Estimating Commonsense Knowledge from a Linguistic Analysis on Information Distribution

Sabrina Mennella[1,2], **Maria Di Maro**[2,3], **Martina Di Bratto**[2,3,4]
[1] University of Catania, Italy
[2] URBAN/ECO Research Center, Italy
[3] University of Naples "Federico II", Italy
[4] Logogramma s.r.l., Italy
sabrina.mennella@phd.unict.it,maria.dimaro2@unina.it,martina.dibratto@unina.it

Abstract

Commonsense Knowledge (CSK) is defined as a complex and multifaceted structure, encompassing a wide range of knowledge and reasoning generally acquired through everyday experiences. As CSK is often implicit in communication, it poses a challenge for AI systems to simulate human-like interaction. This work aims to deepen the CSK information structure from a linguistic perspective, starting from its organisation in conversations. To achieve this goal, we developed a three-level analysis model to extract more insights about this knowledge, focusing our attention on the second level. In particular, we aimed to extract the distribution of explicit actions and their execution order in the communicative flow. We built an annotation scheme based on FrameNet and applied it to a dialogical corpus on the culinary domain. Preliminary results indicate that certain frames occur earlier in the dialogues, while others occur towards the process's end. These findings contribute to the systematic nature of actions by establishing clear patterns and relationships between frames.

Keywords: Commonsense Knowledge, FrameNet, Semantic Annotation

1 Introduction

The development of high-quality Artificial Intelligence hinges on the critical challenge of equipping machines with Commonsense Knowledge (CSK) (McCarthy, 1959). This is essential for implementing systems that can elevate human-machine interaction to a more human-like level. The CSK was described as embodying the fundamental understanding of causal relationships, physical properties, social norms, and cultural references, crucial for effective communication and problem-solving in everyday situations (Cambria et al., 2009). Due to its multifaceted nature, CSK is generally taken for granted and is typically omitted in communication (written or oral) (Grice, 1975), except in cases of ambiguity or when the listener requires clarification (Nguyen et al., 2022). In this regard, the field of Knowledge Representation (KR) has made significant contributions to the acquisition and application of CSK, leading to the design and construction of resources containing such information (Lenat, 1995; Sap et al., 2019). Nevertheless, since it is impossible to represent all human knowledge in one single resource (Brooks, 1991), we posit that the most intriguing aspect of CSK lies in its consideration as a *process*, rather than a static collection of information. Modelling the processes underlying CSK directly from linguistic data presents a more significant challenge compared to representing predefined knowledge. In this work, we propose a three-level analysis model to investigate the structure of CSK. Specifically, we focus on the second level, which entails a semantic analysis grounded in Frame Semantics (Fillmore et al., 1976) and FrameNet (Baker et al., 1998) applied to a dialogic corpus on the culinary domain. By analysing the frames distribution within the dialogues, we aim to extract insights that can, in future work, contribute to a more comprehensive understanding of CSK structure.

The paper is organised as follows: Section 2 provides a general overview of the state of the art and Section 3 outlines the scope of the study. Section 4 discusses the motivation behind selecting the culinary domain and the development of the knowledge base, which underpins the three-level analysis model detailed in Section 5. Section 6 describes the annotation scheme applied to the CookDial dialogical corpus and the methodology used for retrieving frames distribution. Finally, Section 7 presents the results obtained, followed by the Conclusions and Future Work in Section 8.

2 State of the Art

The CSK has been described as embodying the fundamental understanding of the world shared among

individuals, including (i) information about events that occur over time, (ii) the consequences of one's own and others' actions, (iii) the characteristics of physical objects, (iv) their perceptions, (v) their properties, and (vi) its interrelationships (Cankaya and Moldovan, 2009). A universally shared academic definition of CSK does not currently exist. Nevertheless, (Zang et al., 2013) attempted to limit the scope of the investigation by identifying the most representative characteristics that provide a complete description of this type of knowledge, such as (i) sharedness, (ii) fundamentality, (iii) implicitness. From a linguistic perspective, these features appear interesting as they recall some aspects of the Communal Common Ground (CCG) (Clark, 2015), one of the four typologies of Common Ground (CG) (Stalnaker, 2002). Despite the CCG and the CSK involving shared understandings and assumptions, these are essentially distinct concepts: CCG implies a specific connection between an individual and other members of a shared community, emphasising the interaction between the interlocutors; in contrast, CSK concerns an individual's interaction with the world at large, often shared implicitly and unconsciously (Zang et al., 2013). CCG involves active agreement between speakers, establishing shared beliefs and defining a common language for group identities and boundaries (MacWhinney and O'Grady, 2015). On the other hand, CSK does not require explicit agreement, assuming that it is already universally shared among speakers (Zang et al., 2013).

The ongoing need for advancements in equipping AI with robust and adaptable CSK capabilities has provided a significant stimulus for research in KR, which has contributed to the development of large-scale CSK databases (Lenat, 1995; Liu and Singh, 2004; Sap et al., 2019). Although significant progress has been made in this regard (Zhou et al., 2021a,b; Majumder et al., 2020), limitations persist in their ability to capture the open-ended semiotic process, where significance is continuously crafted, contested, and renegotiated within shifting horizons of understanding (Süerdem, 2024).

3 Objectives

Given the vast amount of information that CSK encapsulates and the limitations of aforementioned state-of-the-art approaches, we rather frame it as a *process*. In this case, knowledge is understood as a process that generates structured relationships between actions and entities resulting from recurrent interactions stored in a database, and not as a mere repository of pre-existing facts. Indeed, it is more intriguing to analyse the processes by which this knowledge is formed rather than dwelling on its representation. Our goal is, therefore, to uncover the processes that comprise this knowledge, introducing a three-level analysis designed to extract more detailed information about CSK structure. For the scope of this work, the focus is on the second level, where we aim to identify the distribution of explicit semantic information within the communicative flow in the culinary domain. In future work, this analysis will facilitate the identification and schematisation of implicit information in a given domain. In particular, this will be possible by considering CSK as the result of the analysis of graph patterns and their probability.

4 Data sources

For our investigation, we took into account the culinary domain, guided by two main factors: (i) culinary practices are presumed to be highly familiar due to their everyday nature – most people routinely prepare meals; (ii) the domain exhibits strong action co-occurrences, as individual actions are linked (e.g., the action of *beating eggs* implies the action of *cracking eggs*). To facilitate the identification of the entities and actions involved in recipe instructions along with their relationships and co-occurrences, the initial step involves constructing our knowledge graph. We employed three main resources: the Recipe1M+ dataset (Marın et al., 2021), FlavorDB (Garg et al., 2018), and the Epic-Kitchens dataset (Damen et al., 2018), collectively representing the knowledge base of ingredients, recipe titles with instructions, food flavours, and information about daily activities performed in the kitchen that are not explicitly mentioned in recipe instructions (e.g., *take eggs - crack eggs - throw eggshell into bin*). The domain construction follows the methodology described in (Origlia et al., 2022), where multiple sources were integrated into Neo4J (Webber, 2012). This data organisation facilitates the cross-referencing of information, enabling the establishment of intricate relationships within the domain.

For carrying out the linguistic analysis, the CookDial dialogue corpus (Jiang et al., 2023) is employed. The corpus comprises 260 human-to-human English dialogues based on the culinary

domain, in which an agent, given a recipe document extracted from the RISeC corpus (Jiang et al., 2020), guides the user to cook a dish. Data were collected by applying the experimental *Wizard-of-Oz* method (Fraser and Gilbert, 1991), involving two participants interacting via a live chat platform. The application setup simulated the interaction between a voice assistant (agent) and a user. The agent had full access to the text of the recipe, while the user only knew its title. From this corpus was possible to identify actions relevant to the preparation of dishes, analysing their distribution within the dialogue flow.

5 Analysis Model

The foundation of a good communication is a set of regulative principles that facilitate its success, managing dialogue in accordance with logical and relevant criteria, as well as respecting the principle of cooperation between speakers. The maxim of *quality* (Grice, 1975) states that the contribution to the conversation should be as informative as is required. Therefore, a speaker is not expected to provide an excess or deficiency of information; rather, they will offer only the necessary information. Consequently, people typically *assume* a division of the knowledge they share (Whiting and Watts, 2024). Although some information is explicitly introduced into the discourse, some other is assumed and not explicitly discussed, agreed upon, or questioned (Amaral et al., 2011). Knowing what it can be presupposed and what must be made explicit, in other words, showing communicative competence (Hymes et al., 1972), still represents a challenge for conversational agents.

For this reason, we propose to classify this knowledge into three typologies: *Foreground knowledge*, *Background knowledge* and *Presupposed knowledge*. This classification allows us for a more structured approach to managing information, thereby facilitate a clearer understanding and more effective analysis of the data. We define foreground knowledge as information explicitly expressed in both oral communication and written texts. In contrast, background knowledge refers to basic fundamental information about entities often left omitted. Lastly, following the semantic-pragmatic approach to presuppositions (Stalnaker et al., 1977), we categorise as presupposed knowledge the implicit information automatically inferred by speakers. These typologies are interrelated, as the former facilitates the accurate interpretation of the latter. Though instructions for *whisking the eggs* may not explicitly mention it, we inherently infer essential presupposed knowledge, including prior actions like *egg-breaking* and the use of a tool (e.g., a fork) for the beating process, as long as the background knowledge about the nature of eggs themselves (e.g., eggs are liquid and can be beaten).

To uncover the processes underlying the foreground information, a three-level analysis is presented and summarised in Figure (1).

1. I Level. This level relates to the comprehensive ontological knowledge about entities and actions. This knowledge is represented by sources integrated into the graph database described in Section 4. The action *whisk the eggs* assumes that the knowledge of the entity *egg* is already available for the hearer, regardless of whether it has been explicitly described in the dialogue or not. This assumption is based on the fact that the knowledge of the object is part of the shared understanding of the world. This ontological information can be retrieved by querying the database when necessary (e.g., I need to know the state of an ingredient to perform actions).

2. II Level. The focus is on the action and the entity involved in a foreground event. This level refers to a semantic analysis applied to each sentence of the dialogue, employing an annotation scheme based on FrameNet (Section 6). An example is represented by the sentence *whisk the eggs*, where the action of *whisking* evokes the frame *cause_to_amalgamate* described in FrameNet.

3. III Level. The focus is on the presupposed action and entities that enable the frame identified in the II level to take place. This level pertains to a probabilistic analysis on the Epic Kitchens dataset, containing co-occurrences of nouns and verb classes, ultimately aiming to predict the core action that defines the frame itself. For instance, the action of *cracking the eggs*, which does not appear explicitly in the dialogue, is implied in the action *whisking the eggs*, semantically marked as *cause_to_amalgamate*. By applying the probabilistic calculus on entity relationships within the database, it will be possible to extract the most likely co-occurrences of actions within

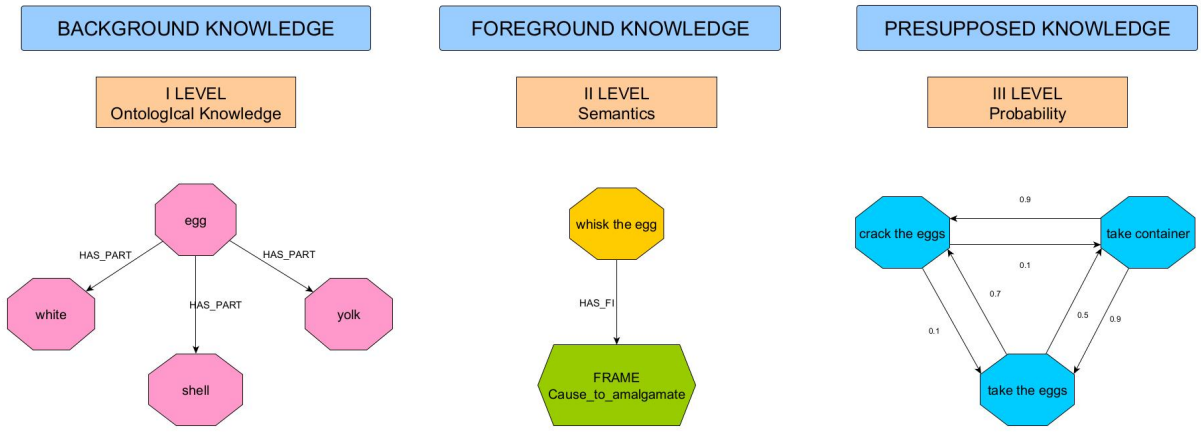

Figure 1: Analysis model with the example of the instruction *whisk the eggs*. At the first level, the model includes ontological knowledge of entities (e.g., *egg*) and their subparts (e.g., *shell*, *yolk*). At the second level, the action *whisk the eggs* invokes the *cause_to_amalgamate* frame. At the third level, the action of *whisking* implies a series of action chains (e.g., *take container*, *crack eggs*), determined by the probability of their occurrences represented as relationship properties.

a given semantic context, while avoiding the verbalisation of presupposed actions.

Due to the complexity of the analysis, the present work will focus only on the second level, exploring the distribution of foreground semantic information within the communication flow. In future work, these results will allow us to deepen the analysis of background and presupposed knowledge.

6 Methodology

To identify the semantic characteristics of foreground information, an annotation scheme was developed using the FrameNet lexical database (Baker et al., 1998) based on *Frame Semantics* (Fillmore et al., 1976) for describing word senses. A semantic frame is defined as a coherent structure of concepts which evokes a situation, an event or a state along with its participants. In FrameNet, each concept (*frame*) is schematised with its definition, its examples, and its *frame elements*, which represents the semantic roles required by the *lexical unit* (LU) evoking the frame. The sentence *Bake the cookies at 350 degrees* corresponds to *Apply_heat* frame described as follows:

A Cook applies heat to *Food*, where the *Temperature_setting* of the heat and *Duration* of application may be specified. A *Heating_instrument*, generally indicated by a locative phrase, may also be expressed.

Cook, *Food*, *Temperature_setting*, *Heating_instrument*, *Duration* are the FEs of the frame. Words such as *fry, bake or boil* represent the LUs evoking the frame. In this work, we identified 29 domain-based frames (defined as Frame Intent, FI) along with their FE, as shown in Table 1. We chose to label frames as FIs as they determine the explicit actions expressed by users. Once the dialogues are annotated, they will be integrated into the database and connected to existing resources.

Frame Intents (FI)	Transcript	Frame Elements (FE)
Taking	take a *knife*	Theme
Cause_change_of_phase	melt 1/4 cup butter *in a medium-size pan*	Container
Cause_to_continue	keep *the chicken warm*	State
Cause_temperature_change	could you preheat your oven *to 400 degrees?*	Temperature_goal
Soaking	marinate it *during the night*	Duration

Table 1: FI example for corpus annotation along with their FE. Due to limited space, only 5 out of 29 FIs and one FE for each are reported.

To gain frame recurrences and their positions within dialogues, we annotated 46 dialogues using Label Studio (Tkachenko et al., 2020-2022) (2), an open-source data labelling platform which facilitates the creation of annotated datasets. Two annotators were engaged to annotate the first ten dialogues, ensuring the annotation agreement. The MASI (Measuring Agreement on Set-valued Items) distance (Passonneau, 2006) was employed as it is particularly useful for handling multiple labels for a single item, ranging from 1 to indicate identical sets, to 0 to indicate completely disjointed sets. Additionally, The Krippendorff's Alpha (Passonneau, 2004) was applied to assess the annotation quality, calculating the metric of weighted agreement. Results show an agreement value of 0.75, confirming the validity of the annotation scheme.

Figure 2: Label Studio interface. Highlighted text segments within the dialogue correspond to the assigned labels.

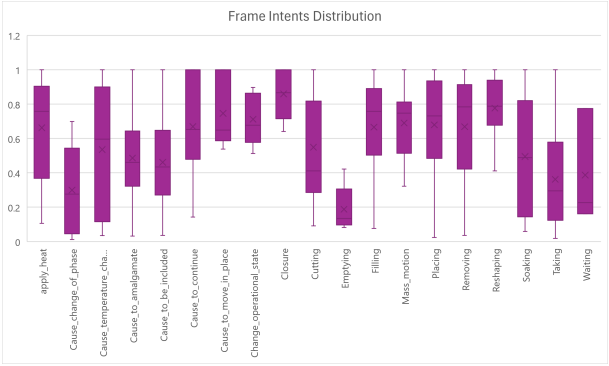

Figure 3: FI distribution within the dialogues. Only 19 out of 29 FI are taken into account for our analysis.

7 Results

Following the completion of the annotation phase, we extracted the dialogues from the platform, executing a Python script to ascertain the FI's distribution within the dialogue stream. For enhanced visualisation purposes, the data was then converted into a graphical representation, as illustrated in (3). From 29 Frames, we identified 19 relevant for our analysis.

Results show that certain FI as *Taking* (e.g. take a bowl), *Soaking* (e.g. soak the chicken), *Emptying* (e.g. drain the turkey), *Cause_temperature_change* (e.g. preheat the oven to 400 degrees), and *Cause_change_of_phase* (e.g. melt 1/4 cup butter) occur earlier, while *Cause_to_continue* (e.g. keep the chicken warm), *Cause_to_move_in_place* (e.g. turn the pancake), *Reshaping* (e.g. roll up each crepes), *Placing* (e.g. put the chicken on plate) and *Closure* (e.g. seal the bag) occur towards the process's end. This distribution reflects the natural flow of a culinary task, where initial steps involve preparing ingredients (Taking, Soaking, Emptying) and manipulating temperature (Cause_temperature_change, Cause_change_of_phase), while later stages focus on cooking food (Cause_to_move_in_place), monitoring progress (Cause_to_continue), modelling the shape (Reshaping) and finalising the process (Placing, Closure). The specific action sequences that frequently occur at particular points in the dialogue enable a deeper investigation into presupposed knowledge and facilitate the extraction of action co-occurrences semantically implied by the foreground knowledge.

8 Conclusions and Future Works

In this paper, we proposed a three-level analysis for deepen the investigation of CSK structure. In particular, we focused on the second level, annotating 46 dialogues extracted from the CookDial corpus to calculate FI recurrences and their positions within dialogues. The analysis revealed that there are FI predominantly appeared in the initial stages of the dialogue and others towards the end of it, reflecting the natural flow of a cooking process. Those results hold significant importance as they contribute to the systematic nature of this information by establishing clear patterns and relationships between frames. A further study is underway on Epic Kitchens, allowing us to identify *presupposed* actions that can be omitted from recipe instructions without impacting completion.

References

Patrícia Amaral, Chris Cummins, and Napoleon Katsos. 2011. Experimental evidence on the distinction between foregrounded and backgrounded meaning. In *Proceedings of the 2011 ESSLLI Workshop on Projective Content*, pages 1–7. Citeseer.

Collin F Baker, Charles J Fillmore, and John B Lowe. 1998. The berkeley framenet project. In *COLING 1998 Volume 1: The 17th International Conference on Computational Linguistics*.

Rodney A Brooks. 1991. Intelligence without representation. *Artificial intelligence*, 47(1-3):139–159.

Erik Cambria, Amir Hussain, Catherine Havasi, and Chris Eckl. 2009. Common sense computing: From the society of mind to digital intuition and beyond. In *Biometric ID Management and Multimodal Communication: Joint COST 2101 and 2102 International Conference, BioID_MultiComm 2009, Madrid, Spain, September 16-18, 2009. Proceedings 2*, pages 252–259. Springer.

Hakki C Cankaya and Dan Moldovan. 2009. Method for extracting commonsense knowledge. In *Proceedings of the fifth international conference on Knowledge capture*, pages 57–64.

Eve V Clark. 2015. Common ground. *The handbook of language emergence*, pages 328–353.

Dima Damen, Hazel Doughty, Giovanni Maria Farinella, Sanja Fidler, Antonino Furnari, Evangelos Kazakos, Davide Moltisanti, Jonathan Munro, Toby Perrett, Will Price, et al. 2018. Scaling egocentric vision: The epic-kitchens dataset. In *Proceedings of the European conference on computer vision (ECCV)*, pages 720–736.

Charles J Fillmore et al. 1976. Frame semantics and the nature of language. In *Annals of the New York Academy of Sciences: Conference on the origin and development of language and speech*, volume 280, pages 20–32. New York.

Norman M Fraser and G Nigel Gilbert. 1991. Simulating speech systems. *Computer Speech & Language*, 5(1):81–99.

Neelansh Garg, Apuroop Sethupathy, Rudraksh Tuwani, Rakhi Nk, Shubham Dokania, Arvind Iyer, Ayushi Gupta, Shubhra Agrawal, Navjot Singh, Shubham Shukla, et al. 2018. Flavordb: a database of flavor molecules. *Nucleic acids research*, 46(D1):D1210–D1216.

HP Grice. 1975. Logic and conversation. *Syntax and Semantics*, 3:43–58.

Dell Hymes et al. 1972. On communicative competence. *sociolinguistics*, 269293:269–293.

Yiwei Jiang, Klim Zaporojets, Johannes Deleu, Thomas Demeester, and Chris Develder. 2020. Recipe instruction semantics corpus (risec): Resolving semantic structure and zero anaphora in recipes. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, pages 821–826.

Yiwei Jiang, Klim Zaporojets, Johannes Deleu, Thomas Demeester, and Chris Develder. 2023. Cookdial: a dataset for task-oriented dialogs grounded in procedural documents. *Applied Intelligence*, 53(4):4748–4766.

Douglas B Lenat. 1995. Cyc: A large-scale investment in knowledge infrastructure. *Communications of the ACM*, 38(11):33–38.

Hugo Liu and Push Singh. 2004. Conceptnet—a practical commonsense reasoning tool-kit. *BT technology journal*, 22(4):211–226.

Brian MacWhinney and William O'Grady. 2015. *The handbook of language emergence*. John Wiley & Sons.

Bodhisattwa Prasad Majumder, Harsh Jhamtani, Taylor Berg-Kirkpatrick, and Julian McAuley. 2020. Like hiking? you probably enjoy nature: Persona-grounded dialog with commonsense expansions. *arXiv preprint arXiv:2010.03205*.

Javier Marın, Aritro Biswas, Ferda Ofli, Nicholas Hynes, Amaia Salvador, Yusuf Aytar, Ingmar Weber, and Antonio Torralba. 2021. Recipe1m+: A dataset for learning cross-modal embeddings for cooking recipes and food images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(1):187–203.

John McCarthy. 1959. Programs with common sense.

Tuan-Phong Nguyen, Simon Razniewski, Julien Romero, and Gerhard Weikum. 2022. Refined commonsense knowledge from large-scale web contents. *IEEE Transactions on Knowledge and Data Engineering*.

Antonio Origlia, Martina Di Bratto, Maria Di Maro, and Sabrina Mennella. 2022. A multi-source graph representation of the movie domain for recommendation dialogues analysis. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 1297–1306.

Rebecca Passonneau. 2004. Computing reliability for coreference annotation.

Rebecca Passonneau. 2006. Measuring agreement on set-valued items (masi) for semantic and pragmatic annotation.

Maarten Sap, Ronan Le Bras, Emily Allaway, Chandra Bhagavatula, Nicholas Lourie, Hannah Rashkin, Brendan Roof, Noah A Smith, and Yejin Choi. 2019.

Atomic: An atlas of machine commonsense for if-then reasoning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 3027–3035.

Robert Stalnaker. 2002. Common ground. *Linguistics and philosophy*, 25(5/6):701–721.

Robert Stalnaker, Milton K Munitz, and Peter Unger. 1977. Pragmatic presuppositions. In *Proceedings of the Texas conference on per˜formatives, presuppositions, and implicatures. Arlington, VA: Center for Applied Linguistics*, pages 135–148. ERIC.

Ahmet Süerdem. 2024. The challenges and opportunities in large language models: Navigating the perils of stochastic and scholastic parrots in artificial understanding and common sense. *AI and Common Sense*, pages 195–212.

Maxim Tkachenko, Mikhail Malyuk, Andrey Holmanyuk, and Nikolai Liubimov. 2020-2022. Label Studio: Data labeling software. Open source software available from https://github.com/heartexlabs/label-studio.

Jim Webber. 2012. A programmatic introduction to neo4j. In *Proceedings of the 3rd annual conference on Systems, programming, and applications: software for humanity*, pages 217–218.

Mark E Whiting and Duncan J Watts. 2024. A framework for quantifying individual and collective common sense. *Proceedings of the National Academy of Sciences*, 121(4):e2309535121.

Liang-Jun Zang, Cong Cao, Ya-Nan Cao, Yu-Ming Wu, and Cun-Gen Cao. 2013. A survey of commonsense knowledge acquisition. *Journal of Computer Science and Technology*, 28(4):689–719.

Pei Zhou, Karthik Gopalakrishnan, Behnam Hedayatnia, Seokhwan Kim, Jay Pujara, Xiang Ren, Yang Liu, and Dilek Hakkani-Tur. 2021a. Commonsense-focused dialogues for response generation: An empirical study. *arXiv preprint arXiv:2109.06427*.

Pei Zhou, Karthik Gopalakrishnan, Behnam Hedayatnia, Seokhwan Kim, Jay Pujara, Xiang Ren, Yang Liu, and Dilek Hakkani-Tur. 2021b. Think before you speak: Explicitly generating implicit commonsense knowledge for response generation. *arXiv preprint arXiv:2110.08501*.