхода к лексической типологии. Автор пробует создать лексический профиль карачаево-балкарского языка на основе количественной оценки нескольких параметров лексической системы. И.Д.Семенова опирается, прежде всего, на метод В.Т.Титова, который предусматривает анализ словаря некоторого языка по четырем параметрам: функциональному (отражающему длину словоформы, являющейся словарным входом, что предположительно коррелирует с активностью использования), синтагматическому (отражающему фразеологическую активность), парадигматическому (отражающему объем синонимических рядов с участием данного слова) и эпидигматическому (отражающему полисемичность слова).

Данный метод интересен тем, что позволяет проводить формализованное сопоставление лексических систем разных языков. Уже обследованы таким образом некоторые германские, романские, славянские языки, новогреческий, а также несколько тюркских языков. Такой метод выглядит интересным и перспективным для типологической оценки лексических систем как целого. Можно приветствовать дальнейшее расширение спектра языков, обследованных с помощью такого подхода.

Тем не менее, сразу обращают на себя внимание некоторые ограничения: метод опирается на работу только с двуязычными словарями, но не с корпусами на данном языке. В результате, функциональный параметр фактически учитывает длину словоформы, тогда как мог бы учитывать реальную частоту употребления в корпусе. Возражение, приводимое на стр. 13 о том, что частотность слов существенно различается по разным частотным словарям, со ссылкой на работу Морковкина 1983 года, не вполне валидно, так как за последние 35 лет был достигнут значительный прогресс в корпусной лингвистике, появились взвешенные по жанрам большие корпуса различных языков и т.п. Синтагматический параметр учитывает входимость слова в разные фразеологические сочетания, тогда как анализ корпуса позволил бы вычислить более общую сочетаемость в рамках устойчивых биграмм и триграмм, которая зачастую проходит мимо составителей словарей. Эпидигматический параметр действительно лучше учитывать с помощью словарей. Парадигматический параметр относится к концептам, тогда как остальные параметры – к конкретным словоформам. Получается, что слово, являющееся малоупотребительным и редким синонимом некоторого важного для языка концепта, автоматически попадает в ядро

языка по данному параметру. Если оно при этом короткое, то оно попадет в ядро языка сразу по двум параметрам.

Было бы естественно дополнить этот метод корпусным анализом. Тем более, что современные методы машинной обработки текстов вполне позволяют при наличии хороших электронных словарей и/или корпусов автоматически рассчитать все четыре используемых параметра и, таким образом, быстро расширить имеющуюся типологическую выборку для типологического анализа лексических систем.

Естественным образом, разные принципы создания словарей, подходы к выделению значений, их переводам и толкованиям, подаче материала все это создает шум, мешающий типологической работе. В этом отношении, работа И.Д.Семеновой представляет собой хорошую попытку пробиться через этот шум, чтобы выделить существенные признаки лексической системы карачаево-балкарского языка и привести их к виду, сопоставимому с другими языками.

В основу работы положены двуязычные карачаево-балкарско-русские словари, Большой (ориентированный на карачаевский вариант языка) и Малый (ориентированный на балкарский вариант). Не вполне ясно, почему в работу не были включены данные трехтомного толкового словаря карачаево-балкарского языка (1996, 2002, 2005 гг). Достаточно очевидно, что, по крайней мере, фразеологические сочетания должны быть лучше представлены в одноязычном словаре. Возможно, стоило потратить некоторые усилия на перевод этого словаря в электронную форму, поскольку его данные могли бы существенно повлиять на результаты.

В целом, работа выполнена на хорошем научном уровне, вносит свой вклад в уточнение методики сопоставления лексических систем, помогает расширить наше понимание устройства лексической системы карачаево-балкарского языка и, несомненно, вполне соответствует квалификационным требованиям. Тем не менее, следует сделать несколько замечаний.

Работе в целом, помогло бы использование современных методов математической статистики и автоматической обработки текстов. Использование Excel в качестве средства работы с данными выглядит архаизмом. Ограничения этой программы не позволяют, в частности, адекватно выстраивать синонимические ряды (основывающиеся при данном подходе на использовании метаслов в толкованиях/переводах). Куда эффективнее было бы проводить кластеризацию лексики с помощью, например, нейросетей.

Методика расчета весов не вполне прозрачна. Так, пытаясь преодолеть, вышеуказанное ограничение, связанное с тем, что замер длины фонетической словоформы лишь косвенно указывает на ее функциональную активность, автор вводит следующую поправку к исходной методологии «Ф-вес присваивался, всем словам, имеющим не нулевой Э-вес, С-вес, или П-вес» (стр. 87). Тем не менее, нигде в диссертации не указано, какой именно вес им присваивался, как он соотносился с их исходным Ф-весом и почему,

несмотря на эту добавку среди функциональных доминант карачаево-балкарской лексики все равно остаются только двухбуквенные слова.

Несколько раз в диссертации утверждается, что достоверность метода подтверждается тем, что параметрические ядра лексики в большом и малом словарях карачаево-балкарского языка совпали на 2/3. Кажется, что здесь слово «достоверность» используется не в терминологическом математическом смысле, а показанный результат совпадений оценивает в действительности не надежность метода, а степень близости словарей по своему качеству и структуре.

Аналогичным образом математический термин «значимость / незначимость» употребляется в описании расхождения тюркских языков по функциональному параметру (стр. 95). Для тофаларского языка рассматриваются только основы (в соответствии с принципом устройства тофаларского словаря В.И. Рассадина), а для прочих тюркских языков полные словоформы. При этом искажение данных объявляется незначимым, поскольку «обработанные множества (звуков) карачаево-балкарского, узбекского и турецкого языков были больше, чем множество (звуков) тофаларского языка». Данная фраза оставляет скорее недоумение, что именно было проделано для уменьшения искажения. Здесь было бы разумно провести стандартную статистическую оценку степени предполагаемого искажения параметра и показать, действительно ли она статистически значима или незначима.

Схожие аргументы применимы и к вычислению «Коэффициента типологической близости», где использование ядер разного размера (карачаево-балкарское ядро в 10 раз превышает по размеру тофаларское ядро) якобы не оказывает «существенного влияния на результаты исследования» (стр. 111). Очевидно, что такая диспропорция должна приводить к тому, что в карачаево-балкарском ядре найдется больше параллелей с любым другим тюркским языком, чем у прочих тюркских языков между собой. В еще большей степени это влияет и на «Индекс периферийности языков», что, впрочем, понимает и сам автор (стр. 119). Индекс периферийности основан на доле уникальной лексики в составе ядра, где уникальность определяется по наличию/отсутствию соответствующего русского переводного эквивалента у слов из ядер других языков. И здесь во весь рост встает проблема установления эквивалентности между русскими словами, использующимися в качестве «метаслов». И здесь для исключения ложных «уникалий» можно было воспользоваться системами вроде WordNet: позволяющими создавать крупные синонимические ряды.

Есть небольшое замечание и к библиографии. Например, автор часто ссылается на работу [Кретов 2011], однако в списке литературы встречается целых три работы А.А. Кретова за 2011 год (под номерами 55, 65 и 69) и остается только гадать, на какую именно работу (или на разные работы в разных местах) ссылается автор диссертации. Напротив, отсутствуют важные работы по лексической многозначности, которая активно обсуждается в

диссертации, например, Зализняк А.А. «Многозначность в языке и способы её представления» 2006.

Автореферат вместе с публикациями отражает содержание диссертации и научную концепцию ее автора.

Все сказанное свидетельствует о достаточной профессиональной подготовке автора диссертации, ее диссертационное исследование «Параметрический анализ лексики карачаево-балкарского языка на тюркском фоне» отвечает требованиям «Положения о порядке присуждения ученых степеней» (№ 842 от 24 сентября 2013 г, п. 9-11, 13-14) и является научно-квалификационной работой, в которой содержится решение задач, имеющих существенное значение для соответствующих отраслей знаний – лексической типологии и тюркской лексикологии.

Индира Джотдаевна Семенова достойна присуждения искомой ученой степени кандидата филологических наук по специальности 10.02.02 – Языки народов Российской Федерации (карачаево-балкарский язык).

старший научный сотрудник
отдела урало-алтайских языков
ФГБУН Институт языкоznания РАН,
кандидат филологических наук по специальности 10.02.20 - «Сравнительно-историческое, типологическое и сопоставительное языкоzнание»
Грунтов Илья Александрович

23 января 2018 года

Грунтов Илья Александрович
Электронная почта: altaica@narod.ru
Рабочий адрес: Москва 125009
Б. Кисловский пер. д.1 стр. 1
тел.: (495) 690-35-85

