

Машинная межъязыковая транскрипция имен собственных
Елкин С.В., Клышинский Э.С., Максимов В.Ю., Аминова С.Н., Мусаева Т.Н.
Институт прикладной математики им. М.В. Келдыша РАН
ipm4info@gmail.com

Актуальность проблемы транскрипции имен собственных ощущается особенно остро при попытке передать фамильно-именные группы, географические названия и названия организаций средствами другого языка. От этого напрямую зависит правильная идентификация объектов.

Традиционно существуют три способа передачи имен собственных на другом языке: перевод, транскрипция и транслитерация. Если в транслитерации используется побуквенная замена, транскрипция подразумевает передачу имен собственных посредством исторически сложившейся орфографической системы языка.

К сожалению, в текстах можно встретить передачу одного и того же имени или названия всеми тремя способами. При этом число вариантов часто не ограничивается тремя. Большие трудности возникают при попытке передать имя собственное иностранного (для языка-оригинала) происхождения, выраженное на каком-либо третьем языке, а также при попытке найти какую-либо информацию на языке оригинала по имеющейся транскрипции иностранного имени собственного.

С целью унификации межъязыковой транскрипции была создана программа автоматической транскрипции «Транскриба», в настоящее время осуществляющая транскрипцию между 27 языками. В это число входят и неевропейские языки. С целью устранения неоднозначностей при транскрипции в качестве языка-посредника использована единая фонетическая таблица (ЕФТ). В итоге в процессе транскрипции с любого языка на любой язык используется 2 базы правил: правила для перевода с языка-оригинала на ЕФТ, и правила для перевода с ЕФТ на выбранный язык. Такая схема при сравнении со стандартной схемой «1 набор правил для одного направления транскрипции в заданной паре языков» дает неоспоримые преимущества: эффективность, сокращение числа баз и затрат на их согласование, снижение числа неоднозначностей при передаче звуков, простота подключения новых языков.

Исторически сложившиеся варианты передачи некоторых имен собственных вводятся в базу исключений, которая постоянно пополняется и расширяется путем включения новых имен собственных. Тестирование программы проводилось на значительных выборках фамильно-именных групп, охватывающих от 1 до 25 тысяч имен. При этом число ошибок не превышало 3%.

По результатам работы выпущено 2 издания монографии «Практическая транскрипция фамильно-именных групп» [1]. В дальнейшем программа автоматической транскрипции «Транскриба» может послужить ядром для создания единой сетевой глобальной системы межъязыковой транскрипции имен собственных.

Список литературы

1. С.Н.Аминова, Н.А.Антонец и др. Практическая транскрипция фамильно-именных групп, изд. 2-е исправленное и дополненное. М.: Наука, 2006.
2. E.Klyshinsky, V.Maximov, S.Yolkeen, Cross-Language Transcription of Proper Names Language Forum (A Journal of Language & Literature V.34 №1 Jan-Jun 2008) pp. 137-152