ORES
Научно-практический электронный журнал
Оригинальные исследования (ОРИС), *ISSN 2222-0402*
www.ores.su
original-research.ru

КОДОВОЕ ПЕРЕКЛЮЧЕНИЕ И ИНТЕРФЕРЕНЦИЯ В АНГЛО-ХИНДИ МЕДИАРЕЧИ: ДИАГНОСТИЧЕСКИЕ КРИТЕРИИ И НОРМАЛИЗАЦИЯ СМЕШАННОЙ РЕЧИ В СУБТИТРАХ

Коростелева Виктория Игоревна,
Старший преподаватель, Дальневосточный Федеральный Университет, Восточный институт - школа региональных и международных исследований. Россия, Хабаровск
miestrellav@gmail.com

Аннотация

В статье предложена многоуровневая диагностика кодового переключения и интерференции в англо-хинди медиаречи. Материал включает фрагменты NDTV, India Today, The Print, Mojo Story, примеры индийского английского, эпизоды с субтитрами. Оценка ведётся на фонетическом, морфологическом, синтаксическом, лексическом, прагматическом, мультимодальном уровнях с учётом грамматической автономности сегмента, матричной рамки, степени адаптации, жанровой функции. Для субтитров систематизированы нивелирование кода, утрата дискурсивов, ошибки в именах и числах, исчезновение просодических границ. Сверка аудио с текстовым каналом рассматривается как обязательная операция корпусной разметки.

Ключевые слова: кодовое переключение, интерференция, англо-хинди медиаречь, Hinglish, субтитры, ASR, мультимодальность

CODE-SWITCHING AND INTERFERENCE IN ENGLISH-HINDI MEDIA SPEECH: DIAGNOSTIC CRITERIA AND NORMALIZATION OF MIXED SPEECH IN SUBTITLES

Victoria I. Korosteleva
Senior Lecturer, Far Eastern Federal University, Oriental Institute-School of regional and international relations, Russia, Vladivostok
miestrellav@gmail.com

ABSTRACT

The article proposes a multilevel diagnostic procedure for code-switching and interference in English-Hindi media speech. The material includes NDTV, India Today, The Print, Mojo Story fragments, examples of Indian English and subtitled audiovisual episodes. Classification combines phonetic, morphological, syntactic, lexical, pragmatic, multimodal criteria with grammatical autonomy, matrix-frame stability, adaptation and genre function. Four subtitle distortions are systematized: code levelling, loss of discourse markers, errors in names and numbers, loss of prosodic boundaries. Audio-to-text verification is treated as a mandatory operation in corpus annotation of mixed-language audiovisual media.

Keywords: code-switching, interference, English-Hindi media speech, Hinglish, subtitles, ASR, multimodality.

Англо-хинди медиаречь объединяет переключения между автономными языковыми сегментами с устойчивыми контактными моделями индийского английского, вследствие чего одинаковая внешняя форма смешения может получать разную лингвистическую квалификацию. В концепции Дж. Гамперца кодовое переключение связано с соположением фрагментов двух грамматических систем в пределах речевого обмена [5], модель матричного языка К. Майерс-Скоттон описывает распределение матричной рамки, включённых элементов [7], типология Ш. Поплак фиксирует теговые, межфразовые, внутрифразовые переходы [8]. Последовательный подход П. Ауэра дополняет структурную диагностику функцией языкового выбора в развёртывании взаимодействия [2]. Для англо-хинди материала совместное применение этих моделей переводит квалификацию единицы из формального признака смешения в процедуру проверки нескольких уровней.

В аудиовизуальном медиатексте звучащая реплика получает текстовую репрезентацию через субтитры, ASR, машинный перевод, где двуязычный фрагмент может сводиться к одному языку, прагматический маркер заменяться функциональным аналогом, имя или число унифицироваться. Мультимодальный анализ связывает речевой канал с экранным текстом, изображением, жестом [1; 3], исследования многоязычного субтитрования фиксируют потерю языкового разнообразия при нормализации [6]. Предмет анализа охватывает содержательное сохранение переключения, временные параметры появления субтитров образуют отдельную задачу.

Цель исследования состоит в формировании критериев разграничения кодового переключения, интерференции с описанием содержательных искажений смешанной речи в субтитрах. Материал включает эпизоды NDTV, India Today, The Print, Mojo Story, реплики индийского английского, англо-хинди конструкции с морфологической, синтаксической, прагматической интеграцией. Для каждой единицы определялись грамматическая автономность, матричная рамка, степень адаптации, устойчивость модели, жанровая функция, соответствие аудио субтитру.

Кодовое переключение квалифицируется через сохранение собственной грамматики соединяемых сегментов, граница перехода обычно читается на уровне клаузы, словосочетания, дискурсивного тега. Интерференция проявляется как перенос структур одной системы в другую через фонетическую адаптацию, морфологическое оформление, калькирование порядка, семантическое расширение, перенос прагматического маркера [4]. Срединная зона включает единичные вставки, смешанные конструкции, формы на пути к закреплённому заимствованию, её квалификация формируется совокупностью признаков.

Уровень	Диагностический признак	Показательный пример
Фонетика	Устойчивая адаптация формы	botal, tren, minit, ретрофлексия, хинди-ударение
Морфология	Основа одного языка получает оператор другого	recommend karũgā, ticket-e, śādī-s
Синтаксис	Перенос порядка, служебной модели, фокусной позиции	Do one thing, He doesn't listen only
Лексика	Стабильное локальное расширение значения	prepone, timepass
Прагматика	Дискурсивный маркер переносится через языковую рамку	yaar, achchha, matlab в английской реплике
Мультимодальность	Аудио, субтитр расходятся по языковому коду	Hinglish-фрагмент передаётся монолингвальным английским

Морфологический пример «Recommend karũgā tumhe…» включает английскую основу recommend в хинди-глагольный комплекс с оператором kar-, что помещает грамматическую обработку основы внутрь хинди-матрицы. Реплика «He doesn't listen only» сохраняет английскую лексику, синтаксическая позиция only воспроизводит локальную фокусную модель индийского английского. Формула «Do one thing» калькирует дискурсивный ход ek kaam karo, переход затрагивает организацию реплики, а в конструкции «Nothing is pakkā» хинди-прилагательное включено в английскую рамку, устойчивость прагматического значения усиливает интерференционную квалификацию [10].

Пограничная реплика «Maine sochā, I should not miss the opportunity» проверяется по границе клаузы, матричной рамке, функции английского отрезка, в то же время полная английская клауза сохраняет собственную структуру, а повторяемая стилистическая функция формулы создаёт прагматический интерференционный признак. Двухуровневая разметка фиксирует основную категорию кодового переключения, дополнительный признак прагматической интерференции.

Жанровое распределение интерференционных признаков

Теледебаты NDTV, India Today концентрируют межфразовые переходы при цитировании лозунгов, правовых формул, политических высказываний, прагматическая интерференция проявляется через yaar, achchha, anyway, этикетные кальки в английской рамке. Подкасты, стримы The Print, Mojo Story дают больше внутрифразовых профессионализмов, light-verb моделей, просодически выделенных вставок. Кино, сериальные диалоги совмещают клаузальные переходы с формами глубокой морфологической интеграции, включая śādī-s, ticket-e. Жанр в этой модели выступает параметром вероятной реализации механизма, квалификация конкретной единицы сохраняет опору на её грамматические признаки.

Нормализация смешанной речи в субтитрах

Сопоставление звучащей реплики с субтитром выявляет четыре повторяющиеся группы содержательных искажений. Нивелирование кода переводит весь смешанный фрагмент в один язык. Потеря дискурсивов устраняет yaar, achchha, matlab либо заменяет их нейтральным англоязычным маркером. Ошибки в именах, числительных, датах меняют участки, где переключение связано с графикой, фонетической адаптацией. Передача просодики через стандартную пунктуацию скрывает паузу, ударное выделение, смену регистра.

Тип искажения	Механизм	Лингвистическое следствие
Нивелирование кода	Смешанный фрагмент получает одну языковую форму	Граница переключения исчезает из текстовой версии
Потеря дискурсивов	Маркер опускается либо заменяется общим эквивалентом	Ослабевает прагматическая маркировка
Имена, числа	ASR унифицирует графику, ошибается в проприетах, числительных	Искажается точка языкового перехода
Просодическая граница	Пауза, ударность, смена регистра не отражаются пунктуацией	Текст маскирует акустический сигнал перехода

Для автоматического распознавания англо-хинди речи проблема имеет независимое корпусное подтверждение. Корпус VITB-HEBiC содержит 7,5 часа Hindi-English речи с представителями 27 штатов Индии, его авторы используют материал для оценки ASR-моделей в условиях кодового переключения [9]. В корпусе HiACC представлены 3318 взрослых, 1858 детских аудиосегментов общей длительностью 5,24 часа, базовые эксперименты дают около 42 % WER для стандартных моделей на тестовой выборке [9]. Эти значения относятся к внешним корпусам, в настоящем исследовании они служат подтверждением повышенной сложности автоматической обработки смешанной речи.

Для корпусного анализа медиатекста содержательная сверка строится на последовательном сопоставлении аудио, субтитра, языковой разметки. Сначала фиксируется звучащая форма с границей перехода, затем определяется матричный язык, после этого проверяется текстовый канал на сохранение переключаемого сегмента, дискурсивов, имён, чисел, просодически значимой пунктуации. Единица с монолингвальной нормализацией в субтитре сохраняет исходную языковую квалификацию по аудио, расхождение получает отдельную мультимодальную метку. Подобная схема предотвращает перенос свойств субтитра на речевое поведение спикера.

Разграничение кодового переключения, интерференции в англо-хинди медиаречи опирается на грамматическую автономность сегмента, матричную рамку, степень фонетической, морфологической интеграции, устойчивость синтаксической модели, прагматическую функцию, мультимодальное соответствие. Light-verb конструкции, локальные позиции only/also, кальки Do one thing, дискурсивы yaar, achchha, matlab представляют разные механизмы контактного взаимодействия. Жанровая среда распределяет ведущие признаки, теледебаты, интервью усиливают прагматический слой, цифровые форматы концентрируют морфологическую, просодическую интеграцию. Субтитровая нормализация создаёт аналитическую ошибку через нивелирование кода, потерю дискурсивов, искажение имён, чисел, стирание просодических границ. Корпусная процедура сохраняет первичность звучащей реплики, текстовый канал кодируется как отдельный слой её репрезентации.

Список литературы:

1. Мокшин С. И. Дискурс противостояния в поликодовых текстах современных медиа // Ученые записки Новгородского государственного университета. 2024. № 1 (52). С. 281–296.

2. Auer P. The pragmatics of code-switching: a sequential approach // One Speaker, Two Languages: Cross-Disciplinary Perspectives on Code-Switching / ed. by L. Milroy, P. Muysken. Cambridge: Cambridge University Press, 1995. P. 115–135. DOI: 10.1017/CBO9780511620867.006.
3. Gambier Y. Audiovisual Translation and Multimodality: What Future? // Media and Intercultural Communication: A Multidisciplinary Journal. 2023. Vol. 1, No. 1. P. 1–16. DOI 10.22034/mic.2023.167451.
4. Gumperz J. J. Discourse Strategies. Cambridge: Cambridge University Press, 1982. 225 p.
5. Jain P., Bhowmick A. VITB-HEBiC: A Bilingual Corpus for Evaluating ASR in Diverse Indian Code-Switching Scenarios // Applied Acoustics. 2024. Vol. 224. Art. 110119. DOI 10.1016/j.apacoust.2024.110119.
6. Moreno García L. D. Subtitling Hong Kong Code-Mixing and Code-Switching: The Case of Netflix English and Spanish Official Subtitles for Hongkonese Audiovisual Creations // Exchanges. 2024. Vol. 11, No. 2. P. 161–187. DOI 10.31273/eirj.v11i2.1155.
7. Myers-Scotton C. Duelling Languages: Grammatical Structure in Codeswitching. Oxford: Clarendon Press, 1993. 285 p. DOI 10.1093/oso/9780198240594.001.0001.
8. Poplack S. Sometimes I'll start a sentence in Spanish y termino en español: toward a typology of code-switching // Linguistics. 1980. Vol. 18, No. 7–8. P. 581–618. DOI 10.1515/ling.1980.18.7-8.581.
9. Singh S., Singh M., Kadyan V. HiACC: Hinglish Adult & Children Code-Switched Corpus // Data in Brief. 2025. Vol. 62. Art. 111886. P. 1–16. DOI 10.1016/j.dib.2025.111886.
10. Thomas S. Code-Switching in Spoken Indian English: A Case Study of Sociopolitical Talk // Journal of Contemporary Philology. 2021. Vol. 4, No. 1. P. 7–40.

References:

1. Mokshin S. I. The Discourse of Confrontation in Polycode Texts of Modern Media // Scholarly Notes of Novgorod State University. 2024. No. 1 (52). pp. 281–296.
2. Auer P. The pragmatics of code-switching: a sequential approach // One Speaker, Two Languages: Cross-Disciplinary Perspectives on Code-Switching / ed. by L. Milroy, P. Muysken. Cambridge: Cambridge University Press, 1995. pp. 115–135. DOI: 10.1017/CBO9780511620867.006.
3. Gambier Y. Audiovisual Translation and Multimodality: What Future? // Media and Intercultural Communication: A Multidisciplinary Journal. 2023. Vol. 1, No. 1. pp. 1–16. DOI 10.22034/mic.2023.167451.
4. Gumperz J. J. Discourse Strategies. Cambridge: Cambridge University Press, 1982. 225 p.
5. Jain P., Bhowmick A. VITB-HEBiC: A Bilingual Corpus for Evaluating ASR in Diverse Indian Code-Switching Scenarios // Applied Acoustics. 2024. Vol. 224. Art. 110119. DOI 10.1016/j.apacoust.2024.110119.
6. Moreno García L. D. Subtitling Hong Kong Code-Mixing and Code-Switching: The Case of Netflix English and Spanish Official Subtitles for Hongkonese Audiovisual Creations // Exchanges. 2024. Vol. 11, No. 2. pp. 161–187. DOI 10.31273/eirj.v11i2.1155.
7. Myers-Scotton C. Duelling Languages: Grammatical Structure in Codeswitching. Oxford: Clarendon Press, 1993. 285 p. DOI 10.1093/oso/9780198240594.001.0001.

8. Poplack S. Sometimes I'll start a sentence in Spanish y termino en español: towards a typology of code-switching // Linguistics. 1980. Vol. 18, No. 7–8. P. 581–618. DOI 10.1515/ling.1980.18.7-8.581.
9. Singh S., Singh M., Kadyan V. HiACC: Hinglish Adult & Children Code-Switched Corpus // Data in Brief. 2025. Vol. 62. Art. 111886. P. 1–16. DOI 10.1016/j.dib.2025.111886.
10. Thomas S. Code-Switching in Spoken Indian English: A Case Study of Sociopolitical Talk // Journal of Contemporary Philology. 2021. Vol. 4, No. 1. P. 7–40.