LANGUAGE MODELS ARE BORROWING-BLIND: A MULTILINGUAL EVALUATION OF LOANWORD IDENTIFICATION ACROSS 10 LANGUAGES

Mérilin Sousa Silva, Sina Ahmadi

University of Zurich

Presented at LREC 2026

MOTIVATION

"Loanword (or lexical borrowing) is here defined as a word that at some point in the history of a language entered its lexicon as a result of borrowing (or transfer, or copying)" (Haspelmath 36)

- What are loanwords?
- Why important?
  - Language preservation
  - Minority languages
  - NLP systems
  - Multilingual models

RESEARCH QUESTION

1. Can LLMs identify loanwords?
2. Do prompts help?
3. Does fine-tuning help?
4. How does performance vary across languages?

DATASET

- ConLoan Dataset - 10 languages:
  - Chinese
  - French
  - German
  - Greek
  - Icelandic
  - Italian
  - Northern Kurdish
  - Portuguese
  - Russian
  - Spanish

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

- Annotated loanword spans
- Example sentence with loanword vs native alternative

Plain Sentence	Annotated Tokens (BIO)	
Podemos vender-te um franchise disto por 3000 dólares.	Podemos	O
	vender-te	O
	um	O
	franchise	B-LOAN
	disto	O
	por	O
	3000	O
	dólares	O

TASK DEFINITION

Loanword identification = **sequence labeling**

BIO tagging:

- O = not loanword
- B-LOAN
- I-LOAN

METHODS

Two approaches:

Approach	Models
Prompting	Gemini, GPT, LLaMA
Fine-Tuning	mBERT, XLM-R, ELECTRA

PROMPTING SETUP

Minimal

You are a loanword detection system. Identify loanwords in: “{sentence}”

Etymological definition

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 entered its lexicon as a result of borrowing (or transfer, or copying). Identify loanwords in: "{sentence}"

Usage-based definition

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 conventionally 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 monolinguals. This is the simplest and most reliable criterion for distinguishing loanwords from single-word switches. Identify loanwords in: "{sentence}"

→ Zero-shot vs Few-shot

MULTILINGUAL ENCODERS

1. Models
   1. mBert - bert-base-multilingual-cased
   2. XLM-RoBERTa - xlm-roberta-base + large
   3. ELECTRA-basemultilingual - google/electra-base-discriminator
2. Configurations
   1. Zero-shot baseline
   2. Fine-tuned

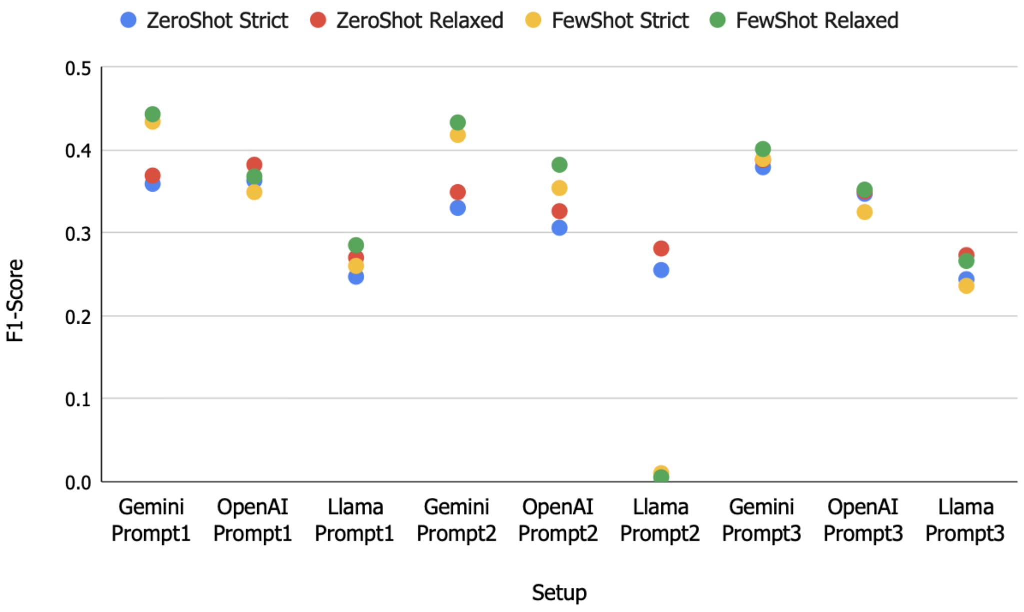
EVALUATION METRICS

Metrics:

- Precision
- Recall
- F1-score
- Strict vs Relaxed evaluation

RESULTS: LLMS

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



RESULTS: FINE-TUNED MODELS

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**

ERROR ANALYSIS

CODE-SWITCHING VS. LOANWORDS

***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?)

*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.)

NAMED ENTITIES AND PROPER NOUNS

*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.)

*[...]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.)

SCIENTIFIC/TECHNICAL AND GRECO-LATIN VOCABULARY

*[...] 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.)

CONCLUSION

- Loanword identification is still a difficult task for NLP models
- LLMs perform poorly (average F1 < 0.5)
- Fine-tuned multilingual models perform much better
  - Best model: XLM-R large (F1 ≈ 0.85)
- Models rely too much on orthography and etymology
- Main errors:
  - Code-switching vs loanwords
  - Named entities
  - Scientific / Greco-Latin vocabulary
- Loanword detection is not a solved problem

FUTURE WORK / LIMITATIONS

1. Limitations:
   1. Models **cannot reliably distinguish loanwords**
   2. Models rely on **surface features** rather than linguistic context
   3. Loanwords are **not binary** → exist on a **continuum of integration**
2. Future Work:
   1. More **fine-grained classification** of loanwords (integration levels)
   2. Models should consider:
      1. Pragmatic context
      2. Speaker intent
      3. Lexical assimilation
   3. Experiments with **controlled vocabulary** (replace loanwords with native alternatives)
   4. Models and dataset will be **released**

REFERENCES

- Ahmadi et al. (2025) — ConLoan: A Contrastive Multilingual Dataset for Evaluating Loanwords
- Alvarez-Mellado & Lignos (2022) — Borrowing or Code-switching?
- Alvarez-Mellado et al. (2025) — Automatic Detection of Anglicisms in Spanish
- Bali et al. (2014) — English-Hindi Code Mixing
- Carling et al. (2019) — The Causality of Borrowing
- Dashti & Dashti (2017) — Morphological Adaptation of English Loanwords
- Gardani (2022) — Contact and Borrowing
- Haspelmath & Tadmor (2009) — Loanword Typology Project
- James et al. (2022) — Language Models for Code-switch Detection
- Mi et al. (2020, 2021, 2023, 2025) — Loanword Identification in Low-Resource Languages
- Miller & List (2023) — Detecting Lexical Borrowings
- Nath et al. (2022) — Multilingual Loanword Detection
- Patro et al. (2017) — Word Borrowing in Social Media
- Poplack & Sankoff (1984); Poplack et al. (1988) — Lexical Borrowing
- Serigos (2022) — Anglicisms in Spanish
- Sokpo et al. (2020) — Lexical Borrowing and Language Endangerment
- Tadmor (2009) — Loanwords in the World's Languages
- Van Hout & Muysken (1994) — Lexical Borrowability
- Winford (2010) — Contact and Borrowing

THANK YOU

Contact: merilin.sousasilva@uzh.ch

Resources:
https://github.com/merilinsilva/LoanwordDetection_LM