[1]URBAN/ECO Research Center — University of Naples “Federico II”
[2]University of Catania
[3]Logogramma s.r.l.

Sabrina Mennella[1,2], Maria di Maro[1], Martina di Bratto[1,3]

# Estimating Commonsense Knowledge from a Linguistic Analysis on Information Distribution

## Introduction

**Commonsense Knowledge (CSK)** is described as a **complex and polyhedric structure**, encompassing a wide range of knowledge and reasoning generally acquired through everyday experiences (Cambria, 2009). It is often **implicit in communication** (written or oral) (Grice,1975), posing a **challenge for NLP systems** that leverage a data-driven approach to simulate human knowledge acquisition information (Ilievski, 2021).

Its multifaceted nature suggests its conceptualisation as a ***process*** rather than a static collection of information. The results of this process are understood as **probabilistic models** that estimate the probability that unobserved facts can be assumed to be true or false.

**GOAL**: **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.

## Three-level analysis model

We classify knowledge into three typologies, facilitating a clearer understanding and effective analysis of the data:

- **Foreground knowledge**: information explicitly expressed in both oral communication and written texts
- **Background Knowledge**: basic fundamental information about entities often left omitted.
- **Presupposed Knowledge**: implicit information automatically inferred by speakers.

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

- **I Level**: Ontological knowledge about entities and actions.
- **II Level**: Semantic analysis applied to each sentence of the dialogue, employing an annotation scheme based on FrameNet
- **III Level**: Probabilistic analysis on the EK dataset, for predicting the core action that defines the frame

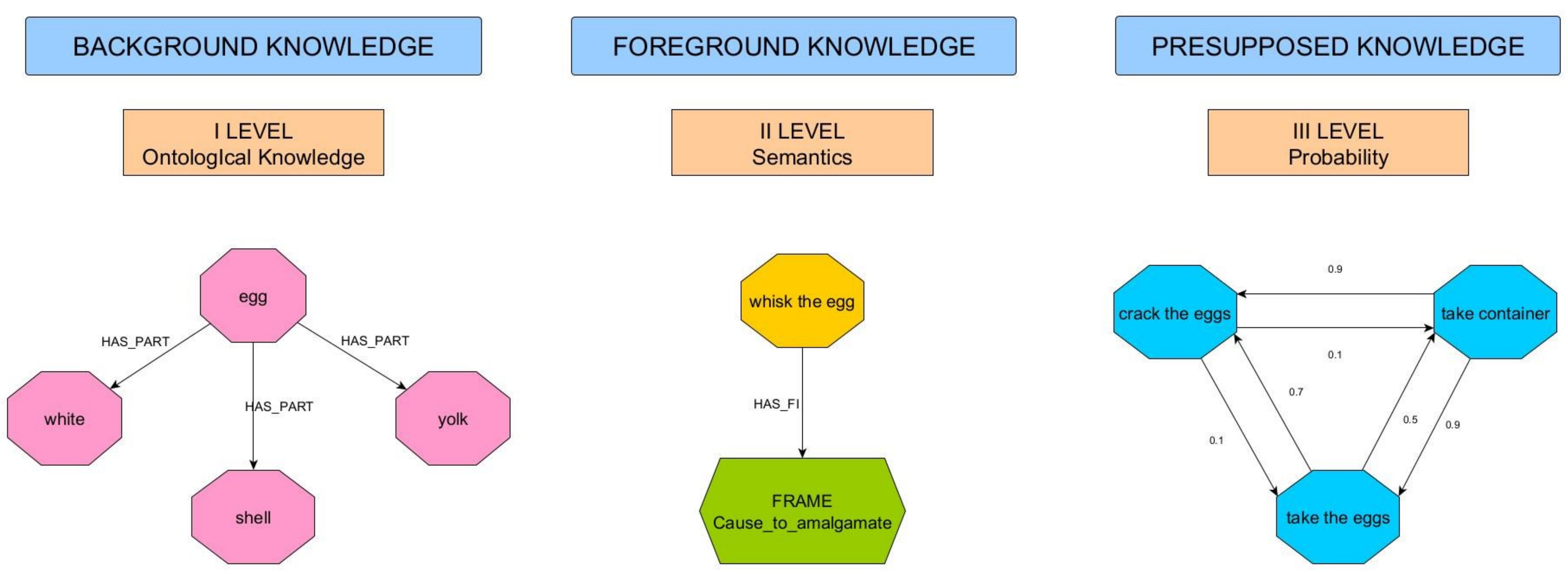


FIG.1: Analysis model with the example of the instruction *whisk the eggs*. At the first level, the model includes ontological knowledge of entities (e.g., *egg*) and their subparts (e.g., *shell, yolk*). At the second level, the action whisk the eggs invokes the cause to amalgamate frame. At the third 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

## Results

**ANNOTATION AGREEMENT:** Two annotators labelled the first 10 dialogues. We employed (i) The **Krippendorff's Alpha** (Passonneau, 2004) to calculate the weighted agreement metric between annotators and (ii) the **MASI distance** (Measuring Agreement on Set-valued Items) as it deals with multiple labels for a single item (it varies from 1 when two sets are identical and 0 when those are disjoint).

⇒ **Agreement value of 0.75,** confirming that the annotation scheme is valid.

**FRAMES DISTRIBUTION:** We extracted dialogues from the platform, executing a Python script to ascertain the FI's distribution within the dialogue stream.

⇒ **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**

⇒ **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**. 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.

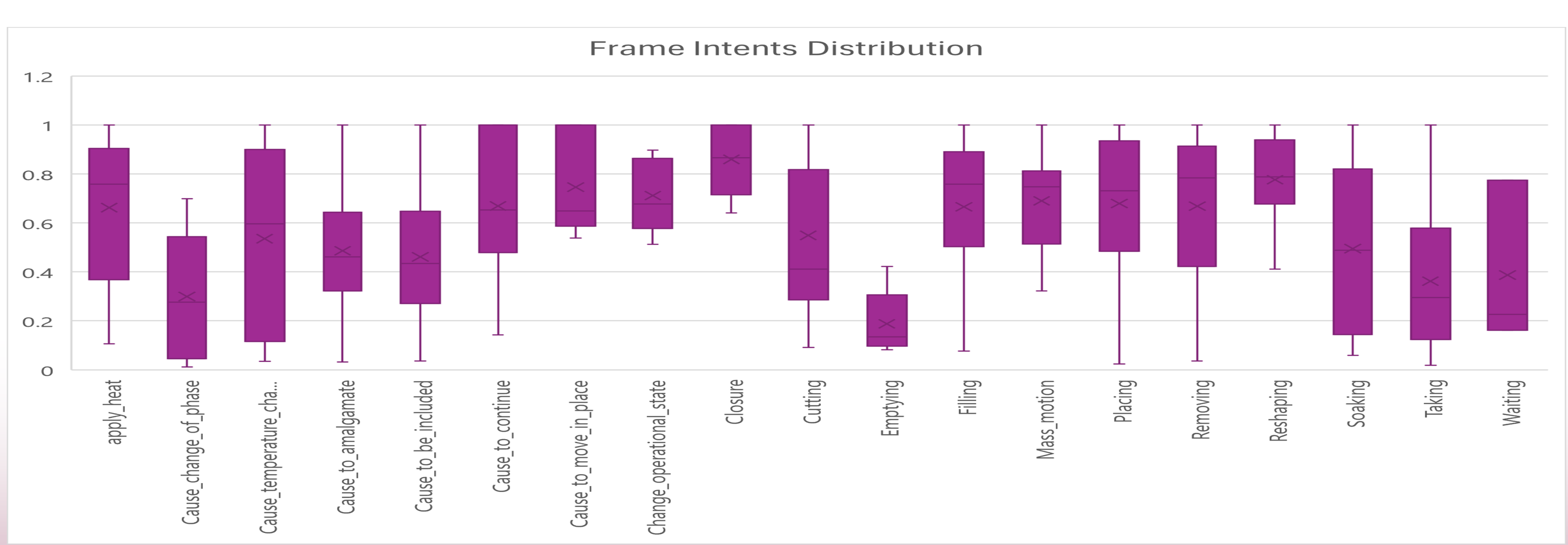


FIG. 3: FI distribution within the dialogues. Only 19 out of 29 FI are taken into account for our analysis.

## Materials

To extract the CS actions, the culinary domain was taken into account. This choice is motivated by two factors:

I. a presumed **high level of familiarity among speakers** due to its integration into daily routines;
II. **individual actions are often interconnected**, which means that one action can influence a subsequent one (e.g. egg-beating necessitates prior cracking).

**KNOWLEDGE GRAPH:**

Three main resources: (i) **Recipe1M+ dataset** (Marın et al., 2021), (ii) **FlavorDB** (Garg et al., 2018), (iii) **Epic-Kitchens dataset** (Damen et al., 2018). Those represent 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*).

**LINGUISTIC ANALYSIS:**

**CookDial dataset** (Jiang et al., 2023): it is a dialogue corpus focused on the culinary domain. It consists of 260 English human-to-human task-oriented dialogues, in which each dialogue is associated with a recipe extracted from the RISeC corpus (Jiang et al., 2020), comprising the annotation of 260 recipe instructions.

**FrameNet** (Baker et al., 1998): it is a lexical database whose word senses descriptions are based on *Frame Semantics* (Fillmore et al., 1976). A semantic frame is defined as a coherent structure of concepts which evokes a situation, an event or a state. In FrameNet, each concept (frame) is schematised, along with its definition and its frame elements, which represents the semantic roles required by the lexical unit (LU) evoking the frame.

**Label Studio:** an open-source data labelling platform which facilitates the creation of annotated datasets.

## Methods

To facilitate the dentification of the entities and actions involved in recipe instructions along with their relationships and co-occurrences, we built our **domain knowledge graph**, taking into account the aforementioned sources. 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 the **development of the annotation scheme**, the FrameNet lexical base (Baker et al., 1998) was employed. We identified **29 domain-based frames** (defined as Frame Intent, FI) along with their Frame Elements (FE), as shown in TAB 1.

Subsequently, CookDial dialogues were loaded into Label Studio for the semantic annotation (FIG.2).

| FRAME INTENT (FI) | TRANSCRIPT | FRAME ELEMENT (FE) |
|---|---|---|
| Cause_to_amalgamate | Combine *1/2 teaspoon of cinnamon* in bowl | Parts |
| | mix the eggs *into the mixture* | Whole |
| | beat an egg *with an egg beater or mixer.* | Means |
| | Combine 1/2 teaspoon of cinnamon *in bowl* | Place |
| | beat the eggs *for 3 minutes*. | Time |
| | mix the eggs *together* | Result |

TAB.1: Frame Intents (FI) example for corpus annotation along with their Frame elements (FE). Due to limited space, only 2 out of 29 FIs are reported.

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

## Conclusions

This work presents a **three-level structured analysis model for deepen the study of CSK** by contrasting linguistic content eliciting actions in humans against actions that were actually performed. In particular, the focus is on the **second level**, where an annotation scheme based on FrameNet is developed and applied to 46 dialogues extracted from CookDial, a dialogic corpus on the culinary domain. 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.