

Are Language Models Borrowing-Blind? A Multilingual Evaluation of Loanword Identification across 10 Languages

Mérlin Sousa Silva

Sina Ahmadi

Department of Computational Linguistics

University of Zurich

{merilin.sousasilva,sina.ahmadi}@uzh.ch

Abstract

Throughout language history, words are borrowed from one language to another and gradually become integrated into the recipient's lexicon. Speakers can often differentiate these loanwords from native vocabulary, particularly in bilingual communities where a dominant language continuously imposes lexical items on a minority language. This paper investigates whether pretrained language models, including large language models, possess similar capabilities for loanword identification. We evaluate multiple models across 10 languages. Despite explicit instructions and contextual information, our results show that models perform poorly in distinguishing loanwords from native ones. These findings corroborate previous evidence that modern NLP systems exhibit a bias toward loanwords rather than native equivalents. Our work has implications for developing NLP tools for minority languages and supporting language preservation in communities under lexical pressure from dominant languages.

Keywords: lexical borrowing, multilingualism, low-resourced languages

1. Introduction

Loanwords or lexical borrowings are words adopted from one language into another through language contact. The English lexicon illustrates this phenomenon vividly: *sugar*, *cushy*, *sky*, and *chocolate* all entered English at different historical periods, reflecting geopolitical, cultural, and commercial exchanges with other languages. Lexical borrowing rates vary greatly among languages, with average borrowing rate at 24.2% based on the Loanword Typology (LWT) project (Haspelmath and Tadmor, 2009). However, these rates also reflect varying degrees of knowledge about each language's history. Despite such measurement challenges, it is clear that borrowing occurs in all languages as every language contains at least some words adopted from contact with other linguistic communities (Tadmor, 2009).

The impact of lexical borrowing extends beyond individual word histories. In bilingual and multilingual communities, the flow of loanwords from dominant to minority languages can accelerate language shift and vocabulary loss (Rabiu et al., 2024; Sokpo et al., 2020). Research has identified two primary drivers of borrowing: need, when languages adopt terms for new concepts, and prestige, when speakers borrow from socially dominant languages even for existing concepts. While all languages borrow regardless of their status, prestige predicts the direction of borrowing as lexical items flow from more powerful to less powerful linguistic communities, with this asymmetry intensifying during periods of significant cultural contact Carling et al. (2019). Understanding and identifying loanwords is there-

fore crucial not only for historical linguistics but also for language preservation efforts, educational initiatives, and developing tools in computational linguistics (CL) and natural language processing (NLP).

Although borrowing has been long studied in linguistics, as in (Poplack et al., 1988; Winford, 2010; Van Hout and Muysken, 1994) to mention but a few, it has received limited attention in CL and NLP where previous studies focus on a handful of languages and are restricted to narrow aspects of lexical borrowing, such as loanword detection based on word lists (Mi et al., 2020; Miller and List, 2023) or analysis of anglicisms in specific languages (Alvarez-Mellado et al., 2025; Mellado, 2020). To further clarify the computational challenges of lexical borrowing, recent work introduced ConLoan (Ahmadi et al., 2025), a contrastive multilingual dataset for evaluating how NLP systems handle loanwords versus native alternatives. That work demonstrated that neural machine translation (NMT) systems and large language models (LLMs) show systematic preferences for loanwords in generation and comprehension tasks, exhibiting higher surprisal for native alternatives and reduced translation quality when loanwords are replaced. Building on these findings, the present study makes two-fold contributions:

1. We directly investigate whether models can identify loanwords when explicitly asked to do so, evaluating various pretrained models on loanword identification across 10 typologically diverse languages using the ConLoan dataset. Our findings show that general-purpose LLMs consistently fail to distinguish borrowed from native vocabulary.

2. We fine-tune several pretrained models and demonstrate, through both quantitative and qualitative analysis, how different architectures and training approaches perform on this task. Our experiments indicate that loanword identification remains far from solved, even with task-specific fine-tuning.

2. Related Work

Loanwords are lexical items borrowed from one language and integrated into another language’s vocabulary, where they become part of the recipient language’s lexicon and are used by its speakers, including monolinguals. Identifying loanwords has proven beneficial across multiple domains, including language contact (Gardani, 2022), language education (Dashti and Dashti, 2017), language technology (Mi, 2023), and speech recognition (Paik et al., 2025; Hirai and Kovalyova, 2024). For language preservation efforts, accurate loanword identification helps document lexical influence patterns and supports native vocabulary development, while in educational contexts, it aids etymology instruction and language history teaching. For NLP systems, understanding lexical borrowing is essential for developing multilingual models that can properly handle diverse linguistic phenomena. The task of loanword identification is closely related to code-switching detection, which has been extensively studied in NLP (Bali et al., 2014; James et al., 2022). Both phenomena involve mixing linguistic material from multiple languages, but they differ fundamentally: code-switching involves alternating between languages within discourse, while borrowing represents lexical adoption into a recipient language’s system (Alvarez-Mellado and Lignos, 2022).

Previous computational approaches to loanword identification have employed diverse feature sets and methodologies. Phonetic, phonological and morphosyntactic features have been widely used, leveraging the observation that loanwords often retain patterns from donor languages (Kpoglu, 2025). Patro et al. (2017) introduced methods to identify borrowing likelihood based on social media content, demonstrating that lexical innovation and borrowing patterns are observable in online language use. More recent work has incorporated semantic similarity metrics from embedding models (Serigos, 2022) and LLMs (Mi et al., 2025). Much of this work has focused on specific language pairs, particularly anglicisms in various recipient languages (Nath et al., 2022). Nevertheless, loanword identification remains a challenging and largely unsolved task, with significant room for improvement particularly for low-resource languages and contexts where borrowed terms are highly integrated (Mi et al., 2021).

Recent work on ConLoan (Ahmadi et al., 2025),

a contrastive multilingual dataset where loanwords are paired with native alternatives across 10 languages, show that modern NLP systems exhibit systematic biases toward loanwords: LLMs (EuroLLM, Llama 2.7 and 3.1) show lower surprisal for sentences containing loanwords compared to those with native alternatives, and NMT models (NLLB) perform less efficiently when translating sentences with native replacements. These findings raise a critical question: **If models prefer loanwords in generation and comprehension tasks, can they identify loanwords when explicitly instructed to do so?** Existing work has not systematically evaluated whether general-purpose pretrained models and LLMs possess this capability, nor explored whether fine-tuning can enable them to distinguish borrowed from native vocabulary. This paper fills this gap by directly testing models’ ability to perform loanword identification across multiple languages and training paradigms.

3. Methodology

We evaluate loanword identification capabilities through two complementary approaches: (1) prompting large language models in zero-shot and few-shot (two-shot) settings, and (2) fine-tuning multilingual token classification models. Both approaches use the ConLoan dataset (Ahmadi et al., 2025), which provides sentence-level annotations with loanword spans across Chinese, French, German, Greek, Icelandic, Italian, Northern Kurdish, Portuguese, Russian, and Spanish. The following is an example in Portuguese where the loanwords ‘franchise’ is replaced by a native alternative ‘franquia’ in the annotation process:

*Podemos vender-te um **franchise** disto por 3000\$.*
(We can sell you a franchise of this for 3,000 dollars.)
*Podemos vender-te um **franquia** disto por 3000\$.*

3.1. Task Formulation

We formulate loanword identification as a sequence labeling task using Inside–outside–beginning (BIO) tagging (O, B-LOAN, I-LOAN). Given a sentence tokenized into words w_1, w_2, \dots, w_n , the task is to assign each word a tag indicating whether it is outside a loanword span (O), begins a loanword span (B-LOAN), or continues a loanword span (I-LOAN). This formulation handles both single-word and multi-word loanwords uniformly.

3.2. LLMs

We evaluate three LLMs: Gemini-2.5-Flash-Lite, GPT-4.1, and Meta-Llama-3-8B-Instruct. Each model is tested under two setups:

- **Zero-shot:** Models receive only task instructions without examples.
- **Few-shot:** Models are provided with the same two in-language exemplars of the target language before processing the test sentence.

How loanwords are defined determines how they are identified and this makes this task particularly compelling. To assess how the internal knowledge of the selected LLMs might (or not) affect the identification task, we evaluate three prompt variants that progressively add more granularity to the definition of loanwords:

- **Prompt 1 (Minimal):** Basic instruction to detect loanwords without definition.

```
You are a loanword detection
system. Identify loanwords in:
"{sentence}"
```

- **Prompt 2 (Etymological):** Adds [Haspelmath and Tadmor \(2009\)](#)'s definition emphasizing historical borrowing.

```
You are a loanword detection
system. Loanword (or lexical
borrowing) is here defined as
a word that at some point in
the history of a language en-
tered its lexicon as a result
of borrowing (or transfer, or
copying). Identify loanwords
in: "{sentence}"
```

- **Prompt 3 (Usage-based):** Extends [Haspelmath and Tadmor \(2009\)](#)'s definition by focusing on conventional use by monolinguals, distinguishing loanwords from code-switching.

```
You are a loanword detection
system. From the point of view
of an entire language (not that
of a single speaker), a loan-
word is a word that can con-
ventionally be used as part of
the language. In particular,
it can be used in situations
where no code-switching occurs,
e.g. in the speech of monolin-
guals. This is the simplest
and most reliable criterion for
distinguishing loanwords from
single-word switches. Identify
loanwords in: "{sentence}"
```

All prompts constrain output to valid Python lists of strings to enable deterministic parsing. In addition, regular expressions are employed as a supplementary method to identify lists within the LLM's responses. These responses are normalized and

cached using MD5 hashing to ensure reproducibility and avoid redundant API calls.

3.3. Multilingual Encoder Models

We evaluate four pretrained multilingual encoders: mBERT¹, base and large variants of XLM-RoBERTa², and ELECTRA-base multilingual³. Each model is tested in two configurations:

- **Zero-shot baseline:** Models are used as frozen feature extractors without task-specific training. A deterministic rule based on subword continuity and token boundaries maps hidden states to BIO tags.
- **Fine-tuned:** Models are fine-tuned with a token classification head on an 80/20 train/test split of ConLoan. We use the following hyperparameters: 10 epochs, batch size 16, AdamW optimizer with weight decay 0.01. Training employs cross-entropy loss with label -100 for special tokens and subsequent subword pieces to exclude them from gradient computation.

3.4. Evaluation Metrics

We compute precision, recall, and F1-score using the seqeval library⁴, which evaluates at the entity span level rather than per-token. This penalizes boundary errors appropriately for the identification task. We report both overall (macro-averaged across languages) and per-language metrics. For LLM evaluation, we employ two scoring protocols:

- **Strict:** Set-based comparison where predicted spans must match gold segmentation exactly. Multi-word loanwords must appear as single items if annotated as such.
- **Relaxed:** Gold multi-word spans are decomposed into their constituent tokens prior to comparison, ensuring that both ['social media'] and ['social', 'media'] are considered equivalent matches. This procedure was adopted because LLMs exhibited difficulty in preserving multi-word loanwords as single tokens. Consequently, this approach accommodates segmentation variability while maintaining strict requirements for correct token coverage.

Language	Prompt-based			Model-based		
	Prompt1	Prompt2	Prompt3	Gemini	Llama	OpenAI
Chinese	0.605	0.521	0.537	0.697	0.429	0.537
French	0.474	0.361	0.453	0.557	0.254	0.476
German	0.141	0.129	0.120	0.168	0.086	0.135
Greek	0.391	0.335	0.327	0.513	0.178	0.362
Icelandic	0.307	0.285	0.299	0.382	0.149	0.359
Italian	0.469	0.384	0.440	0.516	0.259	0.517
Northern-Kurdish	0.438	0.392	0.378	0.489	0.200	0.519
Portuguese	0.325	0.260	0.324	0.350	0.261	0.298
Russian	0.397	0.387	0.412	0.502	0.201	0.493
Spanish	0.432	0.314	0.436	0.486	0.293	0.402
Average	0.3979	0.3368	0.3726	0.466	0.231	0.4098

Table 1: F1 scores per language for each model and prompt (averaged over all prompts and configurations) in loanword identification using LLMs. Not providing explicit definition (Prompt 1) and using Gemini result in higher F1.

4. Experiments

4.1. LLMs Performance

Table 1 presents the performance of each model and prompt configuration. In general, the results reveal that the detection of loanwords remains a difficult task for LLMs; in all settings, the F1 scores stay below 0.70. Among the models tested, Gemini achieved the best overall performance (0.466), followed by OpenAI (0.4098) and Llama (0.231). Adding few-shot examples had mixed effects. In some cases, as in Prompts 1 and 2, it slightly improved results (by 0.01–0.09 points) by helping models achieve higher recall. However, for Prompt 3, performance actually declined (by 0.02–0.04 points), as shown in Figure 1.

A particularly striking anomaly appeared with Llama under the second few-shot prompt, where F1-scores nearly dropped to zero, which is far below its usual average of about 0.35. We suspect this happened because the added examples shifted the model’s internal expectations, causing it to ignore the original instruction format. Instead of returning structured lists as required, Llama began producing free-form text or syntactically broken lists. Since our evaluation relied on strict parsing of list outputs, these deviations were counted as empty predictions, collapsing the F1-score even though the underlying reasoning capability may not have degraded as drastically.

Language Performance: Performance varies notably across languages. Chinese, French and Italian achieve the highest average F1-scores

across both prompt-based and model-based settings, suggesting that loanword boundaries in these languages are comparatively easier for models to identify. In contrast, German, Icelandic, and Portuguese yield the lowest results, likely due to their rich morphological inflection, compounding tendencies, and the greater integration of loanwords into native morphology. Similarly, it can be seen that among the model-based systems, Gemini consistently achieves the strongest results across most languages, followed by OpenAI and Llama.

4.2. Multilingual Encoder Performance

Overall, fine-tuned multilingual encoders substantially outperform their unfine-tuned counterparts across all evaluation metrics and languages. As shown in Table 2, untrained models perform poorly, with average F1-scores below 0.2 across architectures and a maximum of only 0.0197 for Russian

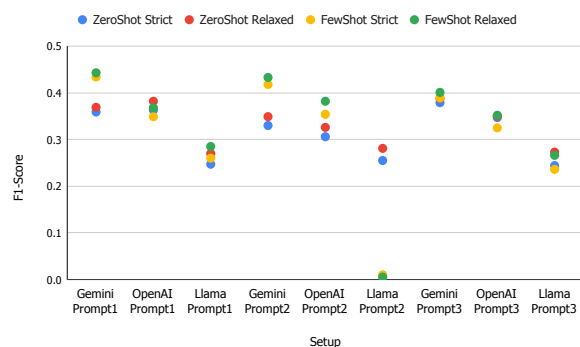


Figure 1: F1-scores of LLMs based on prompt, evaluation protocol, and fine-tuning setup. Overall, providing few shots with relaxed evaluation using Gemini yields higher F1-score for loanword identification.

¹bert-base-multilingual-cased
²xlm-roberta-base/large
³google/electra-base-discriminator
⁴<https://github.com/chakki-works/seqeval>

Language	Zero-shot				Fine-tuned			
	ELECTRA	mBERT	XLM-R _B	XLM-R _L	ELECTRA	mBERT	XLM-R _B	XLM-R _L
Northern Kurdish	0.034	0.034	0	0.028	0.622	0.683	0.705	0.735
Russian	0.031	0.046	0	0.057	0.504	0.691	0.778	0.829
Italian	0.022	0.031	0.009	0.015	0.831	0.888	0.921	0.934
Portuguese	0.015	0.015	0.003	0.019	0.676	0.731	0.755	0.775
French	0.014	0.019	0	0.016	0.924	0.942	0.913	0.935
Spanish	0.011	0.015	0	0.016	0.816	0.914	0.898	0.917
German	0.01	0.018	0	0.008	0.688	0.834	0.856	0.881
Chinese	0.009	0.01	0	0.004	0.244	0.775	0.84	0.815
Greek	0.005	0.005	0	0	0.423	0.783	0.827	0.848
Icelandic	0.003	0.004	0	0.003	0.752	0.818	0.829	0.844
Average	0.0154	0.0197	0.0012	0.0166	0.648	0.8059	0.8322	0.8513

Table 2: F1-scores per language and model in two setups of zero-shot and finetuned. Subscript B and L refer to base and large variants, respectively. Fine-tuned XML-R achieves the highest F-score in loanword identification while the pretrained unfine-tuned models perform poorly on the task. These models also outperform LLMs.

under XLM-R_L. In contrast, fine-tuning leads to drastic improvements, yielding average F1-scores between 0.648 and 0.851, depending on the model. Among fine-tuned models, XLM-R_L achieves the strongest overall performance (0.8513). Our findings show that pretraining on multilingual corpora lays a necessary foundation but does not go far enough for nuanced challenges like loanword identification.

Language Performance: At the language level, fine-tuning particularly benefits languages that previously exhibited low zero-shot performance, such as Chinese, Icelandic, and Northern Kurdish, where F1-scores rise from near-zero to above 0.7–0.8. Romance languages (French, Italian, Spanish, Portuguese) consistently reach the highest scores, reaching or even exceeding 0.8 F1 in most fine-tuned configurations, while German and Greek achieve slightly lower but still strong results (around 0.85 F1 under XLM-R_L).

5. Analysis

We conduct a qualitative analysis of misclassified instances to better understand the underlying reasons for the failures of models and to examine whether fine-tuned models continue to display systematic errors.

5.1. Code-Switching vs. Loanwords

Across the model predictions, both LLMs and fine-tuned models exhibited difficulties in accurately identifying words as non-loanwords, which we would classify as code-switches. Code-switching refers to the practice of alternating between two or more languages, dialects, or accents while speaking. However, code-switches and loanwords are

distinct linguistic phenomena, a distinction that the models generally fail to capture. For instance, LLMs frequently misclassified code-switches as loanwords such as the word *really* in the following sentence in Northern Kurdish:

really ha, nizanim, tiştek nabe tiştek tevli-hev bûbe. Tu çi dikî tu hê li bajarê xwe yî an? (Really, huh, I don't know, nothing's happening, something is confusing. What are you doing; are you still in your city?)

In this context, labeling 'really' as a loanword is not correct, since 'really' is not morphologically or phonologically integrated into Kurdish and is used purely as a code-switching element. However, this pattern did not persist throughout all LLM responses. When presented with Prompt 3, which explicitly distinguishes between loanwords and code-switches, classification precision improved. Nevertheless, certain LLMs, such as OpenAI's models, tended toward the opposite extreme, labeling as code-switches words that merely resemble English orthography, even when these words are conventionally used in monolingual contexts. One such example is the word *deal* in the following sentence in French:

*Il nous appartient, dans la mesure du possible – et je m'y emploie – de faire en sorte que ce qui est globalement un bon **deal** entre les Américains et les Chinois, soit un aussi bon **deal** pour les Européens. (It is up to us, insofar as possible, and I am working on it, to ensure that what is overall a good *deal* between the Americans and the Chinese is also a good *deal* for the Europeans.)*

Although the native alternative would be *accord*, *deal* is commonly used in French monolingual dis-

course and should therefore be classified as a loanword. Despite this, the models prompted under the third setup consistently categorized it as a non-loanword.

As such, both pretrained and fine-tuned models struggle to operationalize the nuanced boundary between code-switching and lexical borrowing. While fine-tuning improves recognition of overt language alternation, models still over-rely on orthographic cues and fail to account for sociolinguistic integration. This suggests that accurate detection of code-switching requires a deeper contextual understanding of speaker intent and linguistic assimilation—dimensions that current models are not yet equipped to capture effectively.

5.2. Named Entities and Proper Nouns

Across models, named entities and proper nouns consistently emerged as a major source for misclassifications. Both zero-shot and few-shot setups frequently mislabeled entities such as country names, organizations, and acronyms (e.g., *NASA*, *ESA*) as loanwords, resulting in a higher false-positive rate. This behavior likely stems from the orthographic or phonological resemblance of such tokens to foreign lexical patterns, which the models interpret as evidence of borrowings. Although fine-tuning led to a notable improvement in recall, these errors persisted. For instance, in the following sentence in German, the compound *PISA-Studie* was incorrectly identified as a loanword despite being a named entity referring to the *Programme for International Student Assessment*:

*An der Spitze der internationalen Rangliste laut der letzten **PISA-Studie** steht der Shanghai-Distrikt von China.* (At the top of the international rankings according to the latest *PISA study* is China's Shanghai district.)

Additional evidence for this pattern is found in the multilingual predictions of XLM-RoBERTa, where tokens such as *Jazz* in Italian or *golfe* in Portuguese were erroneously tagged as loanwords, even though they occur as part of named entities or idiomatic expressions. Similarly, the model occasionally misclassified tokens like 飛船上 (“on the spacecraft”) in Cantonese due to script-based unfamiliarity, which it mistakenly associated with foreign origin rather than contextual meaning.

Few-shot prompting with instruction-tuned LLMs, such as LLaMA, reduced these errors in some cases but introduced new inconsistencies. For example, the LLaMA model frequently failed to distinguish between common nouns derived from proper names (e.g., *rail* in French technical discourse) and genuine borrowings as in the following sentence:

*[...]l'examen des conseillers à la sécurité pour le transport par route, par **rail** ou par voie navigable de marchandises dangereuses.* ([...] the examination of safety advisers for the transport by road, by *rail*, or by inland waterways of dangerous goods.)

In such contexts, capitalization cues alone were insufficient for accurate disambiguation, suggesting that the models rely heavily on surface-level orthographic features rather than semantic or contextual grounding. Taken together, both pretrained and fine-tuned models exhibit a limited capacity to encode the pragmatic and discourse-level properties that distinguish named entities from lexical borrowings.

5.3. Scientific/Technical and Greco-Latin Vocabulary

Another recurrent source of misclassifications involved scientific and technical terminology, particularly words derived from Greco-Latin roots. Such terms exhibit a complex linguistic status: while some are historically borrowed, many are fully assimilated and function as native lexical items within specialized domains. Models fine-tuned on the ConLoan dataset tended to under-detect these terms as loanwords, likely because they appear morphologically regular and orthographically integrated within their host languages. Conversely, LLMs in few-shot settings often exhibited over-detection, labeling scientific or Greco-Latin vocabulary such as *filosofia* in Portuguese as loanwords purely based on etymological cues. In the following sentence in Icelandic, the term *nítröt* (‘nitrates’) was classified as a native word by XLM-RoBERTa_L:

*[...] þar sem **nítröt** geta breyst í **nítrít** og nítrósamín, og hvatti til þess að teknar yrðu upp góðar starfsvenjur í landbúnaði til þess að tryggja eins lágt nítratmagn og kostur er.* ([...] where *nitrates* can turn into nitrites and nitrosamines, and encouraged the adoption of good agricultural practices to ensure the lowest possible nitrate levels.)

Similarly, LLMs like Gemini frequently misclassified scientific expressions such as *mazout* (French) or *national* (Russian context) due to their shared morphological ancestry with Latin-derived words, despite their full lexical assimilation in monolingual usage. These findings highlight the tension between etymological origin and contemporary linguistic integration, an area where models lack clear conceptual grounding.

We believe that scientific and technical terms expose the models' inability to reconcile historical bor-

rowing with synchronic linguistic norms. Whereas fine-tuned models often fail to detect subtle loanword traces, general-purpose LLMs overgeneralize based on etymology, revealing a lack of sensitivity to contextual and disciplinary registers.

6. Conclusion

This paper sheds light on loanword identification in 10 languages using pretrained and language models, widely used in modern NLP. This task has remained relatively underexplored largely because data distinguishing loanwords from native words are limited, even though such knowledge is pivotal for studying language contact and acquisition. Using ConLoan dataset, we evaluate the performance of a few LLMs and pretrained models for this task and show that, despite the differences in performance using different prompts, models and fine-tuning setups, the task is quite challenging for LLMs with an average F-score of less than 0.5. Fine-tuning pretrained models yields higher performance results with XLM-R (large) achieving 0.8513 F-score. While fine-tuning improves precision, both pretrained and instruction-tuned models remain prone to over-reliance on orthographic and etymological cues. Code-switches are often mistaken for borrowings, proper nouns for foreign insertions, and Greco-Latin scientific terms for recent loans. These systematic errors reveal that loanword detection is far from a solved problem.

Limitations The core contribution of the paper lays in the additional empirical results that demonstrate that the studied LLMs in this paper have a proclivity to not detect or differentiate loanwords accurately. In other words, models do not behave as language purists. Future work should consider a more fine-grained identification task where loanwords are categorized based on their integration status or within a continuum (Poplack and Sankoff, 1984). Successful identification requires sensitivity to pragmatic context, speaker intent, and lexical assimilation, dimensions that go beyond surface form. Additionally, creating and analyzing models with controlled vocabulary where loanwords are automatically replaced by native alternatives should be explored. We will release the models upon the acceptance of the paper.

Ethics Statement This research involves computational analysis of publicly available linguistic data and poses no ethical concerns. The ConLoan dataset used contains no personally identifiable information, and our experiments focus solely on linguistic classification tasks without generating new content that could be harmful or biased.

Acknowledgments

Sina Ahmadi gratefully thanks the support of the UZH Postdoc Grant (reference number 269093).

7. Bibliographical References

- Sina Ahmadi, Micha David Hess, Elena Álvarez-Mellado, Alessia Battisti, Cui Ding, Anne Göhring, Yingqiang Gao, Zifan Jiang, Andrianos Michail, Peshmerge Morad, et al. 2025. ConLoan: A contrastive multilingual dataset for evaluating loanwords. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 30070–30090.
- Elena Alvarez-Mellado and Constantine Lignos. 2022. [Borrowing or codeswitching? annotating for finer-grained distinctions in language mixing](#). In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 3195–3201, Marseille, France. European Language Resources Association.
- Elena Alvarez-Mellado, Jordi Porta-Zamorano, Constantine Lignos, and Julio Gonzalo. 2025. Overview of ADoBo at IberLEF 2025: Automatic Detection of Anglicisms in Spanish. *arXiv preprint arXiv:2507.21813*.
- Kalika Bali, Jatin Sharma, Monojit Choudhury, and Yogarshi Vyas. 2014. [“I am borrowing ya mixing ?” an analysis of English-Hindi code mixing in Facebook](#). In *Proceedings of the First Workshop on Computational Approaches to Code Switching*, pages 116–126, Doha, Qatar. Association for Computational Linguistics.
- Gerd Carling, Sandra Cronhamn, Robert Farren, Elnur Aliyev, and Johan Frid. 2019. The causality of borrowing: Lexical loans in Eurasian languages. *PLoS one*, 14(10):e0223588.
- Fatimah Dashti and Abdulmohsen A Dashti. 2017. Morphological adaptation of English loanwords in Twitter: Educational implications. *International Journal of Higher Education*, 6(3):231–248.
- Francesco Gardani. 2022. Contact and borrowing. *The Cambridge handbook of Romance linguistics*, pages 845–869.
- Martin Haspelmath and Uri Tadmor. 2009. The loanword typology project and the world loanword database. *Loanwords in the world's languages: A comparative handbook*, pages 1–34.

- Akiyo Hirai and Angelina Kovalyova. 2024. Speech-to-text applications' accuracy in English language learners' speech transcription. *University of Hawaii National Foreign Language Resource Center*.
- Jesin James, Vithya Yogarajan, Isabella Shields, Catherine I. Watson, Peter Keegan, Keoni Mahelona, and Peter-Lucas Jones. 2022. [Language models for code-switch detection of te reo Māori and English in a low-resource setting](#). In *Findings of the Association for Computational Linguistics: NAACL 2022, Seattle, WA, United States, July 10-15, 2022*, pages 650–660. Association for Computational Linguistics.
- Promise Dodzi Kpoglu. 2025. [Feature-refined unsupervised model for loanword detection](#). *CoRR*, abs/2508.17923.
- Elena Álvarez Mellado. 2020. [An annotated corpus of emerging anglicisms in spanish newspaper headlines](#). In *Proceedings of the The 4th Workshop on Computational Approaches to Code Switching, CodeSwitch@LREC 2020, Marseille, France, May, 2020*, pages 1–8. European Language Resources Association.
- Chenggang Mi. 2023. [Improving the robustness of loanword identification in social media texts](#). *ACM Trans. Asian Low Resour. Lang. Inf. Process.*, 22(4):101:1–101:19.
- Chenggang Mi, Lei Xie, and Yanning Zhang. 2020. [Loanword identification in low-resource languages with minimal supervision](#). *ACM Trans. Asian Low Resour. Lang. Inf. Process.*, 19(3):43:1–43:22.
- Chenggang Mi, Shaoliang Xie, Yu Li, and Zhenghan He. 2025. [Loanword identification in social media texts with extended code-switching datasets](#). *ACM Transactions on Asian and Low-Resource Language Information Processing*, 24(8):1–19.
- Chenggang Mi, ShaoLin Zhu, and Rui Nie. 2021. [Improving loanword identification in low-resource language with data augmentation and multiple feature fusion](#). *Comput. Intell. Neurosci.*, 2021:9975078:1–9975078:9.
- John E. Miller and Johann-Mattis List. 2023. [Detecting lexical borrowings from dominant languages in multilingual wordlists](#). In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2023, Dubrovnik, Croatia, May 2-6, 2023*, pages 2591–2597. Association for Computational Linguistics.
- Abhijnan Nath, Sina Mahdipour Saravani, Ibrahim Khebour, Sheikh Mannan, Zihui Li, and Nikhil Krishnaswamy. 2022. [A generalized method for automated multilingual loanword detection](#). In *Proceedings of the 29th International Conference on Computational Linguistics, COLING 2022, Gyeongju, Republic of Korea, October 12-17, 2022*, pages 4996–5013. International Committee on Computational Linguistics.
- Gio Paik, Yongbeom Kim, Soungmin Lee, Sangmin Ahn, and Chanwoo Kim. 2025. [Hike: Hierarchical evaluation framework for Korean-English code-switching speech recognition](#). 2509.24613.
- Jasabanta Patro, Bidisha Samanta, Saurabh Singh, Abhipsa Basu, Prithwish Mukherjee, Monojit Choudhury, and Animesh Mukherjee. 2017. [All that is English may be Hindi: Enhancing language identification through automatic ranking of the likeliness of word borrowing in social media](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2264–2274, Copenhagen, Denmark. Association for Computational Linguistics.
- Shana Poplack and David Sankoff. 1984. [Borrowing: the synchrony of integration](#).
- Shana Poplack, David Sankoff, and Christopher Miller. 1988. [The social correlates and linguistic processes of lexical borrowing and assimilation](#). Mouton de Gruyter.
- RA Rabiou, OL Abiola, and SO Ayelabola. 2024. [Borrowing as an achilles heel to the vocabulary development and extinction of Yorùbá words: A socio-semantic approach](#). *Tasambo Journal of Language, Literature, and Culture*, 3(2):59–66.
- Jacqueline Serigos. 2022. [Using automated methods to explore the social stratification of anglicisms in Spanish](#). *Corpus linguistics and linguistic theory*, 18(2):391–418.
- Rosaline Mnguhenen Sokpo, Sarah Terwase Shittu, Ph D Titus Terver Udu, and Joseph I Orban. 2020. [Lexical borrowing and language endangerment: A case of the Tiv language](#). *Ahyu: A Journal of Language and Literature*, 3:61–79.
- Uri Tadmor. 2009. [Loanwords in the world's languages: Findings and results](#). *Loanwords in the world's languages: A comparative handbook*, 55:75.
- Roeland Van Hout and Pieter Muysken. 1994. [Modeling lexical borrowability](#). *Language variation and change*, 6(1):39–62.
- Donald Winford. 2010. [Contact and borrowing](#). *The handbook of language contact*, pages 170–187.