Presence or Absence: Are Unknown Word Usages in Dictionaries?

Xianghe Ma[1], Dominik Schlechtweg[2], Wei Zhao[3]

University of Heidelberg[1], University of Stuttgart[2], University of Aberdeen[3]

Paper in a Nutshell

Motivation: The link between computational modeling of semantic change and dictionaries is previously tenuous.

Aim: Present a train-free lexicography system based on LLMs to manage dictionary updates.

Main results: No significant performance gain for GPT-4/Llama3 over GPT-3.5/Llama2.

Research Questions

- Can we detect what word usages are (not) covered by dictionaries?
- Can we automate the process of generating sense glosses for unrecorded word usages?
- How can we evaluate the results of novel sense detection and sense gloss generation?
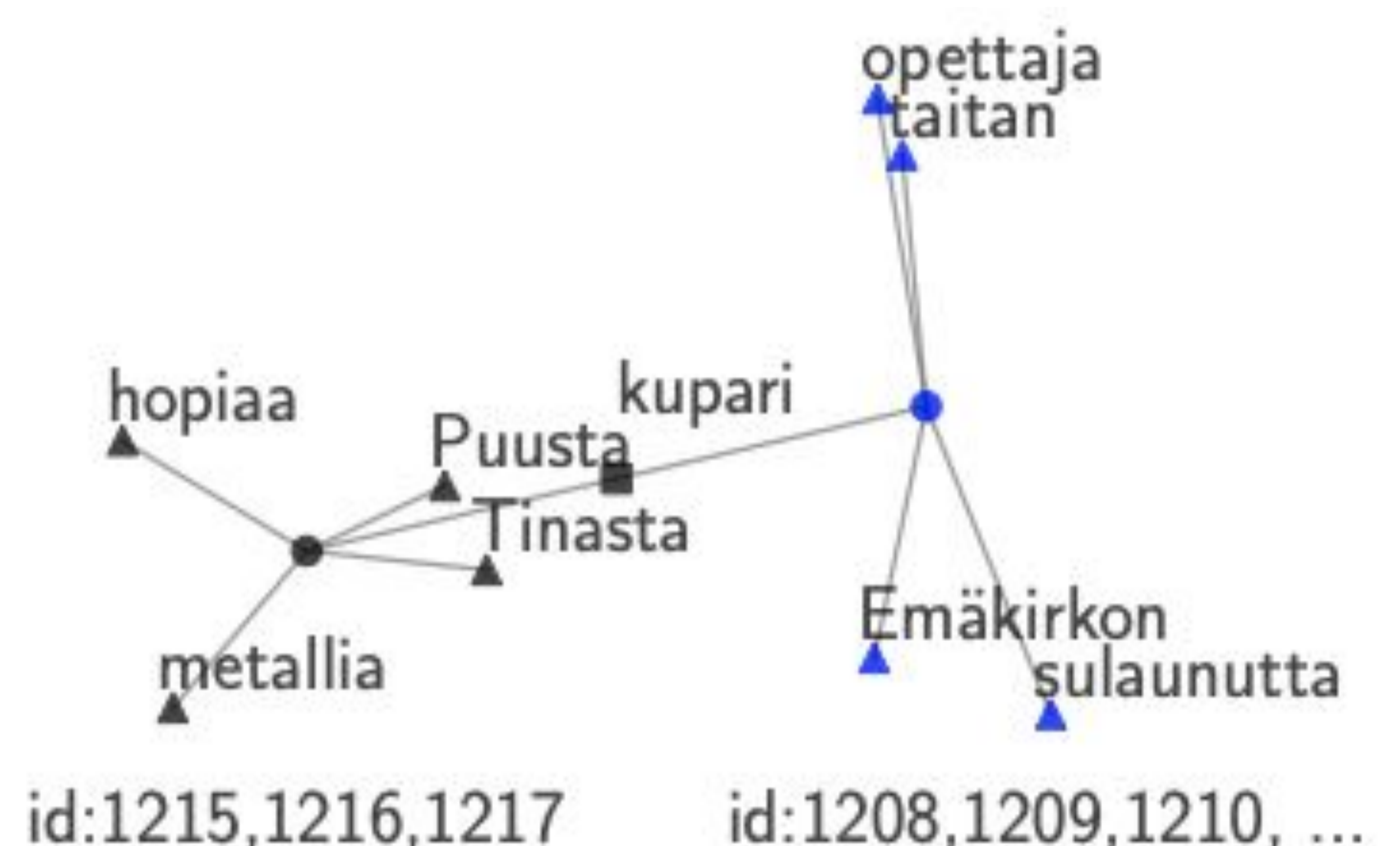
Contributions

- Our system is unsupervised.
- Our system is interpretable.
- A lexicography pipeline that automates the process of novel sense detection and sense gloss generation.
- Investigation of LLMs including GPT-3.5, GPT-4, Llama-2 and Llama-3 on sense gloss generation.

Our system is unsupervised
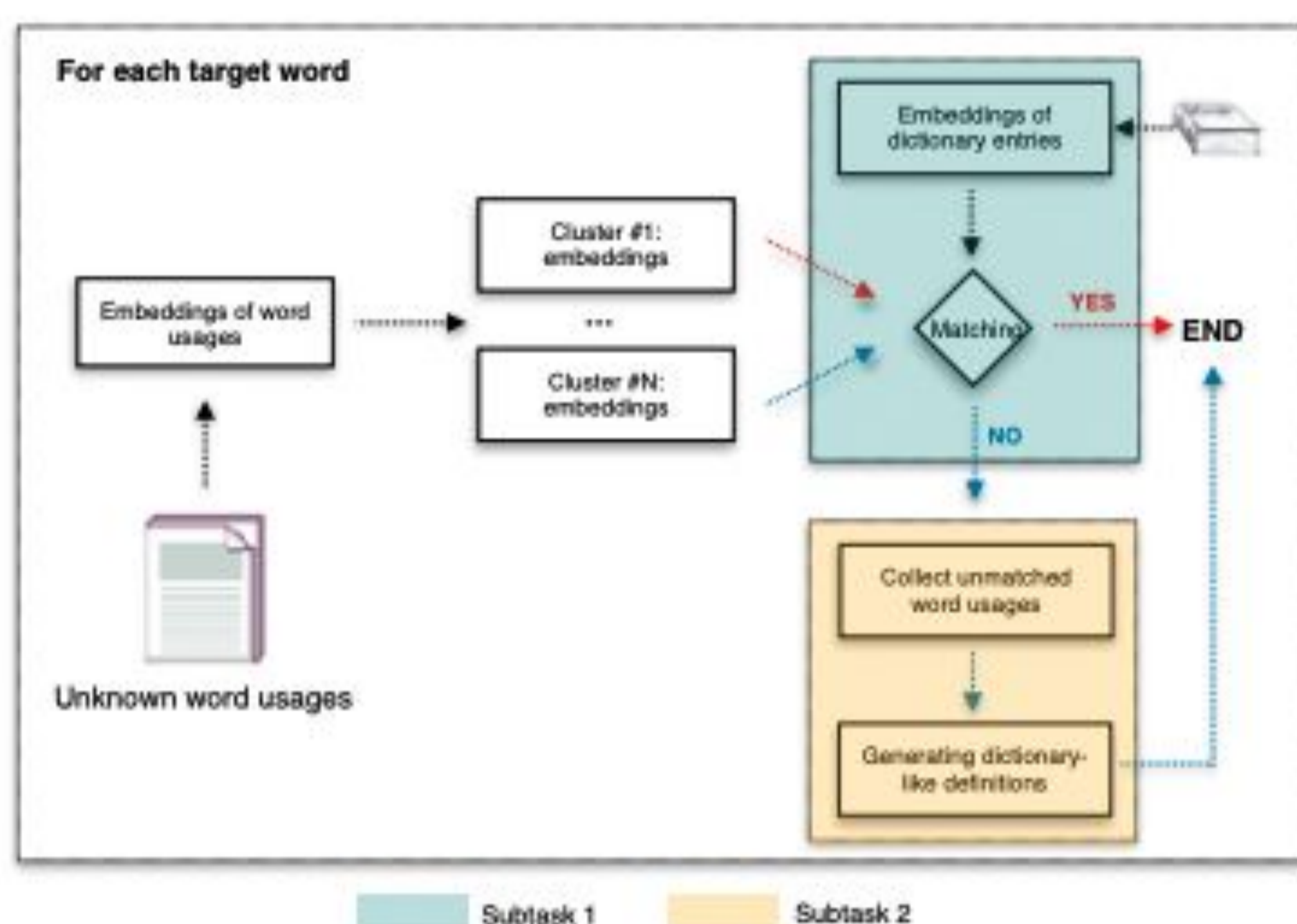
Our system is interpretable

Results – Task 1

Systems	#Entries	Finnish		Russian		German	
		BLEU	BERTScore	BLEU	BERTScore	BLEU	BERTScore
ABDN-NLP (Ours)	3	**0.107**	**0.706**	0.027	0.677	0.000	**0.714**
TartuNLP	1	0.028	0.679	**0.587**	**0.869**	**0.010**	0.630
t-montes	7	0.023	0.675	0.027	0.656	**0.010**	0.650
Baseline	6	0.033	0.403	0.005	0.377	0.000	0.490

Results - Task 2

LLMs	Finnish	
	BLEU	BERTScore
Baseline	**0.248**	0.607
GPT-3.5-turbo	0.022	0.640
GPT-4-turbo	0.025	**0.658**
LLaMA-2-7B	0.013	0.611
LLaMA-3-8B	0.013	0.603

Limitations

- Datasets are small with very few word usages for headword on average - this makes our system less performant.
- BERT is the non-frontier text encoder we use in our system for its flexibility to generate both token and sentence embeddings.
- Data contamination - LLMs may have seen dictionary entries during training.

Paper link

Github link

LChange @ ACL 2024, Bangkok