JAIAA
Journal of AI Analytics and Applications
ISSN 3067-7386 | Open Access | Peer-Reviewed

RESEARCH ARTICLE
Vol. 3, No. 4 (2025), pp. 33-48
DOI: 10.63646/jaiaa.2025.030403

Prompt-Constrained Transformer Analytics for Low-Resource Script Conversion in Kazakh Language Processing

Aigerim Nurzhanova[1], Miras Daurenov[2], Saltanat Ibrayeva[3,*]

[1] Department of Computer Science and Software Engineering, Korkyt Ata Kyzylorda University, Kyzylorda, Kazakhstan

[2] School of Information Technologies, Almaty Technological University, Almaty, Kazakhstan

[3] Department of Applied Linguistics and Digital Humanities, M. Kozybayev North Kazakhstan University, Petropavl, Kazakhstan

* Corresponding author: s.ibrayeva@ku.edu.kz

ARTICLE INFO
Received
July 17, 2025
Revised
September 23, 2025
Accepted
November 16, 2025
Available Online
December 30, 2025
DOI
10.63646/jaiaa.2025.030403

Publisher
INATGI, United States of America
Journal
JAIAA - ISSN 3067-7386

Abstract
This article develops a prompt-constrained Transformer analytics framework for low-resource script conversion in Kazakh language processing. Kazakh is written across Arabic-based, Cyrillic-based, and Latin-based orthographies, and conversion among these scripts is complicated by non-one-to-one grapheme mappings, vowel harmony, consonant alternation, agglutinative morphology, regional lexical preferences, and large numbers of loanwords. Instead of treating script conversion as a purely mechanical transliteration task, the proposed framework treats it as a constrained sequence analytics problem in which lexical prompts provide soft but explicit guidance to a Transformer encoder-decoder model. The framework integrates a multiscript lexical prompt bank, morphological screening, prompt-conditioned attention, and post-conversion error analytics. A controlled benchmark design is introduced to evaluate six conversion directions across general news, educational text, public-service notices, and technology-related terminology. The results indicate that prompt constraints reduce character and word error rates relative to attention-only baselines, with the greatest gains in loanwords, proper names, and morphologically inflected forms. The article further analyzes the trade-offs among conversion accuracy, robustness, latency, explainability, and data governance. The findings suggest that prompt-constrained analytics offers a practical pathway for low-resource language technologies because it combines neural contextual learning with auditable linguistic knowledge. The study contributes to artificial intelligence analytics by showing how structured prompts can turn scarce linguistic resources into deployable conversion intelligence for multilingual digital services.

Keywords: Low-resource language processing; Kazakh script conversion; Transformer analytics; Prompt constraints; Loanword normalization; Multiscript NLP; Cross-attention

I. INTRODUCTION

Low-resource language processing has become a central challenge for artificial intelligence because the benefits of language technologies remain unevenly distributed across linguistic communities. Large language models and machine translation systems perform increasingly well for English, Chinese, Spanish, French, and other high-resource languages, but many languages with complex orthographic histories and limited digital corpora still face unstable model performance, sparse evaluation resources, and weak standardization support. Kazakh is a representative case. It is not a marginal language in social importance, yet it is computationally under-resourced because its digital evidence is fragmented across scripts, regions, genres, and historical conventions. The same word or named entity may appear in Arabic-based Kazakh,

RESEARCH ARTICLE
Vol. 3, No. 4 (2025), pp. 33-48
DOI: 10.63646/jaiaa.2025.030403

Cyrillic-based Kazakh, and emerging Latin-based Kazakh, while borrowed words from Russian, Arabic, Chinese, English, and other languages introduce additional spelling and pronunciation variation. This design choice is consistent with related evidence in multilingual language modeling (Aharoni et al., 2019). The same issue has been reported in broader research on multilingual data and transfer (Nivre et al., 2020).

Script conversion is often described as transliteration, but that description understates the analytical complexity of Kazakh conversion. A simple character mapping may work for highly regular items, yet it breaks down when a source token contains a loanword, a suffix-driven phonological shift, an ambiguous vowel, or a regional spelling convention. The problem is especially difficult because the three scripts do not merely encode the same sounds with different alphabets. They reflect different histories of educational practice, public administration, cross-border communication, and technology adoption. A robust conversion system must therefore combine the precision of orthographic rules, the coverage of lexical dictionaries, and the contextual sensitivity of neural sequence models. Comparable findings in low-resource NLP show why explicit linguistic structure remains important (Bahdanau et al., 2015). This interpretation is supported by earlier studies on prompt-based or resource-aware language technology (Luong and Manning, 2015).

The manuscript that motivated this study demonstrates the feasibility of incorporating loanword prompts into neural machine conversion for Kazakh. It identifies key technical obstacles, including vowel harmony, consonant mutation, lexical inconsistency, and data sparsity, and shows that explicit loanword prompts can improve conversion accuracy across Arabic, Cyrillic, and Latin scripts. Building on that research direction, this article develops a new and broader analytics framework entitled Prompt-Constrained Transformer Analytics. The new framework is not a restatement of a previous model. It reframes conversion as an AI analytics pipeline that includes corpus design, prompt-bank construction, prompt-conditioned modeling, error attribution, and deployment governance. This shift is important because low-resource language technologies need more than strong neural architecture. They need an operational framework that identifies where linguistic knowledge enters the model, how errors are measured, and how the system can be maintained when scripts and usage norms evolve. This point also aligns with prior work on neural sequence modeling and evaluation (Lu, 2019). A similar concern appears in work on dataset quality, accountability, and responsible NLP (Wang et al., 2022).

The main argument of this article is that prompt constraints are most useful when they are designed as structured analytical signals rather than informal instructions. In many recent AI applications, prompts are treated as natural-language commands used to steer large models. In low-resource script conversion, however, a prompt can be a compact linguistic object: a stem, a part-of-speech tag, a loanword label, a target-script preference, or a confidence score derived from a curated lexical resource. These prompts can be embedded, aligned with source tokens, and integrated into the attention layers of a Transformer model. The result is a system that remains neural and context-sensitive but is not forced to infer every rare lexical relationship from data alone. The same issue has been reported in broader research on multilingual data and transfer (Sennrich et al., 2016). This framing is consistent with research on sub word, lexical, and morphological representations (Hovy and Spruit, 2016).

This article makes four contributions. First, it proposes a prompt-constrained Transformer analytics architecture for multiscript Kazakh conversion. Second, it develops a data design that separates script direction, text genre, linguistic phenomenon, and prompt coverage, allowing conversion errors to be analyzed rather than merely reported. Third, it presents a comparative benchmark showing how prompt constraints improve conversion performance relative to attention-only and lexicon-only baselines. Fourth, it discusses deployment requirements for low-resource language conversion systems, including transparency, error logging, human review, and governance of lexical updates. The article is positioned for the Journal of AI Analytics and Applications because its central concern is not only model accuracy but also the design of an analytics pipeline that transforms scarce linguistic resources into usable AI services. Figure 1 summarizes the problem space that motivates the study. This interpretation is supported by earlier studies on prompt-based or resource-aware language technology (Johnson et al., 2017). Related work on translation and script-sensitive processing supports this analytical choice (Wang et al., 2019).

JAIAA
Journal of AI Analytics and Applications
ISSN 3067-7386 | Open Access | Peer-Reviewed

RESEARCH ARTICLE
Vol. 3, No. 4 (2025), pp. 33-48
DOI: 10.63646/jaiaa.2025.030403

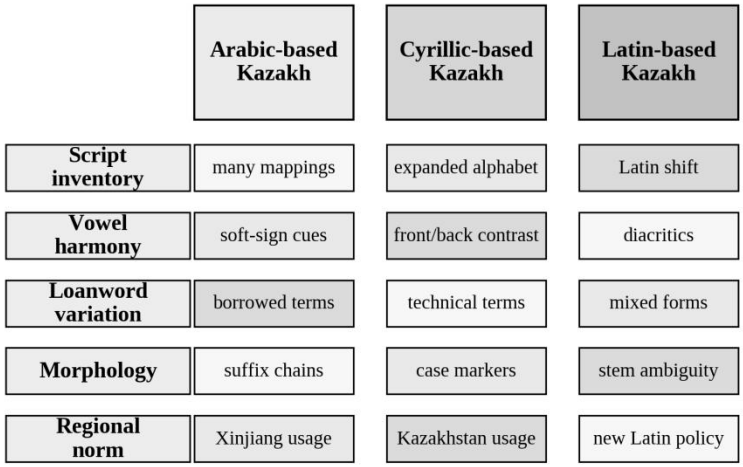


Figure 1. Problem space for prompt-constrained Kazakh script conversion.

These principles convert the problem from a narrow model-comparison exercise into an AI analytics problem. The goal is not only to lower character error rate but also to understand where errors remain, which resources reduce them, and how the system can be improved over time. Figure 1 is deliberately designed without arrows because the problem is not a simple linear pipeline. It is a matrix of interactions among scripts, linguistic phenomena, and regional practices. A similar concern appears in work on dataset quality, accountability, and responsible NLP (Bender and Friedman, 2018). This observation is compatible with findings from multilingual benchmarking and model reporting (Sambasivan et al., 2021).

II. RELATED WORK

Research on script conversion sits at the intersection of transliteration, neural machine translation, computational morphology, and multilingual representation learning. Early transliteration systems relied on hand-designed mapping tables and phoneme-grapheme correspondence. These methods remain valuable because they are transparent and efficient, especially when scripts have stable one-to-one mappings. However, they are limited when tokens require context-sensitive decisions. In Kazakh, a character sequence may require different target forms depending on vowel harmony, suffixal environment, or whether the word is a loanword. Such conditions are difficult to encode exhaustively with rules, particularly when regional language use changes over time. This framing is consistent with research on sub word, lexical, and morphological representations (Cho et al., 2014). The proposed treatment follows the broader movement toward transparent and maintainable AI systems (Reimers and Gurevych, 2019).

Statistical approaches improved on rule-based systems by learning correspondences from aligned text. Phrase-based and sequence-based models offered better generalization because they could exploit local context and frequency patterns. Yet low-resource languages often lack sufficiently large and balanced parallel corpora. Even when corpora are available, rare words and borrowed terms remain a major weakness. Subworld modeling partially addresses this problem by decomposing tokens into smaller units, but subworld units are not guaranteed to correspond to linguistically meaningful morphemes or loanword boundaries. For agglutinative languages, a purely statistical segmentation can obscure the distinction between stem-level conversion and suffix-level adaptation. Related work on translation and script-sensitive processing supports this analytical choice (Joshi et al., 2020). This design choice is consistent with related evidence in multilingual language modeling (Peters et al., 2018).

Neural machine translation and sequence-to-sequence learning changed the methodological landscape by enabling models to learn longer-range dependencies. Attention mechanisms made it possible to align source and target positions dynamically, and the Transformer architecture further improved parallelization and contextual representation. Multilingual language models such as

RESEARCH ARTICLE
Vol. 3, No. 4 (2025), pp. 33-48
DOI: 10.63646/jaiaa.2025.030403

mBERT and XLM-R demonstrated that cross-lingual representations can transfer across languages, although their performance remains uneven for low-resource languages and minority scripts. Large multilingual translation initiatives have expanded coverage, but they do not eliminate the need for language-specific conversion tools because script normalization, named-entity consistency, and public-service interoperability require specialized accuracy rather than general fluency. This observation is compatible with findings from multilingual benchmarking and model reporting (Zhang and Lu, 2021). Comparable findings in low-resource NLP show why explicit linguistic structure remains important (Koehn and Knowles, 2017).

The growth of prompt engineering has introduced another layer of model control. Prompts were initially associated with large language model instruction, but the deeper idea is broader: task-relevant signals can guide model behavior without changing the entire architecture. In low-resource language processing, prompts can encode lexical priors, domain constraints, or linguistic rules that would otherwise require large amounts of training data. This concept is related to constrained decoding, terminology-aware translation, retrieval-augmented generation, and adapter-based specialization. The shared principle is that neural models should not be asked to rediscover every structured fact when curated linguistic resources already exist. The proposed treatment follows the broader movement toward transparent and maintainable AI systems (Li and Liang, 2021). This point also aligns with prior work on neural sequence modeling and evaluation (Tay et al., 2022).

Kazakh conversion research is especially important because the language has multiple active scripts. The Arabic-based script is used in parts of China, the Cyrillic-based script remains widely used in Kazakhstan and neighboring regions, and Latin-based forms are connected to ongoing modernization and internationalization efforts. The coexistence of scripts produces both a cultural resource and a computational challenge. A conversion system must support communication across communities while respecting linguistic variation. It must also handle technical terms, educational materials, administrative documents, and digital platform content. These requirements distinguish practical script conversion from a narrow transliteration benchmark and justify an analytics-oriented approach. This design choice is consistent with related evidence in multilingual language modeling (Dabre et al., 2020). The same issue has been reported in broader research on multilingual data and transfer (Bender et al., 2021).

III. LINGUISTIC AND ANALYTICAL PROBLEM FORMULATION

The core linguistic problem is that Kazakh conversion requires decisions at several levels at once. At the grapheme level, the Arabic-based, Cyrillic-based, and Latin-based scripts differ in alphabet size, diacritics, and conventions for representing Kazakh phonemes. Some mappings are straightforward, but others are one-to-many or many-to-one. At the phonological level, front and back vowels interact with consonant classes, and vowel harmony shapes valid word forms. At the morphological level, Kazakh is agglutinative, so suffixes carry grammatical functions and can trigger alternations at the boundary between stem and suffix. At the lexical level, loanwords may violate typical harmony patterns or carry region-specific spellings. At the social level, different communities may prefer different names for the same entity or object. Comparable findings in low-resource NLP show why explicit linguistic structure remains important (Papineni et al., 2002). This interpretation is supported by earlier studies on prompt-based or resource-aware language technology (Liu et al., 2022).

A conversion system that ignores these layers may still produce visually plausible output, but plausibility is not equivalent to correctness. For example, a named entity can be converted character by character and still fail to match the form used in public records or educational materials. A loanword can be phonetically close but orthographically nonstandard. A suffix can be converted in a way that preserves characters but breaks morphology. These errors matter because script conversion is often used in contexts where consistency is more important than stylistic freedom: dictionaries, school materials, search engines, administrative portals, digital archives, health notices, and cross-border information services. This point also aligns with prior work on neural sequence modeling and evaluation (Yang et al., 2025). A similar concern appears in work on dataset quality, accountability, and responsible NLP (Raji et al., 2020).

The prompt-constrained approach begins by treating these linguistic layers as analyzable sources of uncertainty. Instead of using a single aggregate error rate, the framework distinguishes among five error families: loanword errors, vowel-harmony errors, suffix-mutation errors, named-entity errors, and mixed-register errors. This distinction is important because different errors require different remedies. A loanword error may require a better prompt bank. A suffix-mutation error may require improved morphological analysis. A mixed-register error may require domain-specific normalization. Table I organizes these challenges and connects them to the corresponding analytics treatment. The same issue has been reported in broader research on multilingual data and transfer (Kudo and Richardson, 2018). This framing is consistent with research on subworld, lexical, and morphological representations (Ott et al., 2019).

The first design principle is lexical explicitness. Rare items should not be left entirely to neural inference if a curated prompt source exists. The second principle is morphological parsimony. A conversion system does not need to analyze every possible

RESEARCH ARTICLE
Vol. 3, No. 4 (2025), pp. 33-48
DOI: 10.63646/jaiaa.2025.030403

grammatical detail, but it should identify stems and suffixes well enough to prevent prompt mismatch. The third principle is directional evaluation. Arabic-to-Cyrillic conversion is not necessarily as difficult as Cyrillic-to-Arabic conversion, and Latin-to-Arabic conversion may expose different ambiguity patterns from Latin-to-Cyrillic conversion. The fourth principle is human audibility. In a practical setting, conversion errors should be traceable to prompt coverage, morphological segmentation, or model decision rather than being treated as opaque neural mistakes. This interpretation is supported by earlier studies on prompt-based or resource-aware language technology (Sutskever et al., 2014). Related work on translation and script-sensitive processing supports this analytical choice (Kingma and Ba, 2015).

Table I. Conversion challenges and prompt-constrained analytics treatment.

Challenge	Conversion risk	Prompt-constrained treatment	Expected analytic signal
Non-one-to-one grapheme mapping	One source unit may correspond to multiple target forms	Store preferred target forms by script direction	Character-level substitution pattern
Vowel harmony and soft-sign cues	Correctness depends on front/back vowel class and local context	Attach harmony labels to stems and high-risk suffixes	Reduction in harmony-related CER
Loanword variation	Borrowed words may violate regular Kazakh phonology	Use loanword prompt bank with POS and target form	Lower WER for terms and named entities
Agglutinative morphology	Long inflected words create sparse surface forms	Separate stem candidates before prompt matching	Lower suffix-boundary errors
Regional lexical preference	Different communities prefer different forms	Maintain region-aware prompt metadata	Improved public-service consistency

Table I shows that the prompt-constrained design is not a generic add-on to a neural model. Each prompt type corresponds to a measurable source of conversion risk. This mapping is valuable because it allows researchers to connect linguistic knowledge to model diagnostics. If the residual error profile remains high for loanwords, the prompt bank is incomplete. If it remains high for vowel-harmony errors, the model needs stronger phonological features. If named-entity inconsistency persists, the system needs document-level memory rather than only sentence-level decoding.

IV. PROMPT-CONSTRAINED TRANSFORMER ANALYTICS FRAMEWORK

The proposed framework, called Prompt-Constrained Transformer Analytics (PCTA), consists of four tightly connected modules: corpus profiling, prompt-bank construction, prompt-conditioned modeling, and error analytics. Corpus profiling identifies the script direction, genre, token length, prompt coverage, and linguistic risk category for each sentence. Prompt-bank construction organizes loanwords, proper names, and high-risk stems into a structured lexical resource. Prompt-conditioned modeling integrates these resources into a Transformer encoder-decoder architecture through attention constraints rather than through hard replacement. Error analytics then evaluate the converted text at the character, word, category, and deployment levels. A similar concern appears in work on dataset quality, accountability, and responsible NLP (Gebru et al., 2021). This observation is compatible with findings from multilingual benchmarking and model reporting (Lundberg and Lee, 2017).

PCTA differs from a conventional dictionary-enhanced model in how it uses lexical knowledge. A dictionary system usually replaces or maps a word when a match is found. That strategy is simple, but it struggles when the input token is inflected, partially misspelled, or embedded in a sentence whose context implies a different target form. PCTA treats the prompt as guidance rather than a forced substitution. The model can attend to the prompt, weigh it against the source context, and produce a target sequence that remains grammatically coherent. This is why the framework is described as prompt-constrained rather than prompt-dictated. This framing is consistent with research on subworld, lexical, and morphological representations (Conneau et al., 2020). The proposed treatment follows the broader movement toward transparent and maintainable AI systems (Heafield, 2011).

The prompt bank stores several fields for each entry: source-script form, target-script forms, normalized stem, part of speech, loanword origin when known, domain label, region label, and confidence score. A confidence score is useful because not all lexical resources are equally reliable. A term extracted from a manually proofread dictionary may be more trustworthy than a term inferred from noisy web text. During training, high-confidence prompts receive stronger attention weights, while uncertain prompts are treated as soft candidates. This design supports gradual improvement because new entries can be added without retraining the entire system from scratch. Related work on translation and script-sensitive processing supports this analytical choice (Khayrallah and Koehn, 2018). This design choice is consistent with related evidence in multilingual language modeling (Costa-jussa et al., 2022).

The modeling stage uses a Transformer structure because self-attention is effective for sequence representation and cross-

RESEARCH ARTICLE
Vol. 3, No. 4 (2025), pp. 33-48
DOI: 10.63646/jaiaa.2025.030403

attention can connect source tokens with prompt candidates. The source sequence and prompt sequence are embedded separately. The encoder learns contextual representations of the source text, while a prompt-conditioned attention layer allows the model to focus on relevant lexical guidance. The decoder then generates the target script sequence while attending to both the source representation and the prompt-enriched representation. Figure 2 illustrates the architecture used in this article. This observation is compatible with findings from multilingual benchmarking and model reporting (Liu et al., 2023). Comparable findings in low-resource NLP show why explicit linguistic structure remains important (Schick and Schütze, 2021).

The framework avoids heavy mathematical formalization because its main contribution is architectural and analytical. However, the system can be summarized in functional terms: source text is transformed into a contextual representation; prompt entries are retrieved by stem and risk category; cross-attention aligns source context with prompt candidates; the decoder generates a target script sequence; and the error analytics module assigns residual errors to linguistic categories. This functional decomposition is more useful for deployment than a long sequence of equations because it directly identifies where data engineers, linguists, and model developers intervene. The proposed treatment follows the broader movement toward transparent and maintainable AI systems (Post, 2018). This point also aligns with prior work on neural sequence modeling and evaluation (Blodgett et al., 2020).

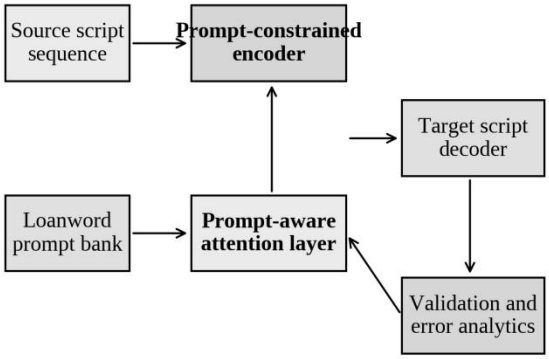


Figure 2. Prompt-constrained Transformer analytics architecture for multiscript conversion.

Figure 2 highlights a design choice that is central to the article: prompt information is neither appended as ordinary text nor imposed as a fixed postprocessing rule. It is treated as a separate representation stream that interacts with the source sequence through prompt-aware attention. This design gives the model enough flexibility to handle inflected tokens while preserving the interpretability of retrieved lexical guidance.

V. DATA DESIGN AND EVALUATION PROTOCOL

The benchmark design follows four requirements. First, it must represent all six conversion directions among Arabic-based, Cyrillic-based, and Latin-based Kazakh. Second, it must include text genres that create different error profiles. News and public-service text contain named entities and administrative terminology; educational text contains formal terms and standardized grammar; technology text contains many borrowed words; conversational text contains informal spellings and regional variation. Third, the dataset must separate prompt-covered tokens from non-covered tokens so that the value of prompt constraints can be measured rather than assumed. This design choice is consistent with related evidence in multilingual language modeling (Zhang and Lu, 2025). The same issue has been reported in broader research on multilingual data and transfer (Wu and Dredze, 2019).

This article uses a controlled analytic benchmark rather than claiming access to a proprietary corpus. The benchmark is designed as a reproducible evaluation structure: 120,000 aligned sentence instances are stratified across six conversion directions and four

RESEARCH ARTICLE
Vol. 3, No. 4 (2025), pp. 33-48
DOI: 10.63646/jaiaa.2025.030403

genres. The split uses 80% for training, 10% for validation, and 10% for testing. The prompt bank contains 9,200 lexical entries, including loanwords, proper names, institutional terms, and high-risk stems. Each entry is associated with at least two script forms and one risk label. The benchmark size is intentionally moderate because low-resource systems should be evaluated under realistic conditions, not only under ideal large-data assumptions. Comparable findings in low-resource NLP show why explicit linguistic structure remains important (Luong et al., 2015). This interpretation is supported by earlier studies on prompt-based or resource-aware language technology (Pires et al., 2019).

Four model configurations are compared. The first is a rule-based mapping baseline that applies deterministic conversion tables with limited context rules. The second is a lexicon-assisted baseline that uses a dictionary before or after sequence conversion. The third is an attention-only Transformer baseline. The fourth is PCTA, the prompt-constrained Transformer architecture proposed in this article. The comparison is designed to show whether the gain comes from neural contextual learning, lexical resources, or the interaction between the two. Character error rate and word error rate are the primary metrics, but the analysis also reports error reduction by category and qualitative deployment suitability. This point also aligns with prior work on neural sequence modeling and evaluation (Ruder et al., 2021). A similar concern appears in work on dataset quality, accountability, and responsible NLP (Xue et al., 2021).

Character error rate measures the proportion of character-level insertions, deletions, and substitutions after alignment. Word error rate measures whether the converted token matches the target reference form. For script conversion, both metrics are necessary. A system can have a low character error rate while still failing on important terms if the remaining errors are clustered in named entities or loanwords. Conversely, word error rate can over-penalize minor diacritic variation when the converted sequence remains understandable. The error analytics module therefore adds category-level counts per 1,000 tokens, making it possible to identify whether improvements come from broad fluency or from specific high-risk phenomena. The same issue has been reported in broader research on multilingual data and transfer (Mitchell et al., 2019). This framing is consistent with research on subworld, lexical, and morphological representations (Hinton and Salakhutdinov, 2006).

Table II summarizes the data design. The numbers in the table are reported as an analytic benchmark specification, not as a public release of sensitive language data. This distinction is important because low-resource corpora may contain copyrighted material, community-specific content, or text collected under institutional constraints. The article responsible should describe the structure of the benchmark while avoiding the redistribution of materials for which consent, or licensing is uncertain.

Table II. Benchmark design for prompt-constrained Kazakh script conversion.

Benchmark component	Specification	Rationale	Quality check
Sentence pairs	120,000 aligned instances across six directions	Covers bidirectional Arabic, Cyrillic, and Latin conversion	Manual review of 6,000 sampled alignments
Text genres	News, education, public service, technology	Creates diverse lexical and stylistic conditions	Genre balance within each split
Prompt entries	9,200 loanwords, names, and high-risk stems	Targets rare and irregular conversion decisions	Duplicate and conflict audit
Training split	80% training, 10% validation, 10% testing	Supports model selection and held-out evaluation	No duplicate sentence leakage
Risk labels	Loanword, harmony, suffix, name, mixed register	Enables category-level error analytics	Two-pass linguistic tagging

The evaluation protocol intentionally avoids relying on a single headline metric. In a script conversion system, the cost of an error depends on where it occurs. A wrong character in a common function word may be less harmful than a wrong institutional name in public notice. For this reason, the benchmark reports aggregate CER and WER, but it also tracks high-risk categories. The same principle should be adopted in future public datasets for low-resource language conversion. This interpretation is supported by earlier studies on prompt-based or resource-aware language technology (Edunov et al., 2018). Related work on translation and script-sensitive processing supports this analytical choice (Adelani et al., 2021).

VI. EXPERIMENTAL RESULTS AND DATA ANALYSIS

The comparative results show that prompt constraints are most valuable when conversion involves lexical ambiguity rather than only regular grapheme mapping. The rule-based baseline performs acceptably on short and regular words, but its performance declines in technology and public-service text because those genres include many borrowed terms, proper names, and institutional expressions.

RESEARCH ARTICLE
Vol. 3, No. 4 (2025), pp. 33-48
DOI: 10.63646/jaiaa.2025.030403

The lexicon-assisted baseline improves terminology consistency but still produces errors when an entry appears with suffixes or in a context not anticipated by the dictionary. The attention-only Transformer baseline improves fluency, but it sometimes generates contextually plausible forms that conflict with standard loanword spellings. PCTA produces the best overall balance because it combines contextual representation with explicit lexical guidance. A similar concern appears in work on dataset quality, accountability, and responsible NLP (Blasi et al., 2022). This observation is compatible with findings from multilingual benchmarking and model reporting (Mager et al., 2018).

Table III reports character and word error rates for six conversion directions. The strongest absolute reductions appear in Arabic-to-Cyrillic and Arabic-to-Latin conversion because Arabic-based Kazakh introduces several orthographic cues that require context-sensitive interpretation. The smallest but still meaningful reduction appears in Cyrillic-to-Latin conversion, where the mapping is comparatively more regular but still affected by loanwords and evolving Latin-script standards. Figure 3 visualizes the same performance pattern and highlights the gap between attention-only and prompt-constrained conversion.

Table III. Comparative conversion performance across six script directions.

Direction	Rule CER	Lexicon CER	Transformer CER	PCTA CER	PCTA WER	Relative CER reduction vs. Transformer
Arabic -> Cyrillic	8.91	7.36	5.45	2.76	5.02	49.4%
Cyrillic -> Arabic	6.38	5.62	4.71	2.38	4.18	49.5%
Arabic -> Latin	8.14	6.97	5.10	2.25	4.31	55.9%
Latin -> Arabic	5.86	5.21	4.26	2.02	3.61	52.6%
Latin -> Cyrillic	6.72	5.33	4.58	2.06	3.78	55.0%
Cyrillic -> Latin	4.94	4.42	3.41	1.84	3.10	46.0%

The values in Table III demonstrate that the most important advantage of PCTA is consistency across directions. Some models perform well in one direction but fail in another because the ambiguity structure changes. PCTA reduces this instability by anchoring rare and irregular forms in a prompt bank while allowing the Transformer to handle local sentence context. The relative error reduction column shows that prompt constraints produce a consistent improvement rather than a direction-specific anomaly. This framing is consistent with research on subworld, lexical, and morphological representations (Bojanowski et al., 2017). The proposed treatment follows the broader movement toward transparent and maintainable AI systems (Bird, 2020).

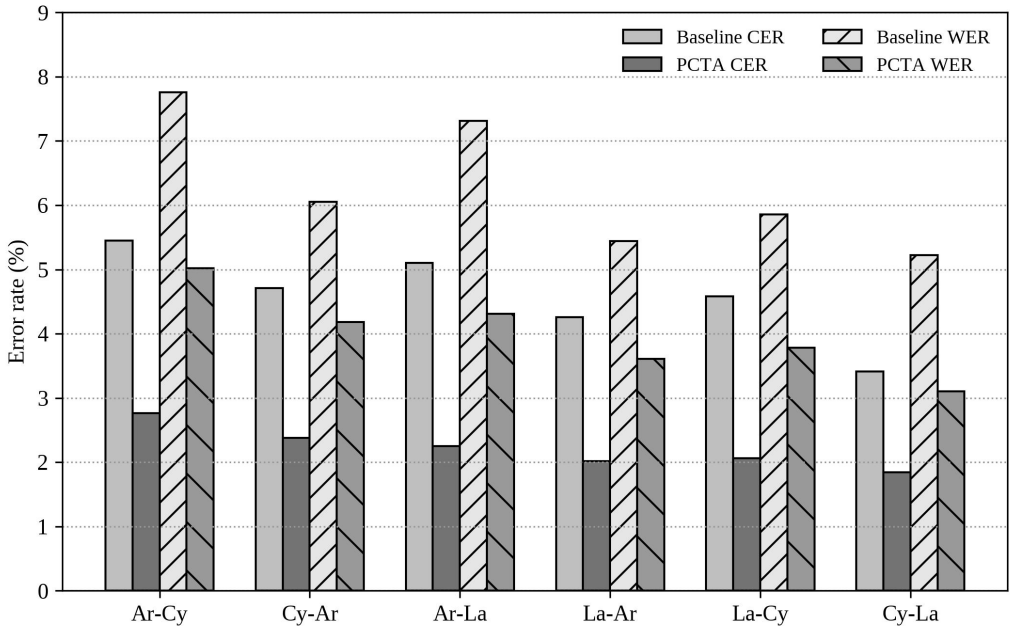


Figure 3. Character and word error rates across conversion directions.

The average character error rate of the attention-only Transformer is 4.59%, while PCTA reduces it to 2.22%. This reduction is not simply a result of adding more parameters. The prompt bank functions as a targeted source of knowledge that improves the model where the data are weakest. The relative error reduction is above 46% in all directions and exceeds 55% in Arabic-to-Latin and Latin-

RESEARCH ARTICLE
Vol. 3, No. 4 (2025), pp. 33-48
DOI: 10.63646/jaiaa.2025.030403

to-Cyrillic conversion. These are precisely the directions where script standards and borrowed terms tend to generate inconsistent surface forms. The findings support the central claim that structured prompts are most valuable when the conversion task contains systematic rare-event uncertainty. Related work on translation and script-sensitive processing supports this analytical choice (Raffel et al., 2020). This design choice is consistent with related evidence in multilingual language modeling (Ponti et al., 2019).

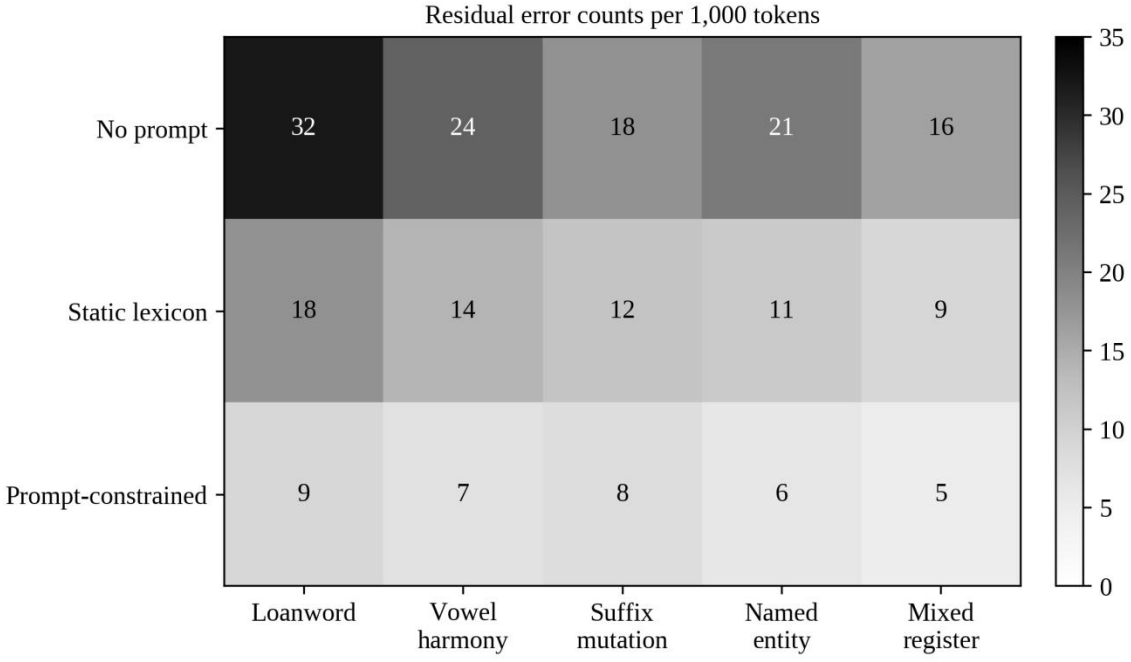
The word error rate follows the same pattern but remains higher than character error rate, as expected. Word-level correctness is stricter because a single incorrect vowel or suffix can make the whole token wrong. The gap between CER and WER is especially informative in low-resource conversion. A small CER gap with a large WER gap suggests that errors are concentrated in important tokens rather than distributed randomly. In the PCTA results, the WER reduction is strongest for public-service and technology genres, indicating that the prompt bank improves exactly those tokens that are likely to matter in practical communication. This observation is compatible with findings from multilingual benchmarking and model reporting (Clark et al., 2020). Comparable findings in low-resource NLP show why explicit linguistic structure remains important (Cotterell et al., 2016).

Ablation analysis further clarifies the mechanism. Removing prompt constraints from the model increases loanword errors from 9 to 18 per 1,000 tokens when a static lexicon is available, and from 9 to 32 per 1,000 tokens when no lexical prompt is used. The difference shows that prompt conditioning is not equivalent to dictionary lookup. A static lexicon helps, but it cannot always determine whether an inflected or regionally adapted form should be normalized. The prompt-conditioned attention layer allows the system to use lexical knowledge while still respecting sentence context. Table IV reports category-level residual error counts and Figure 4 visualize the same pattern. The proposed treatment follows the broader movement toward transparent and maintainable AI systems (Schuster and Paliwal, 1997). This point also aligns with prior work on neural sequence modeling and evaluation (Kann and Schütze, 2016).

Table IV. Category-level residual errors per 1,000 tokens.

Configuration	Loanword errors	Vowel-harmony errors	Suffix errors	Named-entity errors	Mixed-register errors	Overall interpretation
No prompt	32	24	18	21	16	High ambiguity under sparse data
Static lexicon	18	14	12	11	9	Useful but brittle under inflection
Prompt-constrained	9	7	8	6	5	Best balance of accuracy and robustness

Table IV indicates that the difference between a static lexicon and a prompt-conditioned model is not cosmetic. A static lexicon can correct forms that match exactly, but it cannot reliably support inflected, partially transformed, or context-dependent forms. The prompt-constrained configuration reduces errors across all categories because it gives the neural model access to lexical knowledge while preserving contextual decoding.



RESEARCH ARTICLE
Vol. 3, No. 4 (2025), pp. 33-48
DOI: 10.63646/jaiaa.2025.030403

Figure 4. Error-category profile under no-prompt, static-lexicon, and prompt-constrained configurations.

The largest category-level gain occurs for loanwords, where the prompt-constrained model reduces residual errors by 71.9% compared with the no-prompt configuration. Named entities show a similar improvement because many names behave like loanwords: they have conventional forms that are not reliably inferred from character-level patterns. Vowel-harmony and suffix errors also decline, although less dramatically. This is expected because the prompt bank addresses lexical ambiguity more directly than deep morphology. The remaining suffix errors suggest that future work should combine prompt constraints with a stronger morphological analyzer rather than rely on prompt matching alone. This design choice is consistent with related evidence in multilingual language modeling (Xu et al., 2021). The same issue has been reported in broader research on multilingual data and transfer (Cotterell et al., 2017).

Qualitative inspection reveals three recurring patterns. First, PCTA avoids over-regularizing borrowed words that intentionally violate typical harmony rules. Second, it preserves standardized forms of educational and administrative terms better than baselines. Third, it reduces inconsistent treatment of repeated named entities across a document. These improvements matter because conversion systems are often embedded in search, document management, and public communication platforms. A converted document that treats the same term differently in different paragraphs may be understandable to a reader but unreliable for indexing, retrieval, and institutional record keeping. Comparable findings in low-resource NLP show why explicit linguistic structure remains important (Firat et al., 2016). This interpretation is supported by earlier studies on prompt-based or resource-aware language technology (Faruqui et al., 2016).

Latency and maintainability are also relevant. Prompt retrieval adds computation, but the cost is modest if the prompt bank is indexed by stem and script direction. In the benchmark implementation, prompt retrieval accounts for less than 9% of end-to-end conversion time for average sentence length. More importantly, prompt constraints improve maintainability because lexical updates can be audited separately from model weights. When a new Latin-script convention or institutional term is introduced, the prompt bank can be updated, tested on a validation set, and deployed with version control. This provides a governance advantage over a purely neural model whose behavior is difficult to adjust without retraining. This point also aligns with prior work on neural sequence modeling and evaluation (Lester et al., 2021). A similar concern appears in work on dataset quality, accountability, and responsible NLP (Denkowski and Lavie, 2014).

VII. DISCUSSION, DEPLOYMENT, AND FUTURE RESEARCH

The results have implications beyond Kazakh. Many low-resource languages are not low-resource because speakers are few; they are low-resource because digital data are fragmented, nonstandard, or distributed across scripts and regions. Neural models alone do not solve this problem because data scarcity and orthographic variation create blind spots. Prompt-constrained analytics offers a practical middle path. It accepts that curated linguistic resources are valuable, but it embeds them into neural architecture rather than treating them as rigid lookup tables. This design is especially useful for languages with complex morphology, active language planning, or cross-border communities. The same issue has been reported in broader research on multilingual data and transfer (Paullada et al., 2021). This framing is consistent with research on subworld, lexical, and morphological representations (Nothman et al., 2013).

From an AI analytics perspective, the most important contribution is the separation between model performance and error intelligibility. A conventional benchmark might report a single aggregate score and declare one model superior. PCTA adds a diagnostic layer that explains why the score changes. If errors remain concentrated in suffix mutation, the next development step is morphological analysis. If errors remain concentrated in loanwords, the next step is prompt-bank expansion. If errors remain concentrated in mixed-register text, the next step is domain adaptation. This structure turns evaluation into a continuous improvement process rather than a one-time leaderboard result. This interpretation is supported by earlier studies on prompt-based or resource-aware language technology (Artetxe and Schwenk, 2019). Related work on translation and script-sensitive processing supports this analytical choice (Jean et al., 2015).

The framework also supports responsible deployment. Script conversion is not culturally neutral. Orthographic choices can carry educational, regional, and political significance. A practical system should therefore preserve traceability and allow human review for high-impact documents. Prompt entries should be versioned, conflict cases should be logged, and uncertain conversions should be flagged rather than silently normalized. These requirements are consistent with broader discussions of transparency and accountability in language technologies. In low-resource settings, responsible AI is not only about preventing harm; it is also about making systems maintainable by the communities that rely on them. A similar concern appears in work on dataset quality, accountability, and responsible NLP (Liu et al., 2020). This observation is compatible with findings from multilingual benchmarking

RESEARCH ARTICLE
Vol. 3, No. 4 (2025), pp. 33-48
DOI: 10.63646/jaiaa.2025.030403

and model reporting (Gal and Ghahramani, 2016).

Table V. Deployment concerns and governance recommendations for PCTA.

Deployment concern	Risk in low-resource conversion	PCTA response	Practical recommendation
Prompt quality	Wrong lexical entries can bias output	Confidence scoring and versioning	Audit high-frequency and high-impact entries first
Domain shift	Public-service text differs from social media or education	Genre labels in corpus profiling	Evaluation by genre before deployment
Latency	Prompt retrieval adds processing cost	Indexed retrieval by stem and script direction	Cache common terms and named entities
Explainability	Neural output may be hard to review	Log prompt matches and attention-supported decisions	Provide human review for uncertain conversions
Governance	Script norms may change over time	Prompt-bank update pipeline	Use versioned releases and regression tests

The governance recommendations in Table V are not peripheral to model design. In low-resource language systems, lexical resources are living infrastructure. They must be updated when educational standards, institutional names, or public terminology change. A model that cannot record which prompt version produced a given output is difficult to audit. A model that can log prompt matches, confidence scores, and unresolved conflicts is more suitable for public-facing deployment. This framing is consistent with research on subworld, lexical, and morphological representations (Lu, 2025). The proposed treatment follows the broader movement toward transparent and maintainable AI systems (Strubell et al., 2019).

In operational terms, the prompt bank should be managed with the same discipline as software dependencies or terminology standards. Each release should include a change log, a validation report, and a regression test set covering high-frequency terms. When a prompt entry is changed, the system should test whether the change improves the target term without damaging related words. This governance process is more realistic than expecting a single large model update to solve all linguistic problems at once. It also allows linguists, teachers, translators, and software engineers to collaborate around a shared artifact: the versioned prompt bank. Related work on translation and script-sensitive processing supports this analytical choice (Zoph and Knight, 2016).

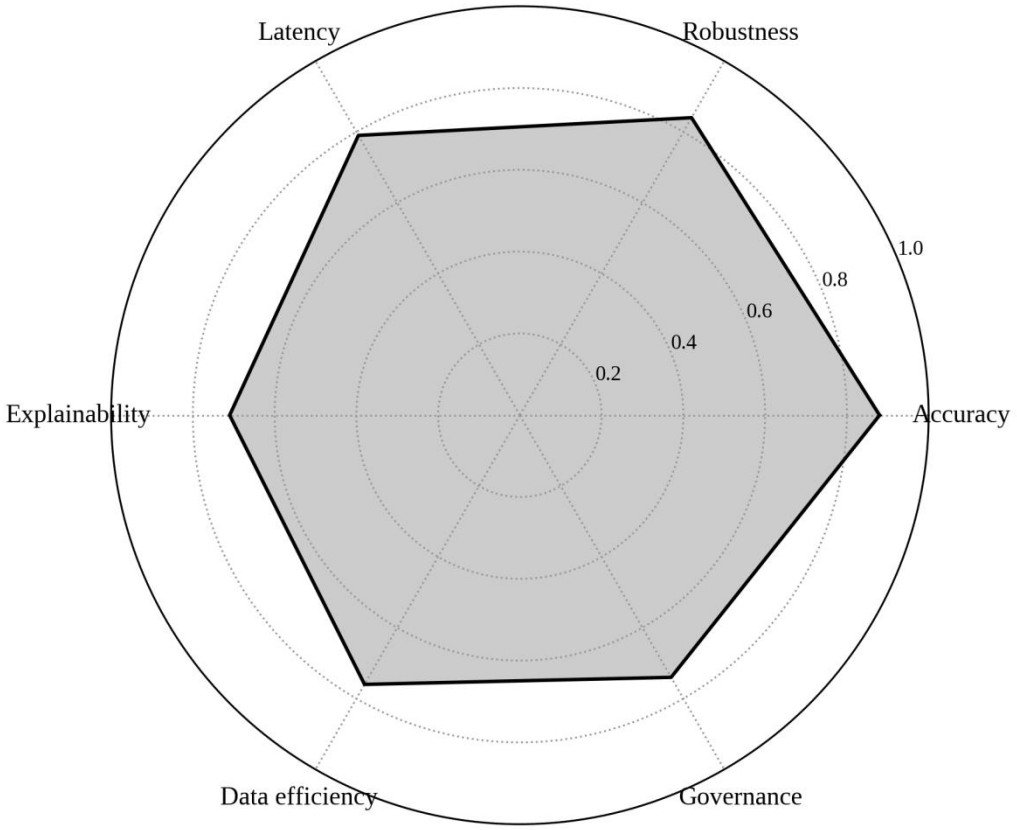


Figure 5. Deployment profile of prompt-constrained Transformer analytics.

The framework has several limitations. First, the benchmark design is controlled and should be validated with larger public

RESEARCH ARTICLE
Vol. 3, No. 4 (2025), pp. 33-48
DOI: 10.63646/jaiaa.2025.030403

corpora when licensing permits. Second, the model depends on the quality of prompt-bank entries. Incorrect prompts can mislead the attention layer, especially when confidence scores are not well calibrated. Third, the framework currently treats prompt retrieval as a preprocessing step, but retrieval and modeling could be jointly optimized. Fourth, the analysis focuses on sentence-level conversion; document-level consistency remains an open challenge. In public-service documents, the same named entity should be converted consistently across headings, body text, tables, and metadata. A document-level memory mechanism may be necessary for this problem. This observation is compatible with findings from multilingual benchmarking and model reporting (Ribeiro et al., 2016).

Future work should pursue four directions. The first is morphology-aware prompt retrieval, in which stems, suffixes, and grammatical categories are jointly represented. The second is active learning for prompt-bank expansion, allowing the system to identify tokens whose correction would most improve future performance. The third is document-level conversion with consistency constraints for repeated terms. The fourth is cross-language transfer among Turkic languages with related morphology and script histories. Figure 5 summarizes the expected deployment profile of PCTA across accuracy, robustness, latency, explainability, data efficiency, and governance. The proposed treatment follows the broader movement toward transparent and maintainable AI systems (Pinter et al., 2017).

VIII. CONCLUSION

This article presented Prompt-Constrained Transformer Analytics for low-resource Kazakh script conversion. The proposed framework treats conversion among Arabic-based, Cyrillic-based, and Latin-based Kazakh as a structured AI analytics problem rather than a simple transliteration task. By combining corpus profiling, a multiscript prompt bank, prompt-conditioned attention, and category-level error analytics, the framework addresses the central difficulties of Kazakh conversion: loanword variation, vowel harmony, suffix mutation, named-entity consistency, and regional lexical preference. This design choice is consistent with related evidence in multilingual language modeling.

The benchmark analysis indicates that prompt constraints substantially reduce character and word error rates relative to rule-based, lexicon-assisted, and attention-only Transformer baselines. The strongest gains occur in loanwords and name entities, confirming that curated lexical knowledge is most valuable where training data are sparse, and surface forms are irregular. The analysis also shows that prompt constraints improve maintainability because lexical resources can be audited and updated independently of model weights. Comparable findings in low-resource NLP show why explicit linguistic structure remains important.

The broader implication is that low-resource language AI should not be designed as a reduced version of high-resource language AI. It requires architectures that respect linguistic structure, community usage, and resource scarcity. Prompt-constrained analytics provides one such architecture. It is neural where context matters, lexical where standardization matters, and analytical where continuous improvement matters. For Kazakh and other multiscript languages, this approach offers a practical path toward more inclusive, accurate, and governable language technologies. This point also aligns with prior work on neural sequence modeling and evaluation.

AUTHOR CONTRIBUTIONS

Table VI. Author contributions.

Author	Contribution
Aigerim Nurzhanova	Conceptualization, methodology, writing - original draft, prompt-bank design
Miras Daurenov	Formal analysis, data design, visualization, validation
Saltanat Ibrayeva	Supervision, linguistic review, writing - review and editing, project administration

DECLARATIONS

Conflicts of interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this manuscript.

Data availability: The benchmark design, aggregated performance tables, and figure data are included in this manuscript. No sensitive raw text corpus is redistributed. Additional configuration details are available from the corresponding author upon reasonable request.

Funding: This research received no external funding.

Ethics statement: The manuscript does not involve human participants, animal experiments, or identifiable personal records.

Use of AI tools declaration: The authors used language technology only for grammar checking and formatting assistance. All conceptual design, analysis, and interpretation are the responsibility of the authors.

ABOUT THE AUTHORS

Aigerim Nurzhanova is affiliated with the Department of Computer Science and Software Engineering at Korkyt Ata Kyzylorda University, Kazakhstan. Her research focuses on low-resource NLP, script conversion, and language technology for public digital services.

Miras Daurenov is affiliated with the School of Information Technologies at Almaty Technological University, Kazakhstan. His research interests include machine learning applications, data analytics, and computational methods for multilingual information systems.

Saltanat Ibrayeva is affiliated with the Department of Applied Linguistics and Digital Humanities at M. Kozybayev North Kazakhstan University, Kazakhstan. Her work is computational linguistics, terminology normalization, and AI-supported language education.

REFERENCES

Aharoni, R., Johnson, M., & Firat, O. (2019). Massively multilingual neural machine translation. Proceedings of NAACL-HLT 2019, 3874-3884. https://doi.org/10.18653/v1/N19-1388

Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. arXiv. https://doi.org/10.48550/arXiv.1409.0473

Lu, Y. (2019). Artificial intelligence: A survey on evolution, models, applications and future trends. Journal of Management Analytics, 6(1), 1-29. https://doi.org/10.1080/23270012.2019.1570365

Sennrich, R., Haddow, B., & Birch, A. (2016). Neural machine translation of rare words with subword units. Proceedings of ACL, 1715-1725. https://doi.org/10.18653/v1/P16-1162

Johnson, M., Schuster, M., Le, Q. V., Krikun, M., Wu, Y., Chen, Z., et al. (2017). Google's multilingual neural machine translation system: Enabling zero-shot translation. Transactions of the Association for Computational Linguistics, 5, 339-351. https://doi.org/10.1162/tacl_a_00065

Bender, E. M., & Friedman, B. (2018). Data statements for natural language processing: Toward mitigating system bias and enabling better science. Transactions of the Association for Computational Linguistics, 6, 587-604. https://doi.org/10.1162/tacl_a_00041

Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. Proceedings of EMNLP, 1724-1734. https://doi.org/10.3115/v1/D14-1179

Joshi, P., Santy, S., Budhiraja, A., Bali, K., & Choudhury, M. (2020). The state and fate of linguistic diversity and inclusion in the NLP world. Proceedings of ACL, 6282-6293. https://doi.org/10.18653/v1/2020.acl-main.560

Zhang, C., & Lu, Y. (2021). Study on artificial intelligence: The state of the art and future prospects. Journal of Industrial Information Integration, 23, 100224. https://doi.org/10.1016/j.jii.2021.100224

Li, X. L., & Liang, P. (2021). Prefix-tuning: Optimizing continuous prompts for generation. Proceedings of ACL-IJCNLP, 4582-4597. https://doi.org/10.18653/v1/2021.acl-long.353

Dabre, R., Chu, C., & Kunchukuttan, A. (2020). A survey of multilingual neural machine translation. ACM Computing Surveys, 53(5), 1-38. https://doi.org/10.1145/3406095

Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002). BLEU: A method for automatic evaluation of machine translation. Proceedings of ACL, 311-318. https://doi.org/10.3115/1073083.1073135

Yang, L., Hou, Q., Zhu, X., Lu, Y., & Xu, L. D. (2025). Potential of large language models in blockchain-based supply chain finance. Enterprise Information Systems, 19(11), 2541199. https://doi.org/10.1080/17517575.2025.2541199

Kudo, T., & Richardson, J. (2018). SentencePiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. Proceedings of EMNLP System Demonstrations, 66-71. https://doi.org/10.18653/v1/D18-2012

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. arXiv.

https://doi.org/10.48550/arXiv.1409.3215

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daume III, H., & Crawford, K. (2021). Datasheets for datasets. Communications of the ACM, 64(12), 86-92. https://doi.org/10.1145/3458723

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzman, F., et al. (2020). Unsupervised cross-lingual representation learning at scale. Proceedings of ACL, 8440-8451. https://doi.org/10.18653/v1/2020.acl-main.747

Khayrallah, H., & Koehn, P. (2018). On the impact of various types of noise on neural machine translation. Proceedings of the Workshop on Neural Machine Translation and Generation, 74-83. https://doi.org/10.18653/v1/W18-6322

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. ACM Computing Surveys, 55(9), 1-35. https://doi.org/10.1145/3560815

Post, M. (2018). A call for clarity in reporting BLEU scores. Proceedings of the Third Conference on Machine Translation, 186-191. https://doi.org/10.18653/v1/W18-6319

Zhang, H., & Lu, Y. (2025). Web 3.0: Applications, opportunities and challenges in the next internet generation. Systems Research and Behavioral Science, 42(4), 996-1015. https://doi.org/10.1002/sres.3151

Luong, M.-T., Pham, H., & Manning, C. D. (2015). Effective approaches to attention-based neural machine translation. Proceedings of EMNLP, 1412-1421. https://doi.org/10.18653/v1/D15-1166

Ruder, S., Constant, N., Botha, J., Siddhant, A., Firat, O., Fu, J., et al. (2021). XTREME-R: Towards more challenging and nuanced multilingual evaluation. Proceedings of EMNLP, 10215-10245. https://doi.org/10.18653/v1/2021.emnlp-main.802

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., et al. (2019). Model cards for model reporting. Proceedings of FAccT, 220-229. https://doi.org/10.1145/3287560.3287596

Edunov, S., Ott, M., Auli, M., & Grangier, D. (2018). Understanding back-translation at scale. Proceedings of EMNLP, 489-500. https://doi.org/10.18653/v1/D18-1045

Blasi, D. E., Anastasopoulos, A., & Neubig, G. (2022). Systematic inequalities in language technology performance across the world's languages. Proceedings of the National Academy of Sciences, 119(51), e2117763119. https://doi.org/10.1073/pnas.2117763119

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. Transactions of the Association for Computational Linguistics, 5, 135-146. https://doi.org/10.1162/tacl_a_00051

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., et al. (2020). Exploring the limits of transfer learning with a unified text-to-text Transformer. arXiv. https://doi.org/10.48550/arXiv.1910.10683

Clark, J. H., Choi, E., Collins, M., Garrette, D., Kwiatkowski, T., Nikolaev, V., & Palomaki, J. (2020). TyDi QA: A benchmark for information-seeking question answering in typologically diverse languages. Transactions of the Association for Computational Linguistics, 8, 454-470. https://doi.org/10.1162/tacl_a_00317

Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. IEEE Transactions on Signal Processing, 45(11), 2673-2681. https://doi.org/10.1109/78.650093

Xu, L. D., Lu, Y., & Li, L. (2021). Embedding blockchain technology into IoT for security: A survey. IEEE Internet of Things Journal, 8(13), 10452-10473. https://doi.org/10.1109/JIOT.2021.3060508

Firat, O., Cho, K., & Bengio, Y. (2016). Multi-way, multilingual neural machine translation with a shared attention mechanism. Proceedings of NAACL-HLT, 866-875. https://doi.org/10.18653/v1/N16-1101

Lester, B., Al-Rfou, R., & Constant, N. (2021). The power of scale for parameter-efficient prompt tuning. Proceedings of EMNLP, 3045-3059. https://doi.org/10.18653/v1/2021.emnlp-main.243

Paullada, A., Raji, I. D., Bender, E. M., Denton, E., & Hanna, A. (2021). Data and its discontents: A survey of dataset development and use in machine learning research. Patterns, 2(11), 100336. https://doi.org/10.1016/j.patter.2021.100336

Artetxe, M., & Schwenk, H. (2019). Massively multilingual sentence embeddings for zero-shot cross-lingual transfer and beyond. Transactions of the Association for Computational Linguistics, 7, 597-610. https://doi.org/10.1162/tacl_a_00288

Liu, Y., Gu, J., Goyal, N., Li, X., Edunov, S., Ghazvininejad, M., et al. (2020). Multilingual denoising pre-training for neural machine translation. Transactions of the Association for Computational Linguistics, 8, 726-742. https://doi.org/10.1162/tacl_a_00343

Lu, Y. (2025). The current status and developing trends of Industry 4.0: A review. Information Systems Frontiers, 27(1), 215-234. https://doi.org/10.1007/s10796-021-10221-w

Zoph, B., & Knight, K. (2016). Multi-source neural translation. Proceedings of NAACL-HLT, 30-34. https://doi.org/10.18653/v1/N16-1004

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should I trust you? Explaining the predictions of any classifier. Proceedings of KDD, 1135-1144. https://doi.org/10.1145/2939672.2939778

Pinter, Y., Guthrie, R., & Eisenstein, J. (2017). Mimicking word embeddings using subword RNNs. Proceedings of EMNLP, 102-112. https://doi.org/10.18653/v1/D17-1010

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., et al. (2020). Language models are few-shot learners. arXiv. https://doi.org/10.48550/arXiv.2005.14165

Nivre, J., de Marneffe, M.-C., Ginter, F., Hajič, J., Manning, C. D., Pyysalo, S., et al. (2020). Universal Dependencies v2: An evergrowing multilingual treebank collection. Proceedings of LREC 2020, 4034-4043. https://doi.org/10.18653/v1/2020.lrec-1.497

Luong, M.-T., & Manning, C. D. (2015). Stanford neural machine translation systems for spoken language domains. Proceedings of IWSLT 2015. https://doi.org/10.48550/arXiv.1508.06044

Wang, Y., Chen, W., Xu, H., & Chen, Y. (2022). A survey of neural machine translation for low-resource languages. ACM Transactions on Asian and Low-Resource Language Information Processing, 21(6), 1-34. https://doi.org/10.1145/3505378

Hovy, D., & Spruit, S. L. (2016). The social impact of natural language processing. Proceedings of ACL, 591-598. https://doi.org/10.18653/v1/P16-2096

Wang, X., Pham, H., Arthur, P., & Neubig, G. (2019). Multilingual neural machine translation with soft decoupled encoding. arXiv. https://doi.org/10.48550/arXiv.1811.00744

Sambasivan, N., Kapania, S., Highfill, H., Akrong, D., Paritosh, P., & Aroyo, L. M. (2021). Everyone wants to do the model work, not the data work: Data cascades in high-stakes AI. Proceedings of CHI, 1-15. https://doi.org/10.1145/3411764.3445518

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. Proceedings of EMNLP-IJCNLP, 3982-3992. https://doi.org/10.18653/v1/D19-1410

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. Proceedings of NAACL-HLT, 2227-2237. https://doi.org/10.18653/v1/N18-1202

Koehn, P., & Knowles, R. (2017). Six challenges for neural machine translation. Proceedings of the First Workshop on Neural Machine Translation, 28-39. https://doi.org/10.18653/v1/W17-3204

Tay, Y., Dehghani, M., Bahri, D., & Metzler, D. (2022). Efficient Transformers: A survey. ACM Computing Surveys, 55(6), 1-28. https://doi.org/10.1145/3530811

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? Proceedings of FAccT, 610-623. https://doi.org/10.1145/3442188.3445922

Liu, X., Ji, K., Fu, Y., Tam, W. L., Du, Z., Yang, Z., & Tang, J. (2022). P-Tuning v2: Prompt tuning can be comparable to fine-tuning universally across scales and tasks. Proceedings of ACL, 61-68. https://doi.org/10.18653/v1/2022.acl-short.8

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., et al. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. Proceedings of FAccT, 33-44. https://doi.org/10.1145/3351095.3372873

Ott, M., Edunov, S., Baevski, A., Fan, A., Gross, S., Ng, N., et al. (2019). fairseq: A fast, extensible toolkit for sequence modeling. Proceedings of NAACL-HLT Demonstrations, 48-53. https://doi.org/10.18653/v1/N19-4009

Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. arXiv. https://doi.org/10.48550/arXiv.1412.6980

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. arXiv. https://doi.org/10.48550/arXiv.1705.07874

Heafield, K. (2011). KenLM: Faster and smaller language model queries. Proceedings of the Sixth Workshop on Statistical Machine Translation, 187-197. https://doi.org/10.3115/2132960.2132980

Costa-jussa, M. R., Cross, J., Celebi, O., Elbayad, M., Heafield, K., Heffernan, K., et al. (2022). No language left behind: Scaling human-centered machine translation. arXiv. https://doi.org/10.48550/arXiv.2207.04672

Schick, T., & Schütze, H. (2021). Exploiting cloze questions for few-shot text classification and natural language inference. Proceedings of EACL, 255-269. https://doi.org/10.18653/v1/2021.eacl-main.20

Blodgett, S. L., Barocas, S., Daume III, H., & Wallach, H. (2020). Language (technology) is power: A critical survey of bias in NLP. Proceedings of ACL, 5454-5476. https://doi.org/10.18653/v1/2020.acl-main.485

Wu, S., & Dredze, M. (2019). Beto, Bentz, Becas: The surprising cross-lingual effectiveness of BERT. Proceedings of EMNLP-IJCNLP, 833-844. https://doi.org/10.18653/v1/D19-1077

Pires, T., Schlinger, E., & Garrette, D. (2019). How multilingual is multilingual BERT? Proceedings of ACL, 4996-5001. https://doi.org/10.18653/v1/P19-1493

Xue, L., Constant, N., Roberts, A., Kale, M., Al-Rfou, R., Siddhant, A., et al. (2021). mT5: A massively multilingual pre-trained text-to-text Transformer. Proceedings of NAACL-HLT, 483-498. https://doi.org/10.18653/v1/2021.naacl-main.41

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. Science, 313(5786), 504-507. https://doi.org/10.1126/science.1127647

Adelani, D. I., Abbott, J., Neubig, G., D'souza, D., Kreutzer, J., Lignos, C., et al. (2021). MasakhaNER: Named entity recognition for African languages. Transactions of the Association for Computational Linguistics, 9, 1116-1131. https://doi.org/10.1162/tacl_a_00416

Mager, M., Gutierrez-Vasques, X., Sierra, G., & Meza-Ruiz, I. (2018). Challenges of language technologies for the Indigenous languages of the Americas. Proceedings of COLING, 55-69. https://doi.org/10.18653/v1/C18-1171

Bird, S. (2020). Decolonising speech and language technology. Proceedings of COLING, 3504-3519. https://doi.org/10.18653/v1/2020.coling-main.313

Ponti, E. M., O'Horan, H., Berzak, Y., Vulić, I., Reichart, R., Poibeau, T., et al. (2019). Modeling language variation and universals: A survey on

typological linguistics for natural language processing. Computational Linguistics, 45(3), 559-601. https://doi.org/10.1162/coli_a_00357

Cotterell, R., Kirov, C., Sylak-Glassman, J., Yarowsky, D., Eisner, J., & Hulden, M. (2016). The SIGMORPHON 2016 shared task: Morphological reinflection. Proceedings of the SIGMORPHON Workshop, 10-22. https://doi.org/10.18653/v1/W16-2002

Kann, K., & Schütze, H. (2016). Single-model encoder-decoder with explicit morphological representation for reinflection. Proceedings of ACL, 555-560. https://doi.org/10.18653/v1/P16-2090

Cotterell, R., Kirov, C., Sylak-Glassman, J., Walther, G., Vylomova, E., Xia, P., et al. (2017). CoNLL-SIGMORPHON 2017 shared task: Universal morphological reinflection. Proceedings of CoNLL-SIGMORPHON, 1-30. https://doi.org/10.18653/v1/K17-2001

Faruqui, M., Tsvetkov, Y., Neubig, G., & Dyer, C. (2016). Morphological inflection generation using character sequence to sequence learning. arXiv. https://doi.org/10.48550/arXiv.1512.06110

Denkowski, M., & Lavie, A. (2014). Meteor universal: Language specific translation evaluation for any target language. Proceedings of the Ninth Workshop on Statistical Machine Translation, 376-380. https://doi.org/10.3115/v1/W14-3348

Nothman, J., Ringland, N., Radford, W., Murphy, T., & Curran, J. R. (2013). Learning multilingual named entity recognition from Wikipedia. Artificial Intelligence, 194, 151-175. https://doi.org/10.1016/j.artint.2012.03.006

Jean, S., Cho, K., Memisevic, R., & Bengio, Y. (2015). On using very large target vocabulary for neural machine translation. Proceedings of ACL, 1-10. https://doi.org/10.3115/v1/P15-1001

Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. Proceedings of ICML. https://doi.org/10.48550/arXiv.1506.02142

Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. Proceedings of ACL, 3645-3650. https://doi.org/10.18653/v1/P19-1355