Open Knowledge Extraction from Dialogue Using In-Context Learning⋆

Kelsey Rook[1,*]

[1]*Rensselaer Polytechnic Institute, 110 8th Street, Troy NY*

Abstract

While Open Information Extraction and Knowledge Graph construction have become viable tasks over formal texts such as research papers and news articles, natural human conversation remains a poorly suited, and under-explored target for knowledge extraction. Dialogue presents unique challenges for information extraction: information is often distributed across multiple dialogue turns and perspectives, and conversational acts such as disagreement, quotation, and hedging complicate the identification of factual assertions. Despite the increasing availability of conversational data, dialogue remains an underutilized source of structured knowledge. In my dissertation research, I aim to investigate how the structural and functional features of dialogue can be leveraged to improve Open Knowledge Extraction using large language models. I propose four core contributions: (1) The formalization of the task of Open Knowledge Extraction from Dialogue, (2) the creation of a dataset towards this task, (3) a perspective-aware ontology of dialogue, and (4) a methodology of in-context learning for Open Knowledge Extraction from Dialogue. I aim to evaluate the performance of my approach against current approaches such as fine-tuning, and demonstrate the utility of dialogue-aligned knowledge graphs and the dialogue ontology on downstream tasks involving machine understanding of human conversation.

Keywords

Knowledge graph construction, In-context learning, Ontology design, Dialogue

1. Introduction

The automated extraction of structured knowledge from natural language is a foundational goal of the knowledge graph (KG) research community, yet structured linked data from dialogue remains underexplored. Dialogue introduces unique challenges due to its unique linguistic, pragmatic, and semantic properties. To address the task of knowledge graph construction from dialogue, I aim to contribute the following:

1. A formal definition of the task of Open Knowledge Extraction from Dialogue
2. A high-quality dataset towards knowledge extraction from dialogue, particularly for the in-context learning setting
3. An ontology to support the modeling of dialogue, as well as metadata and expressed assertions
4. A methodology for knowledge graph extraction from dialogue using a few-shot in-context learning (ICL) strategy to condition large language models (LLMs) on dynamically retrieved examples containing relevant dialogue features.

1.1. Knowledge graphs

Formally, a knowledge graph G can be defined as follows:

1. Let E be a set of concepts and P be a set of predicates
2. Let L be the set of literal values
3. Let T be a set of factual triples, of the form (s, p, o), where $s \in E$, $p \in P$, and $o \in E \cup L$

This triple-based structure aligns with the Resource Description Framework (RDF) model. Knowledge graph construction (KGC) is the task of automatically or semi-automatically mapping a data source onto a knowledge graph structure.

1.2. Dialogue

Understanding how humans communicate through dialogue is essential for developing systems capable of extracting meaning from conversational data. Unlike monologic or expository text, dialogue is multi-party, and typically unfolds in real-time social settings. These characteristics introduce unique challenges for semantic modeling and knowledge extraction, and are the subject of study across linguistic theory. We discuss some features of dialogue that set it apart significantly from formally written text, and that are well-discussed in linguistics and natural language understanding (NLI).

- **Cross-Turn Assertions:** Many facts are implicit, or constructed across multiple dialogue turns and speakers. Yu et. al. [1] demonstrate that the distance between the subject and object in factual triples is higher on average in dialogue data than in textual documents.
- **Anaphora and pronoun density:** Dialogue relies heavily on context-dependent references, making coreference resolution especially important in dialogue-based task settings.
- **Dialogue acts:** Not all utterances assert facts. Dialogue acts are defined as atomic units of conversation that performs specific communicative functions, such as turn management, discourse structuring, and directives [2]. More crucially to dialogue modeling, statements may be of fact, opinion, or suggestion.
- **Perspectives:** Unlike most formal documents, dialogue inherently reflects multiple perspectives. Participants in conversation may express conflicting views, or revise their beliefs over the course of time.

Particularly, dialogue is highly dependent on the context of the conversation; when compared to formal writing, pragmatics must be more largely considered in order to understand the contents of a conversation. These characteristics justify the need for dedicated approaches to dialogue-based knowledge graph construction, as most current methods are concerned with semantics and syntax [3], but not pragmatics.

1.3. Research questions

In this dissertation project, I aim to explore the following questions:

- **RQ1:** What are the limitations of existing dialogue corpora for supporting knowledge extraction tasks, and how can a new dataset better capture the diversity and complexity of real-world conversation?
- **RQ2:** What types of discourse features are necessary to capture for enabling accurate representation of knowledge in conversational contexts?
- **RQ3:** How effective is in-context learning (ICL) with dynamically retrieved examples for relation and knowledge graph extraction from dialogue, compared to traditional fine-tuning and rule-based methods?
- **RQ4:** Can retrieval-based ICL strategies, typically applied at the sentence level, be successfully adapted to dialogue data where relationships span multiple utterances and speakers, and pragmatics must be considered?

2. Related work

2.1. Structured representations of dialogue

While semantics deals with the literal, context-independent meaning of language, pragmatics relates to who is speaking, how it relates to what was said previously, and the communicative goals of

the interlocutors. Language understanding systems, particularly for dialogue, are severely limited if they can't make use of this contextual information to compute meaning [4]. Pragmatics considers phenomena such as dialogue acts, nonverbal communication, and implicature, and while out of scope for this research, aspects such as nonverbal communication and tone of voice are also important context that may augment the meaning of an utterance. Formally modeling the structure, semantic and pragmatic, of dialogue is central to fields such as discourse analysis, dialogue system management, and knowledge representation. Several taxonomies, ontologies, and markup languages exist for various dimensions of dialogue. A foundational contribution is the DIT++ taxonomy [5], a framework for analysis and annotation of dialogue acts. It includes a multidimensional taxonomy of dialogue acts, as well as formal definitions of semantic and pragmatic relations that occur between them. Other dialogue act schema include DAMSL (Dialogue Act Markup in Several Layers), and SWBD-DAMSL, which was constructed from the original DAMSL tag set specifically for annotation of the Switchboard dialogue corpus [6].

Several ontologies and frameworks have been proposed to support dialogue management systems (DMS). The VOnDA (Versatile Ontology-based Dialogue Management Architecture) system combines a domain-independent ontology with symbolic reasoning about dialogue states, user intents, and discourse goals [7]. Teixeira et. al. propose an approach to automatically generate dialogue managers for the health domain by integrating a conversational ontology with AI planning [8]. OntoVPA, a commercial system, integrates ontologies based on speech act theory with reasoning and ontology-based rules for response generation [9]. While the models discussed provide a rich basis for dialogue annotation and dialogue system design, to our knowledge none exist that support association of knowledge with its provenance in human dialogue.

2.2. In-Context learning with large language models

Few-shot in-context learning (ICL) is a prompting strategy by which an LLM is conditioned on a few input examples to perform a new task. ICL presents the advantages of not having to pre-train and fine-tune on large amounts of data. This is useful in scenarios in which relevant large datasets aren't available, or the prerequisite hardware requirements aren't met. Early research has showed success on fixed sets of examples [10] and randomly selected examples [11], and further improvements have been made by using retrieval-based ICL in which examples are selected dynamically based on the input query. ICL with dynamic retrieval has been applied to the relation extraction task on the sentence level [12, 13], but to our knowledge it has not been applied to document-level relation extraction, dialogue relation extraction, or knowledge graph extraction. Additionally, previous methods of dynamic sampling have focused on semantic and syntactic features.

2.3. Knowledge graph construction from dialogue

Knowledge graph construction traditionally builds on Open Information Extraction (OpenIE) pipelines that perform tasks such as Named Entity Recognition (NER), Relation Extraction (RE), and entity linking. Much prior work in this area has focused on sentence-level RE, with more recent work focusing on document-level RE, which involves reasoning across sentence boundaries to infer relations.

Yu et. al. [1] have formalized the Relation Extraction from Dialogue task and generated a dataset from scripts of the sitcom *Friends*. Each dialogue is hand-annotated with (s, p, o) triples with p from 36 social relation types, as well as the minimal span of utterances the triple occurs over. They investigated several RE methods using this dataset, and found that a speaker-aware extension of BERT performed better than the base model. A similar dataset, the CRECIL Corpus [14], is annotated with character relationship triples from the Chinese-language sitcom *I Love My Family*.

Relation extraction and other OpenIE tasks have been well-explored on the document level, to some success in supporting downstream knowledge graph generation. However, as demonstrated in [15], the results of traditional OpenIE are not well-tailored to high-quality linked data generation. Entities from OpenIE are more likely to be noun phrases that can't be directly matched to other phrases describing the same entities, resulting in poorly linked data that doesn't facilitate reuse. Additionally many OpenIE

methods restrict results to a small number of predicates pertaining to limited relationship types. We hypothesize that these limitations also apply to knowledge graph construction from dialogue, and may be amplified by the distinct features of dialogue.

3. Methodology

3.1. Open knowledge extraction from dialogue

As part of my dissertation work, I propose the formalization of the Open Knowledge Extraction from Dialogue (OKE-D). OKE-D extends Open Knowledge Extraction (OKE) as formulated in [15], in which knowledge graph construction is explored as a similar but distinct task from traditional OpenIE. Our preliminary formalization of the OKE-D task is as follows:

Given a dialogue D consisting of a sequence of participant-annotated utterances:

$D = [(u_1, s_1,), (u_2, s_2), \ldots, (u_n, s_n)]$ where u_i is an utterance from speaker s_i, and the utterance u_i directly follows utterance u_{i-1}, Open Knowledge Extraction from Dialogue aims to extract a set of linked knowledge graph triples. The formalization of OKE-D may increase the feasibility of dialogue as a source of knowledge, and can be applied in many domains given the large extent to which human-to-human conversation is used as a source of information exchange.

Initial experiments on the OKE task have used prompt engineering alongside a naive entity linking approach (LOKE-GPT) [15] on the TekGen dataset [16], achieving significant improvement over previous methods. Due to the lack of relevant dataset, which we discuss in Section 3.2, there is no large-scale evaluation of similar techniques on dialogue data. However, the snippet of dialogue in Table 3.1, among other samples of dialogue from the AMI meeting corpus [17], illustrate why the task of D-OKE presents unique challenges. Using a simple zero-shot prompt adapted from LOKE-GPT, Google's Gemini 2.5 Pro, a frontier model at time of publication, produced the linked data in Figure 1. This dialogue is fairly simple, involving a discussion between two participants regarding the engineering of a data browser, with fairly high agreement. It contains coreference, with "it" referring to the development process and the data being browsed multiple times. Most of the extracted triples are about the topic of discussion: the data browser itself. While coreference resolution is handled well, and the triples mostly present accurate information about the system being developed, the type of data desired to be extracted from dialogue is likely to be different than data extracted from formal writing. Many of the extracted triples would not be interesting to stakeholders; rather, the useful information in this snippet emerges only with added context: who thinks that relevant data should be stored in the database, and who thinks that classes should be stored in the database? Is <search *uses* data> currently true, or a suggestion for the future?
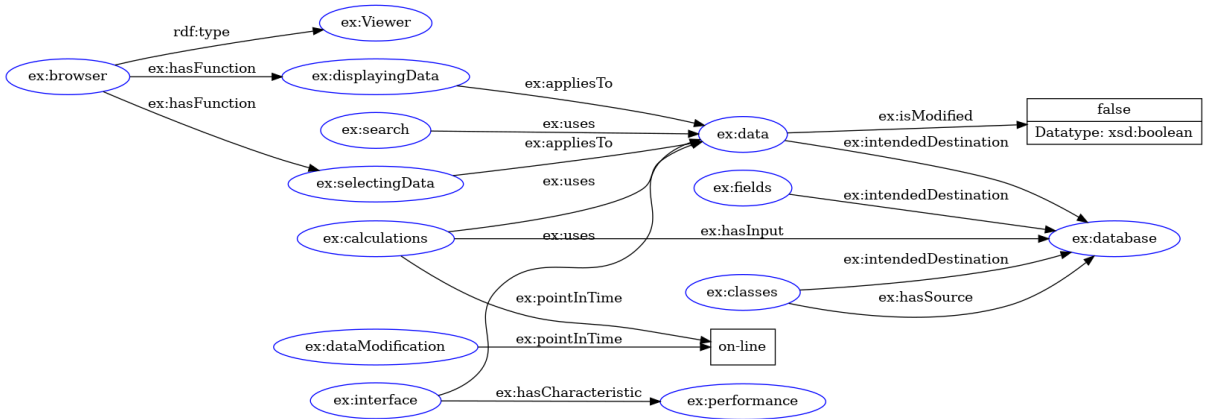

Figure 1: RDF graph generated from the conversation in Table 3.1

S1	Um, but that's still sort of that's good. That means that at least like we don't have the type of situation where somebody has to do like a billion calculations on, on data on-line, 'Cause that would make it a lot more like that would mean that our interface for the data would have to be a lot more careful about how it performs and and everything And nobody is modifying that data at at on-line time at all, it seems nobody's making any changes to the actual data on-line
S2	Don't think so.
S1	So that's actually making it a lot easier. That basically means our browser really is a viewer
S2	Yeah
S1	Mostly which isn't doing much with the data except for sort of selecting a piece piece of it and and displaying it
S2	Are we still gonna go for dumping it into a database?
S1	Hmm
S2	Are we still gonna dump it into a database?
S1	Well some parts relevant for the search yes, I'd say so
S2	'Cause if we are I reckon we should all read our classes out of the database It'll be so much easier
S1	Hmm
S2	Well if we're gonna dump the part of it into a database anyway we might as well dump all the fields we want, into the database calculate everything from there

Table 1
A snippet of a meeting in the AMI Corpus, in which two attendees

3.2. Dataset towards OKE-D

Recent research in discourse analysis has demonstrated the utility of sitcom and movie scripts in dialogue research through linguistic feature analysis [18, 19], revealing remarkable similarities with natural conversation in co-occurrence of specific features such as pronoun types, verb types, and contractions. However, research by Quaglio et. al. [18] notes that the standard deviations for the occurrence of these features are much less than that of face-to-face dialogue corpora, likely due to the limited range of settings and conversation topics in *Friends*. Quaglio et. al. also performed a functional analysis, exposing features that largely differ: face-to-face conversation contains more vague language, with a higher occurrence of hedges ("sort of", "kind of"), coordination tags ("and stuff like that), and the discourse marker "you know". Pilan et. al. observe that scripts from the OpenSubtitles have a significantly lower frequency of communicative feedback including backchannels, acknowledgments, and clarification requests, and Bednarek et. al. [20] observe that dialogue in *Gilmore Girls* and ten other television comedies have a higher occurrence of emotional markers than everyday conversation according to frequency analysis. To mitigate these discrepancies, the creation of a dataset in which these features are represented is well-motivated.

Large-scale dialogue corpora exist across a range of domains and interaction types, including task-oriented, domain-specific, and multimodal conversations. These datasets originate from diverse settings such as customer service interactions, business meetings, and casual chats. Many include annotations such as dialogue acts, sentiment, extractive and abstractive summaries, named entities, and topics. I propose to reuse one or more of these previously existing corpora, selecting among those that are publicly available, well-studied, and already annotated for various linguistic features. I have identified the following candidate corpora:

- **AMI Meeting Corpus:** 100 hours of multi-party meetings; two-thirds elicited using a scenario in which the participants play different roles in a design team, and the remaining from naturally occurring meetings in a range of domains
- **Switchboard Corpus:** over 2,400 two-person phone conversations seeded from about seventy different topics [6]
- **Santa Barbara Corpus of Spoken American English:** Naturally occurring spoken interactions in various settings, predominantly face-to-face conversation; 60 dialogues are freely available

The AMI Meeting Corpus and Switchboard Corpus both include common annotations such as dialogue

acts, topics, and summaries. Dialogue acts in particular are of interest, since they may provide valuable context for properly extracting triples from dialogue. Because this research targets the in-context learning (ICL) setting, in which a small number of examples are selected to guide an LLM for each input, the quality of this dataset should be prioritized over its size; conversely, it needs to be large and diverse enough to represent a large selection of the dialogue features discussed previously. This contribution will include the resulting hand-annotated dataset, as well as a handbook with formal guidelines for labeling entities (individuals and literals) and relations based on quality knowledge graph standards.

3.2.1. Hypotheses

Scripted dialogue exhibits a narrower range of discourse types and linguistic variation than real-world dialogues. Dialogue corpora that includes diverse, naturally elicited conversation provide a stronger foundation for generating high-quality knowledge extraction datasets, and training models for understanding of real-life conversation.

3.3. Perspective-aware dialogue ontology

To support the downstream usage of knowledge extracted from dialogue, the creation of a framework capable of capturing the context and provenance of this knowledge is necessary. Unlike document-based texts, which typically reflect a single, coherent perspective, dialogue is inherently multi-speaker, temporal, and likely to contain subjectivity. As a result, knowledge derived from dialogue will frequently reflect differing viewpoints, evolving beliefs, and varying levels of uncertainty. Without the added context of who made a statement, and when, this knowledge is useless. For example, if two speakers assert contradictory claims, or if one speaker expresses conflicting statements over time, a flat representation of these assertions in a knowledge graph would introduce ambiguity. To address this, I propose the development of an ontology designed to represent dialogue structure and related assertions, with regards to speaker identity and temporal context. A preliminary conceptual diagram of this ontology can be found in Figure 2. Additionally, an example of utterances from Table 3.1, aligned with the ontology's structure, can be seen in Figure 3.
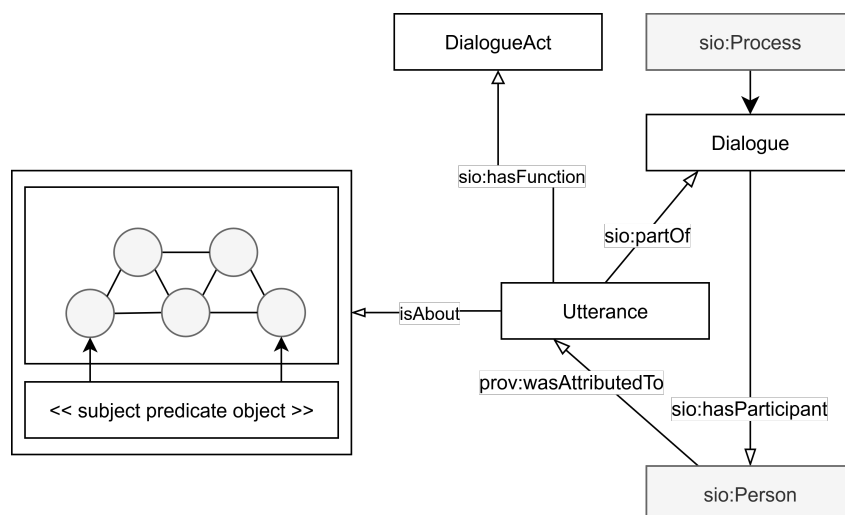

Figure 2: Preliminary Conceptual Diagram. Solid arrows indicate a superclass/subclass relationship. "sio" is the namespace for the Semanticscience Integrated Ontology. "prov" is the namespace for PROV-O.

3.3.1. Hypotheses

An ontology that models speaker identity and utterances will allow structured knowledge extracted from dialogue to be meaningfully contextualized and disambiguated. In addition, alignment of assertions

with speakers and utterances, along with the integration of domain ontologies, can allow for complex reasoning and querying that facilitates many downstream tasks such as dialogue summarization and question answering. Dialogue acts, which are decided by both semantic and pragmatic attributes of an utterance, are also represented as important context for utterances and the associated triples.

3.4. In-context learning for OKE-D

I propose dynamically retrieving relevant examples from our dialogue dataset to use as in-context examples, as seen in recent literature [21]. Rather than fine-tuning or manually annotating a corpus, large language models (LLMs) can be used with in-context learning (ICL), enhanced through semantic similarity-based retrieval of example dialogues and corresponding triples. Compared to fine-tuning strategies, ICL is less resource-intensive, more flexible, and requires less data.

Facilitating this dynamic retrieval requires the segmentation of the training corpus into chunks of n utterances with a sliding window approach to account for cross-sentence relations. Values of n should be determined based on the distribution of subject-object span distances in the training data, and chosen to balance vector database size with the number of triples represented. Lower values of n results in triples with higher distances between subject and object to be thrown out. We obtain embeddings of these segments and apply clustering, and index a sample from each cluster in a vector database in order to obtain a small representative selection of examples from our training dataset. At inference time, the test dialogue is encoded and compared to the index data so that the most relevant examples can be retrieved and used in few-shot prompts.

3.4.1. Hypotheses

Using a semantic search to retrieve similar examples will result in few-shot prompts that reflect relevant dialogue features, allowing LLMs to account for linguistic structures that it would not otherwise be capable of handling. Additionally, clustering the embeddings of training data should reduce computational load without significantly reducing overall performance. Finally, we hypothesize that ICL will provide an efficient alternative to fine-tuning for the OKE-D task, alleviating the need for large-scale datasets.

4. Future Work

4.1. Dataset

The next step in generating a dataset for D-OKE is to identify which combination of one or more candidate dialogue corpora align most closely with our goals of representing diverse dialogue settings, goals, and phenomena. The following step is to create formal guidelines for annotation to serve as documentation and provenance for the generated dataset, as well as to guide the generation of future annotations, should there be a need for a larger volume of data, or annotated dialogues from scenarios not represented in our dataset.

Preprocessing should involve removing content that describes non-verbal information, and standardizing the available orthographic transcriptions to a single format if multiple datasets are used. We then focus on the annotation of relational triples, indicating whether the subject and object entities are individuals or literals. Where applicable, the inverse triple should also be annotated. In addition, entity mentions will be linked to their corresponding Wikidata items, and the minimal supporting span of utterances that justifies each triple will be noted. The annotation guidelines may be refined as needed throughout the process based on annotator feedback or observed edge cases. It would be preferable to involve at least two annotators, to ensure the quality of annotated triples through consensus.

Additionally, I intend to consider a semi-automated annotation strategy using distant supervision. This approach involves aligning dialogue corpora with the Wikidata knowledge graph under the distant supervision assumption [22], resulting in candidate triples to be reviewed by human annotators. For

quality assurance, I intend to adapt procedures from well-established data annotation guidelines such as those from the Automatic Content Extraction (ACE) [23] program.

While the resulting hand-annotated dataset will be relatively small, I anticipate that the distant supervision pipeline could be applied at scale to generate larger, weakly labeled datasets, which would be useful for pretraining or distant supervision in further knowledge extraction experiments.

4.2. Ontology

The next step in realizing an ontology for dialogue knowledge representation is generating use case scenarios and competency questions. The development of this ontology will continue in alignment with upper-level ontologies such as the Semanticscience Integrated Ontology (SIO) [24], and the PROV Ontology (PROV-O) [25] for provenance information.
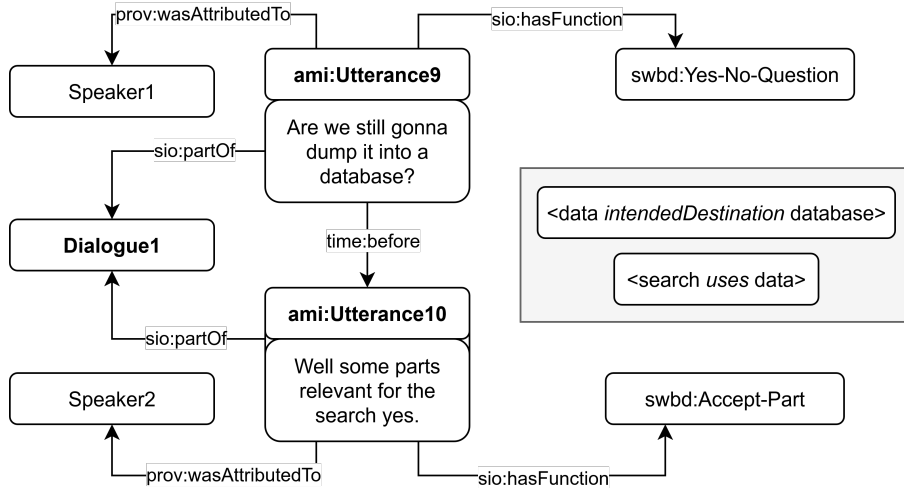

Figure 3: Example of dialogue alignment with ontology. "swbd" is the namespace of the tags in the SWBD-DAMSL dialogue act tag set. "ami" is the namespace for the AMI Meeting Corpus.

4.3. Evaluation

The ontology developed for representing dialogue structure and assertions will be evaluated using established methods for ontology quality and utility assessment. This includes checking basic ontology metrics, and utilizing checkers such as OOPS! (OntOlogy Pitfall Scanner!) [26] to ensure that the ontology doesn't contain any common pitfalls. The core evaluation will center on application of the use case and set of competency questions to evaluate the ontology's coverage of its intended application.

To evaluate the effectiveness of the ICL-based extraction method, I intend to compare its performance to several baselines. These include language models fine-tuned on large-scale, distantly supervised corpora, knowledge graph construction methods based on traditional OpenIE, and prior work on relation extraction from dialogue. Each approach should be assessed using an evaluation set drawn from the proposed dialogue dataset. I plan to adapt the conversational precision and recall metrics as proposed in [1].

To explore the impact of input representation on retrieval quality and OKE-D performance, I propose testing multiple embedding strategies. These include standard BERT-based embeddings, a speaker-aware BERT embedding in which special tokens are added to the input sequence to indicate speaker turn boundaries, and other potential modifications to the BERT input sequence that incorporate dialogue act tags or other dialogue features.

In addition to aggregate metrics, I intend to conduct a fine-grained error analysis by manually inspecting a random sampling of results. This analysis will help identify whether an ICL-based approach is more or less robust to dialogue-specific challenges. Finally, I will demonstrate how extracted triples

produced by the best-performing configuration can be aligned with the dialogue ontology, validating its utility as a schema for organizing and reasoning over structured knowledge derived from conversations.

4.4. Applications

The resources and methods developed in this research are intended to support applications that require understanding of natural, unscripted conversation. We envision that our work will be applicable to dialogue-centric systems to support tasks such as meeting summarization, debate modeling, question answering (QA), and general machine understanding of conversational scenarios. We also envision our contributions being applicable to mixed documents such as literature, in which text is split between an overhead narrator perspective and character dialogue.

Our research primarily targets non-scripted, casual human-to-human conversation, as it exhibits the complex features of dialogue that are absent or less pronounced in typed or written interactions. Consequently, our work is less directly applicable to human-computer interactions, which are usually text-based and more constrained in structure. As a result, evaluation on human-computer dialogue is not a near focus, but we anticipate this line of research becoming increasingly relevant as spoken dialogue systems advance, particularly in their ability to interpret and generate speech that mirrors human conversation in characteristics addressed in our research.

5. Acknowledgment

Declaration on Generative AI

During the preparation of this work, the author(s) used Gemini 2.5 Flash for: Grammar and spelling check. After using this service, the author reviewed and edited the content as needed and takes full responsibility for the publication's content.

References

[1] D. Yu, K. Sun, C. Cardie, D. Yu, Dialogue-based relation extraction, arXiv preprint arXiv:2004.08056 (2020).

[2] A. Stolcke, K. Ries, N. Coccaro, E. Shriberg, R. Bates, D. Jurafsky, P. Taylor, R. Martin, C. V. Ess-Dykema, M. Meteer, Dialogue act modeling for automatic tagging and recognition of conversational speech, Computational linguistics 26 (2000) 339–373.

[3] M. Hofer, D. Obraczka, A. Saeedi, H. Köpcke, E. Rahm, Construction of knowledge graphs: Current state and challenges, Information 15 (2024) 509.

[4] H. Bunt, Dialogue pragmatics and context specification, in: Abduction, belief and context in dialogue: studies in computational pragmatics, John Benjamins Publishing Company, 2011, pp. 81–149.

[5] H. Bunt, The dit++ taxonomy for functional dialogue markup, in: AAMAS 2009 Workshop, Towards a Standard Markup Language for Embodied Dialogue Acts, 2009, pp. 13–24.

[6] J. J. Godfrey, E. C. Holliman, J. McDaniel, Switchboard: Telephone speech corpus for research and development, in: Acoustics, speech, and signal processing, ieee international conference on, volume 1, IEEE Computer Society, 1992, pp. 517–520.

[7] B. Kiefer, A. Welker, C. Biwer, Vonda: A framework for ontology-based dialogue management, arXiv preprint arXiv:1910.00340 (2019).

[8] M. S. Teixeira, V. Maran, M. Dragoni, The interplay of a conversational ontology and ai planning for health dialogue management, in: Proceedings of the 36th annual ACM symposium on applied computing, 2021, pp. 611–619.

[9] M. Wessel, G. Acharya, J. Carpenter, M. Yin, An ontology-based dialogue management system for virtual personal assistants, in: Proceedings of the International Workshop on Spoken Dialogue Systems Technology, Farmington, PA, USA, 2017, pp. 6–9.

[10] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al., Language models are few-shot learners, Advances in neural information processing systems 33 (2020) 1877–1901.

[11] D. Hendrycks, C. Burns, S. Kadavath, A. Arora, S. Basart, E. Tang, D. Song, J. Steinhardt, Measuring mathematical problem solving with the math dataset, arXiv preprint arXiv:2103.03874 (2021).

[12] Z. Wan, F. Cheng, Z. Mao, Q. Liu, H. Song, J. Li, S. Kurohashi, Gpt-re: In-context learning for relation extraction using large language models, arXiv preprint arXiv:2305.02105 (2023).

[13] G. Li, P. Wang, W. Ke, Y. Guo, K. Ji, Z. Shang, J. Liu, Z. Xu, Recall, retrieve and reason: towards better in-context relation extraction, arXiv preprint arXiv:2404.17809 (2024).

[14] Y. Jiang, Y. Xu, Y. Zhan, W. He, Y. Wang, Z. Xi, M. Wang, X. Li, Y. Li, Y. Yu, The crecil corpus: a new dataset for extraction of relations between characters in chinese multi-party dialogues (2022).

[15] J. McCusker, Loke: linked open knowledge extraction for automated knowledge graph construction, arXiv preprint arXiv:2311.09366 (2023).

[16] O. Agarwal, H. Ge, S. Shakeri, R. Al-Rfou, Knowledge graph based synthetic corpus generation for knowledge-enhanced language model pre-training, arXiv preprint arXiv:2010.12688 (2020).

[17] W. Kraaij, T. Hain, M. Lincoln, W. Post, The ami meeting corpus, in: Proc. International Conference on Methods and Techniques in Behavioral Research, 2005, pp. 1–4.

[18] P. Quaglio, Television dialogue and natural conversation: Linguistic similarities and functional differences, in: Corpora and discourse: The challenges of different settings, John Benjamins Publishing Company, 2008, pp. 189–210.

[19] I. Pilán, L. Prévot, H. Buschmeier, P. Lison, Conversational feedback in scripted versus spontaneous dialogues: A comparative analysis, arXiv preprint arXiv:2309.15656 (2023).

[20] M. Bednarek, The language of fictional television: A case study of the 'dramedy' gilmore girls, English Text Construction 4 (2011) 54–84.

[21] J. Liu, D. Shen, Y. Zhang, B. Dolan, L. Carin, W. Chen, What makes good in-context examples for gpt-3?, arXiv preprint arXiv:2101.06804 (2021).

[22] M. Mintz, S. Bills, R. Snow, D. Jurafsky, Distant supervision for relation extraction without labeled data, in: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, 2009, pp. 1003–1011.

[23] G. R. Doddington, A. Mitchell, M. A. Przybocki, L. A. Ramshaw, S. M. Strassel, R. M. Weischedel, The automatic content extraction (ace) program-tasks, data, and evaluation., in: Lrec, volume 2, Lisbon, 2004, pp. 837–840.

[24] M. Dumontier, C. J. Baker, J. Baran, A. Callahan, L. Chepelev, J. Cruz-Toledo, N. R. Del Rio, G. Duck, L. I. Furlong, N. Keath, et al., The semanticscience integrated ontology (sio) for biomedical research and knowledge discovery, Journal of biomedical semantics 5 (2014) 1–11.

[25] P. Groth, L. Moreau, Prov-overview, W3C Working Group Note 1135 (2013) 881–906.

[26] M. Poveda-Villalón, A. Gómez-Pérez, M. C. Suárez-Figueroa, Oops!(ontology pitfall scanner!): An on-line tool for ontology evaluation, International Journal on Semantic Web and Information Systems (IJSWIS) 10 (2014) 7–34.