# INT D 461: Artificial Intelligence Everywhere Capstone

# Information Interview

Donna Mathew

Prabal Mehra

Mohammad Shahriar Hossain

# Problem & Motivation

- Our project explores how **large language models (LLMs)** handle **figurative language** — idioms, proverbs, metaphors across different languages.
- Idioms/proverbs/metaphors are non-compositional (surface words ≠ meaning).
- LLMs and MT often choose the literal sense, producing confident but wrong outputs.
- LLMs often **hallucinate** or "agree" too easily — like a *yes-man*.
- They sometimes **explain nonsense confidently**, showing a lack of real understanding.
- Downstream impact: mistranslation, misunderstanding, cultural bias, and user trust issues.
- We want to test **how well they actually reason** about figurative meaning, not just output correct-looking translations.

# Potential Candidates

We brainstormed researchers in figurative language + NLP

**Dr. Greg Kondrak**

**Bradley Hauer**

**Ning Shi**

# Dr. Greg Kondrak

**Professor, Natural Language Processing & Computational Linguistics**
 **Field:** Sub-word NLP and Lexical Semantics
 **Focus Areas:**

- Letter-phoneme conversion, transliteration, morphology, and word similarity

- Cognate identification and applications in translation and diachronic linguistics

- Lexical semantics: synonymy, polysemy, and word sense disambiguation
 **In essence:** He develops algorithms that capture relationships between words and their forms across languages — foundational for translation, linguistic analysis, and decoding unknown scripts.

# Ning Shi

**Ph.D. Candidate, NLP & Semantics, supervised by Dr. Kondrak**
**Field:** Computational Lexical Semantics (CLS)
**Focus Areas:**

- Multilingual semantic modeling and lexical disambiguation

- Applying algorithmic approaches to improve cross-lingual understanding

- Bridging theoretical semantics with practical language applications
  **In essence:** He builds systems that understand word meaning across contexts and languages — crucial for tasks like translation and question answering.

# Bradley Hauer

**Postdoctoral Researcher, Natural Language Processing (NLP)**
**Field:** Computational Lexical Semantics (CLS)
**Focus Areas:**

- Word sense disambiguation and multilingual semantic relations

- Building and applying semantic knowledge bases (e.g., WordNet, BabelNet)

- Using translation and multilingual data for semantic understanding
  **In essence:** He studies how computers can distinguish and represent word meanings across contexts and languages — advancing translation, semantic search, and question answering.

# Our Goal

- Goal: understand **where LLMs still struggle and why they struggle** along with to understand what would make our project **meaningful in 2025**.

  Questions we brainstormed:

- Where do LLMs still fall short in idiom or figurative language understanding?

- If we built a lightweight system around LLMs, what would add real value?

- What would make a project on idioms stand out today? (e.g., evaluation, multilingual focus, interpretability)

# Key Takeaways from Bradley

- LLMs **still hallucinate** idiom meaning and can't always **justify** why.

- **No true explanation generator** — they can "explain" wrong answers too.

- **Sense retrieval** or external reference systems could add value.

- **Multilingual focus** is promising — current models still **fall short for non-European languages**.

- Average-case performance is good, but **hard cases reveal real weaknesses**.

# Other interesting topics in the Interview

- Retrieval + dictionary definitions → improves understanding.

- BubbleNet-style models still **struggle with non-European idioms**.

- Word sense disambiguation remains **important**.

- Multilingual idiom examples (e.g., Chinese) show **missing metaphors**.

- Small, open-weight models → good for **interpretability** and **testing logic**.
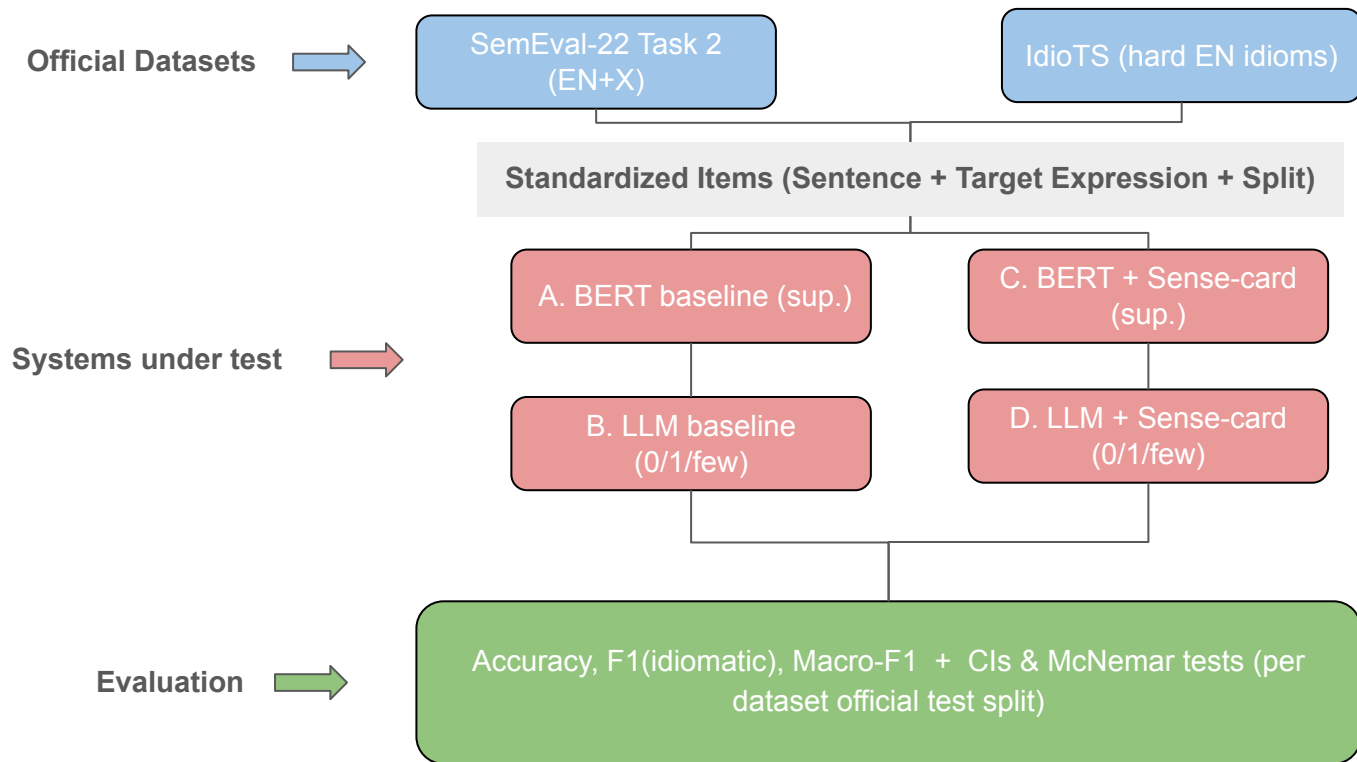
# What We Learned

- Strong **evaluation design** is key — not just accuracy.

- Need to test **reasoning and explanations** behind outputs.

- Multilingual and **word-sense-aware evaluation** could stand out as novel research.

- "Open-weight" models let us inspect the model's logic and are better for experimentation.

# How we are using this information in our project

- Add **prompt engineering** (sense card) to disambiguate the sentence.

- Test **multilingual idiom understanding**.

- Explore **open-weight models** for transparency.

# A flow chart for better understanding

**A big Thanks to Bradley Hauer for taking the time out of his day to talk with us**

# Feedback / Questions?