

СОВРЕМЕННЫЕ ПОДХОДЫ В ЛИНГВИСТИЧЕСКОЙ ТИПОЛОГИИ: ОТ УНИВЕРСАЛИЙ К ВЕРОЯТНОСТНЫМ МОДЕЛЯМ

Пардаева Мохинбону Фазлиддиновна

Узбекский Государственный Университет мировых языков, Ташкент

Преподаватель кафедры «Лингвистика и английская литература»

moxinbonu0520@gmail.com

Аннотация. Типология языков — одно из ключевых направлений в современной лингвистике, цель которого — описать и объяснить структурное разнообразие языков мира, выявить универсалии и корреляции между языковыми признаками. В статье рассматриваются основные концептуальные подходы в лингвистической типологии, включая классические универсалии (Гринберг), количественную типологию, вероятностные и статистические модели, а также современные методы машинного обучения, применяемые к задачам типологии. Особое внимание уделяется проблемам выборки языков, искажениям из-за ареальных и генетических влияний, а также перспективам интеграции типологических данных в системы NLP (Natural Language Processing). В завершение обсуждаются ключевые вызовы и дальнейшие направления исследований в типологии.

Ключевые слова: лингвистическая типология, языковые универсалии, количественная типология, выборка языков, вероятностные модели, NLP.

Введение

Лингвистическая типология занимается сравнением языков на основании их структурных свойств (фонология, морфология, синтаксис, семантика), не фокусируясь на их родстве, а на «типах» и «узорах» языковых конструкций [Oxford Handbook, см. обзор раздела «Linguistic Typology»]. Основная задача типологии — выявление **универсалий** (абсолютных или статистических) и **корреляций** между типологическими признаками, а также объяснение, почему некоторые языковые структуры встречаются чаще, а другие — реже. Типология пересекается как с теоретической, так и с прикладной лингвистикой (например, с NLP и языковыми ресурсами) [Ponti et al., 2018].

В этой статье я излагаю ключевые подходы в типологии, обсуждаю трудности, связанные с выборкой языков и ареальными влияниями, а также рассматриваю современные вероятностные и вычислительные методы. Далее — обзор классических универсалий, введение в количественную типологию, проблемы методологии, современные модели и перспективы.

Основные подходы в типологии

Универсалии Гринберга и импликационные законы

В середине XX века Джозеф Гринберг предложил множество **универсалий** — положений, предполагающих, что если в языке существует признак А, то обязательно существует признак В (импликационные законы) или что большинство языков обладают признаком С (статистические). Например: «если язык имеет порядковую конструкцию А, то он, как правило, имеет порядок В» [Greenberg 1963, цитируется в последующих обзорах].

Такие универсалии дают теоретическую опору для типологии, но они работают не как строгие законы, а как вероятностные тенденции. Они часто пересматриваются под влиянием эмпирических данных и расширения числа исследованных языков.

Количественная типология и статистические модели

Современная типология всё больше ориентируется не на «категорические типы», а на **количественные шкалы и распределения**. Это позволяет фиксировать не просто «наличие / отсутствие признака», но степень выраженности, частоту, корреляции с другими признаками. Такой подход иногда называют **типометрика** (typometrics) [Gerdes, Kahane & Chen et al. 2021].

К примеру, анализ универсальных зависимостей по позициям объекта-в-пространстве (OV / VO) в корпусах Universal Dependencies позволяет измерять процентное распределение объектов слева или справа от глагола и строить «распределения» по языкам, а не жестко классифицировать языки как OV или VO [Gerdes et al., 2021].

Такой подход даёт более гибкие представления о языковом разнообразии и позволяет выявлять **количественные универсалии**, отличающиеся от традиционных импликационных закономерностей.

Вероятностные и вероятностно-генеративные модели

С развитием статистического машинного обучения появились подходы, которые рассматривают типологические признаки как **стохастические величины**. Например, модель «A Probabilistic Generative Model of Linguistic Typology» строит обобщённое вероятностное пространство признаков языков, в котором языки и признаки моделируются совместно [Bjerva et al., 2019].

Другие исследования занимаются **вероятностной типологией**: например, Cotterell & Eisner (2017) используют глубокие генеративные модели для прогнозирования вероятностей в фонетической типологии (например, вероятности наличия определённых гласных).

Также подходы на основе **language embeddings** пытаются обучать представления языков, из которых затем предсказываются типологические свойства (фонетические, морфологические или синтаксические) [Bjerva & Augenstein, 2018]. Таким образом, модель «учится» типологию из текстовых данных, без явного ручного кодирования признаков.

Методологические сложности в типологическом исследовании

Выборка языков и балансировка

При сравнении языков важно, чтобы выборка **не была смещена** по генетическому или ареальному признаку: например, если в выборке преобладают индоевропейские

языки, выводы могут оказаться искаженными. Поэтому в типологических базах (например, WALS) применяют стратифицированные выборки, чтобы обеспечить представительство разных семей и географических зон [Oxford Handbook обзор].

Однако даже такие выборки не гарантируют отсутствие **контактных и ареальных эффектов**. Например, соседние языки могут взаимно влиять и приводить к конвергентности признаков, что затрудняет отделение наследования от заимствований. Для решения этого, в некоторых исследованиях используют **тройные языковые наборы**: фокус-язык, соседний (контактный), и контрольный язык, не контактирующий с ними, чтобы отделить влияние контакта от генетики [Di Garbo & de Souza, 2023].

Качество данных и кодирование признаков

Типологические базы (например, WALS) часто оперируют дискретными, категориальными признаками (например, "есть / нет", "тип А / тип В"). При этом внутри каждой категории могут входить языки с существенными различиями, и тонкие вариации теряются. Это ограничение особенно заметно, когда признаки условно бинарны или сгруппированы по широким классам [Oxford Handbook обзор].

Другая проблема — **пропущенные или неполные данные**: не для всех языков известны все нужные признаки, что требует методов гаперолирования (imputation) или работы с неполными данными.

Корреляции и конфликты признаков

В типологии часто наблюдаются **корреляции между признаками**: наличие признака А статистически связано с признаком В. Но корреляция не означает причинности — нужно аккуратно интерпретировать данные. Кроме того, **конфликты между признаками** (когда один признак «говорит» об одном типе, а другой о другом) требуют комплексных моделей и анализа.

Современные вероятностные подходы позволяют учитывать взаимодействия признаков (ковариации) и учитывать неполные данные [Bjerva et al., 2019].

Современные приложения и направления

Типология и NLP

Типологическая информация всё активнее используется в задачах машинного перевода, анализа синтаксических структур и моделирования языков с низким ресурсом. Например, Ponti et al. (2018) обзорно анализируют, как типологические признаки успешно (но ограниченно) интегрируются в системы многоязыковой обработки — и указывают на ограничения баз данных и необходимость более гибких представлений признаков [Ponti et al., 2018].

Language embeddings используются как способ автоматического вывода типологических признаков, что может помочь обойти ограниченность ручных баз [Bjerva & Augenstein, 2018].

Количественные исследования и выявление новых универсалий

С помощью больших унифицированных корпусов (например, Universal Dependencies) и количественных методов возможно открывать **новые универсалии**, не очевидные в классических подходах. Например, исследование Gerdes et al. (2021)

демонстрирует, как распределения порядка слов могут быть использованы для выявления **количественных универсалий**, отличных от традиционных.

Такие подходы позволяют рассмотреть человеческую языковую вариативность как **континуум**, а не жёсткие классы.

Пример структурного анализа: порядок слов как классический признак

Порядок слов (S, V, O) — один из наиболее изученных типологических признаков. В классическом подходе языки классифицируются как SVO, SOV, VSO и др. — по доминантному порядку.

Но количественные методы показывают, что порядок слов часто не является строго фиксированным: в реальных корпусах языки проявляют «смешанные» варианты. Например, Gerdes et al. анализировали процентное распределение позиции объекта по отношению к глаголу и получили, что языки часто располагаются не чётко в одной категории, а на спектре между OV и VO [Gerdes et al. 2021].

Это позволяет моделировать языки как точки в многомерной пространственной шкале, а не как элемент одной категории.

Кроме того, такие данные можно использовать для проверки классических универсалий Гринберга, но уже через призму статистики и распределений.

Заключение

Лингвистическая типология прошла путь от качественных описаний и универсалий Гринберга к современным количественным и вероятностным моделям. Сегодня типология обращается к большим корпусам, машинному обучению, embeddings и интеграции с NLP.

Ключевые методологические вызовы остаются связаны с выборкой языков, контролем ареального/контактного влияния, неполнотой данных и интерпретацией корреляций. Тем не менее интеграция классических подходов и современных вычислительных методов открывает перспективы для построения более гибких, точных моделей языковой вариативности.

В дальнейшем важно продолжать расширение типологических баз, развитие методов автогенерации признаков и укрепление связи между типологией и обработкой естественных языков.

Список литературы

- Bjerva, J., Augenstein, I. (2018). *From Phonology to Syntax: Unsupervised Linguistic Typology at Different Levels with Language Embeddings*. arXiv preprint.
- Bjerva, J., Kementchedjhieva, Y., Cotterell, R., Augenstein, I. (2019). *A Probabilistic Generative Model of Linguistic Typology*. arXiv preprint.
- Cotterell, R., Eisner, J. (2017). *Probabilistic Typology: Deep Generative Models of Vowel Inventories*. arXiv preprint.
- Di Garbo, F., de Souza, R. N. (2023). *A sampling technique for worldwide comparisons of language contact scenarios*. Linguistic Typology.

- Gerdes, K., Kahane, S., Chen, X. et al. (2021). *Typometrics: From Implicational to Quantitative Universals in Word Order Typology*. Glossa: a journal of general linguistics.
- Ponti, E. M., O’Horan, H., Berzak, Y., Vulić, I., Reichart, R., Poibeau, T., Shutova, E., Korhonen, A. (2018). *Modeling Language Variation and Universals: A Survey on Typological Linguistics for Natural Language Processing*. arXiv preprint.