INT D 461: Artificial Intelligence Everywhere Capstone

Literature Review

Donna Mathew
Prabal Mehra
Mohammad Shahriar Hossain

IdiomSense: Sense-Augmented LLMs for Idioms, Proverbs, and Metaphors

Inference-time “sense cards” to boost detection & translation

We test whether LLMs truly understand figurative language and show that adding a tiny retrieved sense card (from WordNet/BabelNet/IdiomKB) before inference improves idiom detection and idiomatic translation—without fine-tuning.

Sentence → Sense Lookup → Sense Card → LLM → Output

Problem & Motivation

Why figurative language breaks models:

- Idioms/proverbs/metaphors are non-compositional (surface words ≠ meaning).
- LLMs and MT often choose the literal sense, producing confident but wrong outputs.
- Downstream impact: mistranslation, misunderstanding, cultural bias, and user trust issues.

Concrete example (minimal pair):

- **Idiomatic:** “After years of service, he kicked the bucket.” → means died.
- **Literal:** “The child kicked the bucket in the yard.” → means struck a container.

Our goal:

- Build a small, hard evaluation and a simple inference-time sense hint (“sense card”) that helps models pick the right meaning and translate accordingly.

BERT Idiom Detection (MWE'24)

- Fine-tuned BERT to decide if a highlighted span is **idiomatic or literal**, augmenting training with two signals:
- **word cohesion** (idioms "stick together") and **translation drift** (idioms behave oddly under MT).
- Consistently outperforms older baselines on multiple English datasets → small, targeted signals lift performance.
- Our project **keeps the same idea** of adding *light, simple signals*, but instead of training a model with them (like BERT did), you apply them **at inference time** (when the model is already trained).
- We will do this using **sense cards** - short pieces of extra information (like a gloss or definition) that you prepend to the model's input prompt.

"BERT-Based Idiom Detection" — ACL MWE Workshop 2024. https://aclanthology.org/2024.mwe-1.26.pdf

Theme - Data Collection

IdioTS: Hard Idiom Minimal Pairs for LLMs

CoAM: Corpus of all type multiword expressions

- An **expert-curated test suite** built from **minimal pairs**, nearly identical sentences where one usage is idiomatic and the other literal.
- Even modern instruction-tuned LLMs **stumble** on these contrasts → fluency ≠ correct sense selection.
- We **adopt this evaluation style** (EN + one L2) and then test if a brief sense hint flips errors without any fine-tuning.

"IdioTS: A Hard Idiom Test Suite for LLMs." arXiv preprint. https://arxiv.org/pdf/2409.01053
"CoAM: Corpus of all type multiword expressions" [2412.18151] CoAM: Corpus of All-Type Multiword Expressions

Theme - Idiomatic translation & knowledge-base support for idioms

IdiomKB: Retrieval Helps Idiomatic Translation

- **Multilingual idiom knowledge base** (glosses, examples) used at inference to **remind** models of figurative meanings before translating.
- Yields **fewer literal mistranslations**, especially for smaller models, when a short description of the intended sense is included.
- We generalize this into a compact sense card (gloss + micro-example) from **WordNet/BabelNet/IdiomKB** and test it not only for **translation** but also **detection** (and optional paraphrase).

- "IdiomKB: A Knowledge Base for Idioms." arXiv preprint (v2). https://arxiv.org/html/2308.13961v2
- Improving LLM Abilities in Idiomatic Translation - ACL Anthology

Theme - Multilingual figurative language understanding (idioms, similes) in LLMs

Idiom↔Idiom Alignment Approaches

- Rather than explaining a sense, these methods **find an equivalent idiom** in the target language (via embeddings/LLMs) to preserve **style and tone**.
- Works well when a good target idiom **exists**, but offers less help when it **doesn't**, and doesn't solve **idiom vs literal** detection.
- Our approach is **complementary**: a **sense card** still guides the model when no clean target idiom exists, and it directly supports **the first decision**, is the usage idiomatic?

"Idiomatic Translation via Idiom-to-Idiom Alignment." arXiv preprint. https://arxiv.org/html/2405.10579v1
Comparative Study of Multilingual Idioms and Similes in Large Language Models - ACL Anthology

Method & Evaluation

Baselines:

- BERT idiom detector (EN only).
- LLM zero-shot prompts (detect / translate).

Our variant:

- LLM + sense card.

Metrics:

- Detection: accuracy/F1 on held-out pairs.
- Translation (if included): count literal errors; optionally COMETKiwi.
- Ablations: with/without card; small k (1-3); KB source.

Plan & Deliverables

Phases (sequence only)

- **Setup:** choose L2; finalize minimal-pair template; add WordNet lookup.
- **Baselines:** run BERT (EN) and LLM zero-shot; save outputs.
- **Sense Card:** add retrieval → ≤30-token card; run with/without; log coverage.
- **Wrap-up:** tables + a few examples; note limits; finalize slides/report.

Deliverables

- **Mini dataset** (EN + L2 minimal pairs)
- **Scripts** (baselines + sense-card inference)
- **Brief report & slides** (method + observations)

Commonalities

All the related works highlight the same pain point: LLMs are fluent but **semantically shallow** with idioms.

They show that **adding meaning cues**, whether through fine-tuning, retrieval, or data augmentation helps models disambiguate idioms.

Our project continues that trend but explores a **smaller, more practical intervention**: inference-time sense hints that don’t require retraining or new infrastructure.

The gap we are trying to fill

Most past research either focuses on **building new datasets** or **changing the model itself** through fine-tuning.

But there hasn’t been much work on testing whether a **simple meaning hint at inference time**, like our sense card can actually help models understand idioms better.

That’s where *IdiomSense* sits, between detection and translation, combining both into a minimal, model-agnostic framework that could make LLMs more semantically aware using just retrieved sense cards.

Bibliography

1. Khoshtab, Paria, Danial Namazifard, Mostafa Masoudi, Ali Akhgary, Samin Mahdizadeh Sani, and Yadollah Yaghoobzadeh. “Comparative Study of Multilingual Idioms and Similes in Large Language Models.” *Proceedings of the 31st International Conference on Computational Linguistics (COLING)*, Jan. 2025, Abu Dhabi, UAE. Association for Computational Linguistics, pp. 8680–8698.
2. De Luca Fornaciari, Francesca, Begoña Altuna, Itziar González-Dios, and Maite Melero. “A Hard Nut to Crack: Idiom Detection with Conversational Large Language Models.” *Proceedings of the 4th Workshop on Figurative Language Processing (FigLang 2024)*, 21 June 2024, Mexico City (Hybrid). Association for Computational Linguistics, 2024, pp. 35–44.
3. Donthi, Sundesh, Maximilian Spencer, Om B. Patel, Joon Young Doh, Eid Rodan, Kevin Zhu, and Sean O’Brien. “Improving LLM Abilities in Idiomatic Translation.” *Proceedings of the First Workshop on Language Models for Low-Resource Languages (LoResLM)*, Jan. 2025, Abu Dhabi, UAE. Association for Computational Linguistics, pp. 175–181.
4. Li, Shuang, Jiangjie Chen, Siyu Yuan, Xinyi Wu, Hao Yang, Shimin Tao, and Yanghua Xiao. “Translate Meanings, Not Just Words: IdiomKB’s Role in Optimizing Idiomatic Translation with Language Models.” *AAAI*, 2024. (Alternatively: Li, Shuang, et al. “Translate Meanings, Not Just Words: IdiomKB’s Role in Optimizing Idiomatic Translation with Language Models.” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 17, Feb. 2024, pp. …) Note: adjust depending on exact publication details.
5. Ide, Yusuke, Joshua Tanner, Adam Nohejl, Jacob Hoffman, Justin Vasselli, Hidetaka Kamigaito, and Taro Watanabe. *“CoAM: Corpus of All-Type Multiword Expressions.”* arXiv preprint, 24 Dec. 2024, arXiv:2412.18151.
6. Yayavaram, Arnav, Siddharth Yayavaram, Prajna Devi Upadhyay, and Apurba Das. *“BERT-based Idiom Identification using Language Translation and Word Cohesion.”* Proceedings of the Joint Workshop on Multiword Expressions and Universal Dependencies (MWE-UD) @ LREC-COLING 2024, 25 May 2024, Torino, Italia. ELRA & ICCL, pp. 220–230.