

**М. А. Лис**

**ИСПОЛЬЗОВАНИЕ МАШИННОГО ОБУЧЕНИЯ  
ДЛЯ АВТОМАТИЧЕСКОГО ПЕРЕВОДА  
И ТРАНСКРИПЦИИ ГЕОГРАФИЧЕСКИХ НАЗВАНИЙ:  
СОВРЕМЕННЫЕ ПОДХОДЫ И ПЕРСПЕКТИВЫ**

Географические названия, или топонимы представляют собой реалии, которые передают национально-культурные особенности. Это, в свою очередь, отражается не только во внутренней форме имени, но оказывает непосредственное влияние на возможность передачи информации о них теми же средствами, какими происходит перевод апеллятивных единиц. Тесная связь с отражением индивидуальных объектов вынуждает уделить большее внимание звукобуквенному комплексу этой лексической единицы. Поэтому, в процессе адаптации имени собственного на другой язык важно передать эти характерные черты.

Говоря о передаче названий географических объектов с иностранных языков, мы можем выделить несколько принципов:

- 1) транскрибирование;
- 2) транспозиция;
- 3) калькирование;
- 4) смешанный способ (транскрипция и калькирование);
- 5) традиционное соответствие.

**Транскрибирование** представляет собой графическую запись звучания слова, которая позволяет точно передать средствами языка перевода звучание исходного топонима (*Midway* – ‘Мидуэй’). Данный способ является наиболее распространенным, так как при помощи него соблюдается принцип взаимно однозначного соответствия между языком оригинала и графическими соответствиями в языке перевода. Однако может возникнуть случай, когда языковая единица передана при помощи искаженной транскрипции, где запись звуков на языке перевода не соответствует исходной. В некоторых случаях она может сохраняться.

Говоря о переводе имён собственных, нельзя не упомянуть принцип этимологического соответствия, или **транспозицию**. В переводе имен переводчик стремится сохранить смысловое содержание исходного имени, учитывая особенности целевой культуры. Для этого переводчик использует метаперевод, создавая в сознании не только переводческое, но и метапереводческое пространство. В этом и заключается суть данного принципа – использование тех имён собственных, которые разнятся по форме в разных языках, однако имеют одно и то же лингвистическое происхождение: *Michael* – ‘Михаил’, *Eugene* – ‘Евгений’ и др.

Ещё один способ – **калькирование** – создает новое слово посредством копирования структуры исходного (*North Peak* – ‘Северная Вершина’). При использовании данного переводческого приёма происходит процесс изменения формы слова, падежа, порядка, окончания и так далее.

**Смешанным способом** переводчики пользуются в том случае, когда работают с географическими объектами, содержащими в своей структуре указания сторон света и прилагательные.

Также при передаче иноязычных географических имён может сохраняться их устоявшаяся традиционная форма: Navajo – *‘Навахо’* и т. д. Данный переводческий приём носит название **традиционного соответствия**.

Таким образом, только при традиционном способе перевода географических названий могут возникнуть трудности, так как переводчику нужно четко определять, какой или какие из приемов следует применить для того, чтобы перевод географического названия на иностранный язык являлся полным и адекватным с точки зрения языковых, а также культурных особенностей.

Автоматизация перевода и транскрипции топонимов является сложной задачей, требующей учета лингвистических правил, исторических факторов, культурных особенностей и вариативности написания. В последние годы машинное обучение стало мощным инструментом для решения этой задачи, предлагая новые подходы и перспективы. Предельной задачей искусственного интеллекта является понимание естественного языка. При этом в качестве ввода и вывода может использоваться как речь, так и текст на одном из естественных языков, представленный в письменной форме. Но, если мы возьмем для перевода, допустим, имя собственное на английском языке и попытаемся сделать передачу на белорусский с помощью уже существующих систем автоматического перевода, то результат будет не всегда верным.

Современные системы автоматического перевода работают на базе нейронных сетей, которые обучаются на корпусах текстов и другой текстовой информации, находящейся в открытом доступе. Однако для многих языковых пар отсутствует достаточное количество параллельных данных для обучения моделей МО. В примере перевода с английского на белорусский у системы машинного обучения недостаточно данных, чтобы обучиться корректной передаче. Поэтому, целесообразным является обучать нейронную сеть, или систему машинного обучения, не только по корпусам, но также и на правилах, тем самым используя гибридный способ при моделировании системы автоматического перевода.

Стоит отметить, что в Республике Беларусь разработаны правила транслитерации, но они действуют для передачи имен собственных с белорусского или русского языка на латиницу, но не в противоположном направлении. Поэтому в рамках исследования на основе Инструкции по транслитерации фамилий и собственных имен граждан Республики Беларусь при включении их персональных данных в регистр населения [46], а также закона «Абмовах у Рэспубліцы Беларусь» [74] предлагаются правила передачи имен собственных на белорусский язык.

Данная фактическая информация об особенностях передачи имен собственных географических объектов применяется для последующего создания модели, обученной специально на данных по географическим названиям, с учетом их особенностей и характеристик. Для этого правила передачи

преобразуются в программный код, реализующий алгоритм перевода англоязычных топонимов на белорусский язык, который в последующем также используется для моделирования системы машинного обучения для автоматического перевода географических названий. Затем необходимо создать параллельный корпус географических названий на белорусском, русском и английском. Это будет включать в себя уже созданный в рамках исследования автоматический словарь, или базу данных, топонимов, содержащий 1000 вхождений, а также справочники географических названий, веб-страницы и документы, содержащие топонимы как на белорусском, так и на русском и английском языках. Все эти данные необходимо будет формализовать таким образом, чтобы нейронная сеть смогла обучиться. В конечном итоге будет создана система машинного перевода географических названий, которая будет учитывать как языковые, так и культурные особенности географических названий Беларуси.

Подводя итог, хочется отметить, что создание такой системы не только позволит значительно автоматизировать процесс перевода и транскрипции географических названий, повысить его точность и эффективность, но и обеспечит соблюдение языковых норм и культурного наследия Беларуси при работе с геопространственными данными. Данная система станет ценным инструментом для картографов, лингвистов, разработчиков программного обеспечения и всех, кто работает с географической информацией на белорусском языке, способствуя сохранению и популяризации белорусской культуры и языка в цифровом пространстве.