

Clarifying Ambiguities: on the Role of Ambiguity Types in Prompting Methods for Clarification Generation

Anfu Tang*

Sorbonne Université, CNRS, ISIR
F-75005 Paris, France
tang@isir.upmc.fr

Laure Soulier

Sorbonne Université, CNRS, ISIR
F-75005 Paris, France
laure.soulier@isir.upmc.fr

Vincent Guigue

AgroParisTech, UMR MIA-PS
Palaiseau, France
vincent.guigue@agroparistech.fr

Abstract

In information retrieval (IR), providing appropriate clarifications to better understand users' information needs is crucial for building a proactive search-oriented dialogue system. Due to the strong in-context learning ability of large language models (LLMs), recent studies investigate prompting methods to generate clarifications using few-shot or Chain of Thought (CoT) prompts. However, vanilla CoT prompting does not distinguish the characteristics of different information needs, making it difficult to understand how LLMs resolve ambiguities in user queries. In this work, we focus on the concept of ambiguity for clarification, seeking to model and integrate ambiguities in the clarification process. To this end, we comprehensively study the impact of prompting schemes based on reasoning and ambiguity for clarification. The idea is to enhance the reasoning abilities of LLMs by limiting CoT to predict first ambiguity types that can be interpreted as instructions to clarify, then correspondingly generate clarifications. We name this new prompting scheme AMBIGUITY TYPE-CHAIN OF THOUGHT (AT-CoT). Experiments are conducted on various datasets containing humanannotated clarifying questions to compare AT-CoT with multiple baselines. We also perform user simulations to implicitly measure the quality of generated clarifications under various IR scenarios.

CCS Concepts

• Information systems → Information retrieval.

Keywords

Clarifying Question, Dialogue Search System, Ambiguity Type

ACM Reference Format:

Anfu Tang, Laure Soulier, and Vincent Guigue. 2025. Clarifying Ambiguities: on the Role of Ambiguity Types in Prompting Methods for Clarification Generation. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '25)*, July 13–18, 2025, Padua, Italy. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3726302.3729922>

*Corresponding author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
SIGIR '25, July 13–18, 2025, Padua, Italy

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-1592-1/2025/07
<https://doi.org/10.1145/3726302.3729922>

1 Introduction

Ambiguity in information retrieval (IR) is a common factor that could undermine the quality of the retrieved documents. Indeed, real-world users often provide ambiguous queries to initialize a search without further elaboration [6]. The reasons for this ambiguity can vary [6, 44], such as avoiding the effort of typing lengthy queries, uncertainty about information needs, the tip-of-the-tongue phenomenon [4], etc. The ambiguity of natural language itself could also account for this ambiguity in user queries, such as synonyms or polysemies [33]. Regardless of the different causes, ambiguity is essentially a form of uncertainty, i.e. we cannot discern users' real intents by a single query. To better understand the ambiguities underlying user queries, previous studies have investigated ambiguity types (ATs) and proposed different taxonomies [15, 30, 43] to classify ambiguities.

To navigate users through the ambiguity, we need a method that facilitates users in expressing their needs without compromising their user experience. Previous studies seek to achieve this goal by building proactive search-oriented dialogue systems [49], which can take the initiative to provide information or suggestions to help improve the quality of search results, including providing clarifications. Instead of passively receiving a list of documents, users can actively participate in the search by communicating with the proactive dialogue system through conversations. Early studies on clarifications focus on reformulated queries [12, 34], which seek to provide useful suggestions that may meet the user's need, without explicitly exploiting the user's intent. Recent studies focus more on asking clarifying questions [2, 50], which consists of providing a clarifying question and allowing the user to respond freely. Clarification generation methods have evolved with the development of large language model (LLM), from supervised methods that rely on human-annotated data [3, 30], to LLM prompting methods [56, 61], among which Chain of Thought (CoT) prompting [57] is found to generate better clarifying questions [24, 61] compared to prompts with no generated reasoning. However, previous work mostly uses CoT prompts to freely generate reasoning, without explicitly asking LLMs to distinguish different information needs. We argue that understanding and integrating ambiguities into reasoning is important for the clarification process, since humans may first categorize the scenario of ambiguities, and then decide how to clarify the query properly. To simulate how humans handle ambiguous queries, we seek to first analyze the concept of ambiguity from the perspective of ambiguity types, and then integrate them into reasoning for clarification generation. To achieve this, we combine ambiguity types with CoT prompting to build AMBIGUITY TYPE-CHAIN OF THOUGHT (AT-CoT), which prompts LLMs to predict Ambiguity

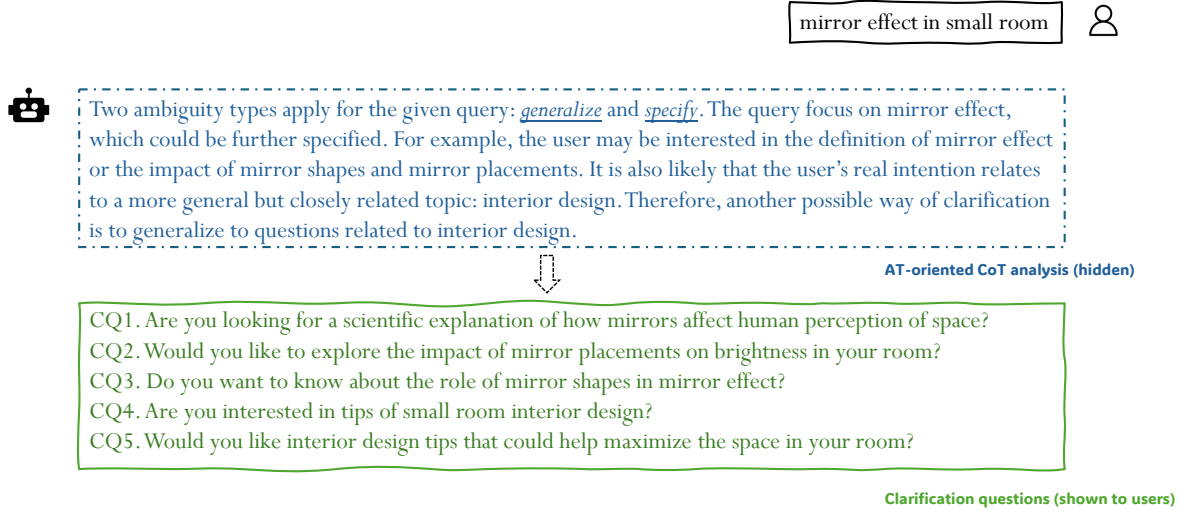


Figure 1: Illustration of AT-CoT: Unlike vanilla CoT, LLM-generated reasoning is limited to predict possible ambiguity types.

Types (ATs) that underlie a given query before generating clarifications correspondingly. To make AT-CoT properly work, we distill an action-based AT taxonomy from existing studies. Each AT in our taxonomy serves not only the purpose of helping LLMs understand ambiguity causes, but can also be interpreted as an instruction for LLMs to generate clarifications. Figure 1 illustrates AT-CoT. Provided a query *mirror effect in small room*, our method first predicts two ATs, then performs the corresponding actions to generate clarifying questions (CQs): CQ4 and CQ5 *generalize* the query; CQ1, CQ2, and CQ3 *specify* the query.

To validate the effectiveness of our method, experiments are carried out on both intrinsic and extrinsic tasks (resp. clarification generation and IR) on numerous datasets including Qulac [2], ClariQ [1], and TREC IR collections [15–18]. For the IR task, following previous work [2, 29, 63], we perform user simulation to generate multi-turn conversations and then transform the generated conversations to reformulated queries. We compare different clarification interaction scenarios such as proposing query reformulations for users to select (*select*) and asking a single clarifying question for users to respond (*respond*). To summarize, our main contribution is as follows:

- We analyze ambiguities from the perspective of ambiguity types, comprehensively investigate the impact of integrating ambiguities and reasoning in LLM prompting methods for clarification.
- We validate the effectiveness of our method through experiments on clarification generation and IR tasks.

2 Related Work

2.1 Ambiguity in User Queries

While there is a lack of a widely accepted taxonomy of ambiguities, ambiguous queries have been long studied in the IR community [15, 59, 61]. Previous work on ambiguity types (ATs) can be categorized into three types. The first group of studies formulates an AT taxonomy by analyzing queries in specific datasets [3, 15, 30, 43]. For

instance, Guo et al. [30] proposed a taxonomy based on ambiguous questions in Abg-CoQA [30] with four ambiguity types: *Coreference resolution* (unclear reference of pronouns), *Time-dependency* (the interpretation of question depends on time), *Answer types* (multiple answer possibilities) and *Event references* (an entity in the question corresponds to multiple events). However, taxonomies in these studies are proposed more for analytical purposes and contain very specific ATs (e.g. *EntityReferences* [43] and *CoreferenceResolution* [30] both correspond to a specific type of semantic ambiguity). Unlike these studies, we seek in this paper to formulate an AT taxonomy containing mutually exclusive ATs that can help LLMs generate better clarifications, rather than analyzing in detail why a query is ambiguous. Another group of studies focuses on the relations between queries by mining query logs [8, 31, 36, 59], mostly based on sampled query reformulations from query logs. Although these studies may not be directly related to clarification generation, their findings provide useful insights into clarification patterns. For instance, two common query reformulation patterns observed [8, 31] are *Generalization* and *Specialization*. While the latter is widely considered in studies related to clarification, the need for generalization is less investigated. We argue that generalization can also help specify the information needs of users in certain scenarios. It corresponds to an important dimension of ambiguity, reflecting the possibility that user queries may fail to accurately convey user intent. The last group consists of studies in the post-LLM era. We notice a recent work [61] that proposed a well-organized taxonomy with a special focus on ambiguities specific to LLMs, such as misaligned interpretations of queries between LLMs and humans. Our work differs from theirs: their taxonomy is used more as a tool to evaluate LLM performances in handling ambiguous queries, while our work focuses on integrating ambiguity types into reasoning for clarification generation. In a nutshell, previous work mostly exploit ATs for analysis, without searching to enhance the reasoning ability of LLM prompting by integrating ATs. To help better understand

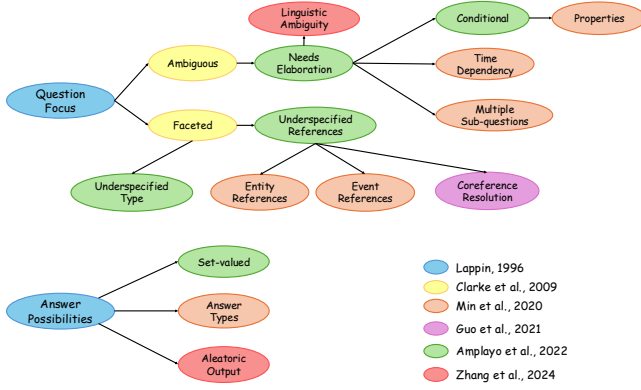


Figure 2: Ambiguity type taxonomies for analytical purposes in previous work.

previous work, we organize ATs in existing taxonomies and present them in Figure 2.

2.2 Clarification in Information Retrieval

Clarification serves the purpose of eliciting the user’s information need [59] by exchanging with the user and exploiting the user’s feedback. The clarification form could be diverse, either by proposing reformulated queries to diversify the retrieval results or by asking clarifying questions to induce users to clarify. Early attempts of clarification generation focus on automatic query expansion [12], whereby users’ original queries are rewritten or augmented. For example, Chirita et al. [13] proposed to expand user queries with terms collected from user data to handle ambiguities of short keyword queries in web searches. Other studies [11, 23, 42] investigate query suggestions by exploring different user-specific data sources such as landing pages, clicks, or hitting time. Both query expansion and query suggestions can be regarded as a form of proposing reformulated queries to users and collecting users’ feedback. Recent studies concentrate more on asking clarifying questions [2, 37, 50, 51, 58]. The common approach consists of using the conversation history as input to a generative language model to generate CQs.

Before the era of LLM, clarification generation methods mostly consist of training sequence-to-sequence neural models (e.g. seq2seq [53]) using labeled data. For example, Guo et al. [30] fine-tune BART [38] to generate clarifying questions with ambiguous questions provided as input; Xu et al. [58] investigated knowledge-based clarifying question generation and concatenated entity texts and the current question as input to a Seq2seq [5] model. Recent studies in the post-LLM era have increasingly focused on LLM prompting methods, such as using few-shot prompting [37, 61], Chain of Thought (CoT) prompting [24, 61]. Our work extends existing studies on LLM prompting methods for clarification generation by integrating ambiguity types into CoT reasoning.

2.3 Conversation Simulation in Search

In IR, user simulation consists of creating artificial conversations based on hypotheses about user behaviors, often used to automatically test the performance of dialogue systems without performing

real user tests [14, 26, 27]. The common approach is to instantiate a user agent to communicate with the dialogue system according to certain strategies [9, 28, 41]. Hypotheses about user behavior are made to control how user agents respond, depending on the purpose of the simulation. For example, to test the quality of reformulation in IR systems, Erbacher et al. [28] assumed that the user agent is greedy and fully cooperative, thus always selecting the reformulations most similar to the user intent. In another work [29], user agents are allowed to only respond ‘yes’ or ‘no’ to augment IR datasets with multi-turn conversations. Some studies involve simulation of more complex user behaviors, in which user agents are initialized with different variables, each corresponding to a specific type of user. Recent studies have increasingly focused on LLM-based conversation simulation. For example, Owoicho et al. [47] built a user simulator for mixed-initiative multi-turn conversation systems by prompting LLMs. Following previous work on LLM-based conversation simulation in IR, in our work, we instantiate our user agent using LLMs and simulate user responses by few-shot LLM prompting.

3 Methodology

We present here the outline of our methodology. We focus on the following research questions:

- RQ1.** What is an appropriate taxonomy of ambiguities for generating clarifying questions that is compatible with LLM prompting methods?
- RQ2.** How to integrate ambiguity and reasoning in LLM prompting methods for clarification?

3.1 Ambiguity Type Taxonomy

To respond to RQ1, we first seek to exploit ambiguity types to concretize the concept of ambiguity. The goal is to establish a taxonomy that can be used to enhance LLM reasoning ability in terms of handling ambiguous queries. Previous work [3, 30, 43] proposed various ambiguity-type taxonomies for the analytical purpose. However, from the perspective of helping LLMs understand ambiguities and better instructing LLMs to generate clarifications, we find existing taxonomies redundant and unsuitable for LLM prompting methods. Firstly, as evidenced by Zhang et al. [61], existing taxonomies were mostly proposed before the era of LLMs, some ATs lack clear definitions and ATs are not mutually exclusive. Secondly, ATs in existing taxonomies can be reduced to two actions that LLMs can take: *Determine the Query Interpretation* or *Further Specify the User Query*. Following Deng et al. [24] who proposed proactive prompting, i.e. making LLMs decide actions to take instead of simply responding to instructions, we propose an LLM action-based taxonomy that encompasses three dimensions, each corresponding to a clarification pattern discovered in previous work [15, 31, 59]:

- *Semantic*: accounts for ambiguity in query interpretations.
- *Generalize*: addresses ambiguity in information needs when users seek relevant yet more general information. It occurs when user queries do not precisely describe real user intents.
- *Specify*: addresses ambiguity in information needs when users seek more specific information. It occurs when user queries lack details and may correspond to a too large search scope. Most ATs in existing taxonomies can be categorized