



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e. g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

Processing Embodied Conversation for Interactive Task Learning

Rimvydas Rubavičius



Doctor of Philosophy
Institute for Language, Cognition and Computation
School of Informatics
University of Edinburgh

2025

Abstract

Lifelong learning is a long-standing goal of human-robot interaction. One approach to achieving this is through Interactive Task Learning (ITL) scenarios, in which the learner uses natural interaction with a user, acting as a teacher, to learn new tasks. One type of natural interaction is embodied conversation, initiated by an instruction such as “move the one red cube in front of a blue cylinder”. A key challenge for ITL is that the learner’s domain conceptualization may entirely lack the concepts necessary for solving the task. In other words, the teacher’s natural language expressions are *neologisms* to the learner (“red” or “cube”, say, in our example) that may denote concepts that are not a part of the learner’s conceptualization of the domain at all. To handle such unforeseen possibilities, the learner must perform *interactive symbol grounding*: that is, they must expand their hypothesis space of possible states to include newly discovered concepts and learn in real-time how to recognize objects denoted by this unforeseen concept to successfully solve the task, using the teacher’s messages as evidence.

This thesis studies three ways in which the *formal semantic analysis* of embodied conversation can aid ITL. Firstly, we study how *processing referential expressions* like “the one red cube” with their logical consequences (e.g. that there is exactly one red cube in the environment) aids interactive symbol grounding. Secondly, we look into designing *dialogue strategies under unawareness* by quantifying the value of asking questions which require the teacher’s effort vs. risking solving tasks with current beliefs, which will be costly if wrong. Our unique contribution is that our learning models cope with an ever-expanding hypothesis space of possible states and actions that arise in ITL. Finally, we consider *corrections of the agent’s execution actions*, which arise when the ITL agent attempts to solve the task but performs a sub-optimal action in the environment (e.g., picking a red cylinder rather than a red cube), exposing its false beliefs and so prompting the teacher to express the source of the error (e.g., “No, this is a cylinder”). Such corrective feedback triggers belief revision, which, by exploiting the semantic consequences of the fact that it’s a correction (in this case, that the picked object is not a cube), complements and further improves the learning process.

We study these three facets by developing and evaluating neuro-symbolic methods for interactive symbol grounding and policy learning. Through our experiments, we conclude that agents who use formal semantic analysis when processing embodied conversation, outperform learners lacking these capabilities.

Lay Summary

In a collaborative teacher-learner setting, the learner’s objective is to acquire knowledge via interaction with the teacher, a domain expert, to help solve a novel task that has not been performed before. The teacher can provide information to the learner in a timely and relevant manner. One natural form of such interactions is embodied conversation, consisting of natural language utterances and gestures forming discourse that includes, for example, the learner asking a question and the teacher providing the answer by pointing to some objects in the environment. Natural language can provide compact descriptions of the scene that aid learning. However, it also introduces the possibility that the teacher can utter words that describe concepts about which the learner was unaware but is expected to know to solve the task correctly. For example, the teacher may ask the learner to “put the two granny smiths inside the basket”, but it is the first time the learner hears the words “granny smiths”, so it is uncertain to which objects this phrase refers. To handle situations in which the learner is exposed to such unforeseen possibilities, they can engage in embodied conversation to ground the novel concepts by, for instance, asking questions and the teacher’s response to their actions in the environment. Even though such an interactive task learning scenario is natural for humans to learn novel tasks, it is challenging for contemporary methods of artificial intelligence: to engage in such a natural interaction requires learning to be incremental (learning piecemeal, from one message at a time); it also requires dealing with an increasing number of concepts in the domain and have mechanisms to cope with false beliefs which must be revised when exposed by the teacher as part of the extended interaction.

In this thesis, we develop a method for processing embodied conversation to interactively learn new tasks. The unique contribution of our method is that it uses linguistic insights and reasoning, particularly about what various messages implicate beyond their surface linguistic form. We propose algorithms that exploit such insights both for learning new concepts from the conversation and for learning decent strategies for engaging in the conversation with the teacher to acquire the information the learner needs from them. We evaluate our methods by comparing them with learners who lack such capabilities and do not engage in such reasoning when faced with an interactive task learning scenario.

Acknowledgements

First, I would like to thank my supervisor team: Alex Lascarides and Subramanian (Ram) Ramamoorthy. Since our first interaction during my undergraduate, Alex has been an inspiration and a role model in pursuing academic excellence. Supervision with Ram has been a great compliment to Alex, as he has always encouraged me to think outside my immediate research area and be interested in broad problems in making robots act in the real world. I immensely appreciate their support and guidance.

I would also like to thank my examiners: Mohan Sridharan and Ann Copestake (for viva), Mirella Lapata and N. Siddharth (for annual reviews), as well as other members of the School of Informatics who have taken the time to look at various iterations of my research and provided their perspective and valuable criticisms. These interactions helped me to reassure and make me excited about the research I have been conducting.

I thank my fellow members at the Institute for Language, Cognition and Computation, Institute of Perception, Action and Behaviour as well as at the Centre for Doctoral Training in Natural Language Processing. By being part of these intellectual clusters I have benefited from various perspectives, especially in trying to bridge the gap and interests between these seemingly diverse yet related communities.

Lastly, thanks to my family and friends for supporting me and reminding me of the beauty of the world outside of research. As my doctorate journey draws to an end, new adventures are beckoning on the horizon.

To my family

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(Rimvydas Rubavičius)

Table of Contents

Notation	13
Abbreviations	17
1 Towards Embodied Lifelong Agents	23
1.1 Motivation	24
1.2 Contributions	29
1.3 Task Overview and Scope	30
2 System Requirements	33
2.1 Interactive Task Learning Capabilities	34
2.2 Embodied Conversation Capabilities	36
2.3 Symbol Grounding Capabilities	39
2.4 Reasoning Capabilities	42
2.5 Summary	45
3 Base Architecture	47
3.1 Overview	48
3.1.1 Teacher’s Competence	48
3.1.2 System Description	49
3.2 Knowledge Representation and Reasoning	51
3.2.1 The Domain Model	52
3.2.2 The Agent’s Belief State	53
3.2.3 Logic of Sentences	55
3.2.4 Logic of Referential Expressions	59
3.2.5 Reasoning about Beliefs	62
3.3 Summary	65

4	Processing Referential Expressions	67
4.1	Introduction	68
4.2	Reference Resolution under Unawareness	70
4.3	Interactive Neuro-symbolic Grounding	71
4.3.1	The Grounding Model	72
4.3.2	An Illustrative Example	75
4.4	Experiments	81
4.4.1	Experimental Setup	81
4.4.2	Results and Discussion	85
4.5	Summary	93
5	Dialogue Strategies under Unawareness	95
5.1	Introduction	96
5.2	Interactive Reference Resolution under Unawareness	98
5.3	Interaction Planning	99
5.3.1	Decision Problem	100
5.3.2	Policy	101
5.4	Experiments	105
5.4.1	Experimental Setup	105
5.4.2	Results and Discussion	106
5.5	Summary	111
6	Processing Corrective Feedback	113
6.1	Introduction	114
6.2	Rearrangement under Unawareness	116
6.3	Corrections on Execution Actions	120
6.3.1	Discourse Coherence	120
6.3.2	Formal Semantic Analysis	121
6.4	Experiments	123
6.4.1	Experimental Setup	123
6.4.2	Results and Discussion	127
6.5	Summary	134
7	Conclusions	135
7.1	Findings	135
7.2	Future Work	137

<i>TABLE OF CONTENTS</i>	11
Appendix A Prompts	139
Bibliography	143

Notation

Domain.

$\mathcal{E} \in \mathbb{E}$ environment.

$o \in \mathbb{O}$ object in the set of objects in the environment.

$p \in \mathbb{V}$ predicate symbol from the vocabulary of the domain.

$I: \mathbb{V} \mapsto \mathbb{O}^*$ interpretation function from symbols to denotations.

$\mathcal{M} \in \mathbb{M}$ finite first-order domain model.

$\mathcal{H}_{\mathcal{M}} \subseteq \mathcal{H}$ set of atoms that are true in \mathcal{M} .

$b \in \mathbb{B}$ belief state.

Logic.

$x \in \mathbb{L}_{\text{var}}$ variable.

$o \in \mathbb{L}_{\text{const}}$ constant.

$p \in \mathbb{V}$ predicate symbol.

$\mathbb{L}_{\text{conn}} = \{\wedge, \vee, \neg, \leftarrow, \leftrightarrow\}$ boolean operators and logical connectives.

$Q \in \mathbb{L}_{\text{quant}}$ generalized quantifier.

$A \in \mathcal{H}$ atom in Herbrand base constructed from n -place predicate and n constants.

$\phi \in \mathbb{L}_{\text{snt}}$ logical form of sentence in language \mathbb{L}_{snt} .

$\Delta \subseteq \mathbb{L}_{\text{snt}}$ domain theory (set of sentences that are true in the environment).

$\lambda x. \phi$ λ -expression for sentence $\phi \in \mathbb{L}_{\text{snt}}$.

$\text{RSTR} \subseteq \mathbb{O}, \text{BODY} \subseteq \mathbb{O}$ restrictor and body sets.

$Q: \mathbb{O} \times \mathbb{O} \mapsto \{0, 1\}$ generalized quantifier condition.

$g: \mathbb{L}_{\text{var}} \mapsto \mathbb{L}_{\text{const}}$ variable assignment function.

$[[\cdot]]^{\mathcal{M}, g}: \mathbb{L}_{\text{snt}} \mapsto \{0, 1\}$ valuation function.

\models entailment (semantic consequence).

$\langle \text{Qx}.\phi \rangle \in \mathbb{L}_{\text{ref}}$ logical form of a referential expression in language \mathbb{L}_{ref} .

$\sigma: \mathbb{M} \times \mathbb{L}_{\text{snt}} \times \mathbb{L}_{\text{var}} \mapsto \mathbb{M}$ domain model projection.

$\Phi(\cdot)^{\mathcal{M}}: \mathbb{N}_{\text{ref}} \mapsto 2^{|\mathbb{O}|}$ Referent computation.

Probability and Information Theory.

$\text{Pr}: \Omega \mapsto [0, 1]$ probability measure over sample space Ω .

\mathcal{B} Bernoulli random variable.

$\mathbb{H}_{\mathcal{B}}[\cdot]: [0, 1] \mapsto \mathbb{R}_{>0}$ entropy of the \mathcal{B} .

$\mathbb{H}: \mathbb{B} \mapsto \mathbb{R}^+$ Entropy of the belief state.

$\mathbb{E}[\cdot]$ Expected value of a random variable.

$w_b: \mathcal{H} \mapsto [0, 1]$ base weights function (prior beliefs).

$w_g: \mathcal{H} \mapsto [0, 1]$ grounded weights function (integrated evidence).

$WMC_w: \mathbb{L}_{\text{snt}} \mapsto [0, 1]$ weighted model count with weights w .

$CON_w(\cdot | b): \mathbb{L}_{\text{snt}} \mapsto [0, 1]$ conditional probability with weights w and belief state b .

$MAP_w(\cdot): \mathbb{B} \mapsto \mathbb{M}$ maximum a posteriori with weights w .

Embodied Conversation.

$\Phi: \mathbb{N} \mapsto \mathbb{L}$ semantic parser from natural language expressions to logical forms.

$t \in \mathbb{N}_{\text{task}}$ task instruction.

$r \in \mathbb{N}_{\text{ref}}$ referential expression.

$\mathcal{R} \subseteq 2^{\mathbb{O}}$ referent.

$q \in \mathbb{N}_{\text{quest}}$ question.

$c \in \mathbb{N}_{\text{corr}}$ corrective feedback.

$\xi: \mathbb{N}_{\text{ref}} \times 2^{\mathbb{O}} \mapsto \mathbb{L}_{\text{snt}}$ semantic analysis of referential expression and referent.

$\zeta: \mathbb{P} \times \mathbb{N}_{\text{task}} \times \mathbb{N}_{\text{corr}} \mapsto \mathbb{L}_{\text{snt}}$ semantic analysis of execution action and correction.

Symbol Grounding.

$\mathbf{x} \in \mathbb{X}$ embedding of size d for each object $o \in \mathbb{O}$ in the environment $\mathcal{E} \in \mathbb{E}$.

$\mathbf{y} \in [0, 1]^{\mathbb{V}}$ semantic vector.

$(\mathbf{x}, \mathbf{y}) \in \mathbb{S}$ embedding-semantic vector pair in the support.

$\omega_b: \mathbb{E} \mapsto [0, 1]^{\mathbb{V}}$ grounding model under belief state $b \in \mathbb{B}$.

Update: $\mathbb{B} \times \mathbb{L}_{\text{snt}} \mapsto \mathbb{B}$ belief update procedure following Algorithm 1.

$\mathbb{S}_p^{+/-}$ positive/negative support for predicate symbol $p \in \mathbb{V}$.

$\mathbf{z}_p^{+/-}$ positive/negative prototype for predicate symbol $p \in \mathbb{V}$.

Decisions.

$a \in \mathbb{A}$ agent's action.

$R: \mathbb{A} \mapsto \mathbb{R}$ reward function.

$C: \mathbb{N}_{\text{quest}} \mapsto \mathbb{R}_{>0}$ question cost function.

$act \in \mathbb{A}$ agent's decision to solve the task using current beliefs.

$\mathbb{P} = \{pick, place, complete\}$ execution actions to manipulate environment.

$\phi \sim Result(a)$ ϕ constructed from teacher's response as the result of action $a \in \mathbb{A}$.

$I: \mathbb{B} \times \mathbb{A} \mapsto \mathbb{R}_{\geq 0}$ Expected information gain of taking action a in belief state b .

$Q: \mathbb{B} \times \mathbb{A} \mapsto \mathbb{R}$ state-action value function.

$\theta = [\theta_1, \theta_2]^{\top}$ interpolation constants.

$h: \mathbb{B} \times \mathbb{A} \mapsto \mathbb{R}^n$ functional approximation for state-action value function.

$\pi: \mathbb{B} \mapsto \mathbb{A}$ policy.

$\epsilon \in \mathbb{R}$ exploration factor for ϵ -greedy policy.

$\alpha \in \mathbb{R}$ learning rate.

$\gamma \in (0, 1]$ discount factor.

Abbreviations

ITL Interactive Task Learning.

AI Artificial Intelligence.

CON Conditional Probability Query.

FSA Formal Semantic Analysis.

LML Lifelong Machine Learning.

MAP Maximum a Posteriori Query.

ML Machine Learning.

NLP Natural Language Processing.

POMDP Partially Observable Markov Decision Processes.

PRM Probabilistic Relational Model.

PSDD Probabilistic Sentential Decision Diagrams.

RL Reinforcement Learning.

RSA Rational Speech Acts.

TAMP Task and Motion Planning.

TTR Type Theory with Records.

WMC Weighted Model Counting.

List of Figures

1.1	Interactive task learning involving the neologism “granny smith”	23
1.2	Confidence-awareness matrix adapted from (Saravanan, 2021).	26
2.1	Personal robot that could be used to clean the house.	33
3.1	Base architecture for an embodied lifelong agent that is designed to solve tasks under unawareness by interacting with the user through embodied conversation.	47
4.1	Embodied Conversation for the sequence of reference resolution under unawareness tasks.	67
4.2	Visualization of Blocks tabletop environment.	82
4.3	The learning curves for solving the task of reference resolution under unawareness. Each curve is plotted with its 95% confidence interval.	86
4.4	The difference in performance between π_{sem} and π_{base} , comparing learning curves when the number of tasks that the agents face in each environment in the experiments is 1, 2 or 3 tasks.	87
4.5	The difference in performance between π_{sem} and π_{base} , comparing learning curves (with a 95% confidence interval) when the number of objects in each environment across all trials is 4–6, or 6–8, or 8–10 objects.	89
4.6	The difference in performance between π_{sem} and π_{base} , comparing learning curves (with a 95% confidence interval) when the probability that the task instruction features a quantifier that leads to logical consequeness is 0.5, or 0.7, or 0.9.	91
5.1	Embodied conversation for interactive reference resolution task.	95

5.2	The learning curves for solving the task of interactive reference resolution under unawareness. Each curve is plotted with its 95% confidence interval.	107
5.3	Environment in which agents interact to solve interactive reference resolution under unawareness given task instruction “show me the two starry green cubes” with embodied conversations given in Table 5.2. .	109
6.1	Embodied conversation with teacher’s corrective feedback.	113
6.2	Learning curves for solving move rearrangement under unawareness for $\pi_{\text{sem}}, \pi_{\text{base}}$, and π_{corr} given the default setup. Each curve is plotted with a 95% confidence interval.	128
6.3	Learning curves for solving move rearrangement under unawareness for pessimistic agents: $\pi_{\text{sem}}^{\text{pes}}, \pi_{\text{base}}^{\text{pes}}$, and $\pi_{\text{corr}}^{\text{pes}}$. Each curve is plotted with a 95% confidence interval.	130
6.4	Learning curves for solving move rearrangement under unawareness for optimistic agents: $\pi_{\text{sem}}^{\text{opt}}, \pi_{\text{base}}^{\text{opt}}$, and $\pi_{\text{corr}}^{\text{opt}}$. Each curve is plotted with a 95% confidence interval.	131

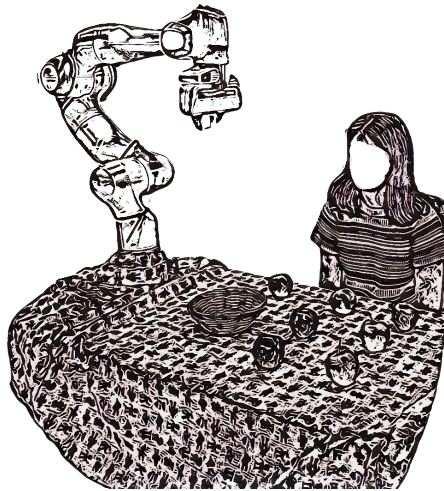
List of Tables

3.1	Generalized quantifiers (column 1), their surface forms (column 2), and truth-conditions Q between the restrictor set \mathbb{RSTR} and body set \mathbb{BODY} , used to evaluate well-formed formulas of the form $Q x.(\phi, \psi)$ (column 3).	57
3.2	Generalized quantifiers (column 1), surface form (column 2), and referent constructor used to evaluate the logical form of referential expressions for the domain model \mathcal{M} (column 3).	60
4.1	Comparison of different probabilistic queries: CON_{w_b} and MAP_{w_b} for different belief states b_1, b_2 and b_{23} containing different domain theories: Δ_1, Δ_2 and Δ_{23} , respectively. The symbol “–” denotes the unawareness about that predicate symbol for that particular stage of the interaction. This may change: e.g. upon the teacher conveying r_3 , the vocabulary $\mathbb{V} \leftarrow \mathbb{V} \cup \{\text{blue}\}$ expands, as the agent becomes aware of the concept <code>blue</code> . Thus for b_{23} “–” gets replaced with an appropriate value. In MAP_{w_b} columns, 1 means that an atom is in the estimated domain model $\hat{\mathcal{M}}$; 0 means it is not.	77
4.2	Comparison of different support embeddings, prototypes, grounded weights, and domain models (MAP_{w_g} query) constructed with different belief states b_1, b_2 , and b_{23} . Semantic vectors \mathbf{y} are omitted as they are equivalent to CON_{w_b} queries in Table 4.1. Grounded weights w_g are given in the same order for atoms as in Table 4.1.	80
4.3	P-values for comparing the agent’s learning curves for the initial setup.	85
4.4	P-values for comparing the agent’s learning curves in sensitivity analysis with * indicating no statistically significant difference.	88

5.1	P-values for comparing the agent’s learning curves for the different agents using different dialogue strategies and referential expression processing with * indicating no statistically significant difference.	106
5.2	Sample embodied conversations and rewards received from the teacher’s embodied conversation messages by different agents when solving tasks under unawareness in the environment in Figure 5.3.	110
6.1	Visualisation of apple manipulation environments, their ground-truth labels, and initial grounding DINO predictions <i>before</i> interaction with the environment and the user via embodied conversation.	124
6.2	P-values for comparing the agent’s learning curves for default, optimistic, and pessimistic base weights with * indicating no statistically significant difference.	127
6.3	Learner behaviour traces for in two consequent environments showing initial belief state, embodied conversation with rewards given in angle brackets and the updated belief state.	133
6.4	Real-world experiment results with 95% confidence interval with * indicates a statistically significant difference between other agents.	134

Chapter 1

Towards Embodied Lifelong Agents



(T1) Teacher: Put the two granny smiths inside the basket.

(L1) Learner: Before that, is this a granny smith? (designates a Bramley)

(T2) Teacher: No.

(L2) Learner: Let's try to execute the task. (picks golden delicious)

(T3) Teacher: No. This is a golden delicious.

(L3) Learner: Sorry. Let's try again. (returns picked object to the original location)

(L4) Learner: Before that, show me the two granny smiths.

(T4) Teacher: Here they are. (designates the two granny smiths)

(L5) Learner: Let's try to execute the task. (places the two granny smiths inside the basket)

(L6) Learner: I have finished the task.

(T5) Teacher: You have finished the task correctly.

Figure 1.1: Interactive task learning involving the neologism "granny smith".

1.1 Motivation

This thesis is about interaction—how to reason about it and how to learn from it. It aims to develop a framework that processes embodied conversation for interactive task learning. The motivation for developing such a framework comes from the necessity of using all learning opportunities to augment the incomplete domain specification.

Consider a general-purpose robot that assists humans in their daily activities, e.g. cleaning their dwelling, making a meal, or buying groceries. During factory development, this robot will have been trained to perform various skills such as picking, placing, pouring, and folding. To solve such complex multi-step tasks, it will have a policy that chooses which of these actions to perform given its sensory observations and task instruction (Garrett et al., 2021). Humans can interact with such an agent using embodied conversation (Clark, 1996; Cassell, 2001) to instruct and teach new tasks (Ahn et al., 2022; Brohan et al., 2023). In this pursuit, models for vision, language, and action have been developed to design *generalist robotic agents* (Reed et al., 2022; Zitkovich et al., 2023; Octo Model Team et al., 2024).

Using such agents in the context of human-robot interaction (Bartneck et al., 2020) requires additional mechanisms to align them with human preferences (Fürnkranz and Hüllermeier, 2005) and their domain conceptualisation. To illustrate, consider a scenario where the human user utters the task instruction “put the two granny smiths into the basket”. This task can be solved by locating objects that are denoted by the words “granny smiths” and “basket” in the environment and then manipulating the environment, with the agent’s sensorimotor skills for picking and placing, to achieve the desired goal configuration. But what happens if the user’s natural language instruction cannot be understood by the agent because, before deployment, it was not trained on it and cannot correctly associate the words “granny smith” with objects in the environment? What if this phrase is a *neologism* to the agent—in other words, it is not a part of the agent’s vocabulary—and the concept that “granny smith” denotes is not a part of its domain specification? In other words, the agent cannot distinguish granny smiths from other objects, including other kinds of apples. Doing such fine-grained categorisation (Wei et al., 2022) with limited experience (zero-shot or few-shot learning) might not be sufficiently captured in advance (Udandarao et al., 2024). For instance, contemporary models for vision and language, like grounding DINO (Liu et al., 2023), can accurately detect a variety of objects in the environment, including an

excellent ability to localize fruits in the discussed scenario. Nevertheless, even as an open-set object detector, in our experience, this model does not work out-of-the-box when required to localize granny smiths (presumably because it completely lacked any examples labelled granny smiths during training). To cope with this situation, the agent must update the *structure* of its domain conceptualisation with this unforeseen concept, thereby expanding the set of possible domain states, which in turn demands adapting the learner’s policies.

Coping with *unforeseen possibilities* is not a robot-specific issue. Many real-world tasks have to be solved under *unawareness*: in other words, the agent starts out lacking a complete and accurate understanding of the possible domain states, with new and unforeseen factors being introduced over time, which in turn are critical to solving its task. Examples of such situations include discovering new factors for disease diagnosis (Park et al., 2020) or emerging new categories of species (Karbstein et al., 2024). The relationship between the domain estimation confidence and awareness about concepts in the domain is captured in the confidence-awareness matrix (see Figure 1.2). It distinguishes four degrees of knowledge. Firstly, it is the *known knowns* region with well-understood domains for which the agent is designed and knows how to act to solve a variety of tasks. On the opposite side is the *unknown unknown* region, consisting of real-world domains where the agent has to cope with and acknowledge the existence of concepts that go beyond the agent’s current understanding of the domain. For the agent to transition from one region to the other, it first becomes aware of the existence of unforeseen possibilities, which, for instance, in our example above, happens when the user utters a task instruction that features a neologism. By becoming aware, the agent must first resolve structural uncertainty about a recently introduced *poorly-known unknown* that was not a part of its domain conceptualisation before the user’s utterance by updating its domain structure. After this, parametric uncertainty is resolved for now *well-known unknowns* by gathering evidence to update parameters for more accurate state estimation. Finally, orthogonal to all this, the agent may be in the *unknown known* region, in which the agent is confident but wrong about certain factors about the domain. Such situations require the agent to revise its beliefs about the domain and move through the same understanding improvement process. In this thesis, we consider agents faced with learning scenarios that traverse this matrix.

When solving tasks under unawareness, the agent must learn before it is deployed, and after it is deployed—i.e., it must be a *lifelong* learner (Chen and Liu, 2018). What is

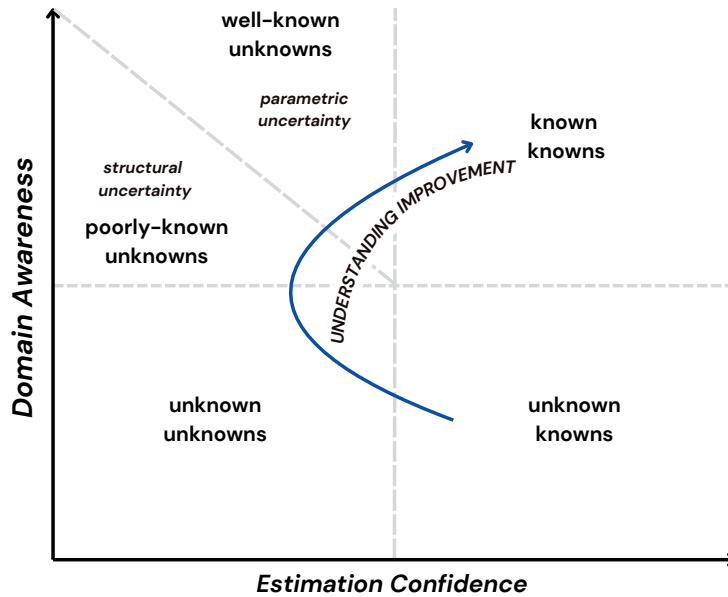


Figure 1.2: Confidence-awareness matrix adapted from (Saravanan, 2021).

unique about robotics is its physical interaction with the environment—*embodiment*—requiring the robot to adapt based on the evidence it observes in the environment. A truly *embodied lifelong agent* (Thrun and Mitchell, 1995; Mendez and Eaton, 2022) deployed in the real world thus needs to cope with unforeseen possibilities and the need to update its domain understanding via evidence that is observed via interaction in (and with) the environment.

The framework of Interactive Task Learning (ITL) (Laird et al., 2017) aims to handle such situations. In this setup, the embodied lifelong agent takes the role of the *learner* while the human, being a domain expert, takes the role of the *teacher*. Both of them can use natural interaction to establish contextually relevant knowledge that helps the learner to solve new tasks. Natural interaction is essential from the social perspective to make the interaction pleasant for the human (Tanevska et al., 2020), as well as to enable humans to interact with the agent without knowing its underlying hardware or software.

One type of natural interaction is *embodied natural language conversation*. Using it, robotic agents can interactively learn how to solve their task, as illustrated in Figure 1.1. This type of interaction includes both the learner querying the teacher to request exemplars of the neologism that the teacher just introduced into the conversa-

tion, as well as executing the task to the best of the agent’s ability, given its current beliefs, with the teacher providing corrective feedback that exposes misalignment between the estimated and the ground-truth domain conceptualisation. This thesis studies these kinds of embodied conversations in service of ITL, leading to the first hypothesis of the thesis:

Thesis hypothesis I : An ITL agent that engages in embodied conversation can discover and exploit unforeseen possibilities to solve tasks under unawareness.

A principal way to process embodied conversation is to parse it into a logical form in a symbolic logic (Blackburn and Bos, 2005). *Formal semantics* (Gamut, 1991a) is the study of analysing natural language signals as formal meaning representations—logical forms—expressed in symbolic logic. Being in symbolic logic, logical forms support *automated inference* and notions of satisfaction, validity and entailment. Formal semantics focuses on how the meaning of natural language signals contribute to valid arguments, and thus its focus is on linguistic constructions that affect valid argument structure—for example, quantifiers (e.g., “the two” in L5), as well as discourse-level logical concepts like coherence/rhetorical relations (e.g., T3 is a correction of L2). This thesis studies how reasoning about the logical consequences of these logical forms can enhance learning about the domain, contributing benefits to ITL. We call the process of going from linguistic and non-linguistic expressions to logical forms supporting automated inference and being truthful to the logical consequences of these expressions as Formal Semantic Analysis (FSA).

There are *prima facie* reasons for thinking such reasoning will help. For example, if a teacher points to a set of objects and says “all granny smiths”, then from FSA, the logical consequence of “all” allows the learner not only to acquire positive exemplars for learning to recognise granny smiths (i.e., those objects the teacher pointed to) but also negative examples (all those objects in the scene that the teacher did not point to). Furthermore, FSA can enable the agent to better quantify the value of certain queries. In the example conversation, “is this a granny smith?” (L1 in Figure 1.1) is a query learner decides to ask the teacher which is cheaper than “show me the two granny smiths?” (L5 in Figure 1.1) query in terms of the teacher’s effort: roughly speaking, it obliges the teacher to point to fewer things, and on average the teacher will be able to respond with a quicker search for correct denotations within the visual scene. However, on the other hand, the answers to L1 are less informative than those of L5: a response to L1 will provide the learner with only one positive exemplar, while a response to L5

will provide several positive and negative exemplars. Finally, FSA can help the learner to understand the consequences of the teacher’s corrective feedback. For instance, recognising that T3 is a *correction* of the agent’s latest move implicates that the picked object is *not* a granny smith. In other words, by recognising this type of speech act, the agent can infer not only the linguistically explicit content (the picked object is a golden delicious) but also linguistically implicit content that is necessary for the coherent use of this type of speech act (that it is not a granny smith). Overall, then, reasoning about the logical consequences of the conversation, at both the sentence level and speech act or discourse coherence level, will, on average, provide additional evidence for the learner to exploit as the agent attempts to estimate an accurate domain model and solve its task.

Processing embodied conversation using such FSA brings many opportunities and challenges for ITL. On the one hand, processing embodied conversation allows one to move beyond supervised learning involving object-label pairs and enables the learner to utilize incidental supervision signals (Roth, 2017) from reasoning about the context in which the information is provided and the agent’s current background knowledge about the domain. Fully utilising such supervision signals is essential for ITL because the teacher inherently expects the learner to exploit all knowledge and cues they provide. On the other hand, embodied conversations like the one in Figure 1.1 are not handled by contemporary methods of Natural Language Processing (NLP) in a fully principled way, taking advantage of both: (a) symbolic inference (e.g., the entailments supported by words like “the two” in L4 of Figure 1.1 and the constraints on satisfaction imposed by coherent corrective moves like that in L3); and (b) quantitative reasoning (e.g., the likelihoods that two objects are denoted by the same symbol, given how similar their observed visual features are). The development of appropriate methodologies stems from many areas of Artificial Intelligence (AI): it requires the use of data-driven representation learning methods (Bengio et al., 2013) to enable robustness when processing observational data to conceptualize the domain and the use of rich knowledge about patterns of symbolic and probabilistic reasoning (Holyoak and Morrison, 2005) to fully exploit the learning signal and to make strong generalizations.

The quest to take advantage of both FSA and neural methods for learning about object similarity and denotations of open class words when processing embodied conversation for ITL leads to the second hypothesis of the thesis:

Thesis hypothesis II: An ITL agent that processes embodied conversation using FSA masters tasks more efficiently than those that do not use it.

The tasks to be mastered include both the ability to perform the task that the teacher instructed, with their guidance throughout the learning process and performing the task without the teacher monitoring the agent's performance. This thesis develops computational models to evaluate and answer these hypotheses using qualitative and quantitative evidence.

1.2 Contributions

This thesis develops an ITL agent to process embodied conversation by exploiting FSA. The main contributions of this thesis are as follows:

1. Development of an interactive neuro-symbolic grounding model, which can process referential expressions like “the one red cube” and its designation in the environment, even if the phrase features a neologism that denotes a concept that is entirely absent from the agent's current model of the domain. This grounding model is the first one to our knowledge that combines the following two features, which make it particularly suitable for ITL: (a) it learns incrementally to ground unforeseen concepts; and (b) it utilizes FSA (in particular, the logical consequences of quantifiers and negation) to extract as much information as possible from the teacher's signals.
2. Design a method for learning an effective dialogue strategy that copes with and overcomes the agent's unawareness of concepts critical to achieving its task in the environment. This capacity to learn when (and what) to ask the teacher vs. when (and what) to do in the domain enables the ITL agent to engage in *active learning*, in a scenario where the domain is non-stationary thanks to the ever-expanding hypothesis space of possible states and actions due to the information exchange that occurs in the dialogue, which triggers the discovery of unforeseen concepts (via neologisms).
3. Deployment and evaluation of an ITL agent that can engage in conversation before executing actions in the domain, using its models for grounding and decision making, together with the ability to process the teacher's corrective feedback during execution of its plan, utilising the FSA of the coherent corrective move.

The work presented in this thesis has been published and is under review as part of the following papers:

Rubavicius, R. and Lascarides, A. (2022). Interactive symbol grounding with complex referential expressions. In Carpuat, M., de Marneffe, M.-C., and Meza Ruiz, I. V., editors, *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4863–4874, Seattle, United States. Association for Computational Linguistics

Rubavicius, R., Lascarides, A., and Ramamoorthy, S. (in press). Learning neuro-symbolic dialogue strategies for interactive symbol grounding. *Linguistic Issues in Language Technology*

Rubavicius, R., Fagan, P. D., Lascarides, A., and Ramamoorthy, S. (2025). SECURE: Semantics-aware embodied conversation under unawareness for lifelong robot learning. In Eaton, E., Liu, B., Mahmood, R., Rannen-Triki, A., Chandar, S., and Pascanu, R., editors, *Proceedings of The 4th Conference on Lifelong Learning Agents*, Proceedings of Machine Learning Research (PMLR)

1.3 Task Overview and Scope

When an embodied intelligent agent acts in the environment, it decides what to do by *planning*. There are several kinds of planning an agent engages in. Firstly, there is *motion planning* (LaValle, 2006): that is, planning joint poses and control to produce desired torques to move and manipulate the environment, taking into account considerations like collision avoidance and trajectory generation. Studying motion planning is outside the scope of this thesis: in our experiments that require environment manipulation (Chapter 6), it is manipulated using industrial motion planners (Chitta, 2016).

The second kind of planning is *task planning* (Haslum et al., 2019): this is planning a sequence of actions to take an agent from its current state (or, in partially observable environments, its beliefs about the current state) to a goal state, using its model of the domain and a library of high-level actions like picking and placing objects.¹ Studying task planning is also outside the scope of this thesis. In our experiments, the task

¹Domain model is often referred to as world model in the related literature. This thesis explicitly uses domain models to emphasize the connection with task-specific knowledge, as opposed to the dynamics of the world, a.k.a. transition model. Furthermore, we aim to avoid the confusion with the notion of a world in possible worlds semantics used for modal logics (Blackburn et al., 2007).

plans are either trivial (e.g., showing the referent in Chapters 4 and 5) or very simple (Chapter 6). The combination of Task and Motion Planning (TAMP) (Garrett et al., 2021) provides a robust approach for solving long-horizon planning problems, requiring multiple task-level and motion-level actions to be executed in the environment.

In ITL scenarios, the agent has an additional dimension of planning because it faces a choice at any given point: to engage in an interaction with the teacher—and if so, to choose what to say, such as choosing a particular question to ask—or to act in the environment. We refer to this type of planning as *interaction planning*. Such planning is a core decision-making problem studied in this thesis. It includes both the decision on whether to interact with the teacher to learn more about the task or instead to use the agent’s current domain understanding and risk executing a plan to achieve the task. If the decision is to interact, then this has to be accompanied by a decision on what that interaction should be. This is particularly important in our scenario as we are studying tasks under unawareness in which agent has to update their hypothesis space over their lifecycle. This thesis only considers embodied conversation as the means of natural interaction in ITL. Thus, our interaction plan boils down to having a *dialogue strategy* to engage with the teacher.

This thesis makes several idealised assumptions and constraints to evaluate the utility of FSA of embodied conversation messages in isolation from other potential confounds:

- **Domain conceptualisation:** we consider a first-order domain model, represented as a set of objects in the environment, a set of concepts in the domain, a.k.a. the vocabulary, and an interpretation function mapping symbols in the agent’s vocabulary to their denotations, for 1-place predicates (properties) it is a set of objects, for 2-place predicates (relationships) it is a set of pairs of objects, or in general for an n -place predicate it is a set of n -tuples. ITL agent can sense and correctly identify objects and their corresponding visual features but has no clue about the interpretation function: the learner does not know the denotation of each symbol, but perhaps more profoundly, starts the learning process not knowing the vocabulary of symbols either.
- **Perfect semantic parsing:** ITL agent has a perfect mapping from natural language utterances and pointing gestures to their symbolic representation—i.e., the logical forms and gesture-denoted objects in the domain. The process for map-

ping referential expressions to their logical forms is specified using in-context learning and provides a robust method to parse the (short) utterances studied in this thesis, while accurate designations come from object detection and no occlusion in the environment.

- **Teacher’s knowledge and behaviour:** teacher in ITL is assumed to know perfectly the underlying domain model and is cooperative to with the learner giving them truthful guidance —i.e., there is no deception nor accidental inaccuracies in their assertions. Moreover, the teacher’s feedback is incremental and interactive (one message at a time), and there is an assumption that these messages are conveyed in a timely fashion, making them relevant to the current state of the environment. When modelling teacher-learner interaction we make *no* assumptions about their pragmatic behaviour such as Gricean maxims (Grice, 1975), and we are not exploring how the teacher can be strategic in providing information in a pedagogical manner (Shafto et al., 2014) with all interaction initiative residing with the learner.

The rest of the thesis is structured as follows. Chapter 2 discusses the requirements and desiderata for processing embodied conversation for ITL, highlighting the problem we tackle and the novel approach this thesis takes. This concludes with essential and desired capabilities of ITL agent for our experiments. Chapter 3 discusses a generic embodied lifelong agent’s architecture that can draw on both logical and probabilistic reasoning during interactive task learning. It also details how we represent the agent’s knowledge and the reasoning that the agent has access to. Chapter 4 describes experiments with learning to ground complex referential expressions (e.g., “every red square to the left of the one blue triangle”), in contexts where the agent starts its learning process unaware of all open class words (e.g., “red”, “square”, “blue”, “triangle”) and the concepts they denote. We show in these experiments that using the insights of fFSA makes learning more data efficient. Building on it, Chapter 5 discusses how the learner can use the insights of FSA to deploy active learning, i.e., to learn a policy on when and what to ask the teacher to solve the task more efficiently. Chapter 6 studies teacher’s corrective feedback and describes experiments that study the value to the ITL agent of interpreting such feedback using FSA. Finally, Chapter 7 summarizes the thesis findings and gives future research directions in processing embodied conversation for ITL.

Chapter 2

System Requirements

This chapter surveys the related literature of the key themes of this thesis to motivate specific requirements for designing an embodied lifelong agent that processes embodied conversation for ITL. We highlight the problem space, the requirements we must impose on the learning capacity of an agent that copes with the kinds of scenarios we described in Chapter 1, and the uniqueness of our approach. For illustration, this chapter uses the design of a personal robot for cleaning a house as a running example.

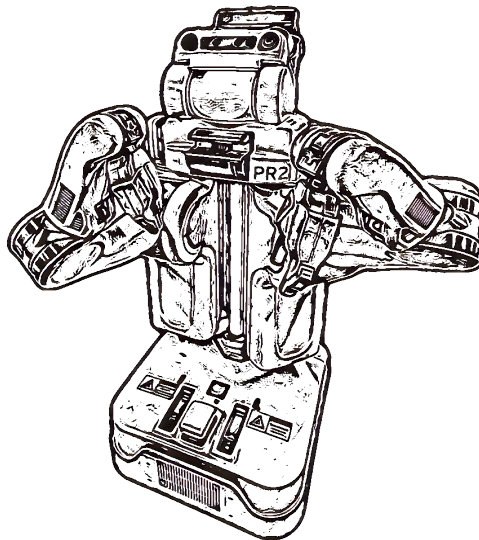


Figure 2.1: Personal robot that could be used to clean the house.

2.1 Interactive Task Learning Capabilities

Intelligent autonomous agents are designed to perceive the environment using sensors and act using actuators in a way that (on average) maximizes some performance measure to solve various tasks (Russell and Norvig, 2020). For example, the personal robot uses sensors (e.g., cameras, LiDAR, GPS) to navigate the rooms and manipulate objects to clean the home, which might involve several intermediate tasks like organizing the dishes in the kitchen. Reinforcement Learning (RL) is a way of designing such intelligent agents. It supports learning from experience, in particular observations about the consequences of its actions (or the actions of others), given its observations about the state in which those actions were executed. This experience is used to acquire a *policy*—a function from its beliefs about the current state, given its sensors, to an action—that is optimal in the rational sense: the chosen action maximises the agent’s expected benefits, given its beliefs (Sutton and Barto, 1998).

For instance, a cleaning robot could be trained using offline RL (Levine et al., 2020), utilising a replay buffer of the experience gathered by physical interaction with the real world or by interacting with a simulator which additionally requires addressing reality gap (Zhao et al., 2020). Interactive Task Learning (ITL) (Laird et al., 2017) considers environments that include another agent that is a domain expert that assumes the role of the *teacher*, providing additional experience to the learner through natural interaction. For a personal robot, the user guides the robot on how the house should be cleaned with task instructions like “put all but two mugs inside the dishwasher”.

Home cleaning is a complex endeavour involving various tasks and various domains (i.e., different homes), in which the personal robot learns after deployment: e.g., the home may contain strange and unfamiliar objects, and the robot will need to tailor its behaviour to the user’s preferences like some users will want all dusting in all rooms before they are vacuumed, while others might prefer the robot to clean each room in turn. Lifelong Machine Learning (LML) (Chen and Liu, 2018) considers agents that can learn many tasks over their lifecycle. They effectively retain and reuse the experience they have acquired so far (both before and after deployment) and use it to solve new tasks in new domains (Silver et al., 2013). It has motivated research in transfer learning (Pan and Yang, 2010) in how to transfer the experience between tasks, multitask learning (Caruana, 1997) in using the experience from multiple tasks to bootstrap learning and meta learning (Vilalta and Drissi, 2002) of learning to learn

new tasks.

When evidence is gathered through interaction with a teacher, then a key feature of LML is that learning must be *incremental*: the agent has to update beliefs and behaviour in real-time (online), as and when they have a new experience. This contrasts setups in which the experience is *batched*, and the agent is trained offline (Mohri et al., 2012). The necessity for incremental learning is fundamental for personal robots: the user will expect the robot’s behaviour to be affected immediately by any information that the user imparts as and when they impart it. Following the van de Ven et al. (2022) taxonomy, ITL is a class-level incremental learning problem, as the agent is learning both the task and how to represent the domain.

Experience in ITL can lead to a change to the agent’s representation of the domain with two possible domain updates. The first is *parameter-level*, in which the underlying structure of the domain representation stays the same, but its parameters are tuned to better match the domain distribution. The second *structure-level*, in which the domain representation itself changes: acquiring skills and concepts that were not previously a part of the domain at all, thereby expanding the hypothesis space of possible domain states and actions. Generalist agents (Reed et al., 2022) that are trained offline via batch learning are designed to support broad domain coverage and parameter-level updates during interaction with the user. They offer tangible and immediate gains for designing ITL agents (Wray et al., 2021) but are not sufficient: there are natural interaction scenarios where a robot will not have been exposed in any way to a particular concept that is critical for decision making in its current environment. For instance, the personal robot could be asked to discriminate household items like a Chardonnay glass and a Burgundy glass, but their (visual) sensory processing that benefited from pretraining did not include knowledge that supports distinguishing different apple varieties, or different types of wine glasses, at all. In this case, to meet the user’s needs, the agent will need to perform structure-level domain updates to learn about poorly-known unknowns: they must expand the hypothesis space of possible domain representations and adapt what it has learned so far to this newly expanded space of possible states. To deal with structure-level updates, the domain representation should be *compositional* to support systematic generalization and to enhance the reusability of prior experience to inform current learning (Hupkes et al., 2020; Mendez and Eaton, 2022).

The scenarios and situations elicited above indicated the following ITL capabilities for the agent: LML should support class-level incremental learning and compositional

domain representation so as to tackle structure-level domain updates.

2.2 Embodied Conversation Capabilities

When a user utters “put all but two mugs inside the dishwasher”, they establish a communication channel of embodied conversation (Cassell, 2001) with the personal robot, which enables active learning (Settles, 2012): the learner can query the teacher via this natural interaction to control the evidence that it observes next via the teacher’s response. The decision of whether to query the teacher, or aim to solve the task without guidance is the agent’s dialogue strategy, or generally an interaction plan. In ITL, the purpose of this communication channel is to achieve common ground (Clark, 1996). In particular, the teacher and agent must be able to resolve the references of various linguistic expressions in their shared embodied environment, as is necessary for the robot to understand and execute the user’s instructions and subsequent instructions.

When studying embodied conversation in experimental settings, an image or a simulation has been used as the environment, and the conversation includes referring expressions, occasionally accompanied by pointing gestures towards objects in the environment (Das et al., 2017; Haber et al., 2019; Kottur et al., 2019; Loáiciga et al., 2021). In such settings, both the sensory and verbal context have to be used to resolve dialogue-specific phenomena such as co-reference (Kottur et al., 2018) in dialogue messages like “the rest should stay on the table”, where “the rest” is an anaphoric expression, whose interpretation requires identifying its antecedent; if it is uttered subsequent to “put all but two mugs inside the dishwasher” then it resolves to two mugs, and to carry out the instruction the robot must choose which two mugs in the environment it plans to leave on the table. In this thesis, however, we avoid problems of co-reference and anaphoric resolution between embodied conversation messages. We treat reference resolution as the task of identifying for each referential expression in the embodied conversation the objects in the environment it denotes, thereby making our task correspond to the symbol grounding problem (Harnad, 1990).

Signalling games (Fudenberg and Levine, 1998) are often used to model the pragmatics of embodied conversation, with various definitions of game equilibria used to predict what speakers choose to say and how their interlocutors interpret the speaker’s (ambiguous) signals (DeVault et al., 2005; Thompson and Kaufmann, 2010; Caelen and Xuereb, 2011). The probabilistic pragmatic framework of Rational Speech Acts

(RSA) (Frank and Goodman, 2012; Goodman and Stuhlmüller, 2013; Goodman and Frank, 2016) provides a game-theoretic view of such interactions by explicitly defining the speaker and listener behaviours recursively (Andreas and Klein, 2016; Monroe et al., 2017; Zariëß and Schlangen, 2019; White et al., 2020; Fried et al., 2021). One kind of signalling game models pedagogical aspects of the teacher-learner interaction (Shafiq et al., 2014; Rafferty et al., 2016)—in these games, the initiative is given to the teacher to transfer the desired knowledge to the learner, utilising various cues and biases (Csibra and Gergely, 2009; Jara-Ettinger et al., 2016). For instance, Liu et al. (2016); Chai et al. (2018); Gu et al. (2024) considers machine teaching scenarios in which ITL is achieved with a mixture of physical demonstrations (e.g., how to perform physical placement to the dishwasher) and teaching about new objects using their attributes (e.g. “mugs are made of ceramic and have one handle.”). In this thesis, however, we are considering the opposite scenario: namely, one where the initiative resides with the learner to choose a query to address to the teacher, whose response is maximally likely (given the learner’s current beliefs) to help in solving the task. Most teacher-learner scenarios between humans involve initiative on both sides, and it would be interesting to explore equilibria in situations where both the teacher and the learner can take the initiative in conversation, but that is beyond the scope of this thesis.

The RSA framework has modelled cooperative communication in general (Wang et al., 2020a; Hao et al., 2023) with theoretical and empirical results aligning with established observations of pragmatic implicatures and other context-sensitive aspects of discourse (Grice, 1975; Sperber and Wilson, 1986). Nevertheless, modelling embodied conversation between the learner and a teacher in an ITL setting as a collaborative game that draws on formal game theory is insufficient. By design, formal game theory assumes perfect knowledge about the hypothesis space: all agents share perfect conceptualisation about what the possible domain states are and what the possible (dialogue) actions are. However, having this level of knowledge is not always reasonable. Asher and Lascarides (2013) argue that it is too strong an assumption in certain strategic conversations. Thanks to the information exchange that occurs in dialogue, an interlocutor may learn of domain-level possibilities (e.g., the existence of the concept “mug”) that it was unaware of before the conversation—in effect, the player discovers they are playing a game with a different hypothesis space of possibilities than they assumed when they started. Once the unforeseen possibility is discovered, e.g., by the teacher uttering task instruction, the agent should refine and extend its hypothesis space

of domain states accordingly—in other words, it calls for a structure-level update. For RSA, such scenarios are not permitted. Being built on game theory, a conversation that entertains a different (or larger) set of possible domain states and/or signals is treated as a different and unrelated game. But in ITL (and more generally in lifelong learning), the agent, on discovering that their hypothesis space of possible states is deficient, should retain at least some of what they have learned so far and use it to influence actions in the ‘larger’ game, even within the current embodied conversation.

Indeed, scenarios in which the learner has a deficient domain model (unknown unknowns) are a key area of interest in this thesis. Because of this, the computational models developed are related to but do not simply reuse RSA. There are approaches of dialogue modelling that do not explicitly model the conversation as a game but make the speaker *listener-aware*, to modulate their decision making in inference (DeVault et al., 2005; Vedantam et al., 2017) and learning (Mao et al., 2016; Yu et al., 2017) or implicitly capture well-known patterns inefficient communication in human-to-human conversation like reference reuse (Takmaz et al., 2020), descriptiveness (Takmaz et al., 2022), information-density (Giulianelli et al., 2021) and discourse context (Giulianelli and Fernández, 2021). This thesis aims to develop a framework in which learning and inference are *semantics-aware*, meaning that the learner can perform and exploit FSA when processing the embodied conversation.

At the same time, this thesis uses artificial embodied conversations, rather than humans conversing ‘in the wild’, so that we retain the control that is necessary for quantifying the difference in performance between ITL agents that utilize FSA vs. agents that do not. How often such insights help in real human-human conversation depends heavily on how the task at hand influences that dialogue exchange. For the personal robot, usage of “all but two” has a logical consequence that there are at least two mugs in the environment, while if instead the user instructed “put a mug inside the dishwasher”, such a implication would not be valid. Note that both RSA-style game-theoretic approaches and listener-aware dialogue modelling approaches discussed above have been effective in showing how speakers can adapt in interaction to misaligned beliefs and asymmetry in knowledge for parameter-level model update (Wang et al., 2016; Wang, 2017; Hawkins et al., 2020; Takmaz et al., 2023), but we want to support structure-level domain updates, not just parameter-level updates.

Overall, this thesis aims to develop an agent with the following embodied conversation capabilities: the ability to engage in active learning using embodied conversation,

perform structure-level domain updates, and be semantics-aware, which is the ability to perform FSA.

2.3 Symbol Grounding Capabilities

Symbol grounding is the problem of learning a mapping from symbols—e.g., the symbol mug corresponding to the word “mug”—to their denoted objects, given sensory observations. To understand and execute instructions, the agent must perform symbol grounding (Harnad, 1990). Symbol grounding has been extensively studied (Hu et al., 2016a,b; Du et al., 2021; Chandu et al., 2021), including in designing autonomous robotic agents (Matuszek, 2018). But for the task this thesis is interested in, many of the existing grounding models are insufficient because they do not meet all the desired features required when processing embodied conversation for ITL.

Firstly, from the Machine Learning (ML) perspective, we are tackling *multilabel classification with an expanding hypothesis space*. It might be tempting to formalise the problem as a multiclass classification problem in which, given object sensory observations, a single symbol is predicted (Krishna et al., 2016). However, such a modelling approach assumes mutual exclusivity between symbols, which might be false. For example, two symbols mug and container are not mutually exclusive because the former is a hyponym of the latter (“is a” relation). Because of this, a *multilabel classification* approach needs to be deployed in which, for each object in the environment, multiple symbols can be predicted. Furthermore, the key feature of the task we are interested in is that the teacher may utter a novel symbol (or neologism); that is, a symbol that was not a part of the learner’s vocabulary at all. In this case, that new symbol may refer to an existing concept (i.e., it is synonymous with an existing symbol for which a mapping to sensory observations is already acquired, which is a further motivation for a multilabel classification problem), or it may denote an entirely new, and unforeseen concept. Thus in this thesis, a symbol grounding model must be able to handle an *expanding hypothesis space* of domain-level concepts and what they denote.

Secondly, the symbol grounding model should support learning that is *online*. Many grounding models (Ye et al., 2019; Datta et al., 2019) heavily utilize offline learning for extensive fine-tuning and batch learning, which results in data regularization for the stochastic optimization procedure. Even though these design decisions lead to increased performance, such grounding models are unsuitable for learning from *inter-*

action. The knowledge is acquired piecemeal during the embodied conversation from the sequentially uttered messages. This knowledge has to influence the agent’s current beliefs—the user will expect the agent to change behaviour based on and when they utter their guidance. Thus the agent must learn *online*. Note that in many real-world scenarios, *occasional* offline updates are desirable. As an agent’s experience is expanding over time, maintaining it is unreasonable over the potentially long lifecycle of the agent. To bound this complexity, the experience from interaction could be integrated into offline learning by fine-tuning (Yosinski et al., 2014) using batch learning techniques. Such learning is not lossless and can lead to catastrophic forgetting (McCloskey and Cohen, 1989), but does bound inference time and space complexity. Nevertheless, such offline updates are out of this thesis’s scope.

Thirdly, due to data scarcity, learning should happen in a *few-shot* manner. Even when extensively utilising the learning signal from the embodied conversation, it is unreasonable to expect that sufficient data is acquired through interaction to deploy data-intensive learning methods. Hence, few-shot learning approaches that generalize from a few learning exemplars (Wang et al., 2020b) should be used.

Finally, the method has to be able to *reason about beliefs*. Because of the strong assumptions that are necessary for efficient knowledge acquisition, some of the agent’s assumptions will likely bring inconsistencies in the beliefs about the underlying ground-truth domain model (unknown knowns). For example, the user’s latest message is inconsistent with the agent’s previously made predictions. Because of this, the method of active knowledge acquisition should be able to detect inconsistencies and be able to revise and repair beliefs accordingly (Hansson, 2022).

Grounding models are key components for more complex tasks like visual question answering (Antol et al., 2015) or manipulation (Alomari et al., 2017a). For these tasks, grounding models are often composed by exploiting the principle of compositionality (Mao et al., 2019; Wang et al., 2023)—i.e., that the meaning of a phrase is a function of the meaning of its parts and how they are put together. This compositionality enables the grounding models to exhibit systematic generalization (much like the LML scenario). They acquire a grounded natural language lexicon (Mao et al., 2021) and/or grammar rules (Alomari et al., 2017a,b) from evidence. In this thesis, we do not learn mappings from natural language to logical form. But we exploit compositionality in another way: namely, unpacking the truth conditions of arbitrarily complex logical forms is compositional. In this way, this thesis explores how the valid consequences

of the logical forms of natural language expressions assist generalization during learning, enhancing data efficiency. Such notions of compositionality are not at odds with exploiting it to aid grounded semantic parsing, but complementary.

From the symbol grounding perspective, the grounding model in this thesis is most similar to [Alomari et al. \(2022\)](#). They perform incremental updates of visual features to arrive at a set of concepts that are (roughly) clusters of visually similar situations. But when their system processes language, those incrementally acquired domain concepts, based on visual processing, are fixed. Consequently, symbol grounding is restricted to mapping words to that fixed set of domain concepts. This means that they cannot reliably distinguish lexical relationships, particularly synonyms and hyponyms, in those situations where words denote visually similar concepts: e.g., burgundy glass vs. chardonay glass vs. wine glass. In contrast, in this thesis, we aim to develop computational models that incrementally learn from both vision and language *simultaneously* so that the set of domain concepts is not determined solely by visual information but also by the way those concepts are described and referred to by the teacher. Further, because [Alomari et al. \(2022\)](#) does not support incremental learning from language, their models are not suited to ITL via embodied conversation, as discussed above.

Finally, [Dobnik et al. \(2022\)](#) offers a careful study of meaning representations of embodied conversation in the context of symbol grounding. They advocate using probabilistic Type Theory with Records (TTR) ([Cooper et al., 2015](#); [Cooper, 2023](#)) to model symbol grounding from both visual percepts ([Larsson, 2013](#); [Larsson et al., 2021](#)) and definitions ([Larsson, 2021](#); [Noble et al., 2022](#); [Noble and Ilinykh, 2023](#)). TTR offers a rich framework for representing natural language utterances. However, this thesis does not aim to commit to a specific framework for representing natural language meaning. Rather, we aim to design a general processing framework that can later be adopted by different communities when they design more robust natural language interfaces, handling richer interactions as part of the embodied conversation from a linguistic and perceptual point of view. For this reason, we utilise classical first-order logic as the framework for expressing formal semantic representations of natural language utterances.

To summarize, this thesis develops a grounding model that has the following capabilities: (a) performs multilabel classification with an expanding hypothesis space; (b) learns few-shot, incrementally, and online from evidence coming from language and sensory observations; and (c) reasons about its beliefs and revise them when new con-

tradicting evidence is acquired.

2.4 Reasoning Capabilities

In the early days of AI, the design of intelligent agents was dominated by relational reasoning (van Harmelen et al., 2008), which utilizes symbolic knowledge representations and their underlying logics to engineer and encode domain-level knowledge. For instance, for a personal robot, the symbolic knowledge could capture the topology of the house—e.g., which rooms are connected to each other—or facts like “to get an object from a drawer, you first need to open it”. Inference in relational reasoning is achieved by using domain-independent axioms such as *modus ponens* to synthesize new knowledge. One of the thesis hypotheses is that domain-independent axioms of symbolic reasoning, as captured in symbolic logic, are key when processing embodied conversation for ITL, because it enables the learner to make non-trivial inferences from the FSA of the teacher’s utterances, as alluded to in Section 1.1.

Relational reasoning has (at least) four limitations, however. Firstly, it supports reasoning about relations among open class words like “mug” only if someone specifies relevant domain-level rules that capture those relations; thus, it requires knowledge engineering. Second, classical symbolic logic on its own does not fully capture the defeasibility of such relations and the unbounded number of exceptions to any given rule.¹ For example, the statement “mugs are stored in the cupboard” is only partially true; there are exceptions, such as dirty mugs, which are in the dishwasher, or the mug that just broke, which is in the bin. Thirdly, symbolic logic by itself does not quantify the extent to which an agent may be uncertain about what is currently true or false due to partial observability (e.g., what mugs are already in the dishwasher?), non-determinism (e.g., placing a mug inside the dishwasher may not succeed) and adversaries (e.g., someone might remove a mug from the dishwasher without the agent observing it). Finally, it is hard to integrate into the symbolic representations that the logic supports information that can be inferred from sensory observations, like the personal robot’s camera feed, which yields estimations of the current state that are inherently dense and noisy. Reasoning used for processing embodied conversation in ITL has to overcome these limitations to scale to real-world tasks.

¹There are symbolic approaches to defeasible reasoning (Reiter, 1980; Asher and Morreau, 1991), but these do not quantify the *extent* to which a given rule or generalisation is defeasible.

To overcome the second and third limitations of relational reasoning, statistical relational reasoning (Getoor and Taskar, 2007) aims to integrate probabilistic semantics into logic, leading to different types of epistemological commitments. In logic, facts are evaluated as true, false or unknown in some three-valued logics (Post, 1921), while probabilistic semantics allows one to express the degree of belief about a particular proposition. Such probabilistic semantics in relational settings yields a Probabilistic Relational Model (PRM) (Friedman et al., 1999; Koller and Friedman, 2009)—examples of PRMs include Markov Logic (Richardson and Domingos, 2006), probabilistic logic programming (Raedt et al., 2007; Baral et al., 2009), answer set programming (Lifschitz, 2008), and probabilistic circuits (Choi et al., 2020). Inference in PRMs is achieved via queries like marginal and conditional probabilities, expectation or maximum a posteriori. These queries are computed by counting while complying with constraints of valid domain configurations, otherwise known as Weighted Model Counting (WMC) (Chavira and Darwiche, 2008). Computation of such queries, in general, is an NP-hard problem, but for many queries of practical interest, there are tractable algorithms of computation in poly-time (Choi et al., 2020) or anytime variations (Dubray et al., 2024). When processing embodied conversation for ITL this thesis uses PRM representation of Probabilistic Sentential Decision Diagrams (PSDD) (Kisa et al., 2014) to define distributions over domain models given propositional knowledge but in principal does explicitly use this representation but just use it as inference engine for WMC with other representation mentioned above could have been used instead as viable alternative.²

To overcome the first and fourth limitations of relational reasoning, methods of neuro-symbolic reasoning (Garnelo and Shanahan, 2019) have been emerging in recent years that aim to equip statistical relational reasoning approaches with the advantages of robustness and generalization that are inherent properties of neural networks (Marra et al., 2021; Manhaeve et al., 2021) when applied to processing sensory observations. There are many approaches to integrating symbolic and neural approaches (Sarker et al., 2021), including using neural network predictions as parameters for probabilistic logic programming (Manhaeve et al., 2018) (bottom-up) or using logic as a constraint on the learning process (Xu et al., 2018; Innes and Ramamoorthy, 2020) (top-down).

Overall, this thesis aims to develop an agent that processes embodied conversation for ITL in real-world settings that has neuro-symbolic reasoning capabilities: it utilizes the

²PSDD was chosen due to available open source software for performing desired queries.

strength of symbolic reasoning and can overcome its weaknesses by utilising probabilistic interpretation of logic (statistical relational reasoning) and the predictive power of neural networks for processing sensory observations of the environment.

2.5 Summary

Here we summarize system requirements—desiderata—for embodied lifelong agents to process embodied conversation for ITL based on the elicited capabilities.

Firstly, we distinguish the following *essential* features:

1. *Interactive symbol grounding*: the agent has a grounding model that performs multi-label classification with an expanding hypothesis space, supports learning that is class-level incremental (non-stationary set of tasks and domains), few-shot (learning from small amount of exemplars), and online (learn now, not in some point in the future), as well as able to reason about its beliefs and in case of them being false (unknown knowns) and revise accordingly.
2. *Compositional domain representation*: the agent maintains an internal representation of the environment—a domain model—that is compositional and enables the agent to perform structure-level updates so that when the user issues an utterance that features a neologism (an unforeseen concept), the hypothesis space of possible domain states can be built by augmenting the existing domain representation with the new concept.

Secondly, we distinguish the following *desired* features:

1. *Interaction planning*: the agent can decide whether to engage in embodied conversation and general active learning by querying the teacher about uncertain concepts to use current beliefs to solve the task instructed.
2. *Neuro-symbolic reasoning*: the agent can process both natural language and sensory observations (the former being discrete and the latter being dense) and also employ reasoning to form beliefs about the environment. This reasoning should utilize the strengths of logical and probabilistic reasoning and accept multiple modalities via the predictive power of neural networks.
3. *Formal Semantic Analysis (FSA)*: the agent is semantics-aware, by which we mean that they can utilize the logical consequences arising from various linguistic expressions (e.g., quantifiers and negation) and dialogue moves (e.g., correction) to contribute to their reasoning and learning in the ITL scenarios.

The features above will be enforced when designing an agent to process embodied conversation for ITL.

Chapter 3

Base Architecture

This chapter outlines an architecture for an embodied lifelong agent that processes embodied conversation for ITL, taking into account the system requirements established in Chapter 2 and with a particular focus on knowledge representation and reasoning outlined in Section 3.2. To illustrate architectural features, this chapter uses examples from the canonical blocks domain, which contains objects of various 3D shapes and colours placed on the tabletop workspace.

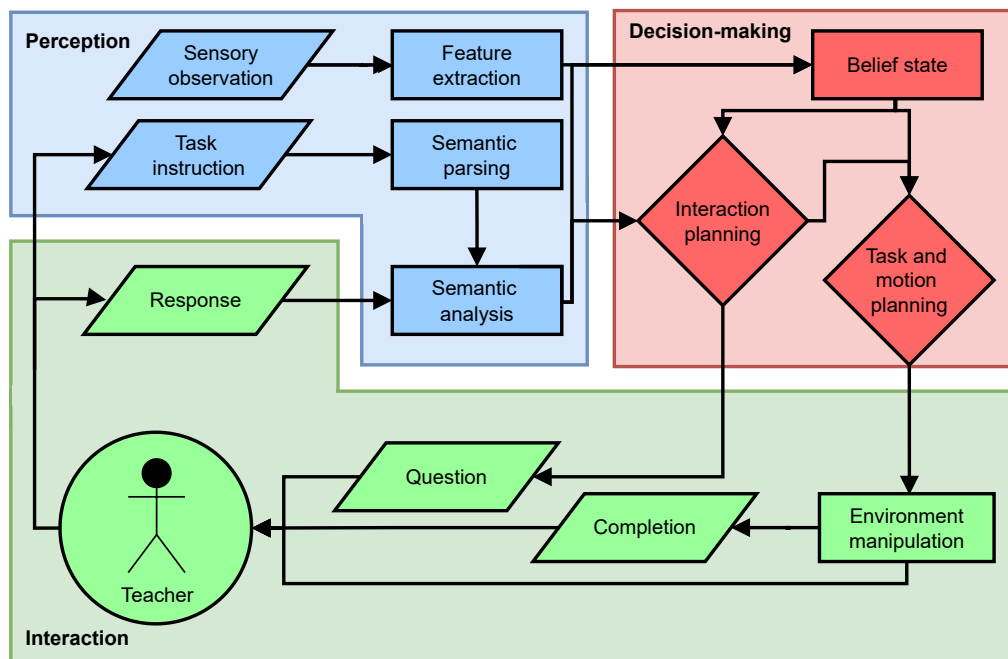


Figure 3.1: Base architecture for an embodied lifelong agent that is designed to solve tasks under unawareness by interacting with the user through embodied conversation.

3.1 Overview

This section gives a high-level overview of assumptions about the teacher’s competence and the system description used for the base architecture.

3.1.1 Teacher’s Competence

To evaluate our thesis hypotheses and conduct our experiments in controlled settings, we make several assumptions about the competence of the teacher. These assumptions simplify the semantic analysis that is integral to our agent’s architecture.

Recall that in Section 1.3, we highlighted key assumptions, particularly concerning the type of knowledge and behaviour exhibited by the teacher. Here, we elaborate on these points. In our setup, the teacher is *cooperative*, meaning that the teacher’s embodied conversation messages are always accurate and relevant. Relevancy is particularly important when interpreting task instructions, as we assume that the teacher’s instructions imply actions to be performed in the environment to reach the goal state. This is important when interpreting universal quantifiers in task instructions like “move every cube to the left of a cylinder” which would be logically valid to say if there are no cubes in the environment, but that would violate cooperativity which has logical consequences that there is at least one object in the environment that is a cube.

In our domain conceptualisation, we also assume that objects in the domain are known and shared between interlocutors. This has implications for grounding and the scope of various linguistic expressions. In our architecture, we assume that the scope of a quantifier is with respect to the *environment* (current state of the world the agents are in) and holds unless the state of the environment changes (e.g., adding or removing objects). This may not always be true, as the statements involving quantifiers may refer not only to the particular environment but to the entire domain (multiple environments with objects having shared concepts). For example, the statement “all cups have handles” intuitively applies not only to the current state of affairs but potentially to many other environments. Detecting such a scope for the use of quantifiers is beyond the scope of this thesis. Our teacher and learner adopt conservative estimation, assuming the scope to be limited to the environment.

When interpreting quantifiers (the main logical expressions studied in this thesis), both the learner and the teacher use Russell’s analysis of definite descriptions (Russell, 1917). Russell argued that definite descriptions, such as “the red cube,” imply the exis-

tence and uniqueness of the described object. In our architecture and particular Formal Semantic Analysis (FSA), we adopt this approach to the interpretation of quantifiers and the logical consequence of their use in the environment. This interpretation allows us to construct logical consequences within the context of a single linguistic expression. There are richer theories of interpretation of quantifiers that take into account not only individual sentences but the overall discourse between interlocutors (Kamp and Reyle, 1993). However, in this thesis, we do not attempt to model how the overall discourse affects reference interpretation and instead use a simpler and more conservative theory for the semantic analysis of quantifiers. To alleviate the possibility of such ambiguities when Russell’s analysis may not be sufficient, our teacher always utters “the one” instead of “the” in their embodied conversation messages.

3.1.2 System Description

We now describe an architecture that aims to satisfy the requirements that we stipulated in Chapter 2. In particular, we focus here on how we process embodied conversation. The architecture depicted in Figure 3.1 acts as a blueprint for the agents that we develop throughout the rest of the thesis. It has three modules: the perception module is used to process and interpret evidence acquired from the environment and interaction; the decision-making module is used to make decisions on what actions the agent should take (given its current beliefs); and the interaction module encapsulates changes in the environment that result from interaction—these changes can result from manipulating the environment and from issuing an utterance in the conversation.

The perception module is responsible for interpreting the observations from the environment and the interaction with the teacher. The input for this module is three-fold: (a) a task instruction, which is expressed in natural language; (b) a sensory observation, which for this thesis is an RGB image; and (c) a response the teacher provides to the learner’s actions in the environment—an embodied message consisting of a natural language expression, as well as in some cases pointing gestures to objects in the image.

The perception module has three components that process these inputs. The semantic parsing (Mooney, 2007) component is used to convert embodied conversation messages¹ to their meaning representations in symbolic logic. The agent uses an off-

¹The semantic parser is used to parse *all* embodied conversation messages, not only task instructions depicted in Figure 3.1, but for simplicity and clarity dependencies to question and completion blocks are not explicitly represented.

the-shelf semantic parser to obtain logical forms of these messages, particularly their natural language expressions. The feature extraction component is used to process sensory observations. As we require a compositional domain representation (which we argued was an essential feature in Section 2.5), this architecture uses object-centric visual processing. In this approach, for each object in the environment, a region of interest is localised as a patch from which an embedding is extracted. Each object has a one-to-one correspondence to an embedding. The agent uses off-the-shelf object detection, localization, and feature extraction models for visual processing.

The semantic component is used to interpret the teacher’s response to the learner’s actions. This analysis takes the logical form produced by the semantic parser and reasons about what follows from the assumption that this logical form is true in the current model or image—i.e., we assume the teacher is competent and sincere, so everything they say is true, and we compute the consequences of this. To quantify the value of computing the logical consequences of the teacher’s message to learning to solve new tasks, our experiments vary the extent to which the learner computes and exploits consequences, ranging from a full Formal Semantic Analysis (FSA) to a more rudimentary base semantic analysis, or (in the extreme) no analysis at all. Details of these variations are described in the experiments in Chapters 4, 5, and 6.

The decision-making module enables the agent to plan using a belief state that encapsulates the agent’s current (probabilistic) beliefs about the state of the environment.² The belief state maintains both certain knowledge constructed from interaction with the teacher and defeasible knowledge that stems from the visual similarity among objects.

The belief state is used for two types of planning components. The interaction planning component is used to decide whether to engage in the embodied conversation by asking the teacher a question (and if so, which question to ask) or to execute the task using the current belief state (i.e., to execute a plan to reach a goal state, which according to the agent’s current belief state is most likely to succeed). As we will be discussing in Chapter 5, interaction planning can take into account the FSA (i.e., the logical consequences) of the teacher’s utterances, including in particular the expected benefits of asking certain questions, based on how the teacher is likely to respond. The Task and

²A belief state is often operationalised as the distribution of possible states the agent is in (Mykel J. Kochenderfer and Wray, 2022), while in our setting a belief state is richer, since our belief states separate certain knowledge about the world (a well-formed formula that is guaranteed to be true) from likely or unlikely information, and thus a more appropriate term would be an epistemic belief state (Bochman, 2007). But for simplicity, we just use belief state as a shorthand for epistemic belief state.

Motion Planning (TAMP) component estimates the most likely domain model using the belief state. This domain model determines the agent's (potential) plan for manipulating the environment to reach a goal state: these plans consist of moving objects on the tabletop via picking and placing actions.

The interaction module summarizes the agent's interaction with both the environment and the teacher. This interaction could be two-fold: (a) engaging in the embodied conversation with the teacher, either by uttering a question or by declaring that the goal state has been reached; and (b) manipulating the environment, with the teacher correcting the learner whenever the agent executes a sub-optimal manipulation move or the learner declares the goal state has been reached when they have not. For both of these corrective scenarios, the teacher, a domain expert, provides a response that can be used by the perception module and, in particular, the semantic analysis component to update the learner's beliefs about the environment.

This architecture aims to fulfil the essential features of an ITL agent from Section 2.5. As we'll show in the subsequent sections of this chapter, and in Chapters 4 to 6, the semantic analysis of embodied conversation (and vision) that we deploy supports calculating the probability of arbitrary propositions and these probabilities are informed by and informs, the interpretation of both language and vision during learning and inference. In other words, this architecture supports a fully interactive neuro-symbolic approach to interpreting language and vision. Further, as we'll see in section 3.2, the agent's belief state incorporates compositional representations with the capacity for structural updates (in other words, changing the hypothesis space of possibilities) without sacrificing already established knowledge and relations. The upcoming Section 3.2 will give the full definitions of the domain model and the belief state and show how the desired neuro-symbolic reasoning and semantic analysis features are achieved.

3.2 Knowledge Representation and Reasoning

This thesis uses compositional representations of the domain model and belief state to represent and reason about the knowledge acquired through the agent's lifecycle. In this section, we define the domain model and belief state and their knowledge representation, using logic and reasoning with probabilistic semantics.

3.2.1 The Domain Model

The domain model is used to encapsulate the environment $\mathcal{E} \in \mathbb{E}$ and to abstract away from innocuous (visual) variations between environments (and even within the same environment)—by innocuous, we mean that the variation does not impact the tasks that the agent must achieve, and through the life cycle the agent has to learn which visual variations matter, and which do not.

Informally, the domain model representation used in this thesis is a finite first-order model from classical logic (Gamut, 1991a). More formally, a domain model $\mathcal{M} \in \mathbb{M}$ is a triple $(\mathbb{O}, \mathbb{V}, I)$ consisting of:

- \mathbb{O} : a set of objects, where each $o \in \mathbb{O}$ is present in the environment $\mathcal{E} \in \mathbb{E}$;
- \mathbb{V} : a set of predicate symbols (predicates) a.k.a. the vocabulary of the domain;
- $I: \mathbb{V} \mapsto \mathbb{O}^*$: an interpretation function mapping predicates $p \in \mathbb{V}$ to denotations: for 1-place predicates it is a set of objects, for 2-place predicates it is a set of object pairs, or in general for n -place predicates it is a set of object n -tuples.

This thesis, without a loss of generality, assumes that the vocabulary \mathbb{V} contains a special predicate symbol `object` that denotes all objects in the domain: $I(\text{object}) = \mathbb{O}$. This domain model representation is compositional, which is essential for the desired system, as argued in Section 2.5. Compositionality of the domain representation is realised by the symbolic structure of the domain model, and updating the domain model can be structural: in other words, an update may extend the vocabulary \mathbb{V} and hence also the domain of the interpretation function I , without affecting other elements in the representation.

This thesis is concerned with ITL agents that start their learning process entirely unaware of concepts (domain vocabulary) that are critical to solving their task, and so they must discover and then learn to exploit these through interaction. Therefore, in the algorithms we develop, and in the experiments on those algorithms, the agent starts out knowing only a part of the full domain model: i.e., the agent knows the objects \mathbb{O} , but both \mathbb{V} and I are completely unknown. In other words, the agent's visual sensors are good enough that they have an accurate estimation of where objects are in the environment \mathcal{E} , but they have no knowledge about what denotations each symbol maps to, and perhaps more profoundly do not know the symbols themselves either.

3.2.2 The Agent's Belief State

The agent's belief state maintains (certain and probabilistic) knowledge that the agent has acquired in their lifecycle from several sources of information: perception, experience, and prior beliefs.

Knowledge from perception comes from the agent's perception module and represents the environment \mathcal{E} and interaction with it. It includes a set of objects \mathbb{O} and their embeddings \mathbb{X} , which are identified through sensory observations. Additionally, it encompasses a domain theory Δ , which is a set of well-formed formulas that hold to the current environment. This domain theory is built from natural interaction to the specific environment.³

Knowledge from experience comes from previous interactions: this includes the interaction so far in the current environment and those from prior environments within the same domain. Knowledge from experience constitutes both the vocabulary \mathbb{V} of known symbols of the domain, as well as the support \mathbb{S} , which is a set of vector pairs consisting of an embedding $\mathbf{x} \in \mathbb{R}^d$ and the accompanying semantic vector $\mathbf{y} \in [0, 1]^{|\mathbb{V}|}$. Each element y_p of the semantic vector indicates the degree of belief that an object with an embedding \mathbf{x} is denoted by the symbol $p \in \mathbb{V}$. Intuitively, these pairs in the support \mathbb{S} give exemplars for the agent of objects that are denoted by, and not denoted by, certain symbols. This support can, in turn, be used to estimate symbols for a new object (we'll see how shortly). These pairings are obtained from both vision and the interactions so far: in particular, any of the teacher's moves in which they utter a referential expression and point to objects in the image that are denoted by that expression. In Chapter 4, and in particular Section 4.3.1, we will discuss in detail how the support is constructed as the embodied conversation proceeds and also how it informs interactive symbol grounding.

Knowledge of prior beliefs about the denotations of symbols is represented using base weights $w_b: \mathcal{H} \mapsto [0, 1]$. These weights can come from either the initial classification based on sensory observation or, in the case of pure unawareness about the relationship between the object and the symbol, we capture complete ignorance about denotation with the *default weight* of 0.5. The domain of these weights is the set of ground

³The domain theory can be seen as knowledge coming from experience and, in particular, previous interactions. We choose to associate this with perception instead of stressing that the domain theory is not retained between the environments (because the objects in different environments differ), while the experience (and what is learned from it about mappings from symbols to visual features) is retained.

atoms \mathcal{H} , described in detail in Section 3.2.3.

Using knowledge stored in the belief state, the agent can reason about the underlying domain model and, in turn, make decisions based on what they believe is the most likely domain model. To do this, evidence from perception, experience, and prior beliefs are combined into *grounded weights* $w_g: \mathcal{H} \mapsto [0, 1]$.⁴

More formally, the belief state $b \in \mathbb{B}$ as a 7-tuple $(\mathbb{O}, \mathbb{X}, \mathbb{V}, \mathbb{S}, \Delta, w_b, w_g)$ consisting of:

- \mathbb{O} : a set of objects in the environment $\mathcal{E} \in \mathbb{E}$;
- \mathbb{X} : a set of embedding with one-to-one correspondence to object in the environment $\mathcal{E} \in \mathbb{E}$;
- \mathbb{V} : a set of *known* predicate symbols (vocabulary, which may be a subset of the true vocabulary in the teacher's domain model);
- \mathbb{S} : support which includes the set of exemplars about symbols and their denotations in the form of embedding and binary labels. These labels encode whether the object with this embedding is denoted by a particular symbol or not (see Section 4.3.1 for details). This support acts as the explicit memory of the knowledge the agent has gathered over their lifecycle.
- Δ : a set of well-formed formulae (domain theory) known to be true for the current environment $\mathcal{E} \in \mathbb{E}$;
- $w_b: \mathcal{H} \mapsto [0, 1]$: base weights function defining prior beliefs about symbol denotations;
- $w_g: \mathcal{H} \mapsto [0, 1]$: grounded weights function giving distribution over domain models, taking into account evidence from perception, experience, and prior beliefs.

For the rest of this section, we will examine how to represent certain knowledge derived from embodied conversation using logic and how to reason with the belief state.

⁴Compared to the domain model representation, belief state representation is not compact as grounded weights w_g are derived from other elements of the belief state. It is, nevertheless, represented explicitly in the belief state to make it a useful abstraction to talk about the belief as a whole and in different granularity, depending on the context.

3.2.3 Logic of Sentences

This section presents the formal meaning representations of sentences: their logical forms and their evaluation against a model using truth-conditional semantics. Our exposition is based on the theory of generalized quantifiers (Barwise and Cooper, 1981) but it is used unusually to be able to talk about references and not only truth-conditions as motivated and explained in detail in Section 3.2.4.

3.2.3.1 Logical Forms

We are interested in natural language sentences or statements like “every cube is a cuboid” that can be represented in first-order logic. The syntax of the language of first-order logic is constructed recursively: non-logical predicate symbols followed by the appropriate number of terms (i.e., a variable $x \in \mathbb{L}_{\text{var}}$ or a constant $o \in \mathbb{L}_{\text{const}}$) are *well-formed formulae*, and these combine with boolean operators and connectives $\mathbb{L}_{\text{conn}} = \{\wedge, \vee, \neg, \rightarrow, \leftrightarrow\}$ and the quantifiers \forall and \exists in the usual way to create more complex well-formed formulae. First-order languages have been used extensively to capture natural language semantics (Blackburn and Bos, 2005): for example, “every cube is a cuboid” has the logical form $\forall x.\text{cube}(x) \rightarrow \text{cuboid}(x)$.⁵

In this thesis, we draw on generalized quantifier theory to represent the meanings of natural language determiners such as “the one” and “both” (Barwise and Cooper, 1981).⁶ Generalized quantifier theory extends the classical first-order language with quantifiers additional to \forall and \exists , and it is designed to express the meanings of other quantitative statements about the domain: e.g., *at least two*, *all but one*, *the one*, *both*, and so on.

Syntactically, a well-formed formula involving a generalized quantifier $Q \in \mathbb{L}_{\text{quant}}$ is of the form $Q x.(\phi, \psi)$, where Q binds the variable $x \in \mathbb{L}_{\text{var}}$, and the restrictor ϕ and body ψ are both well-formed formulae in which x is free or possibly absent in case of a restrictor. Very roughly, the restrictor ϕ expresses the descriptive content of the noun phrase that introduced the determiner whose meaning is Q , and the body ψ expresses the content of ‘the rest’ of the sentence (typically the verb phrase). For example, the statement “the one red cube is o_1 ”, or more explicitly “the one red cube is object o_1 ”,

⁵Given our restricted use case we simplify meaning representations, for instance ignoring tense.

⁶This thesis does not consider vague quantifiers (Channell, 1994) like “most” or “few”, which implicate soft constraints on objects in the model, as a consequence of pragmatic principles of cooperative conversation (Grice, 1975). The interpretation of these quantifiers in cooperative conversation is outside our scope but, in principle, could be captured by the agent following the same architecture.

has the following logical form:

$$\text{the_one_} _q \ x.(\text{red}(x) \wedge \text{cube}(x), \text{object}(o_1)) \quad (3.1)$$

The full list of quantifiers considered in this thesis is given in Table 3.1 column 1.

Formally, the set of syntactically well-formed formulae—the language of sentences \mathbb{L}_{snt} —is defined recursively as follows:

1. If p is an n -place predicate symbol and t_1, t_2, \dots, t_n are terms, then $p(t_1, t_2, \dots, t_n)$ is well-formed formula;
2. If ϕ is well-formed formula, then $\neg\phi$ is a well-formed formula;
3. If ϕ and ψ are well-formed formulas, then $\phi \wedge \psi$, $\phi \vee \psi$, $\phi \rightarrow \psi$, $\phi \leftrightarrow \psi$ are all well-formed formulas;
4. If ϕ and ψ are well-formed formulas, known as *restrictor* and *body*, and ϕ and ψ either do not feature the variable x or x is free in these formulae, and if Q is a quantifier, then $Q \ x.(\phi, \psi)$, is well-formed formula. For this kind of well-formed formula, the variable x is a *bound variable* that is bound by Q .

This thesis restricts attention to predicates that can take one or two terms—in other words, they denote properties and relationships, respectively. For convenience, we also assume that for every object $o \in \mathbb{O}$ in the domain, there is a unique constant o that denotes it. Well-formed formulae constructed from predicates followed with *only* constant terms $p(o_1, o_2, \dots, o_n)$ are known as propositional atoms, or just simply *atoms*, denoted as A .

The set of atoms that can be constructed from the vocabulary \mathbb{V} and objects in the domain $o \in \mathbb{O}$ is known as the set of ground atoms or the Herbrand base \mathcal{H} . We will shortly use ground atoms to define the base and grounded weights $w: \mathcal{H} \mapsto [0, 1]$, which, as mentioned earlier, are part of the belief state (Section 3.2.2). These weights are crucial for probabilistic reasoning about the domain, as described in Section 3.2.5.

Using this formal language to express the logical forms for natural language sentences is convenient because the form of these formulae is close to the surface natural language forms. However, in practice, such logical forms are not directly used for automated inference. Instead, the quantified formulae are first converted via universal instantiation and skolemisation (Bundy and Wallen, 1984) to variable-free representations in propositional logic, containing only atoms and logical operators. For instance,

Quantifier Q	Surface form	Truth condition Q
<code>_exactly_n_q</code>	exactly n	$ \mathbb{RSTR} \cap \mathbb{BODY} = n$
<code>_at_most_n_q</code>	at most n	$ \mathbb{RSTR} \cap \mathbb{BODY} \leq n$
<code>_at_least_n_q</code>	at least n	$ \mathbb{RSTR} \cap \mathbb{BODY} \geq n$
<code>_a_q</code>	a/an	$ \mathbb{RSTR} \cap \mathbb{BODY} \geq 1$
<code>_every_q</code>	all/every	$ \mathbb{RSTR} \cap \mathbb{BODY} = \mathbb{RSTR} \wedge \mathbb{RSTR} = 1^7$
<code>_the_n_q</code>	the n	$ \mathbb{RSTR} \cap \mathbb{BODY} = n \wedge \mathbb{RSTR} = n$
<code>_both_q</code>	both	$ \mathbb{RSTR} \cap \mathbb{BODY} = 2 \wedge \mathbb{RSTR} = 2$
<code>_all_but_n_q</code>	all but n	$ \mathbb{RSTR} \cap \mathbb{BODY} = \mathbb{RSTR} - n \wedge \mathbb{RSTR} > n$
<code>_n_of_the_m_q</code>	n of the m	$ \mathbb{RSTR} \cap \mathbb{BODY} = n \wedge \mathbb{RSTR} = m$

Table 3.1: Generalized quantifiers (column 1), their surface forms (column 2), and truth-conditions Q between the restrictor set \mathbb{RSTR} and body set \mathbb{BODY} , used to evaluate well-formed formulas of the form $Q x.(\phi, \psi)$ (column 3).

for a domain with two objects $\mathbb{O} = \{o_1, o_2\}$, the formula in Eq. 3.1 is equivalent to the following:

$$\text{red}(o_1) \wedge \text{cube}(o_1) \wedge \text{object}(o_1) \wedge \neg(\text{red}(o_2) \wedge \text{cube}(o_2) \wedge \text{object}(o_2)) \quad (3.2)$$

The negative conjunct in this formula follows from the truth conditions of “the one”, specified in Table 3.1, as we’ll see in the next section.

3.2.3.2 Truth-conditional Semantics

Elements of \mathbb{L}_{snt} that are well-formed expressions in generalized quantifier theory are assigned a truth-conditional semantics with respect to a domain model \mathcal{M} and a variable assignment function g , which maps variables to individuals in the model (Tarski, 1931). The truth-conditional semantics define a valuation function $\llbracket \cdot \rrbracket^{\mathcal{M}, g} : \mathbb{L}_{\text{snt}} \mapsto \{0, 1\}$, which in turn specifies when a domain model variable assignment pair (\mathcal{M}, g) satisfies a formula $\phi \in \mathbb{L}_{\text{snt}}$, written $\mathcal{M}, g \models \phi$:

$$\mathcal{M}, g \models \phi \text{ if and only if } \llbracket \phi \rrbracket^{\mathcal{M}, g} = 1 \quad (3.3)$$

⁷If a teacher utters an embodied conversation message with the universal quantifier like “show me all blocks” there is an implicit expectation given the cooperative and collaborative nature of the teacher that there is at least one object in the environment denoted by the word “block”. This does not match the standard logic interpretation of the universal quantifier.

If there are no free variables in $\phi \in \mathbb{L}_{\text{snt}}$ then $\llbracket \phi \rrbracket^{\mathcal{M},g} = \llbracket \phi \rrbracket^{\mathcal{M},g'}$ for all g, g' and without a loss of generality such situation can be expressed as:

$$\mathcal{M} \models \phi \text{ if and only if } \forall g \llbracket \phi \rrbracket^{\mathcal{M},g} = 1 \quad (3.4)$$

To define valuation function $\llbracket \cdot \rrbracket^{\mathcal{M},g}$, λ -expressions of the form $\lambda x.\phi$ are used in which x is a variable that is free or absent from ϕ . Note that such λ -expressions are well-formed expressions but are not in \mathbb{L}_{snt} . Here, they are simply used for defining the truth conditions of formulae that feature generalized quantifiers.

Well-formed formulas in truth-conditional semantics are either evaluated as being true (1) or false (0). This evaluation is given by a standard recursive definition (Gamut, 1991a):

$$\begin{aligned} \llbracket a \rrbracket^{\mathcal{M},g} &= \begin{cases} I(a) & \text{if } a \text{ is a predicate symbol} \\ g(a) & \text{if } a \text{ is a variable or a constant}^8 \end{cases} \\ \llbracket p(t_1, t_2, \dots, t_n) \rrbracket^{\mathcal{M},g} &= 1 \text{ if } (\llbracket t_1 \rrbracket^{\mathcal{M},g}, \llbracket t_2 \rrbracket^{\mathcal{M},g}, \dots, \llbracket t_n \rrbracket^{\mathcal{M},g}) \in \llbracket p \rrbracket^{\mathcal{M},g} \\ \llbracket \neg \phi \rrbracket^{\mathcal{M},g} &= 1 \text{ if } \llbracket \phi \rrbracket^{\mathcal{M},g} = 0 \\ \llbracket \phi \wedge \psi \rrbracket^{\mathcal{M},g} &= 1 \text{ if } \llbracket \phi \rrbracket^{\mathcal{M},g} = 1 \text{ and } \llbracket \psi \rrbracket^{\mathcal{M},g} = 1 \\ \llbracket \phi \vee \psi \rrbracket^{\mathcal{M},g} &= 1 \text{ if } \llbracket \phi \rrbracket^{\mathcal{M},g} = 1 \text{ or } \llbracket \psi \rrbracket^{\mathcal{M},g} = 1 \\ \llbracket \phi \rightarrow \psi \rrbracket^{\mathcal{M},g} &= 1 \text{ if } \llbracket \phi \rrbracket^{\mathcal{M},g} = 0 \text{ or } \llbracket \psi \rrbracket^{\mathcal{M},g} = 1 \\ \llbracket \phi \leftrightarrow \psi \rrbracket^{\mathcal{M},g} &= 1 \text{ if } \llbracket \phi \rightarrow \psi \rrbracket^{\mathcal{M},g} = 1 \text{ and } \llbracket \psi \rightarrow \phi \rrbracket^{\mathcal{M},g} = 1 \\ \llbracket \lambda x.\phi \rrbracket^{\mathcal{M},g} &= \{o \in \mathbb{O} \mid \llbracket \phi \rrbracket^{\mathcal{M},g[x/o]} = 1\} \\ \llbracket Q x.(\phi, \psi) \rrbracket^{\mathcal{M},g} &= Q(\llbracket \lambda x.\phi \rrbracket^{\mathcal{M},g}, \llbracket \lambda x.\psi \rrbracket^{\mathcal{M},g}) \end{aligned} \quad (3.5)$$

where $g[x/o]$ is a variable assignment function just like g except that $g[x/o](x) = o$ and Q is a truth-condition (Table 3.1 column 3) used to evaluate well-formed formulas that feature generalized quantifiers. The value of these truth conditions depends on two sets defined by the λ -expression, commonly referred to as the restrictor set \mathbb{RSTR} and the body set \mathbb{BODY} .

To illustrate, consider the well-formed formula in Eq. 3.1 and the domain model \mathcal{M} with $\mathbb{O} = \{o_1, o_2\}$, vocabulary $\mathbb{V} = \{\text{red, cube, object}\}$ and the interpretation function $I(\text{object}) = \mathbb{O}$, $I(\text{red}) = \{o_1, o_2\}$, $I(\text{cube}) = \{o_1\}$. It is evaluated as being true

⁸As mentioned earlier, we assume a unique constant denotes each object, and so without loss of generality, we can assume $g(a) = a$, where $a \in \mathbb{O}$ is the unique object denoted by variable $a \in \mathbb{L}_{\text{var}}$.

by computing restrictor and body sets and using `_the_one_q` truth-condition:

$$\begin{aligned}
 \text{RSTR} &= \llbracket \lambda x. \text{red}(x) \wedge \text{cube}(x) \rrbracket^{\mathcal{M}} = \{o_1\} \\
 \text{BODY} &= \llbracket \lambda x. \text{object}(o_1) \rrbracket^{\mathcal{M}} = \{o_1, o_2\} \\
 Q_{\text{the_one_q}}(\text{RSTR}, \text{BODY}) &= (|\text{RSTR} \cap \text{BODY}| = 1 \wedge |\text{RSTR}| = 1) \\
 &= (|\{o_1\} \cap \{o_1, o_2\}| = 1 \wedge |\{o_1\}| = 1) \\
 &= \top \wedge \top \\
 &= \top
 \end{aligned} \tag{3.6}$$

3.2.4 Logic of Referential Expressions

In this section, we discuss referential expressions: their logical forms and evaluation using reference semantics.

3.2.4.1 Logical forms

Linguistic expressions that denote objects in the environment are known as referential expressions r . They are expressed by noun phrases: a noun phrase can consist of a determiner and a noun with pre- and post-modifiers (e.g., “the one red cube that is to the left of every green sphere”); a pronoun or demonstrative (e.g., “it”, “this”, “that”), or proper name (e.g., “Edinburgh”, “Kim”, “ Mr. Smith”). This thesis focuses on referential expressions that are of the first kind: i.e.; they include a determiner and a noun with pre-and/or post-modifiers. We denote the logical form of referential expression r as $\Phi(r)$ and the language of these well-formed expressions as \mathbb{L}_{ref} .

Since r is a noun phrase and not a full sentence, $\Phi(r)$ consists of a generalized quantifier Q (corresponding to r 's determiner), its restrictor ϕ (corresponding to r 's adjectives and nouns) but the body of the quantifier is ‘missing’. In semantic analysis, this is traditionally captured using λ -calculus with logical forms like $\lambda \psi. Q x. (\phi, \psi)$ but this is not what we want to capture to process embodied conversation. The λ -term denotes the set of properties satisfied by the objects that satisfy ϕ , with quantifier Q imposing constraints on the relationship between the denotations of ϕ and ψ (and sometimes also on ϕ itself as well). In our scenario, however, we need to identify the set of referent(s) that are denoted by the referring expression r , not the properties satisfied by those referents. Because of these requirements, we express the logical form $\Phi(r)$ of the referential expression r another way and interpret it using the theory of generalized quantifiers (Barwise and Cooper, 1981).

Quantifier Q	Surface form	Referent constructor $\langle Q \rangle^{\mathcal{M}}$
<code>_exactly_n_q</code>	exactly n	$\{A \subseteq \mathbb{O} \mid A = n\}$
<code>_at_most_n_q</code>	at most n	$\{A \subseteq \mathbb{O} \mid A \leq n\}$
<code>_at_least_n_q</code>	at least n	$\{A \subseteq \mathbb{O} \mid A \geq n\}$
<code>_a_q</code>	a/an	$\{A \subseteq \mathbb{O} \mid A = 1\}$
<code>_every_q</code>	all/every	$\{A \subseteq \mathbb{O} \mid A = \mathbb{O} \wedge \mathbb{O} = 1\}$
<code>_the_n_q</code>	the n	$\{A \subseteq \mathbb{O} \mid A = \mathbb{O} \wedge \mathbb{O} = n\}$
<code>_both_q</code>	both	$\{A \subseteq \mathbb{O} \mid A = \mathbb{O} \wedge \mathbb{O} = 2\}$
<code>_all_but_n_q</code>	all but n	$\{A \subseteq \mathbb{O} \mid A = \mathbb{O} - n \wedge \mathbb{O} > n\}$
<code>_n_of_the_m_q</code>	n of the m	$\{A \subseteq \mathbb{O} \mid A = n \wedge \mathbb{O} = m\}$

Table 3.2: Generalized quantifiers (column 1), surface form (column 2), and referent constructor used to evaluate the logical form of referential expressions for the domain model \mathcal{M} (column 3).

Specifically, we represent the referential expression “the one red cube” with the following logical form:

$$\langle \text{_the_one_q } x.\text{red}(x) \wedge \text{cube}(x) \rangle$$

More generally, the logical forms of referential expressions are of the form $\langle Q x.\phi \rangle$, where $\phi \in \mathbb{L}_{\text{snt}}$ is a well-formed formula with only one free variable x . Intuitively, this formula captures the content of the described objects and Q captures the content of the determiner. The notation $\langle \cdot \rangle$ is a convenient way to differentiate these logical forms from well-formed formulae and in particular, the logical forms of sentences.

3.2.4.2 Reference Semantics

The logical form of a referential expression $\Phi(r) \in \mathbb{L}_{\text{ref}}$ is evaluated for the domain model \mathcal{M} to yield the referent \mathcal{R} . An element of \mathcal{R} is a *set* of objects because r may denote more than one object (e.g., any denotation of “two cubes” consists of two objects). \mathcal{R} itself is a set of (potentially) more than one element because \mathcal{M} may have more than one set of objects that are denoted by the referring expression r : e.g., if there is more than one object in \mathcal{M} that is a cube, then “a cube” has more than one referent in \mathcal{M} . For example, given the domain model \mathcal{M} with objects $\mathbb{O} = \{o_1, o_2\}$ and interpretation function $I(\text{cube}) = \mathbb{O}$, the expression “all cubes” should denote $\{\{o_1, o_2\}\}$ (there is only one set denoted by “every cube”, and that is the set of all

objects in the domain that are cubes), while “a cube” denotes $\{\{o_1\}, \{o_2\}\}$ (i.e., there are two sets of objects that are denoted by this referential expression). Thus more formally, \mathcal{R} is always a subset of the power set of \mathbb{O} : $\mathcal{R} \subseteq 2^{\mathbb{O}}$.

These examples reveal two factors that affect the referent of $\Phi(r)$. First, objects in a denotation of $\Phi(r)$ satisfy the restrictor of generalized quantifier.⁹ Secondly, the generalized quantifier imposes its conditions on \mathcal{R} , in particular on the cardinality of each set in \mathcal{R} that is a denotation, and for some quantifiers there is also a constraint on the relationship between a denotation and all the objects \mathbb{O} , which is equivalent to a constraint on the cardinality of \mathcal{R} itself. For instance, each set in a referent for “at least two cubes” must have a cardinality of at least two; for “exactly two cubes” the cardinality must be equal to 2; and “the two cubes” and “both” impose the additional constraint that not only should each denotation have cardinality 2, but also this denotation is *unique* (i.e., \mathcal{R} itself has cardinality 1). Further, the referent for “every cube” is unique because it is the (unique) maximal set of objects in \mathcal{M} that satisfy the restrictor (again, \mathcal{R} must have cardinality 1). In general, constraints imposed by different quantifiers in definite descriptions lead to different logical consequences (Russell, 1917). Note that such constraints are not English-specific: Finnish “molempi” has the same constraint as English “both”.

With this in mind, we obtain a formal definition of the reference semantics for an arbitrary logical form of the referential expression $\Phi(r) = \langle \mathbb{Q} x.\phi \rangle$ by first defining a projection of \mathcal{M} onto a smaller domain model $\sigma(\mathcal{M}, \phi, x)$, which consists of *all and only* those objects $o \in \mathbb{O}$ that satisfy $\phi[x/o]$ (i.e., the formula ϕ with each occurrence of x substituted with the unique constant o that denotes $o \in \mathbb{O}$):

$$\begin{aligned} \sigma(\mathcal{M}, \phi, x) &= \sigma((\mathbb{O}, \mathbb{V}, I), \phi, x) = (\mathbb{O}', \mathbb{V}, I') \\ \mathbb{O}' &= \{o \in \mathbb{O} \mid \mathcal{M} \models \phi[x/o]\} \\ I' &= I \downarrow \mathbb{O}' \text{ (i.e., } I \text{ projected onto } \mathbb{O}') \end{aligned} \tag{3.7}$$

We then define the reference semantics of $\Phi(r)$ for this projected domain model. As discussed before, the truth-conditional semantics of generalized quantifiers are defined using the restrictor \mathbb{RSTR} and body \mathbb{BODY} . But in our scenario, consisting only of referring expressions, the body \mathbb{BODY} is not present for truth-conditional evaluation. However, by evaluating the content of the quantifier for the smaller projected model

⁹In this thesis, we ignore group and mass nouns like “committee” or “water”, respectively.

and making $\text{BODY} = \mathbb{O}'$ (i.e., the objects in that projected model), the reference semantics can focus solely on how they constrain the *cardinalities* of these sets to achieve the above-desired effects for constructing the referent \mathcal{R} . Table 3.2 column 3 shows how the referent is defined using set constructors. These constraints and the definition of model projection in Eq. 3.7 yield the reference semantics (Eq. 3.8) for the logical form $\Phi(r)$ with the desired properties:

$$\mathcal{R} = \Phi(r)^{\mathcal{M}} = \langle \mathbb{Q} \ x.\phi \rangle^{\mathcal{M}} = \langle \mathbb{Q} \rangle^{\sigma(\mathcal{M}, \phi, x)} \quad (3.8)$$

where $\langle \mathbb{Q} \rangle^{\mathcal{M}}$ is a referent constructor, utilising the condition specific to the (generalized) quantifier (and defined in Table 3.2 column 3).

To illustrate reference semantics, consider the domain model used in Section 3.2.3.2 with objects $\mathbb{O} = \{o_1, o_2\}$, vocabulary $\mathbb{V} = \{\text{red}, \text{cube}, \text{object}\}$ interpretation function $I(\text{object}) = \mathbb{O}$, $I(\text{red}) = \{o_1, o_2\}$, $I(\text{cube}) = \{o_1\}$. The referent of the referential expression “the one red cube”, whose logical form is $\langle _the_1_q \ x.\text{red}(x) \wedge \text{cube}(x) \rangle$ is computed according to Eq. 3.7: the projected domain model consists of the following objects:

$$\mathbb{O}' = \{o \in \mathbb{O} \mid \mathcal{M} \models \text{red}(o) \wedge \text{cube}(o)\} = \{o_1\} \quad (3.9)$$

which in turn leads to the following referent:

$$\begin{aligned} \mathcal{R} &= \{\mathbb{A} \subseteq \mathbb{O}' \mid |\mathbb{A}| = |\mathbb{O}'| \wedge |\mathbb{O}'| = 1\} \\ &= \{\{o_1\}\} \end{aligned} \quad (3.10)$$

If there had been another object in the domain model, o_3 say, that is a “red cube” that is $\llbracket \text{red}(o_3) \wedge \text{cube}(o_3) \rrbracket^{\mathcal{M}} = 1$, then $\mathcal{R} = \emptyset$. In other words, there is a referential failure because the Russellian uniqueness condition triggered by “the one” (see Table 3.2) is violated. By similar symbolic reasoning, the referent for “a cube” is $\{\{o_1\}, \{o_2\}\}$ and the referent for “all cubes” is $\{\{o_1, o_2\}\}$.

3.2.5 Reasoning about Beliefs

We now turn to probabilistic reasoning, given the belief state defined in Section 3.2.2. We begin by defining a set-theoretic representation of the domain model $\mathcal{H}_{\mathcal{M}}$ in terms of the ground atoms \mathcal{H} :

$$\mathcal{H}_{\mathcal{M}} = \{\mathbb{A} \in \mathcal{H} \mid \mathcal{M} \models \mathbb{A}\} \quad (3.11)$$

where A is an atom constructed from predicate symbols and constants. In this representation, truth (\top) is equivalent to an atom being part of this set $A \in \mathcal{H}_{\mathcal{M}}$, and falsehood (\perp) otherwise.

In reasoning about beliefs, we treat each atom as a Bernoulli random variable. The degree of belief that an atom is part of $\mathcal{H}_{\mathcal{M}}$ is quantified by the weight within the interval $[0, 1]$, representing probability that an atom is part of the underlying domain model. In this context, weights $w: \mathcal{H} \mapsto [0, 1]$ is a function mapping each ground atom to a corresponding weight/probability. In the belief state $b \in \mathbb{B}$, we have two kinds of weights: base weights w_b for initial beliefs that do not take into account any evidence gathered over the lifecycle of the agent; and grounded weights w_g that integrate this evidence. Since the weights capture the probabilities of atom being true or false in the underlying domain, one can exploit the truth conditions of complex formulae in the underlying symbolic logic to answer the probabilistic query for any arbitrarily complex formula $\phi \in \mathbb{L}_{\text{Snt}}$. In this thesis, we use a general-purpose inference procedure of Weighted Model Counting (WMC) (Chavira and Darwiche, 2008):

$$\begin{aligned} WMC_w(\phi) &= \sum_{\mathcal{M}: \mathcal{M} \models \phi} \prod_{A: A \in \mathcal{H}} w(A)[A \in \mathcal{H}_{\mathcal{M}}] + (1 - w(A))(1 - [A \in \mathcal{H}_{\mathcal{M}}]) \\ [A \in \mathcal{H}_{\mathcal{M}}] &= \begin{cases} 1 & \text{if } A \in \mathcal{H}_{\mathcal{M}} \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (3.12)$$

If $\forall A. w(A) = 0.5$, WMC reduces to the model counting otherwise known as the #SAT problem (Valiant, 1979). Intuitively, WMC estimates the probability of ϕ by counting and weighting events (particular domain models \mathcal{M}) that are valid under constraints (namely, the logical entailments of ϕ). For illustration, let's consider a blocks domain with $|\mathcal{O}| = 3$ objects and the vocabulary of two one-place predicates $\mathbb{V} = \{\text{red}, \text{cube}\}$. #SAT for an arbitrary object being a “red cube” is counted by including domain models that include one red cube, two red cubes, etc. yielding the total of 37 possible domain models.¹⁰ If instead we know that there should only be one “red cube” it reduces to 3 possible domain models. The examples in Tables 4.1 illustrate the effects of (monotonic) constraints on the domain model reduction on overall domain likelihoods. In general, the more monotonic constraints there are, the more the overall entropy of the belief state is reduced.

WMC is used to compute the Conditional Probability Query (CON): the probability

¹⁰There are $2^{|\mathbb{V}| \cdot |\mathcal{O}|} - (2^{|\mathbb{V}|} - 1)^{|\mathcal{O}|}$ domain models where at least one object satisfies all predicates.

that a formula ϕ is true, given some other formulae are true. The conditioning formulae can be constructed from the belief state $b \in \mathbb{B}$ —in particular, the conjunction of sentences in the domain theory $\tilde{\Delta} = \bigwedge_{\phi \in \Delta} \phi$ (recall that Δ captures certain knowledge, guaranteed to be true). So given a belief state b , CON is computed as follows:

$$CON_w(\phi | b) = \frac{WMC_w(\phi \wedge \tilde{\Delta})}{WMC_w(\tilde{\Delta})} \quad (3.13)$$

To connect the probabilistic reasoning and truth-conditional semantics, we can estimate the most likely domain model $\hat{\mathcal{M}}$ given the belief state by performing Maximum a Posteriori Query (MAP). This is achieved by computing the probability of the domain model, which is computed by considering atom values in each domain model:

$$MAP_w(b) = \arg \max_{\mathcal{M}} CON_w\left(\bigwedge_{A \in \mathcal{H}} \Psi(\mathcal{M}) | b\right) \quad (3.14)$$

where $\Psi(\mathcal{M})$ is a well-formed formula that is a conjunction of ground atoms \mathcal{H} , constructed using domain model $\mathcal{H}_{\mathcal{M}}$:

$$\Psi(\mathcal{M}) = \left(\bigwedge_{A \in \mathcal{H}_{\mathcal{M}}} A\right) \wedge \left(\bigwedge_{A \notin \mathcal{H}_{\mathcal{M}}} \neg A\right) \quad (3.15)$$

The estimated domain model $\hat{\mathcal{M}} = MAP_w(b)$ can be used for evaluation of well-formed formulas using truth-conditional semantics (Section 3.2.3.2) and logical forms of referential expressions using reference semantics (Section 3.2.4.2).

3.3 Summary

This chapter has outlined the base architecture for an embodied lifelong agent that processes embodied conversation in service of ITL. We also provided a particular, more detailed discussion of how we represent the agent’s belief state and how, over the lifecycle of interactions with the teacher and the environment, they will compute the likelihood of propositions being true, given what it knows for certain. We have not yet explained how weights w_g are learned: in subsequent chapters, we’ll see that this depends on a combination of visual perception and the teacher’s embodied utterances (in particular, when the teacher points to objects and describes them during their conversation).

The presented architecture satisfies the two essential requirements and two desired features we stipulated in Chapter 2. We now move into the experiments with this architecture: we aim to evaluate the value of Formal Semantic Analysis (FSA) when processing embodied conversation for ITL, where in particular, the agent faces scenarios where they must overcome their unawareness of concepts that are critical to understanding and solving the task the teacher wants the agent to do.

Chapter 4

Processing Referential Expressions

This chapter studies how performing Formal Semantic Analysis (FSA) on referential expressions like “the one red cube” can be utilised for interactive symbol grounding. This, in turn, serves to solve the task of reference resolution under unawareness—that is, the task of identifying a referent in the current visual scene of a given referring expression in a context where the agent starts the learning process unaware of the vocabulary that will feature in the referring expressions that it will have to interpret, and a domain model that lacks the denotations of the words in that vocabulary. As explained in Chapter 2, an ITL agent that must learn new tasks in a context where it is unaware of concepts that are necessary for solving it requires interactive symbol grounding.



(T1) Teacher: show me the one red cube.

(L1) Learner: I think you are asking for this. (designates not a red cube)

(T2) Teacher: No, here it is. (designates the one red cube)

(T3) Teacher: show me the one yellow cube.

(L2) Learner: I think you are asking for this. (designates a yellow cube)

(T4) Teacher: Correct.

Figure 4.1: Embodied Conversation for the sequence of reference resolution under unawareness tasks.

4.1 Introduction

The goals of this thesis are to quantify the value of FSA when processing embodied conversation in service to ITL, as well as to evaluate the general utility of embodied conversation to enable agents to cope with unforeseen possibilities. We begin answering these questions by considering individual embodied conversation messages, and a task of reference resolution under unawareness. We aim to show how the FSA of these messages makes learning a grounding model, through interaction, more data efficient.

As established in the previous chapters, when processing embodied conversation for ITL, we aim to perform interactive symbol grounding (an essential feature in Section 2.5), which arises due to the user uttering task instructions that feature concepts the agent is unaware of but is expected to use when solving the task. To cope with this, this thesis considers an agent that engages in embodied conversation so as to gather relevant evidence and so learn online how to recognize objects denoted by the newly discovered concept (or neologism).

The embodied conversation features an exchange of messages with natural language expressions and non-linguistic elements (e.g., pointing gestures, environment manipulation, etc.) via the established communication channel between the interlocutors. Referential expressions are the fundamental building blocks of natural language expressions in ITL. When considering knowledge representation and reasoning in the base architecture (Section 3.2), we looked at referential expressions that have a rich interpretation from the point of view of generalized quantifier theory (Barwise and Cooper, 1981). In this chapter, we take a closer look into processing such referential expression using this theory in the context of interactive symbol grounding. We develop an interactive neuro-symbolic grounding method that satisfies the desired features of neuro-symbolic reasoning and FSA that we argued are system requirements (see Section 2.5).

Referential expressions in the context of interactive symbol grounding have been studied before (Hristov, 2021), but in these settings, they have been *simple*, taking the form of bare nouns like “cube” or very short phrases like “red cube” or “sphere above cube”. For such referential expressions, FSA using the theory of generalized quantifiers has little to no value. They do not feature a quantifier, and their syntactic structure is relatively flat. This thesis studies referential expressions from the point of view of computational linguistics and considers *complex* referential expressions like “the one

red cube above every sphere”, which both feature a quantifier and have an embedded prepositional phrase in their construction. The motivation for studying complex referential expressions comes from experimental evidence in which humans are observed to use complex referential expressions even when a simple referential expression would successfully designate the referent (Dale and Reiter, 1995).

Chapter 1 discussed briefly how FSA allows us to infer additional information not present in the surface form of referential expressions. The embodied conversation in the blocks domain given in Figure 4.1 illustrates this. The message T1 contains a referential expression r = “the one red cube”, which as discussed in Section 3.2.4.1 has logical form $\Phi(r) = \langle \text{_the_1_q } x.\text{red}(x) \wedge \text{cube}(x) \rangle$. The referent provided in T2 of this referential expression has an additional value from the point of view of the FSA, as the usage of the quantifier _the_1_q in the definite description has a logical consequence that entails that objects that are not designated with this message are *not* red cubes. A similar inference, that there is exactly one yellow cube, is inferrable from the subsequent instruction T3.

Using the machinery established in Section 3.2, we can formally describe the difference between inferring the logical consequences, and not doing so. Let $\xi: \mathbb{N}_{\text{ref}} \times 2^{\mathbb{O}} \mapsto \mathbb{L}_{\text{snt}}$ be a Formal Semantic Analysis (FSA) of a referential expression $r \in \mathbb{N}_{\text{ref}}$ and the referent $\mathcal{R} \subseteq 2^{\mathbb{O}}$. Without a loss of generality, let’s say that the object $o \in \mathbb{O}$ was designated in T2, forming a referent $\mathcal{R} = \{\{o\}\}$ for the referential expression discussed above from Figure 4.1. When performing FSA ξ_{sem} , the following well-formed formula is constructed:

$$\xi_{\text{sem}}(r, \mathcal{R}) = (\text{red}(o) \wedge \text{cube}(o)) \wedge \bigwedge_{o' \in \mathbb{O} - \{o\}} \neg(\text{red}(o') \wedge \text{cube}(o')) \quad (4.1)$$

If, instead, the logical consequences of T2 are ignored, then the (base) semantic analysis ξ_{base} constructs the following, less informative, well-formed formula (less informative in that it is satisfied by more possible domain models):

$$\xi_{\text{base}}(r, \mathcal{R}) = \text{red}(o) \wedge \text{cube}(o) \quad (4.2)$$

The difference between these two semantic analyses leads to the following hypothesis:

Experiment hypothesis I An ITL agent that performs FSA when processing referential expressions during learning masters the task of reference resolution under unawareness more efficiently than an agent that does not use it.

For the rest of this chapter, we formalize the task we use to evaluate the difference between exploiting FSA vs. not doing so, introduce a framework for interactive neuro-symbolic grounding and conduct experiments to test the above empirical hypothesis.

4.2 Reference Resolution under Unawareness

To test our hypothesis, we consider a modified version of classical reference resolution—in contrast to existing tasks, our agent has a deficient conceptualisation of the domain, by which we mean it has a deficient hypothesis space of possible models because it has missing vocabulary and their perception of the environment does not distinguish the denotations of its missing words. The agent must, therefore, perform interactive symbol grounding to fix these deficiencies. We refer to this task as *reference resolution under unawareness*.

In this task, the agent takes the role of a learner while the user is a teacher and provides a task instruction $t \in \mathbb{N}_{\text{task}}$ in the context of the environment $\mathcal{E} \in \mathbb{E}$ of the following sort:

Teacher: Show me $\underbrace{\text{the one red cube}}_r$.

where $r \in \mathbb{N}_{\text{ref}}$ is a referential expression. In the restrictive task of reference resolution that we study in this chapter, the learner can only perform one action: to designate the referent (in later chapters, we expand the learner’s options to include asking questions). We refer to this action as $act \in \mathbb{A}$, forming the singleton action space for the agent.

As described in Section 3.1, using the semantic parser, the agent obtains the logical form of the referential expression $\Phi(r) = \langle \text{_the_q } x.\text{red}(x) \wedge \text{cube}(x) \rangle$. From the task instruction, the agent also constructs a logical form $\Phi(t)$ which encodes information that there are objects in the environment denoted by this referential expression. For the example above that would correspond to the following well-formed formula $\Phi(t) = \text{_the_q } x.(\text{red}(x) \wedge \text{cube}(x), \text{object}(x))$, where object is a special predicate for which $I(\text{object}) = \mathbb{O}$ as introduced in Section 3.1. This formula is part of the initial domain theory Δ for the environment \mathcal{E} and is part of the agent’s current belief state $b \in \mathbb{B}$ that is $\Delta = \{\Phi(t)\}$.

At the start of performing a sequence of reference resolution under unawareness tasks, the learner does not know the underlying vocabulary \mathbb{V} and the interpretation function

$I: \mathbb{V} \mapsto \mathbb{O}^*$. Throughout the agent’s lifecycle, it gather evidence and updates its belief state $b \in \mathbb{B}$, which helps them to estimate the domain model $\hat{\mathcal{M}}$ by MAP with grounded weights MAP_{w_g} (Eq. 3.14), as discussed in Section 3.2.5. Using $\hat{\mathcal{M}}$, the agent computes the referent $\hat{\mathcal{R}} = \Phi(r)^{\hat{\mathcal{M}}}$, as outlined in Section 3.2.4.2. This referent is used by the learner to produce the following embodied conversation message as a result of $act \in \mathbb{A}$:

Learner: Here it is. (points to $\hat{\mathcal{R}}$)

Upon receiving this message, the teacher utters the following response:

$$\text{Teacher: } \begin{cases} \text{Correct.} & \text{if } \hat{\mathcal{R}} = \Phi(r)^{\mathcal{M}} \\ \text{No. Here it is. (points to } \Phi(r)^{\mathcal{M}} \text{)} & \text{otherwise} \end{cases}$$

Based on the teacher’s response, the learner receives the reward $R: \mathbb{A} \mapsto \mathbb{R}$, which for the reference resolution task is binary:

$$R(a) = \begin{cases} 1 & \text{if reference was designated successfully} \\ -1 & \text{otherwise} \end{cases} \quad (4.3)$$

We’ll use this reward to evaluate the learner’s performance in our experiments.

The teacher’s response enables the learner to construct a well-formed formula $\phi = \xi(r, \mathcal{R})$, using either a FSA or base semantic analysis (see Eq. 4.1 vs. Eq. 4.2 for an example of the difference between the two). This sentence ϕ holds in the environment \mathcal{E} and can be used to update the belief state and, in particular, expand the domain theory Δ .

The interaction then proceeds with another task instruction t' from the teacher, in the same environment \mathcal{E} or in a new environment \mathcal{E}' . This new reference resolution task may feature some of the same symbols but may also feature further unforeseen symbols or neologisms, too. In the case where the new environment \mathcal{E}' (i.e., a new image) is observed, the domain theory in the belief state is reset to ($\Delta = \emptyset$) (since Δ consists of propositions that are true of the *current* environment or image), but the existing grounding model, the acquisition of which is detailed in the next section, is retained.

4.3 Interactive Neuro-symbolic Grounding

In this section, we outline how well-formed formulas, constructed from the FSA of embodied conversation messages, can be used to update a belief state. This involves

using and updating the learner’s current grounding model, which, among many things, can be used to estimate the domain model. We provide an illustrative example to highlight the features of the proposed approach. Note that even though we will be using the task of reference resolution under unawareness, the neuro-symbolic grounding approach introduced here is general, and in subsequent chapters, we will use it to address other tasks (in particular, manipulation of objects in the environment to achieve a goal configuration). Furthermore, we restrict attention to grounding one-place predicates (properties) only. However, the presented grounding model is general and can, in principle, be extended to grounding predicates of arbitrary arity. For instance, for grounding two-place predicates like spatial relationships, we could consider *relationship* embeddings constructed from the pair of the embeddings of objects participating in this relationship. As discussed in Section 3.2.2, a belief state with such embeddings is out of the scope of this thesis.

4.3.1 The Grounding Model

A grounding model is a function that includes sensory observations and predicates of the domain. In this thesis, it is operationalised as a probabilistic classifier $\omega_b: \mathbb{R}^d \mapsto [0, 1]^{|\mathbb{V}|}$ —given the d -dimensional embedding $\mathbf{x} \in \mathbb{X}$, corresponding to object $o \in \mathbb{O}$, which is extracted from sensory observations, this mapping predicts a $|\mathbb{V}|$ -dimensional semantic vector $\hat{\mathbf{y}} \in [0, 1]^{|\mathbb{V}|}$. Each element of this semantic vector \hat{y}_p is a probability that the atom $p(o)$ is satisfied by the domain model \mathcal{M} . This corresponds to the probability that $p(o)$ is a member of $\mathcal{H}_{\mathcal{M}}$ —in other words, the probability that that $p(o) \in \mathcal{H}_{\mathcal{M}}$, or that $p(o)$ is true in the ground truth model \mathcal{M} .

The grounding model predictions depend on the belief state $b \in \mathbb{B}$:

$$\begin{aligned} \hat{\mathbf{y}} &= \omega_b(\mathbf{x}) \\ \hat{y}_p &= \Pr(p(o) \in \mathcal{H}_{\mathcal{M}} \mid b) \quad o = \text{Const}(\text{Obj}(\mathbf{x})) \end{aligned} \tag{4.4}$$

where $\text{Const}(\text{Obj}(\mathbf{x}))$ is the constant that denotes the object with the embedding $\mathbf{x} \in \mathbb{X}$.

For interactive symbol grounding, this thesis uses a non-parametric memory-based grounding model of prototype networks (Yang et al., 2019; Cano Santín et al., 2020). They perform grounding using evidence in the belief state and, in particular, the support \mathbb{S} of the belief state. As introduced in Section 3.2.5, the support of the belief state consists of embedding–semantic vector pairs, which record the evidence of what objects are believed to be denoted by certain symbols and what objects are not. The

unique approach of our interactive neuro-symbolic grounding and this thesis lies in using FSA to construct this support.

As explained above, FSA of the teacher’s signals yield well-formed formulas ξ , which go beyond the surface form of the teacher’s utterance, thanks to the logical entailments of quantifiers (and negation). During the embodied conversation, several of these well-formed formulas can be gathered, which is used to build a domain theory Δ : a set of well-formed formulae that are known to hold in the environment— $\mathcal{M} \models \Delta$ — thanks to being constructed using evidence from the teacher that is a domain expert and provides information in a cooperative and timely manner. This domain theory Δ is used to interactively update semantic vectors in the support, by estimating each element of the semantic vector using Conditional Probability Query (CON) (see Section 3.2.5 Eq. 3.13). Algorithm 1 gives the Update: $\mathbb{B} \times \mathbb{L}_{\text{snt}} \mapsto \mathbb{B}$ procedure of how the evidence in the form of a well-formed formula can be integrated into the belief state: to both update the support \mathbb{S} as well as grounded weights w_g . We now look at how grounded weights can be constructed using the grounding model with the updated support.

Algorithm 1 Update (Belief state update with well-formed formula)

Require: Belief state $b = (\mathbb{O}, \mathbb{X}, \mathbb{V}, \mathbb{S}, \Delta, w_b, w_g) \in \mathbb{B}$, well-formed formula $\phi \in \mathbb{L}_{\text{snt}}$.

```

 $\mathbb{S}' \leftarrow \emptyset$  ▷ Initialize new support
 $\Delta \leftarrow \Delta \cup \{\phi\}$  ▷ Update domain theory
for  $o, \mathbf{x}$  in Zip( $\mathbb{O}, \mathbb{X}$ ) do ▷ Iterate over corresponding object-embedding pairs
     $o \leftarrow \text{Const}(o)$  ▷ Get constant of an object  $o \in \mathbb{O}$ 
     $y \leftarrow \text{HashMap}()$  ▷ Hash-map from symbols to predicted symbol probability
    for  $p$  in  $\mathbb{V}$  do ▷ Iterate over predicate symbols
         $A \leftarrow p(o)$  ▷ Construct atom from symbol and constant
         $y[p] \leftarrow \text{CON}_{w_b}(A \mid b)$  ▷ Conditional Probability Query (CON) (Eq. 3.13)
     $\mathbf{y} \leftarrow \text{Vectorize}(y)$  ▷ Construct semantic vector of size  $|\mathbb{V}|$  using hash-map
     $\mathbb{S}' \leftarrow \mathbb{S}' \cup (\mathbf{x}, \mathbf{y})$  ▷ Add exemplar to new support
 $\mathbb{S} \leftarrow \mathbb{S}'$  ▷ Update support
for  $o, \mathbf{x}$  in Zip( $\mathbb{O}, \mathbb{X}$ ) do ▷ Iterate over corresponding object-embedding pairs
    for  $p$  in  $\mathbb{V}$  do ▷ Iterate over predicate symbols
         $w_g(p(o)) \leftarrow \omega_b(\mathbf{x})_p$  ▷ Grounding model prediction Eq.4.9
return  $b$ 

```

The grounding model, in the form of a prototype networks, uses support \mathbb{S} of the belief state to construct positive and negative support for each symbol $p \in \mathbb{V}$, denoted as \mathbb{S}_p^+

and \mathbb{S}_p^- , respectively. Whether a pair goes into one of these support sets is decided by: (a) the value of y_p (b) the entropy $\mathbb{H}_{\mathcal{B}}$ of a Bernoulli random variable \mathcal{B} parametrised by y_p , and (c) the threshold τ for the overall noise level in the support construction. The support sets are constructed as follows:

$$\begin{aligned}\mathbb{S}_p^+ &= \{(\mathbf{x}, \mathbf{y}) \in \mathbb{S} \mid y_p > 0.5 \wedge \mathbb{H}_{\mathcal{B}}[y_p] \leq \tau\} \\ \mathbb{S}_p^- &= \{(\mathbf{x}, \mathbf{y}) \in \mathbb{S} \mid y_p < 0.5 \wedge \mathbb{H}_{\mathcal{B}}[y_p] \leq \tau\}\end{aligned}\quad (4.5)$$

where the entropy of the Bernoulli random variable is computed in the closed form:

$$\mathbb{H}_{\mathcal{B}}[x] = -x \log(x) - (1-x) \log(1-x) \quad (4.6)$$

Using these support sets, positive and negative prototypes, denoted as \mathbf{z}_p^+ and \mathbf{z}_p^- respectively, are computed by taking the weighted average of embeddings of the corresponding support sets, the weight is the likelihood that the relevant object is (or respectively is not) denoted by p (i.e., the values y_p and $(1-y_p)$ respectively):

$$\begin{aligned}\mathbf{z}_p^+ &= \frac{1}{|\mathbb{S}_p^+|} \sum_{(\mathbf{x}, \mathbf{y}) \in \mathbb{S}_p^+} y_p \mathbf{x} \\ \mathbf{z}_p^- &= \frac{1}{|\mathbb{S}_p^-|} \sum_{(\mathbf{x}, \mathbf{y}) \in \mathbb{S}_p^-} (1-y_p) \mathbf{x}\end{aligned}\quad (4.7)$$

Note that if due to a lack of evidence (so far) $\mathbb{S}_p^{+/-}$ are empty—in other words, there is not enough evidence (yet) to deem any of the observed embeddings to be *good-enough* exemplars for positive/negative support—then $\mathbf{z}_p^{+/-}$ defaults to support elements with the largest/smallest entropy $\mathbb{H}_{\mathcal{B}}$ as the best guess of what exemplars are suitable for these support sets.

We can now use cosine similarity $\cos: \mathbb{R}^d \times \mathbb{R}^d \mapsto [0, 1]$ to compare how (visually) similar an embedding for the objects of interest is to positive and negative prototypes for each predicate $p \in \mathbb{V}$. This, in turn can be used to estimate the likelihood that the object is denoted by the symbol p : namely, one uses this (cosine) similarity metric as activation of a sigmoid function:

$$\hat{y}_p = \frac{1}{1 + \exp(-\cos(\mathbf{z}_p^- - \mathbf{z}_p^+, \mathbf{x}))} \quad (4.8)$$

The prototype network, which assigns to each object a (binary) probability distribution for each symbol, handles unawareness in the following way. Suppose that the agent

observes a new symbol or neologism p^* from an embodied conversation message uttered by the teacher. The agent becomes aware of p^* by extending its vocabulary $\mathbb{V} \leftarrow \{p^*\} \cup \mathbb{V}$ in the belief state and by constructing new support following Algorithm 1—equivalently, the agent extends \mathbf{y} with a new element that is assigned the value corresponding to its base weight (by default, this weight is 0.5, equivalent to complete ignorance about its denotations).

The agent can update grounded weights w_g in the belief state b using grounding model predictions for all objects in the environment \mathcal{E} :

$$w_g(p(o)) = \omega_b(\mathbf{x})_p \quad (4.9)$$

which is part of the update procedure in Algorithm 1.

4.3.2 An Illustrative Example

To illustrate the belief update for interactive neuro-symbolic grounding, we consider an illustrative example in the blocks domain that has the following ground-truth domain model:

$$\begin{aligned} \mathcal{M} : \\ \mathbb{O} &= \{o_1, o_2, o_3, o_4\} \\ \mathbb{V} &= \{\text{blue}, \text{red}, \text{sphere}, \text{cube}\} \\ I &= \{\text{blue} : \{o_1, o_3, o_4\}, \text{red} : \{o_2\}, \text{sphere} : \{o_1, o_3\}, \text{cube} : \{o_2, o_4\}\} \end{aligned} \quad (4.10)$$

with each object o_i having the following embedding \mathbf{x}_i :

$$\mathbb{X} = \{\mathbf{x}_1 = [0.7, 0.2]^\top, \mathbf{x}_2 = [0.1, 0.7]^\top, \mathbf{x}_3 = [0.6, 0.1]^\top, \mathbf{x}_4 = [0.9, 0.8]^\top\} \quad (4.11)$$

Note that this example is discussed for illustrative purposes only: we are using here an extremely small feature space with only $d = 2$ dimensions, but the proposed interactive neuro-symbolic grounding approach for combining evidence from the embodied conversation and sensory observations can and is designed to cope with arbitrarily complex feature spaces. The experiments in this chapter and subsequent chapters use feature spaces with much bigger dimensions (e.g. embeddings from Grounding DINO (Liu et al., 2023) are of size $d = 256$).

We first discuss the belief update under different domain theories Δ using base weights w_b ; in other words, belief updates when Δ alone changes, without integration of in-

formation from sensory observations (we'll integrate sensory observations in a subsequent illustration). To do this, we consider three separate domain theories involving the following three referential expressions r and referents \mathcal{R} :

$$\begin{array}{ll} r_1 = \text{“a red cube”} & \mathcal{R}_1 = \{\{o_2\}\} \\ r_2 = \text{“the one red cube”} & \mathcal{R}_2 = \{\{o_2\}\} \\ r_3 = \text{“the one blue cube”} & \mathcal{R}_3 = \{\{o_4\}\} \end{array}$$

Our three contrasting domain theories are then defined as follows:

- $\Delta_1 = \{\xi(r_1, \mathcal{R}_1)\}$: built from r_1 and the teacher pointing to o_2 ;
- $\Delta_2 = \{\xi(r_2, \mathcal{R}_2)\}$: built from r_2 and the teacher pointing to o_2 ;
- $\Delta_{23} = \{\xi(r_2, \mathcal{R}_2), \xi(r_3, \mathcal{R}_3)\}$: built from r_2 and the teacher pointing to o_2 as well as from r_3 and the teacher pointing to o_4 .

Each of them yield different belief states: b_1 from Δ_1 , b_2 from Δ_2 , and b_{23} from Δ_{23} , while everything else stays the same.

Table 4.1 shows probabilistic queries using w_b resulting from being in these different belief states. Between b_1 and b_2 , there is additional information inferred about the domain, similarly as shown in Eq. 4.1 vs. Eq. 4.2. Crucially, this additional information affects the probabilities of atoms that are not a part of Δ_1 or Δ_2 : e.g., the atom $\text{red}(o_1)$ is less likely to be true given b_2 compared with its likelihood given b_1 , because $\text{CON}_{w_b}(\text{red}(o_1) \wedge \text{cube}(o_1)|b_2) = 0$, and this is not the case for b_1 .

When comparing b_2 and b_{23} , observe that $\text{blue}(o_2)$ and $\text{red}(o_4)$ are false thanks to the (classical) logical reasoning from the logical formulae in Δ_{23} . Reasoning with Δ yields soft belief changes as well. For example, the neologism *blue* has just been introduced, but one logical consequence of Δ_{23} is that for all objects other than o_4 , the probability that they are both a cube and blue is 0. This reduces the likelihood that o_1 and o_3 are blue for similar reasons to those we mentioned for $\text{red}(o_1)$ given b_2 . In Δ_{23} we have equivalent evidence for symbols *red* and *blue*, coming from r_2 and r_3 , respectively leading to the *same* probability assigned for atoms $\text{red}(o_1)$, $\text{red}(o_3)$, $\text{blue}(o_1)$, $\text{blue}(o_3)$ while it is different and smaller for symbol *cube*, because we have evidence about this symbol from both r_2 and r_3 . For MAP, there is no qualitative difference between b_1 and b_2 because MAP_{w_b} sets the truth value of A to true if, and only if, $\text{CON}_{w_b}(A|b) > 0.5$. For b_{23} we infer additional atoms to be true in the domain due to the addition of new knowledge from r_3 .

Atoms A	$CON_{w_b}(A b_1)$	$CON_{w_b}(A b_2)$	$CON_{w_b}(A b_{23})$	$MAP_{w_b}(b_1)$	$MAP_{w_b}(b_2)$	$MAP_{w_b}(b_{23})$
red(o ₁)	.50	.33	.40	0	0	0
red(o ₂)	1	1	1	1	1	1
red(o ₃)	.50	.33	.40	0	0	0
red(o ₄)	.50	.33	0	0	0	0
blue(o ₁)	—	—	.40	—	—	0
blue(o ₂)	—	—	0	—	—	0
blue(o ₃)	—	—	.40	—	—	0
blue(o ₄)	—	—	1	—	—	1
sphere(o ₁)	—	—	—	—	—	—
sphere(o ₂)	—	—	—	—	—	—
sphere(o ₃)	—	—	—	—	—	—
sphere(o ₄)	—	—	—	—	—	—
cube(o ₁)	.50	.33	.20	0	0	0
cube(o ₂)	1	1	1	1	1	1
cube(o ₃)	.50	.33	.20	0	0	0
cube(o ₄)	.50	.33	1	0	0	1

Table 4.1: Comparison of different probabilistic queries: CON_{w_b} and MAP_{w_b} for different belief states b_1, b_2 and b_{23} containing different domain theories: Δ_1, Δ_2 and Δ_{23} , respectively. The symbol “—” denotes the unawareness about that predicate symbol for that particular stage of the interaction. This may change: e.g. upon the teacher conveying r_3 , the vocabulary $\mathbb{V} \leftarrow \mathbb{V} \cup \{\text{blue}\}$ expands, as the agent becomes aware of the concept blue. Thus for b_{23} “—” gets replaced with an appropriate value. In MAP_{w_b} columns, 1 means that an atom is in the estimated domain model $\hat{\mathcal{M}}$; 0 means it is not.

To illustrate the change in beliefs when integrating (visual) evidence from the environment gathered over the lifecycle of the agent, we consider the same belief states b_1 , b_2 , and b_{23} as we used in Table 4.1, combined with the example 2-dimensional vectors for our toy example that are shown in Eq. 4.10. Together, these produce grounded weights w_g , following the grounding model from Section 4.3.1. For this example, we set the threshold τ to 0.7, which corresponds to a positive (or negative) exemplar being added to the support, thereby affecting the representation of positive (or negative) prototypes, if CON_{w_b} for an atom is true is above 0.55 (or for the negative case, below 0.44). This comparison is given in Table 4.2.

Consider first b_1 . \mathbf{x}_2 is in the positive support for red and cube thanks to CON_{w_b} (entries in Table 4.1, column 2 are above 0.55 for these atoms). In contrast, \mathbf{x}_1 is chosen at random to be assigned to the negative support for these predicate symbols, because no exemplars CON_{w_b} is below the 0.44 margin (see Table 4.1, column 2): since the possible candidates ($\{\mathbf{x}_1, \mathbf{x}_3, \mathbf{x}_4\}$) all have the same entropy, one is chosen without a loss of generality.

Using positive and negative support, prototypes for symbols red and cube is $\mathbf{z}_{\text{red/cube}}^+ = \mathbf{x}_2$ and $\mathbf{z}_{\text{red/cube}}^- = 0.5\mathbf{x}_1$ (0.5 being CON_{w_b} in Table 4.1 that o_1 is red/cube). Similarly, computation between these prototypes and the embedding for each object acts as an activation for the sigmoid in Eq. 4.8, which yields grounding weights (Table 4.2, column 6) that show that o_4 is more likely to be red and cube than o_1 or o_3 , even though no evidence is provided about that in the domain theory Δ . This difference stems entirely from the relevant cosine similarities on embeddings vs. the positive (and negative) prototypes.

When reasoning about the domain with b_2 , the ordering of the likelihoods persists: o_4 is more likely to be red and cube than o_1 and o_3 , but the weights are overall lower because the domain theory now has the information stemming from the logical consequences of using “the one” instead of “a”.

In the belief state b_{23} , we see the difference when using base weights w_b and grounded weights w_g , in particular for atoms $\text{red}(o_1)$, $\text{red}(o_3)$, $\text{blue}(o_1)$ and $\text{blue}(o_3)$. The evidence coming from sensory observations make $\text{blue}(o_1)$ and $\text{blue}(o_3)$ more probable than $\text{red}(o_1)$ and $\text{red}(o_3)$, thanks to the visual similarity of \mathbf{x}_1 and \mathbf{e}_3 to the positive prototypes for blue. These differences lead to a different and more accurate estimated domain model for Δ_{23} when compared to using base weights (MAP_{w_b} in Table 4.1 col-

umn 7) additionally inference with w_g suggest that $\text{blue}(o_1)$ and $\text{blue}(o_3)$ holds in the domain, in contrast to inference with w_b (Table 4.1 column 7).

This illustrative example shows the value of the proposed interactive symbol grounding approach with possible non-trivial inferences drawn from the updated belief state. We now turn to quantify the value of performing FSA when constructing sentences in the context of the task of reference resolution under unawareness.

Belief state	Positive support embeddings	Negative support embeddings	Positive prototypes \mathbf{z}^+	Negative prototypes \mathbf{z}^-	Grounded weights w_g	Domain model $\mathcal{H}_{\mathcal{M}}$
b						
b_1	red: $\{\mathbf{x}_2\}$	red: $\{\mathbf{x}_1\}$	$\mathbf{z}_{\text{red}}^+ = \begin{bmatrix} .10 \\ .70 \end{bmatrix}$	$\mathbf{z}_{\text{red}}^- = \begin{bmatrix} .35 \\ .10 \end{bmatrix}$	$\begin{bmatrix} .36, 1., .33, .47 \\ -, -, -, - \\ -, -, -, - \\ .36, 1., .33, .47 \end{bmatrix}$	$\{\text{red}(u_2), \text{cube}(u_2)\}$
	blue: $-$	blue: $-$				
	sphere: $-$	sphere: $-$	$\mathbf{z}_{\text{cube}}^+ = \begin{bmatrix} .10 \\ .70 \end{bmatrix}$	$\mathbf{z}_{\text{cube}}^- = \begin{bmatrix} .35 \\ .10 \end{bmatrix}$		
	cube: $\{\mathbf{x}_2\}$	cube: $\{\mathbf{x}_1\}$				
b_2	red: $\{\mathbf{x}_2\}$	red: $\{\mathbf{x}_1, \mathbf{e}_3, \mathbf{e}_4\}$	$\mathbf{z}_{\text{red}}^+ = \begin{bmatrix} .10 \\ .70 \end{bmatrix}$	$\mathbf{z}_{\text{red}}^- = \begin{bmatrix} .49 \\ .24 \end{bmatrix}$	$\begin{bmatrix} .27, 1., .26, .31 \\ -, -, -, - \\ -, -, -, - \\ .27, 1., .26, .31 \end{bmatrix}$	$\{\text{red}(u_2), \text{cube}(u_2)\}$
	blue: $-$	blue: $-$				
	sphere: $-$	sphere: $-$	$\mathbf{z}_{\text{cube}}^+ = \begin{bmatrix} .10 \\ .70 \end{bmatrix}$	$\mathbf{z}_{\text{cube}}^- = \begin{bmatrix} .49 \\ .24 \end{bmatrix}$		
	cube: $\{\mathbf{x}_2\}$	cube: $\{\mathbf{x}_1, \mathbf{e}_3, \mathbf{e}_4\}$				
b_{23}	red: $\{\mathbf{x}_2\}$	red: $\{\mathbf{x}_1, \mathbf{e}_3, \mathbf{e}_4\}$	$\mathbf{z}_{\text{red}}^+ = \begin{bmatrix} .10 \\ .70 \end{bmatrix}$	$\mathbf{z}_{\text{red}}^- = \begin{bmatrix} .56 \\ .33 \end{bmatrix}$	$\begin{bmatrix} .29, 1., .28, 0 \\ .41, 0., .42, 1 \\ -, -, -, - \\ .20, 1., .19, 1 \end{bmatrix}$	$\{\text{red}(u_2), \text{cube}(u_2), \text{cube}(u_4), \text{blue}(u_1), \text{blue}(u_3), \text{blue}(u_4)\}$
	blue: $\{\mathbf{x}_4\}$	blue: $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$	$\mathbf{z}_{\text{blue}}^+ = \begin{bmatrix} .90 \\ .80 \end{bmatrix}$	$\mathbf{z}_{\text{blue}}^- = \begin{bmatrix} .29 \\ .29 \end{bmatrix}$		
	sphere: $-$	sphere: $-$				
	cube: $\{\mathbf{x}_2, \mathbf{x}_4\}$	cube: $\{\mathbf{x}_1, \mathbf{e}_3\}$	$\mathbf{z}_{\text{cube}}^+ = \begin{bmatrix} .50 \\ .75 \end{bmatrix}$	$\mathbf{z}_{\text{cube}}^- = \begin{bmatrix} .52 \\ .12 \end{bmatrix}$		

Table 4.2: Comparison of different support embeddings, prototypes, grounded weights, and domain models (MAP_{w_g} query) constructed with different belief states b_1 , b_2 , and b_{23} . Semantic vectors \mathbf{y} are omitted as they are equivalent to CON_{w_b} queries in Table 4.1. Grounded weights w_g are given in the same order for atoms as in Table 4.1.

4.4 Experiments

In this section, we describe our experiments to quantify the value of processing referential expressions using FSA to inform the proposed interactive neuro-symbolic grounding approach.

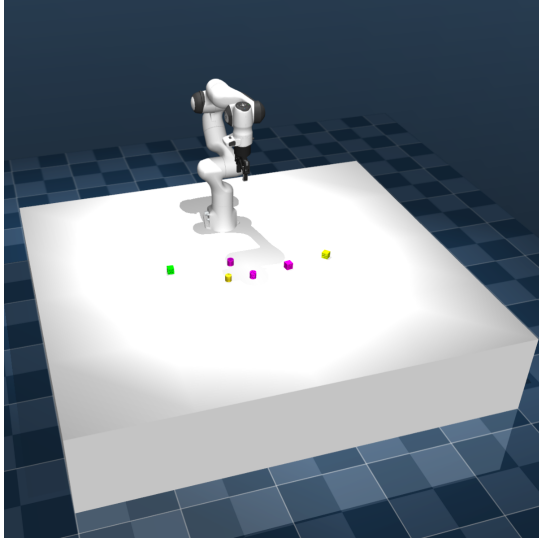
4.4.1 Experimental Setup

We consider ITL agents following the base architecture presented in Chapter 3, which in the current experiments have no decision-making module: the task, for now, is simply reference resolution under unawareness as presented in Section 4.2, and this has only a single action $act \in \mathbb{A}$ (to attempt to identify a referent of the teacher’s referential expression).

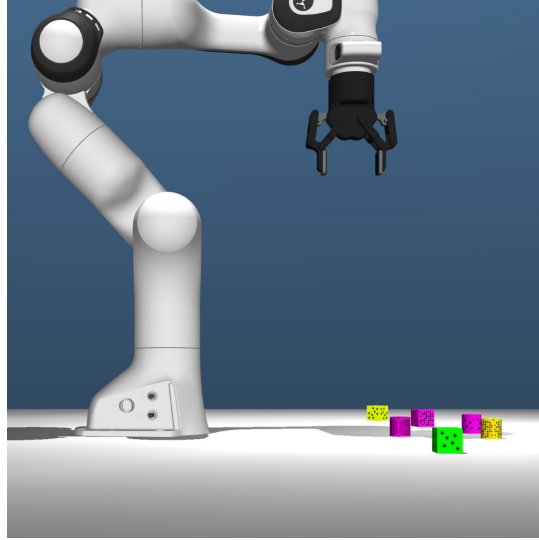
To test our hypothesis, we conduct simulation experiments, following [Kuhnle and Copestake \(2017\)](#) methodology for generating a variety of environments and valid tasks in a block domain. Instances of the environment are simulated using the MuJoCo ([Todorov et al., 2012](#)) (see Figure 4.2 for visualization). Each environment instance consists of a tabletop with a Franka Emika robot and 6-8 rigid bodies (blocks) in front of it. The vocabulary of this domain consists of seven colour categories: red, blue, green, yellow, magenta, cyan, grey, three shapes: cube, cylinder, rectangle, and three textures: plain, dotted, starry. So there are 13 predicate symbols in total. The agent does not know any of these symbols in advance.

Sensory observations for each of these predicate symbols have variability: for colours, its hues leading to different shades; for shapes, its size and orientation; and for texture, its density, leading to different patterns per texture. The blocks domain is used to evaluate the multimodal language understanding abilities of ITL agents in this thesis: being a simulated environment, this allows us to easily control all aspects of the artificially generated data ([Kuhnle and Copestake, 2018](#)), allowing flexible and reusable evaluation. Because of this, the goal of these experiments is not to build agents that solve tasks in this environment *per se* (this could be done with, e.g., offline RL) but to evaluate their capabilities and, in particular, the value of FSA.

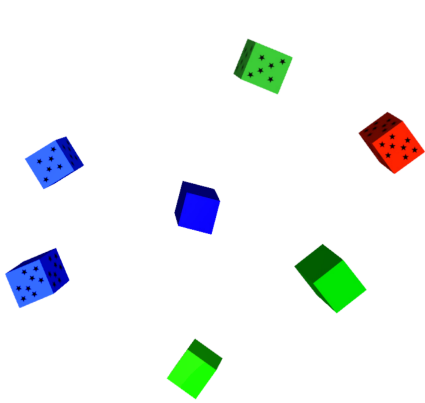
Each environment (i.e., an instance of the blocks domain) comes with the sensory observation in the form of a top-down view, and no robots or other objects are occluding the view of each object in the tabletop (see Figure 4.2c for an example). DI-NOv2 ([Oquab et al., 2023](#)) is used to extract embeddings $\mathbf{x} \in \mathbb{X}$ of size $d = 384$ for



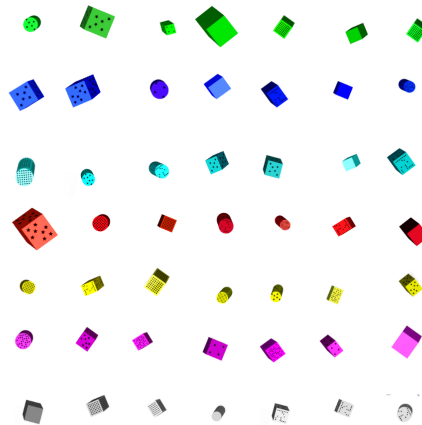
(a) Bird-eye view.



(b) Side view.



(c) Top-down view.



(d) Variability of objects.

Figure 4.2: Visualization of Blocks tabletop environment.

each object $o \in \mathbb{O}$ from the top-down view of the workspace, using patches obtained from the ground-truth object locations provided by the simulator.

To generate task instructions, we use the template-based referential expression generation in which we first sample a subset of predicate symbols from the ground-truth domain model and then sample quantifiers for the referential expression. For these experiments, we consider 3 different quantifiers, namely existential “a/an”, and two quantifiers that carry logical consequences in their use in the environment: universal “every/all” and uniqueness “the n ”. In the quantifier selection, we biased the sampling process and encouraged a quantifier with a probability of 0.7 to be chosen to be universal of a uniqueness quantifier. To ensure the correct order of words in referential expressions (e.g., “a red cube” is a valid referential expression in English while “a cube red” is not), we prompt CodeLlama 7B (Rozière et al., 2023) (see Appendix A). Additionally, we run a post-processing routine to ensure agreement by number (e.g., “the two red cube” is not grammatically correct while “the two red cubes” is). To obtain the logical form of referential expressions like the ones described in Section 3.2.4, we again prompt CodeLlama 7B with several exemplars of referential expression and their logical form pairs (see Appendix A). Note that for the exemplars given for CodeLlama 7B, we do not use predicate symbols in our block domain. The generated referential expression and its corresponding logical form are validated to successfully refer to the environment by using the ground-truth domain model, and if needed, a new referential expression is generated by repeated sampling. For each environment, we generate 2 task instructions.

This setup allows evaluating agents that can correctly identify objects in the environment as well as process and generate embodied conversation messages when interacting with the teacher. What ITL agent lacks is the knowledge of the ground-truth domain model which the agent estimates by learning a grounding model from the interaction with the teacher. In our experiments, the teacher is a simulated oracle that responds to embodied conversation messages using the ground-truth domain model.

For quantifying the value of FSA when processing referential expression, we consider two agents, each deployed with a different policy on how to process referential expressions:

- π_{sem} : its processing policy in which, upon receiving referential expression r and its referent \mathcal{R} , it is processed using full FSA constructing well-formed formulae

$\xi_{\text{sem}}(r, \mathcal{R})$ like given in Eq. 4.1. This formula is added to the belief state (in particular, the domain theory) to update the grounding model;

- π_{base} : its processing policy in which, upon receiving referential expression r and its referent \mathcal{R} , it is processed ignoring the benefits of full FSA constructing well-formed formulae $\xi_{\text{base}}(r, \mathcal{R})$ like given in Eq. 4.2. This formula is added to the belief state (in particular, the domain theory) to update the grounding model.

These agents are presented in our experiments with the same environments and task instructions in the same sequence. For forming positive/negative support, the threshold is $\tau = 0.65$ for both processing policies.

To evaluate the performance difference between the two processing policies, in each experiment consisting of $N = 60$ task, we record the trace of the received rewards, predicted referents and ground-truth referent. Using it for each referent prediction scores for the Precision, Recall, and F1 can be recorded:

$$\begin{aligned} \text{Precision}(\mathcal{R}, \hat{\mathcal{R}}) &= \frac{|\mathcal{R}' \cap \hat{\mathcal{R}}'|}{|\hat{\mathcal{R}}'|} \\ \text{Recall}(\mathcal{R}, \hat{\mathcal{R}}) &= \frac{|\mathcal{R}' \cap \hat{\mathcal{R}}'|}{|\mathcal{R}'|} \\ \text{F1}(\mathcal{R}, \hat{\mathcal{R}}) &= 2 \cdot \frac{\text{Precision}(\mathcal{R}, \hat{\mathcal{R}}) \cdot \text{Recall}(\mathcal{R}, \hat{\mathcal{R}})}{\text{Precision}(\mathcal{R}, \hat{\mathcal{R}}) + \text{Recall}(\mathcal{R}, \hat{\mathcal{R}})} \end{aligned} \quad (4.12)$$

where in the specification of Precision and Recall, we can exploit that for the referential expressions studied in these experiments, the referent is a singleton set, which is not the case, for example, for a referential expression “at most two cubes” in the environment with three cubes. Using this observation, we can just consider the set inside the predicted and ground-truth referent, denoted \mathcal{R}' and $\hat{\mathcal{R}}'$, respectively. To illustrate the computation of these scores, consider $\mathcal{R} = \{\{o_1, o_2\}\}$ and $\hat{\mathcal{R}} = \{\{o_1, o_3\}\}$ for which $\mathcal{R}' = \{o_1, o_2\}$ and $\hat{\mathcal{R}}' = \{o_1, o_3\}$ leading to $\text{Precision}(\mathcal{R}, \hat{\mathcal{R}}) = \frac{1}{2}$, $\text{Recall}(\mathcal{R}, \hat{\mathcal{R}}) = \frac{1}{2}$ and $\text{F1}(\mathcal{R}, \hat{\mathcal{R}}) = 0.5$.

To test experiment hypothesis I, we consider two performance measures: the learning curves of the cumulative reward $cR = \sum_i^N R_i$ encoding intrinsic performance (performance for the task of reference resolution) and the mean F1 score $mF1 = \frac{1}{N} \sum_i^N \text{F1}(\mathcal{R}_i, \hat{\mathcal{R}}_i)$ encoding extrinsic performance (performance on reference prediction, giving partial credit for partially correct referents). These curves are computed by 5 repeated runs of the experiment to capture the performance variability due to different

agents \ performance measure	cR	mF1
π_{sem} vs. π_{base}	$7.6 \cdot 10^{-13}$	$4.3 \cdot 10^{-13}$

Table 4.3: P-values for comparing the agent’s learning curves for the initial setup.

sequences of the stimuli (both environment and task instructions). Note that the absolute values for the cR and mF1 are not good reflectors of the agent’s performance in a particular ITL scenarios as we are averaging metrics over diverse scenarios, but is valid in our experiments when we are controlling and giving the same stimuli for both agents. This leads to the following evaluation protocol:

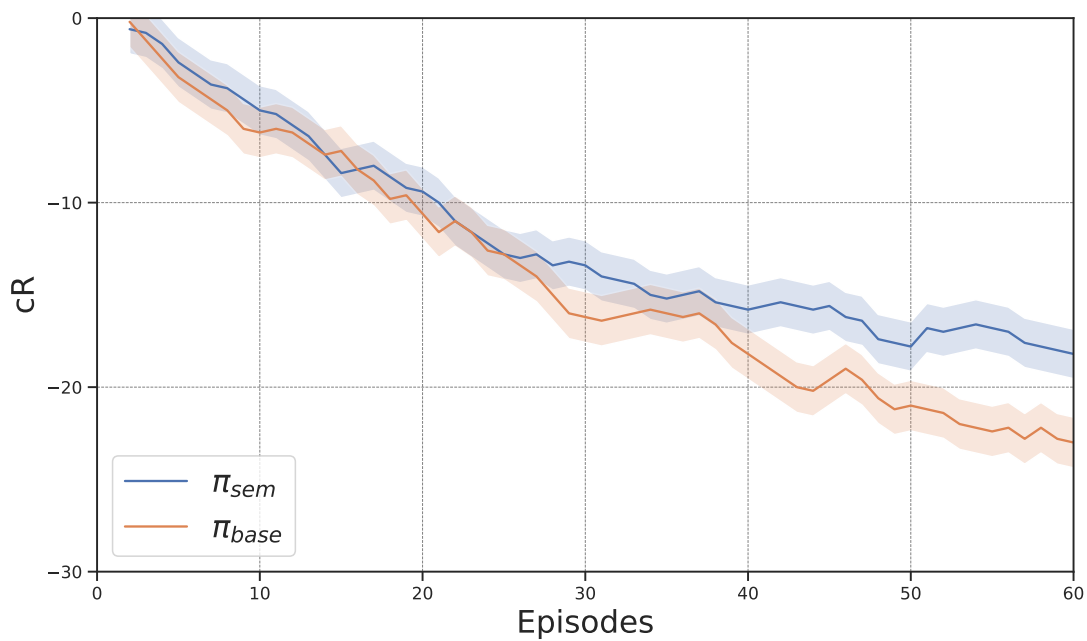
Experiment hypothesis I evaluation protocol: If FSA has benefits for processing embodied conversation when solving reference resolution under unawareness, then π_{sem} learning curves for cR and mF1 should be above and be statistically significantly different (paired t-test p-value on learning curves < 0.05) to π_{base} .

4.4.2 Results and Discussion

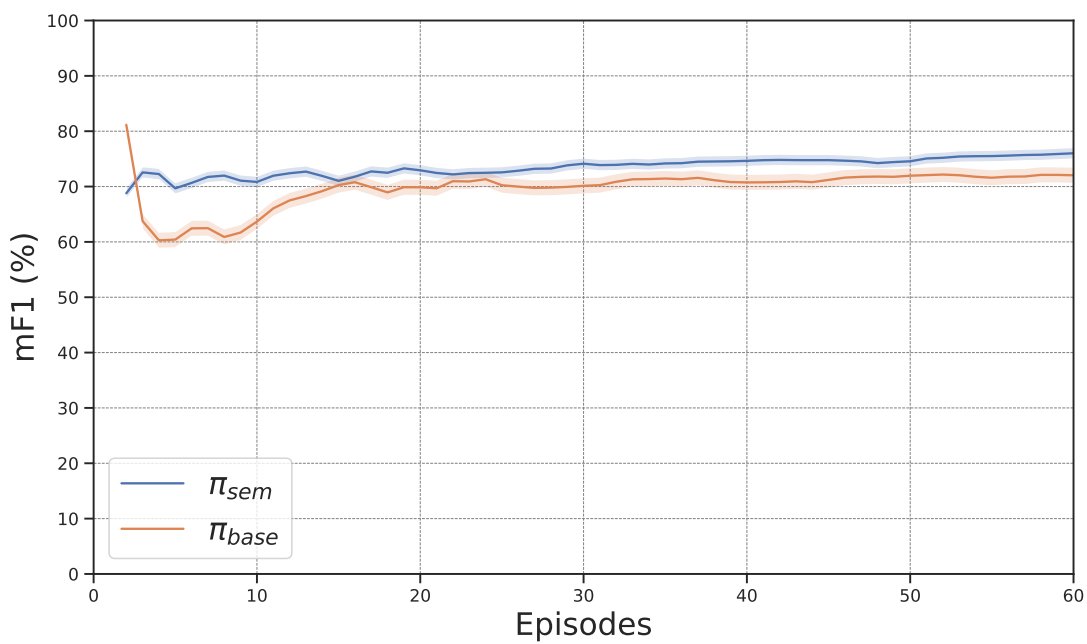
Figure 4.3 shows the learning curves for the initial setup described in Section 4.4.1. For the cumulative reward cR, the learning curve takes a downward trajectory because agents, on average, make more mistakes than choose correct referents (Eq. 4.3). When comparing π_{sem} and π_{base} performance, we observe that π_{sem} performs better on average than π_{base} for both extrinsic and intrinsic performance measures. Indeed, the cumulative reward curve for π_{sem} flattens out during the final tasks more than that for π_{base} does, showing it is making fewer mistakes towards the final trials and so has learned a more accurate grounding model. Moreover, the difference between π_{sem} and π_{base} is statistically significant (see Table 4.3). These quantitative results suggest that processing referential expressions by utilising their logical consequences is beneficial for learning to solve the task of reference resolution under unawareness.

To inspect the relationship between the stimuli and the performance of agents using π_{sem} and π_{base} processing, we conduct a sensitivity analysis (Saltelli et al., 2000). In the initial experiments, the agents faced two tasks in each environment. We now explore what happens to performance when the number of tasks in each environment is 1, 2 or 3 tasks. To quantify the difference between π_{sem} and π_{base} , we measure the *difference* in their cumulative reward $\text{cR}_{\pi_{\text{sem}}} - \text{cR}_{\pi_{\text{base}}}$ and mean F1 score $\text{mF1}_{\pi_{\text{sem}}} - \text{mF1}_{\pi_{\text{base}}}$.

Figure 4.4 shows the results. Firstly, when comparing the two types of agents, the

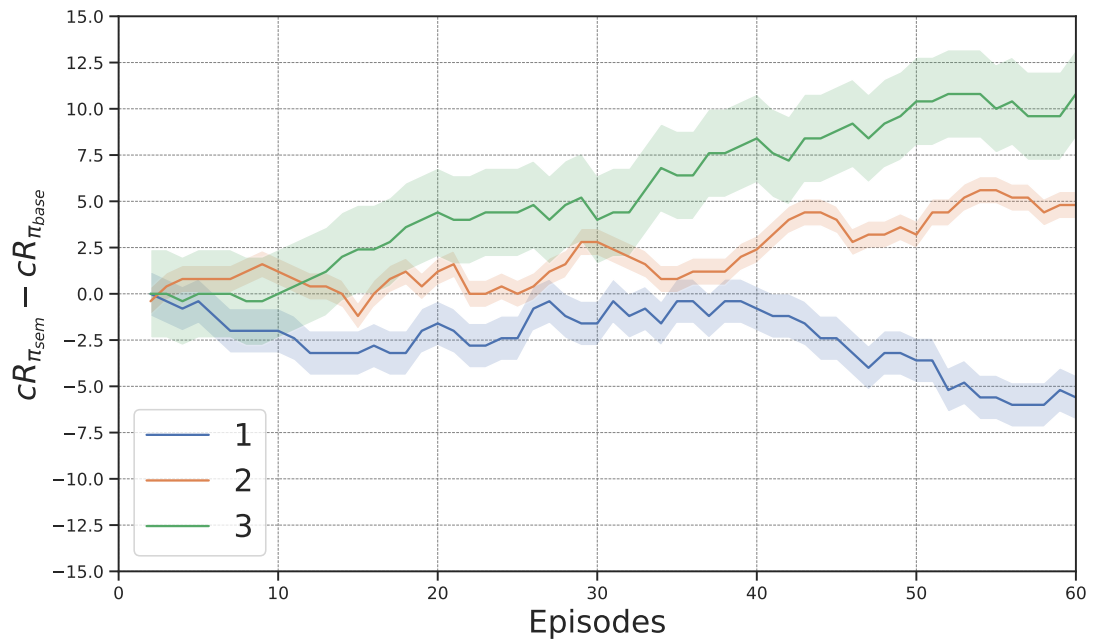


(a) Cumulative reward

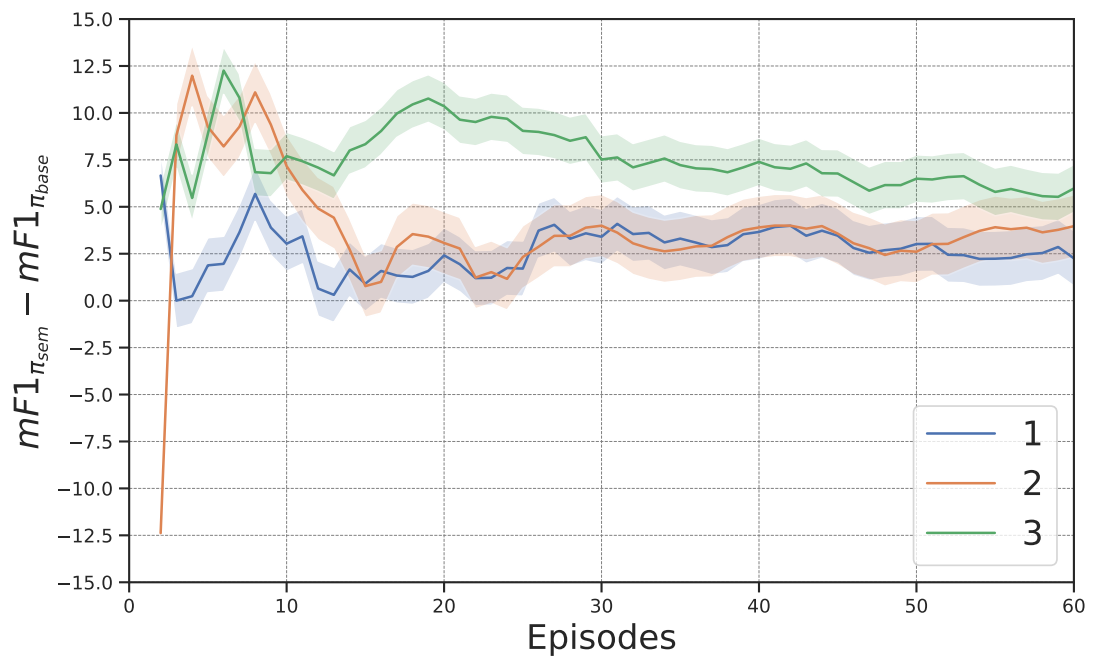


(b) Mean F1 score

Figure 4.3: The learning curves for solving the task of reference resolution under unawareness. Each curve is plotted with its 95% confidence interval.



(a) Cumulative reward



(b) Mean F1 score

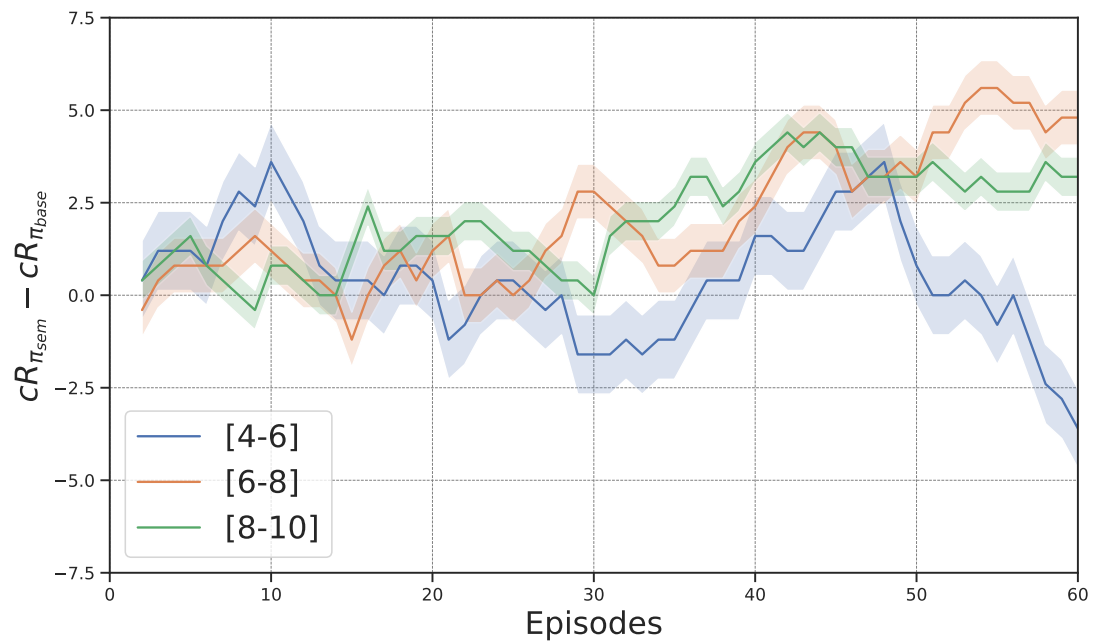
Figure 4.4: The difference in performance between π_{sem} and π_{base} , comparing learning curves when the number of tasks that the agents face in each environment in the experiments is 1, 2 or 3 tasks.

	Number of tasks			Number of objects			Quantifier probability	
	cR	mF1		cR	mF1		cR	mF1
1	$2 \cdot 10^{-16}$	$2 \cdot 10^{-24}$	[4-6]	$7 \cdot 10^{-2}$	$4 \cdot 10^{-1*}$	0.5	$2 \cdot 10^{-12}$	$4 \cdot 10^{-19}$
2	$8 \cdot 10^{-13}$	$4 \cdot 10^{-13}$	[6-8]	$7.6 \cdot 10^{-13}$	$4 \cdot 10^{-13}$	0.7	$8 \cdot 10^{-13}$	$4 \cdot 10^{-13}$
3	$7 \cdot 10^{-17}$	$3 \cdot 10^{-43}$	[8-10]	$7 \cdot 10^{-18}$	$1 \cdot 10^{-2}$	0.9	$7 \cdot 10^{-20}$	$8 \cdot 10^{-9}$

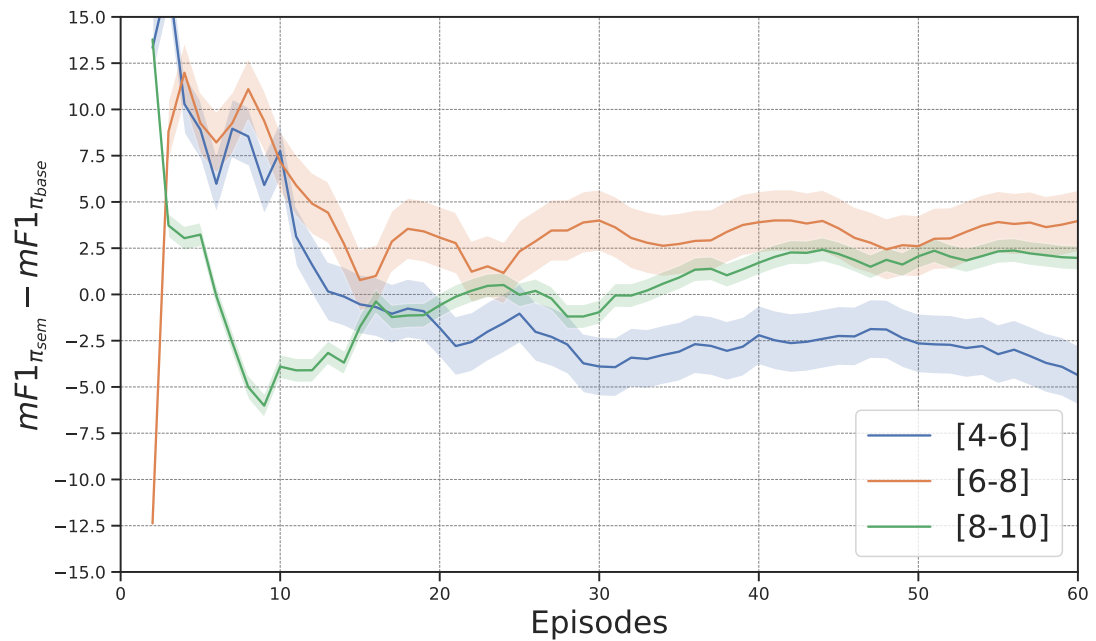
Table 4.4: P-values for comparing the agent’s learning curves in sensitivity analysis with * indicating no statistically significant difference.

difference between their performance is still statistically significant (see Table 4.4). Secondly, The learning curves show that as the number of tasks per environment increases, so does the performance difference between π_{sem} and π_{base} . The main reason to attribute this difference is that having multiple tasks in the same environment enables the π_{sem} to build a much richer domain theory via the embodied conversation compared with that built by π_{base} . This phenomenon was illustrated in Section 4.3.2 (compare the probabilistic beliefs stemming from b_2 vs b_{23}). Given these results, we can conclude that processing referential expressions using FSA is particularly useful when multiple tasks must be solved in the same environment. This is especially valuable in the real-world domain, in which an agent is expected to have an extended lifecycle in the environment that may not drastically change over time (e.g., a robot in a household).

We further study how the number of objects in the environment influences the performance difference. In the initial setup, the number of objects per environment ranged between 6 and 8 objects. To analyse this factor, we consider environments with a lower (4-6 objects) and higher (8-10) number of objects than in the initial setup. The results for these experiments are shown in Figure 4.5. When comparing the policies, there is no statistically significant difference between the agents for the lower object scenario, but this difference is statistically significant for the higher object scenario (see Table 4.4). This can also be observed when analysing Figure 4.5. Scenarios with more objects are relatively more beneficial for agents that engage in FSA: when the environments have fewer objects (4-6 objects in the Figure), the cR and mF1 differences are both lower. However, the relationship between the increasing number of objects and the difference in performance between the two agents is not linear: there is no similar size difference between π_{sem} and π_{base} for experiments in which environments contain



(a) cumulative reward



(b) Mean F1 score

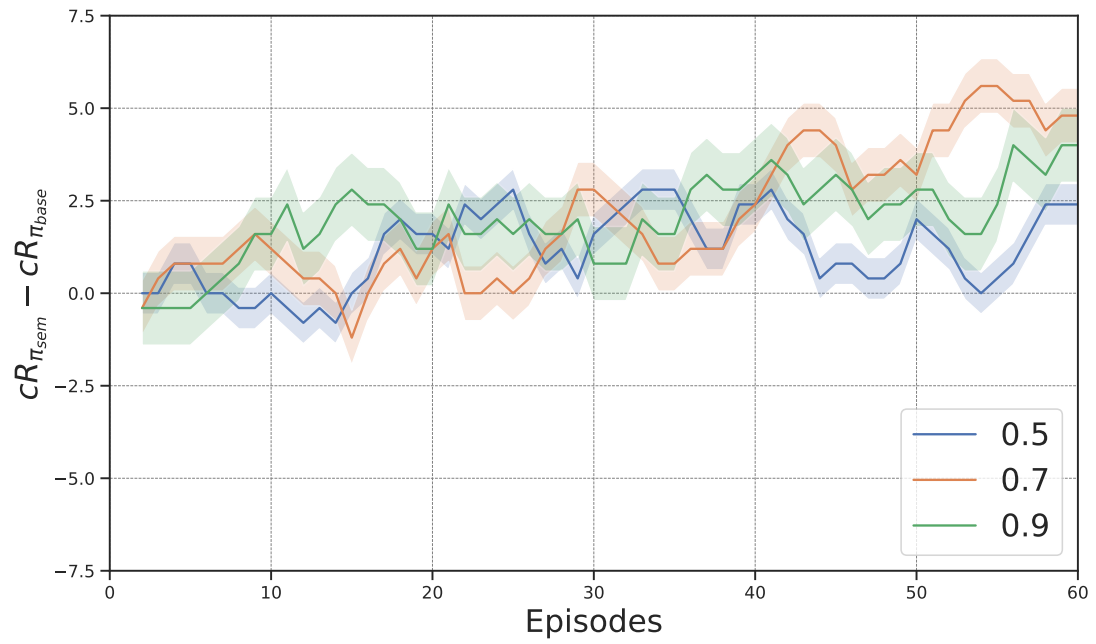
Figure 4.5: The difference in performance between π_{sem} and π_{base} , comparing learning curves (with a 95% confidence interval) when the number of objects in each environment across all trials is 4–6, or 6–8, or 8–10 objects.

more objects. We attribute this to the fact that in both the initial and the higher number objects environment setup, the agent acquires enough of the exemplars to have a good estimate of what is *not* denoted by the predicate symbols while positive exemplars are still driven by the explicit designation by the oracle.

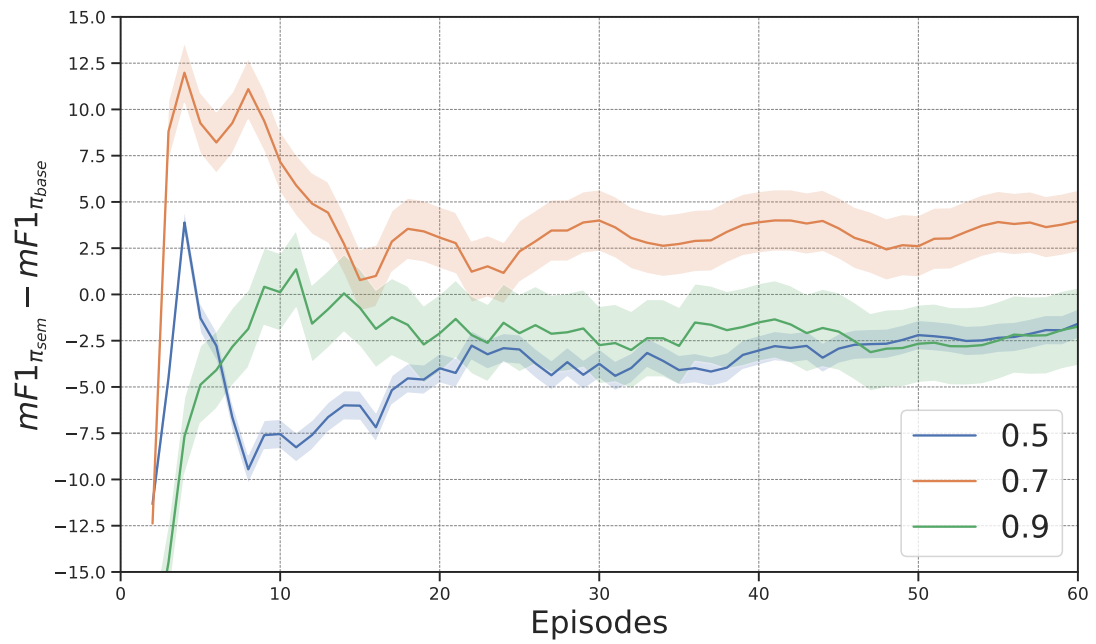
Finally, we consider the effect of the frequency of referential expressions that support relatively strong logical consequences about the environments: a.k.a. the quantifier in the referential expression has a logical consequence about the environment. In these experiments, we vary the probability of sampling a quantifier that has logical consequences about the domain to 0.5, 0.7 (which was the initial setup), and 0.9. The results of these experiments are shown in Figure 4.6. When conducting these experiments, we observe statistical difference between π_{sem} and π_{base} in all scenarios (see Table 4.4). Nevertheless, the overall task performance difference, measured by cR, is not significantly different across the different quantifier distributions. That is, no quantifier makes performing the task easier or harder for the two agents. For the intrinsic metric of mF1, too many exemplars coming from the frequent tasks with quantifiers that have logical consequences about the environment leads to lower reference estimation accuracy for the semantic agent than the base agent. We attribute this to the observation that in such scenarios, the agent acquires too many degenerate examples that may be noisy (due to the relatively low threshold τ for adding the examples), leading to a loss in performance. Because of this, we argue that to ensure good performance over the agent lifecycle, knowledge management should be deployed, too, for instance, limiting the number of exemplars stored in the support.

Besides the abovementioned experiments, we consider additional experiments conducted in [Rubavicius and Lascarides \(2022\)](#) for insights into the value of FSA when processing referential expressions. In these experiments, we extended the tasks to identifying a referent of referential expressions that are more complex than the ones reported here, in that they feature prepositional phrase: e.g., “the one red square above every magenta star”.¹ This task is challenging because the feedback on the agent’s incorrect estimate points to the referents (a “red square” in this example). However, the neologism may be in the prepositional phrase (“magenta” or “star”) and thus not designated. Nevertheless, the key takeaway from these experiments was

¹In [Rubavicius and Lascarides \(2022\)](#) instead of blocks domain, we used a shape world domain from [Kuhnle and Copestake \(2017\)](#). For generating and parsing referential expressions, we used the semantics from the English Resource Grammar ([Copestake and Flickinger, 2000](#)) and their application building resources ([Copestake et al., 2016](#)). For the grounding model, instead of prototype networks, we used matching networks ([Vinyals et al., 2016](#)).



(a) cumulative reward



(b) Mean F1 score

Figure 4.6: The difference in performance between π_{sem} and π_{base} , comparing learning curves (with a 95% confidence interval) when the probability that the task instruction features a quantifier that leads to logical consequences is 0.5, or 0.7, or 0.9.

that processing the full referential expressions is beneficial instead of just the head (i.e., a baseline processing policy that treats “the one red square above every magenta star” and “the one red square” as equivalent). Nevertheless, processing the full referential expression comes with additional computational complexity considerations. As an illustrative example, consider a referential expression $r =$ “a cube to the left of the one sphere”. This referential expression has a logical form $\Phi(r) = \langle _a_q\ x1._the_1_q\ x2.(sphere(x2), cube(x1) \wedge left(x1, x2)) \rangle$ and a corresponding referent $\mathcal{R} = \{\{o\}\}$. Using them, FSA constructs the following well-formed formula $_the_1_q\ x2.(sphere(x2), cube(o) \wedge left(o, x2))$. Processing such well-formed formulas is challenging as probabilistic inference using arbitrary first-order logic is intractable (den Broeck, 2011). Active research addresses this problem using lifted inference, in which computational complexity of WMC scales in polynomial time with the size of the domain. Recently, Kuzelka (2021) have shown that WMC in first-order logic formulas that contains up to two variables with counting quantifiers (which include all quantifiers in Table 3.1) is domain-liftable (has a polynomial time solution with the size of the domain) and Malhotra and Serafini (2022) have provided a closed-form solution. In our setting, such solutions would be valid for a single well-formed formula constructed via semantic analysis, but as the domain theory in the belief state grows, so does the number of variables for inference: Beame et al. (2015) shows that inference with a well-formed formula containing 3 variables can be $\#P_1$ -complete.² Note that this is exact inference, while there are a variety of approximate inference methods for addressing the problem of supporting inference at scale (van Bremen and Kuzelka, 2020) but analysing the effects on the performance of FSA might not be measurable in approximate computation.

Given the empirical results, we can conclude that FSA is valuable when processing referential expressions. This value is mostly prominent when task instructions have frequent usage of quantifiers that carry logical consequences about the environment and several tasks in an environment with many objects. Nevertheless, although beneficial, processing complex referential expressions may not be computationally feasible as the belief state and domain theory expand over the agent’s lifecycle.

²The $\#P_1$ -complete complexity class consists of the hardest counting problems that involve counting unique solutions to decision problems in NP, where any problem in $\#P_1$ can be reduced to these in polynomial time. An example of $\#P_1$ -complete problem would be unique, satisfying assignments in SAT.

4.5 Summary

This chapter has presented a novel and first-of-its-kind interactive neuro-symbolic grounding model utilising the belief state with a logical component. This approach satisfies several of the essential and desired features for an embodied lifelong agent that processes embodied conversation for ITL as elicited in Section 2.5:

- *Interactive symbol grounding*: The grounding model is a multilabel classifier which can cope with and expand the hypothesis space, as prediction for each predicate does not depend on other predicates (and the number of them), and thus can be reasoned about independently, given the certain knowledge Δ , as we'll see shortly. The grounding model also performs class-level incremental learning upon receiving each well-formed formula, constructed throughout the embodied conversation. The belief state is updated without a dependency on the particular domain and the task; further, the belief state supports reasoning about the possible domains and tasks, which expands as new vocabulary emerges in the extended conversation. The grounding model also works online, not requiring offline updates. Finally, when the grounding agent maintains a belief state with multiple possible hypotheses about the underlying domain models with different likelihoods, upon receiving new evidence, the likelihoods can be recomputed and previously unlikely domain models could become probable;
- *Compositional domain representation*: because of the symbolic nature of the domain model and the structure of the belief state (maintaining an explicit record of the vocabulary) our domain model is compositional and can grow over the agent's lifecycle by the expansion of the concepts the agent becomes aware of which addition does not influence the predictions of concepts that are already known by the agent.
- *Neuro-symbolic reasoning*: The grounding model can be seen as an instance of neuro-symbolic learning and reasoning, as it has both neural and symbolic components. Neural networks are used to process and extract features from high-dimensional sensory observations (in particular vision), while logical forms are used for statistical relational reasoning and, in particular, to perform 'labelling' of the support upon receiving new evidence. While many neuro-symbolic models utilise an entire symbolic component before learning as an injective bias, our model supports learning the symbolic component piecemeal, which in turn

influences visual perception. Finally, inference of the most probable domain model utilizes MAP, which is a symbolic computation that ensures monotonic constraints that result from the teacher’s assertions (and that are captured in Δ , which is a part of the agent’s belief state) are respected;

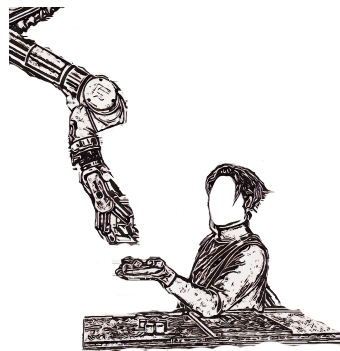
- *Formal Semantic Analysis (FSA)*: Belief state update with Algorithm 1 utilizes sentences (well-formed formulas), and as explained in Chapter 3, these formulae can be constructed to capture the logical consequences of the logical form of the teacher’s moves in the embodied conversation, or they can ignore those logical consequences (see also Section 4.1). This choice is made possible by the fact that our interactive neuro-symbolic grounding procedure supports updates with arbitrary sentences in propositional logic, and the logical forms that utilize generalised quantifiers can achieve these required propositional forms through universal instantiation and skolemisation.

The experiments have shown that the ITL agent that engaged in FSA obtained a better grounding model and performed better on the task of resolving referential expressions under unawareness than an agent that lacks reasoning with the logical consequences about the environment, confirming the experimental hypothesis I following our evaluation protocol. We explored the limitations and distilled the scenarios in which our grounding model excels (many objects, many questions per environment, many quantifiers with logical consequences) while acknowledging computational challenges in scaling our object-centric approach for grounding to dense object environments.

Chapter 5

Dialogue Strategies under Unawareness

This chapter studies the learner's decision-making process regarding interaction with the teacher when solving the task of interactive reference resolution under unawareness and how Formal Semantic Analysis (FSA) of possible questions to the teacher can help the agent learn to make better strategic decisions.



- (T1) Teacher: show me the one red cube.
(L1) Learner: Before that, show me a red object.
(T2) Teacher: Here it is. (designates a red object)
(L2) Learner: I think you are asking for this. (designates red sphere)
(T3) Teacher: No, here it is. (designates red cube)

Figure 5.1: Embodied conversation for interactive reference resolution task.

5.1 Introduction

In this thesis, we aim to develop an embodied lifelong agent that can deal with ITL scenarios in which a task instruction uttered by the user reveals to the agent unforeseen possibilities: the user utters a neologism that denotes a concept that is not part of the agent’s conceptualisation of the domain (in other words, the agent’s hypothesis space of possible states, and potentially actions as well, is wrong). In the thesis, all the initiatives in the interactions we have studied have been with the teacher. But it would be better if, whenever the agent feels very unconfident about how to fulfil the user’s needs, it could ask a pertinent question that the agent believes the user has answers to, drastically reducing their uncertainty about what to do. Doing this extends the agent’s options to those we have considered so far: the agent’s set of available actions includes attempting an action in the environment that solves the user’s task and (before or even during such an attempt) asking the user a question. Ideally, the agent should *learn* a strategy of interacting with the user effectively, given their experience so far. We address that learning task in this chapter.

We discussed the need for an ITL agent to deploy a dialogue strategy in the desiderata we argued for in Section 2.5. In particular, having a dialogue strategy is a significant tool for solving tasks under unawareness because, with it, the agent gets to control the information acquired via embodied conversation—in other words, it is an opportunity to deploy a form of *active learning*. Acquiring such a dialogue strategy (or, more generally, planning the interaction) naturally admits to the classical exploration-exploitation trade-off (Berger-Tal et al., 2014). An ITL agent may *explore* and so learn from its experience by observing the consequences of various actions in various (belief) states, or it may *exploit* its beliefs by performing what agent currently believes to be the optimal action. What the agent is aiming to learn is an optimal trade-off between asking questions, which inherently has the cost of the teacher’s effort but also the potential benefit of providing beneficial information, and risking the agent’s current beliefs to perform the task, which comes at a high cost if done incorrectly. This chapter implicitly studies this trade-off.

When considering the types of questions an ITL agent may ask the user, it is essential to reason about the information acquired from particular questions and the value this information creates for solving tasks under unawareness. In particular, in this chapter, we want to explore the effects of FSA for processing and interaction planning. As

alluded to in Chapter 1, different questions may have different values: it all depends on the likely answers the teacher will give to a question (and that depends on a combination of the question and the agent’s current beliefs about the answer the teacher will provide to the question) and how much effort the teacher puts into answering (which again, as we’ll see, can be different for different questions).

To illustrate, consider an embodied conversation in the blocks domain depicted in Figure 5.1 in which an ITL agent is asked to solve several tasks of interactive reference resolution under unawareness—the task we will study in this chapter. With T1, the learner is asked to designate “the one red cube” but is uncertain about the denotations of symbols *red* and *cube* as they are not part of the ITL agent’s domain conceptualisation before deployment. To cope with this, the learner makes a strategic dialogue move: it asks the teacher to provide an exemplar of the symbol *red* from the objects in the environment. The user’s answer reduces the agent’s uncertainty about the domain state but (unfortunately) does not correctly identify the referent, resulting in the teacher’s message (T3), indicating a wrong referent prediction. If instead, however, in this situation, the ITL agent had deployed a different dialogue strategy and had uttered “Before that, show me the one red cube” instead of L1, then the teacher’s response would have enabled the agent to complete the task successfully. What is more, the information acquired from such a question is potentially more valuable for future grounding tasks involving red objects and cubes: as we saw in Chapter 4, processing such referential expressions allows us to infer additional exemplars for the grounding model to bootstrap interactive symbol grounding using neuro-symbolic reasoning.

A key challenge in deciding what questions to ask is not knowing what questions could be asked before the task instruction, as observing neologism leads to the non-stationary and task-specific action space. To cope with this, this chapter develops the procedure to evaluate the expected information gain of asking particular questions (or not) and how this gain can influence dialogue strategy by designing the information-gathering interaction policy, capturing the exploration-exploitation trade-off.

Overall, building on top of the interactive neuro-symbolic grounding with referential expressions developed in Chapter 4, we now consider how FSA can influence interaction planning to lead to the following experimental hypothesis:

Experiment hypothesis II: An ITL agent whose dialogue strategy uses FSA masters interactive reference resolution under unawareness more effectively than an agent without it.

The rest of this chapter formalises the task we study to answer the hypothesis: the information-gathering learner for decision-making to quantify the value of various questions and the experiments to evaluate the value of FSA both for grounding and interaction planning.

5.2 Interactive Reference Resolution under Unawareness

In this chapter, to study dialogue strategies under unawareness, we consider an extension of the reference resolution under the unawareness task from Chapter 4. The only difference is that in this chapter, we add the option(s) that the agent can ask the teacher questions. We call this task *interactive reference resolution under unawareness*.

In the original version of the task from Chapter 4, given the teacher’s request to show the referent of a referential expression, the agent provided it—an action we denoted as $act \in \mathbb{A}$. The agent did *not* have the option to ask a question $q \in \mathbb{N}_{\text{quest}}$ to the user before attempting to identify a referent. Interaction, then, was the form where the teacher did all the talking! Either the teacher issued a task instruction or corrected a wrong move and gave the correct one (e.g., “No, here is the one red cube”). We now enrich the interaction to one where the agent can also contribute verbal actions. Specifically, the agent can query the teacher, resulting in an embodied conversation like the one in Figure 5.1. In effect, this extends the agent’s action space: now, $\mathbb{A} = \mathbb{N}_{\text{quest}} \cup \{act\}$.

For querying, this chapter considers two types of questions. The first type is what we’ll call *exploration questions*: to ask for exemplars for a particular referential expression. The following is an example of this type of question in the blocks domain:

Learner: Before that, show me $\underbrace{\text{all red cubes}}_{r_{\text{exp}}}$.

r_{exp} must include a subset of r ’s non-logical symbols (where r was the referential expression in the task instruction), plus a quantifier that is restricted to guarantee a valid reference. E.g., where r is “the one green cylinder”, r_{exp} could be “the one green cylinder”, “a cylinder”, “one green object”, “a green cylinder”, “every cylinder”, and so on. But “the one cylinder” and “two green cylinders” are forbidden: the former carries a Russellian uniqueness condition (there is exactly one cylinder) that might be false, and the latter violates the uniqueness condition entailed by r (see column 3 of

Table 3.2 in Section 3.2.4.2). The above query results in a response where the teacher designates a all objects in the referent \mathcal{R} , like this:

Teacher: Here it is. (points to $\Phi(r_{\text{exp}})^{\mathcal{M}}$)

This response enables the learner to construct a logical form via semantic analysis $\xi(r_{\text{exp}}, \mathcal{R})$, which, as in the reference resolution task, updates its belief state as described in Algorithm 1 from Section 4.3.1.

The second type of question the learner can ask are *clarification questions*: *yes/no*-questions about a particular object, e.g.:

Learner: Is this $\underbrace{\text{a cube?}}_{r_{\text{cla}}}$ (points to $o_{\text{cla}} \in \mathbb{O}$)

where r_{cla} is just like r_{exp} but uses quantifier “a”, “an” or “the one” to ensure that denotations are single object o_{cla} . The teacher’s response to this question is as follows:

Teacher: $\begin{cases} \text{Yes. (if } o \in \Phi(r_{\text{cla}})^{\mathcal{M}}) \\ \text{No. (otherwise)} \end{cases}$

From this exchange of messages, a different well-formed formula could be constructed via semantic analysis (both base and formal): $\xi(r_{\text{cla}}, \{\{o_{\text{cla}}\}\})$, which in turn can be used to update the belief state.

If the agent chooses action $act \in \mathbb{A}$, it use the estimated domain model $\hat{\mathcal{M}}$ to predict the referent $\hat{\mathcal{R}} = \Phi(r)^{\hat{\mathcal{M}}}$ leading to the following embodied exchange:

Learner: Here it is. (points to $\hat{\mathcal{R}}$)

The teacher responds as follows:

Teacher: $\begin{cases} \text{Correct.} & \text{if } \hat{\mathcal{R}} = \Phi(r)^{\mathcal{M}} \\ \text{No. Here it is. (points to } \Phi(r)^{\mathcal{M}}) & \text{otherwise} \end{cases}$

As before, the agent constructs a well-formed formula from the teacher’s response via semantic analysis $\xi(\Phi(r), \mathcal{R})$, which again triggers an update to its belief state.

5.3 Interaction Planning

In this section, we present and formulate a decision problem and a policy for interaction planning when solving tasks under unawareness.

5.3.1 Decision Problem

When given the ability to engage in active learning by having the option to ask a question to the teacher, the agent needs to consider the inherent cost of the teacher's effort $C: \mathbb{N}_{\text{quest}} \mapsto \mathbb{R}_{>0}$. In our decision problem, we approximate the inherent cost of answering the question using two quantities: the number of objects $Obj: \mathbb{N}_{\text{quest}} \mapsto \mathbb{R}_{>0}$ in the referent, which is determined by the head quantifier for a referential expression and approximates the pointing effort; and the number of symbols $Sym: \mathbb{N}_{\text{quest}} \mapsto \mathbb{R}_{>0}$ in the logical form of the referential expression that needs to be resolved by the teacher to answer the question, which approximates the effort to search for referents. These quantities are weighted by unit pointing cost c_{point} and unit symbol reference cost c_{ref} , resulting in the overall question cost defined as follows:

$$C(q) = c_{\text{point}}Obj(q) + c_{\text{ref}}Sym(q) \quad (5.1)$$

Different questions in different environments lead to different costs. For instance, an exploration question $q = \text{"before that, show the one red cube"}$ which is answered by the teacher designating the referent always has the cost of $c_{\text{point}} + 2c_{\text{ref}}$ as q features two symbols, and the referent always has a single object. For clarification questions like $q = \text{"Before that, is this a red object?"}$, its cost is $c_{\text{point}} + c_{\text{ref}}$ as answering this question involve resolving one symbol and one pointing gesture. Note that the resolution of the word "object" does not carry any cost as it is true for all objects in the domain (as established in Section 3.2.1). Finally, the cost of exploration question with a universal quantifier like "before that, show me every red cube" depends on the environment and the number of objects that, in this case, are red cubes, which, when solving tasks without knowing the ground-truth domain model, is not known by an agent before asking the question.

The action act 's reward depends on task success and is not observed by the agent until it attempts to solve the task. As in the reference resolution task form Chapter 4, If act results in the correct referent identified, the reward is 1, while an incorrect referent leads to the reward of -1 . Overall, this leads to the following reward function $R: \mathbb{A} \mapsto \mathbb{R}$ for the task under unawareness:

$$R(a) = \begin{cases} 1 & \text{if } a = act \text{ and if task executed successfully;} \\ -C(a) & \text{if } a \in \mathbb{N}_{\text{quest}} \text{ (question to the teacher);} \\ -1 & \text{otherwise.} \end{cases} \quad (5.2)$$

To quantify the benefits of asking the teacher various questions or choosing action $act \in \mathbb{A}$, resulting in a well-formed formula $\phi \in \mathbb{L}_{\text{snt}}$ constructed, the agent considers the information various actions may give. To do this, the agent measures an *entropy of the belief state* $\mathbb{H}: \mathbb{B} \mapsto \mathbb{R}_{>0}$ as the quantity of the information content of the belief state. Recall from Section 3.2.2 that the belief state contains elements that are known, like the set of objects and currently established vocabulary, which does not change for the task instructed due to the way we constructed the action space in Section 5.2 by not allowing question with symbols not in the task instruction. What captures the uncertainty of the belief state are the *grounded weights* w_g representing Bernoulli random variable \mathcal{B} for each atom $A \in \mathcal{H}$. Because of this, the entropy of the belief state is defined as the sum of the entropies of Bernoulli random variables, each parameterised by its grounded weights¹:

$$\mathbb{H}[b] = \sum_{A \in \mathcal{H}} \mathbb{H}_{\mathcal{B}}[w_g(A)] \quad (5.3)$$

Using entropy, we can quantify the information benefits of $\phi \in \mathbb{L}_{\text{snt}}$ from by computing *information gain* $\mathbb{H}[b] - \mathbb{H}[\text{Update}(b, \phi)]$, where $\text{Update}: \mathbb{B} \times \mathbb{L}_{\text{snt}} \mapsto \mathbb{B}$ is the belief update procedure as defined in Algorithm 1.

5.3.2 Policy

For an agent to make strategic decisions about what actions to take in the environment, it use a state-action value function $Q: \mathbb{B} \times \mathbb{A} \mapsto \mathbb{R}$ that takes into account the expected reward and expected information gain of each action.

The expected reward $\mathbb{E}_b[R(a)]$ from taking certain actions, which is given as follows:

$$\mathbb{E}_b[R(a)] = \begin{cases} -\tilde{C}(a) & \text{if } a \in \mathbb{N}_{\text{quest}} \\ 2 \cdot \text{CON}_{w_g}(\Psi(\text{MAP}_{w_g}(b)) \mid b) - 1 & \text{otherwise} \end{cases} \quad (5.4)$$

where Ψ as in Eq. 3.15 constructs a well-formed formula from a conjunction of ground atoms, based on the domain model, which in this equation is estimated using MAP (Eq.3.14) and \tilde{C} is the expected cost of asking a question which is just like C from Eq. 5.1 except in the case of the universal quantifier in which case the agent do not known $\text{NumObj}(q)$ which is approximated by the average number of objects in the

¹We do not use base weights as they only capture prior beliefs about the domain state, while grounded weights integrate the perceptual knowledge and (previous) experience in interaction giving a holistic view about agent's beliefs.

environment $\frac{|\mathcal{O}|-1}{2}$. Intuitively, the second case of Eq. 5.4 evaluates the probability of $\hat{\mathcal{M}}$ and rescales it to range $[-1,1]$ to match the range of the reward function (Eq. 5.2).

The expected information gain considers all possible well-formed formulas constructed from taking an action $a \in \mathbb{A}$ and their benefits. To operationalize this, let Result be a random variable over actions $a \in \mathbb{A}$ that encapsulates a well-formed logical formula $\phi \sim \text{Result}(a)$ constructed from the response provided by the teacher, and the semantic analysis on the learner's message (see Section 5.2 on how particular actions leads to different formulas). This formula can be used for the $\text{Update}: \mathbb{B} \times \mathbb{L}_{\text{SNT}} \mapsto \mathbb{B}$ as defined in Algorithm 1. Using it, agent can estimate the expected information gain $\mathbb{I}: \mathbb{B} \times \mathbb{A} \mapsto \mathbb{R}_{\geq 0}$ from being in particular belief state $b \in \mathbb{B}$ and taking an action $a \in \mathbb{A}$:

$$\mathbb{I}[b, a] = \mathbb{H}[b] - \mathbb{E}_{\phi \sim \text{Result}(a)}[\mathbb{H}[\text{Update}(b, \phi)]] \quad (5.5)$$

The second term of \mathbb{I} is conditional entropy (MacKay, 2003) quantifying the expected future entropy of the new belief state upon receiving ϕ . The expected information gain is equivalent in the predictive power to other information-theoretic measures like mutual information, Kullback–Leibler divergence, or value of information (Shannon, 1948; Cover and Thomas, 2005). The expectation term in Eq. 5.5 can be computed by considering possible well-formed formulas observed, weighted by their likelihood given by the Weighted Model Counting (WMC) (See Section 3.2.5 and Eq. 3.12):

$$\begin{aligned} \mathbb{E}_{\phi \sim \text{Result}(a)}[\mathbb{H}[\text{Update}(b, \phi)]] &= \sum_{\phi} \Pr(\text{Result}(a) = \phi \mid b) \mathbb{H}[\text{Update}(b, \phi)] \\ &= \sum_{\phi} \text{CON}_{w_g}(\phi \mid b) \mathbb{H}[\text{Update}(b, \phi)] \end{aligned} \quad (5.6)$$

Note this term could be computed exactly or approximately by sampling possible answers using the Monte Carlo method (Metropolis and Ulam, 1949).

The agent uses these two quantities to trade off exploration (increasing information about the domain) and exploitation. A linear interpolation in the state-action value function expresses this relation:

$$Q(b, a) = \theta_1 \mathbb{I}[b, a] + \theta_2 \mathbb{E}_b[R(a)] \quad (5.7)$$

where $\theta = [\theta_1, \theta_2]^\top$ are interpolation parameters.

Using state-action value function Q from Eq. 5.7, the agent can take a greedy action yielding the agent's dialogue strategy for interaction planning:

$$\pi(b) = \arg \max_{a \in \mathbb{A}} Q(b, a) \quad (5.8)$$

The presented exploration-exploitation policy has several unique features compared to conventional state-action value functions used in RL (Sutton and Barto, 1998). Firstly, because of the compositional nature of the belief state and, in particular, entropy computation, the agent that uses this function to plan can cope with expanding hypothesis space. In particular, introducing new concepts in the domain by assigning some base weight w_b , which, upon receiving more evidence, allows the agent to compute grounded weight w_g . Secondly, the computation of each action in Q does not depend on the values of other actions (e.g., there is no normalisation between action values). Because of this, policy utilising this state-action value function can deal with non-stationary action space which is crucial as in different tasks, different questions are considered (both because of the referential success as well as the symbols that the agent is concerned about in general). The exploration-exploitation policy in Eq. 5.8 depends only on the interpolation parameters $\theta = [\theta_1, \theta_2]^\top$, which can be optimised offline. In particular, we use semi-gradient SARSA (Rummery and Niranjan, 1994) as an algorithm for policy optimization, treating $\mathbb{I}[b, a]$ and $\mathbb{E}_b[R(a)]$ as features for the functional approximation $h(b, a) = [\mathbb{I}[b, a], \mathbb{E}_b[R(a)]]^\top$; that is:

$$Q(b, a) = \theta_1 \mathbb{I}[b, a] + \theta_2 \mathbb{E}_b[R(a)] = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \cdot \begin{bmatrix} \mathbb{I}[b, a] \\ \mathbb{E}_b[R(a)] \end{bmatrix} = \theta^\top h(b, a) \quad (5.9)$$

Algorithm 2 outlines the overall optimisation procedure for the dialogue strategy. It is conceptualised as the decision-making problem for tasks under unawareness as a single-shot decision-making problem. This makes the learner ‘myopic’: when choosing actions, the learner does not speculate about the effects that the teacher’s response to a query might have on the expected value of future actions. This modelling choice allows the policy to be tractable. Treating this as a single-shot decision problem contrasts with, for instance, Partially Observable Markov Decision Processes (POMDP), which supports sequential decision-making, trading off immediate rewards against longer-term returns (Mykel J. Kochenderfer and Wray, 2022). However, as well as the vastly increased complexity of solving a sequential decision problem, our problem cannot be modelled, even in principle, by a POMDP as they do not support an expanding hypothesis space of possible states and actions and provide no mechanism about utilising experience and learning so far on discovering an unforeseen possibility. While decision-making in our model is myopic, it does not exclude the option of asking a sequence of questions before acting in the environment (see examples of such cases that are attested in our experiments in Section 5.4.2).

Algorithm 2 Dialogue Strategy Optimisation using Episodic Semi-gradient SARSA

Require: belief state $b \in \mathbb{B}$, state-action value function Q parameterised by θ , number of tasks per environment m , learning rate α , discount factor γ , epsilon ϵ .

```

1: while not converged do
2:    $\mathcal{E} \leftarrow \text{GetEnvironment}()$  ▷ Initialise an environment
3:   for  $t$  in  $\text{GetTasks}(\mathcal{E}, m)$  do ▷ Get  $m$  tasks in environment  $\mathcal{E}$ 
4:      $b \leftarrow \text{Update}(b, \Phi(t))$  ▷ Update belief state (Algorithm 1)
5:      $\mathbb{N}_{\text{quest}} \leftarrow \text{GetQuestions}(t)$  ▷ Questions as defined in Section 5.2
6:      $\mathbb{A} \leftarrow \{\text{act}\} \cup \mathbb{N}_{\text{quest}}$  ▷ Construct action space
7:     while  $True$  do ▷ Training loop
8:        $a \leftarrow \arg \max_{a \in \mathbb{A}} Q(b, a)$  ▷ Greedy action
9:        $R, \phi \leftarrow \text{Act}(\mathcal{E}, a)$  ▷ Act in  $\mathcal{E}$  and construct  $\phi$  via semantic analysis
10:       $b' \leftarrow \text{Update}(b, \phi)$  ▷ Update belief state (Algorithm 1)
11:      if  $|R| = 1$  then ▷ Terminating state
12:         $\delta \leftarrow R - Q(s, a)$  ▷ Temporal-difference error for terminal state
13:         $\theta \leftarrow \theta + \alpha \delta \nabla_{\theta} Q(s, a)$  ▷ Parameter update by gradient ascent
14:        break
15:      else
16:         $a' \leftarrow \begin{cases} \arg \max_{a \in \mathbb{A}} Q(b', a) & \text{with prob. } 1 - \epsilon \\ \text{Choose}(\mathbb{A}) & \text{otherwise} \end{cases}$  ▷  $\epsilon$ -greedy action
17:         $\delta \leftarrow R + \gamma Q(b', a') - Q(b, a)$  ▷ Temporal-difference error
18:         $\theta \leftarrow \theta + \alpha \delta \nabla_{\theta} Q(b, a)$  ▷ Parameter update by gradient ascent
19:       $b \leftarrow b'$  ▷ Update belief state

```

5.4 Experiments

This section describes experiments we use to quantify the value of FSA for dialogue strategies under unawareness.

5.4.1 Experimental Setup

To test the experimental hypothesis II in this chapter, we conduct simulation experiments in the blocks domain following the setup in previous experiments (Section 4.4.1). The key difference now is generating an action space for the agent that includes questions, described in Section 5.2. We generate the possible questions using template-based natural language generation: this manipulates the string representation of the referential expression in the task instruction and (potentially) changes the quantifier used for exploration questions as well as to form a list of valid adjectives and nouns to ask for in the case of clarification questions. All these referential expressions and their logical forms are validated in the generation process, and if an error occurs, a new referential expression for the task is sampled.

What is unique for interactive reference resolution under unawareness is the cost function C for particular questions: i.e., $C: \mathbb{N}_{\text{quest}} \mapsto \mathbb{R}_{>0}$ (see Eq. 5.1). In our experiments, we specify this cost function with the following values: $c_{\text{point}} = c_{\text{ref}} = 0.1$.

Similarly to the experiments from Chapter 4, we compare an agent that uses FSA with an agent that does not exploit it for both grounding and decision-making. In particular, we consider the following types of agents:

- π_{sem} : uses FSA when grounding (same as Chapter 4) as well as when deciding on what action to take (in particular in expected information gain \mathbb{I} in Eq. 5.5 computation);
- π_{base} : uses base semantic analysis (not exploiting logical consequences) when grounding (same as Chapter 4) as well as when deciding what action to take (in particular in expected information gain \mathbb{I} in Eq. 5.5 computation);

For the two agents, we optimise their interpolation constants using the task in the simulated blocks domain that uses a subset of predicate symbols for task instructions in training ($\mathbb{V}_{\text{train}} = \{\text{plain, cube, red, green, blue}\}$) while for the evaluation, we consider using all task instruction that involves all 13 predicate symbols of the blocks domain. In policy optimization, we obtain that $\theta_{\text{sem}} = [0.2, 1.5]^\top$ and $\theta_{\text{base}} = [0.2, 1.0]^\top$.

agents \ performance measure	cR	mF1
π_{sem} vs. π_{base}	$3.7 \cdot 10^{-18}$	$4.2 \cdot 10^{-20}$
π_{sem} vs. π_{mix1}	$9.3 \cdot 10^{-21}$	$4.3 \cdot 10^{-28}$
π_{sem} vs. π_{mix2}	$1.1 \cdot 10^{-2}$	$2.3 \cdot 10^{-23}$
π_{base} vs. π_{mix1}	$1.2 \cdot 10^{-1*}$	$1.7 \cdot 10^{-5}$
π_{base} vs. π_{mix2}	$2.6 \cdot 10^{-21}$	$8.8 \cdot 10^{-9}$
π_{mix1} vs. π_{mix2}	$1.1 \cdot 10^{-24}$	$8.7 \cdot 10^{-1*}$

Table 5.1: P-values for comparing the agent’s learning curves for the different agents using different dialogue strategies and referential expression processing with * indicating no statistically significant difference.

Additionally, we consider two mixture agents to evaluate the isolated effects of FSA for both grounding and decision-making in an ablation study:

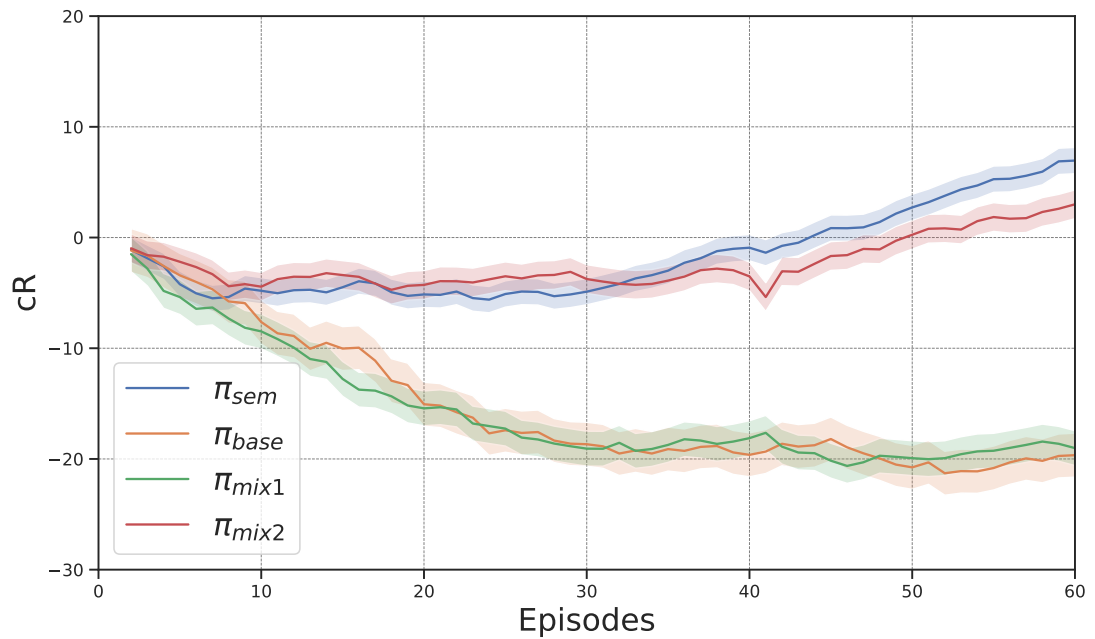
- π_{mix1} : performs grounding like π_{base} (base semantic analysis without using logical consequences), but decision-making like π_{sem} (using FSA).
- π_{mix2} : performs grounding like π_{sem} (using FSA), but decision-making like π_{base} (base semantic analysis without using logical consequences);

As in experiments in Section 4.4, we use the learning curves of the cumulative reward cR and mean F1 score mF1 as intrinsic and extrinsic performance measures leading to the following evaluation protocol:

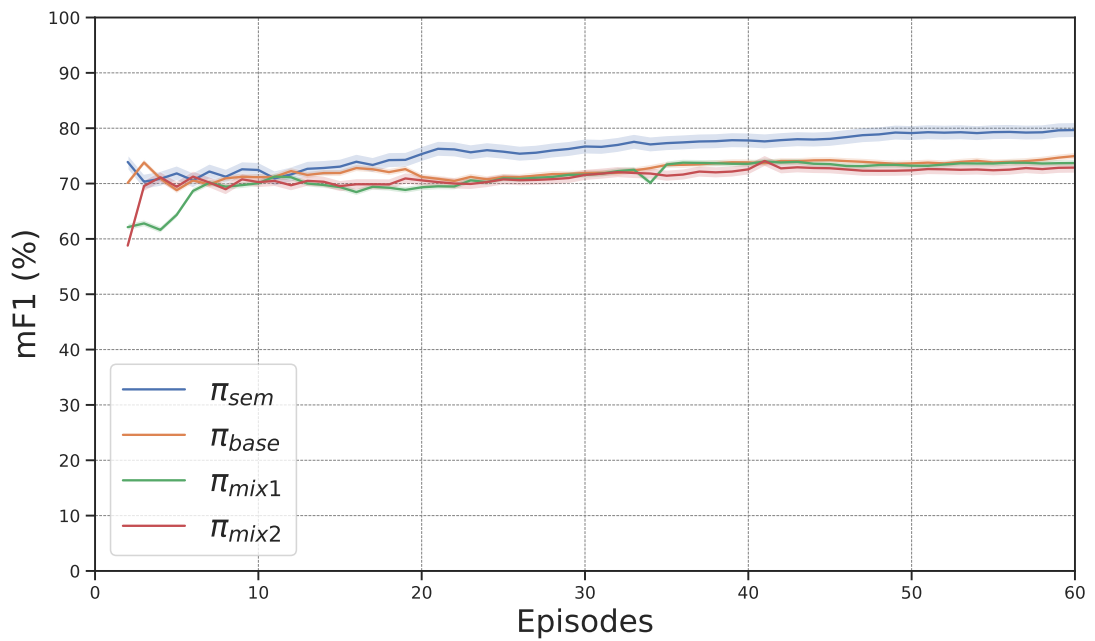
Experiment hypothesis II evaluation protocol: If FSA has benefits for the decision-making that are additional to FSA for processing referential expressions, then π_{sem} learning curves for cR and mF1 should be above and statistically significantly different (paired t-test p-value on learning curves < 0.05) to π_{base} , π_{mix1} and π_{mix2} .

5.4.2 Results and Discussion

The learning curves for each type of agent are given in Figure 5.2, and the p-values for agent comparison are given in Table 5.1. When analysing the results, we observe that π_{sem} performs statistically significantly better than π_{base} and two ablated agents (π_{mix1} and π_{mix2}) for both cR and mF1. By the end of the learning process, π_{sem} and π_{mix2} have gathered enough evidence through interaction to overcome unawareness



(a) Cumulative reward



(b) Mean F1 score

Figure 5.2: The learning curves for solving the task of interactive reference resolution under unawareness. Each curve is plotted with its 95% confidence interval.

and predict correct referents more often than not (on average). This increased cumulative reward to above 0 was not the case with the agents from Chapter 4, which did not benefit from the active learning that asking strategic questions that the teacher affords. This contrasts with the dialogue strategies of π_{base} and π_{mix1} , which are *not* using FSA in calculating the expected value of the teacher’s answer to questions. These results indicate that active learning from the embodied conversation is an effective method for solving tasks under unawareness, and especially so when the adopted strategy is computed in a way that takes FSA into account.

mF1 achieved by the agent π_{sem} , which utilises FSA for *both* decision-making and grounding, performs significantly better than the other three agents. Note that π_{mix1} —an agent that makes its decision using FSA—performs similarly to the agents that do not engage in this analysis and with no statistically significant difference to π_{base} , signalling that the benefits of the FSA for this agent are not utilised.

To better understand the difference between the behaviour of different agents and their emerging dialogue strategies, we analyse conversations produced by different agents with their representative examples given in Table 5.2 with respect to the situation depicted in Figure 5.3. For π_{sem} , we observe that whenever the task instruction affords the ability to infer logical consequences from the exploration question, the agent would choose to ask this question to acquire both the necessary information to solve the task (by asking the teacher to demonstrate the task) and rich knowledge for grounding. This contrasts with the typical conversations of π_{base} : the agent tends to ask clarification questions, presumably because for this agent, they are often cheaper than exploration questions (clarification questions only require resolving a single symbol, while an exploration question could require the resolution of several symbols in the blocks domain—e.g. “before that, show me a dotted red cube” required to resolve 3 symbols). The consequence of this penchant for clarification questions over alternatives is that several questions are given to the teacher with compounding costs, reflected in the cumulative reward curve. This dialogue strategy means that the agent denies itself from observing exemplars, which would have enabled the agent to exploit FSA when reasoning about grounding.

For π_{mix1} , which has the grounding of (π_{base} and dialogue strategy of π_{sem} , we observe that often its first question is (typically) the same as with π_{sem} (an exploration question with the referential expression in the task instruction), but the rest of the conversation diverges, as π_{mix1} continues to engage in the conversation with follow-up clarification

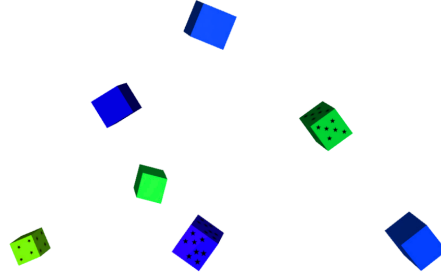


Figure 5.3: Environment in which agents interact to solve interactive reference resolution under unawareness given task instruction “show me the two starry green cubes” with embodied conversations given in Table 5.2.

questions. This can be attributed to the inferior evidence (using just a base semantic analysis in grounding) that was acquired from an exploration question, with much uncertainty remaining about the domain that π_{mix1} aims to resolve with clarification questions.

Finally, π_{mix2} 's conversation uses a mixture of clarification questions and exploration questions that include quantifiers like “the one”, which has the same cost as “a” (for both, the expected cost of using such quantifier is the same). Because of this, the agent is willing to ask this question, and significant benefits ($|\textcircled{0}| - 1$ negative exemplars acquired) for the grounding using FSA can be realised as observed in the cR behaviour, yet not as much as for π_{sem} . π_{sem} benefits from more learning opportunities by gathering more positive exemplars than π_{mix2} when the task instruction suggests that there are multiple objects to be designated, as in the dialogue depicted in Table 5.2 for the situation depicted in Figure 5.3.

In light of the experiments and their results, we can conclude that for an ITL agent, using FSA both in decision-making and in grounding is beneficial, and those benefits are complementary.

Agent	Embodied conversation	Rewards
π_{sem}	Teacher: Show me the two starry green cubes.	
	Learner: Before that, show me the two starry green cubes.	
	Teacher: Here it is (points to the two starry green cubes).	-0.5
	Learner: I think you are looking for this. (points to target referent)	
	Teacher: Correct.	1
π_{base}	Teacher: Show me the two starry green cubes.	
	Learner: Before that, is this a green object (points to a green object)	-0.2
	Teacher: Yes.	
	Learner: Before that, is this a starry object. (points to the starry object)	
	Teacher: Yes.	-0.2
	Learner: I think you are looking for this. (points to two cubes)	
	Teacher: No. Here it is. (points to target referent)	-1
π_{mix1}	Teacher: Show me the two starry green cubes.	-0.5
	Learner: Before that, show me the two starry green cubes.	
	Teacher: Here it is. (points to the two starry green cubes)	
	Learner: Before that, is this a cube. (points to a cube)	-0.2
	Teacher: Yes.	
	Learner: Before that, is this a starry object. (points to a starry object)	-0.2
	Teacher: Yes.	
	Learner: I think you are looking for this. (points to target referent)	
	Teacher: Correct.	1
π_{mix2}	Teacher: Show me the two starry green cubes.	
	Learner: Before that, show me a starry green cube.	
	Teacher: Here it is (points to a starry green cube).	-0.5
	Learner: Before that, is this a green object (points to a green object)	
	Teacher: Yes.	-0.2
	Learner: Before that, is this a starry object (points to a non starry object)	
	Teacher: No.	-0.2
	Learner: I think you are looking for this. (points to target referent)	
	Teacher: Correct.	1

Table 5.2: Sample embodied conversations and rewards received from the teacher's embodied conversation messages by different agents when solving tasks under unawareness in the environment in Figure 5.3.

5.5 Summary

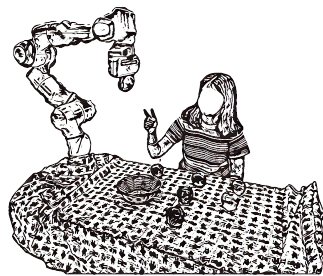
In this chapter, we designed and evaluated dialogue strategies that agents can deploy to help them learn the task of reference resolution under unawareness. The unique feature of our modelling approach is the ability to cope with non-stationary action space and at the same time utilize our neuro-symbolic grounding model. This was achieved by learning dialogue strategies that take into account the FSA of questions and possible answers, in particular, the expected value of information to be gained. We conducted a series of simulation experiments to validate and contrast strategies that use FSA and those that do not. We showed the value of FSA for decision-making and grounding, with an ablation study revealing compounding benefits. This confirms our experimental hypothesis II following our evaluation protocol. Our evaluation for different dialogue strategies is nevertheless limited as we considered only the exact inference method which is not realistic in more complex scenarios but was required to correctly measure the value of FSA.

Coming back to our elicited system features in Section 2.5, in this Chapter we built on top of our modelling in Chapter 4 and now in addition we satisfy the requirement of interaction planning which was the focus in this Chapter. With these components, our agent is now able to solve tasks under unawareness adhering to all the desiderata.

Chapter 6

Processing Corrective Feedback

This chapter continues our study of using Formal Semantic Analysis (FSA) to process embodied conversation in service to teaching an agent new tasks in scenarios where the agent is unaware of words and the domain-level concepts they denote but which are critical to solving the task. In this chapter, we study a particular teacher move: namely, correcting the agent's latest action in the environment. We focus on how using FSA insights from coherence-based theories for interpreting embodied conversation can enhance embodied lifelong agents. We centre this study on a task of rearrangement under unawareness, of the kind illustrated in Figure 6.1.



(T1) Teacher: Put the two granny smiths inside the basket.

(L1) Learner: Let's try to execute the task. (picks golden delicious)

(T2) Teacher: No. This is a golden delicious.

(L2) Learner: Sorry. Let's try again. (returns golden delicious to original location)

Figure 6.1: Embodied conversation with teacher's corrective feedback.

6.1 Introduction

This thesis has so far designed an interactive neuro-symbolic grounding model and an interaction planning algorithm that utilizes Formal Semantic Analysis (FSA). Together, these enhance the data efficiency of embodied lifelong agents when they face an ITL scenario in which the user gives a task instruction containing a concept the agent was unaware of before deployment. In this chapter, we consider further ways in which FSA could be integrated into the learning algorithms for ITL: specifically, the interpretation of the teacher’s embodied conversation messages, given in response to mistakes the agent has just made in manipulating the environment. In other words, this chapter focuses on exploiting the truth conditions of the teacher’s corrective feedback.

The beginning of the thesis (Section 1.1) motivated the notion of tasks under unawareness by considering the confidence-awareness matrix (see Figure 1.2). The algorithms developed in the previous two chapters for interactive neuro-symbolic grounding and interaction planning enable the agent to move from the unknown unknowns region to known knowns by gathering evidence via embodied conversation to reduce the uncertainty about the underlying domain model. We now look at the region of *unknown knowns*: in other words, contexts in which the agent is confident but wrong. One way to become aware that you’re wrong is for a competent teacher to tell you so via a speech act of correction. That speech act can not only express you’re wrong with the linguistic cue “No.” but provide further information about the nature of the mistake; see utterance T2 in Figure 6.1 for an example. In this scenario, the agent that follows the base architecture described in Chapter 3 chooses within their interaction planning module *not* to engage in the embodied conversation—in other words, it is sufficiently confident about their estimate of the domain model to act in the environment rather than acquire more information about the model. However, their estimated domain model is wrong in a way that makes the agent’s action wrong because it is not a part of an optimal plan to achieve a goal state. In particular, an agent predicts an object being a “granny smith” when it is not.

Such a situation is entirely plausible: even when the agent is aware of the existence of the concept “granny smith”, it may not have a sufficiently good sample to date for distinguishing fruits that are “granny smith” from those that are not, but thinks it does! For instance, the agent has possibly acquired many examples of granny smiths, but only in environments where the other objects were not apples. In that case, it may well

confuse granny smiths with other (similar looking) apples at inference time, but be confident about classifying granny smiths. In this chapter, we argue that when faced with such a contingency, embodied conversation—as the universal means of natural interaction—can help solve tasks under unawareness due to these unforeseen possibilities, as postulated by one of the thesis hypotheses.

To handle this contingency, we now endow the teacher with the option of giving task instruction answering questions and monitoring the agent’s execution actions for reaching the goal state and, when such an action is not optimal, providing corrective feedback in the contextually relevant manner. Such corrective feedback is part of the embodied conversation, potentially contributing to providing useful evidence for solving the task. For instance, in the example from Figure 6.1, the teacher is proactive in observing the agent’s mistake—picking golden delicious when such an action is not required to reach the goal state. The teacher informs the agent of its mistake with the following embodied conversation message “No. This is a golden delicious (points to the picked object)”.

This corrective feedback can have several effects on the agent’s beliefs. Firstly, the agent observes the referential expression “a golden delicious” with its referent (i.e., the picked object), and using the techniques detailed in Chapter 4 and 5, it can update their grounding model. In particular, this updates their positive support for golden delicious (i.e., it learns something about the visual features of objects denoted by “golden delicious”). Secondly, specifically from the fact that this is a *corrective* move, the agent can also infer that the picked object is *not* a granny smith—if it were, then the truth conditions of being a corrective move would not be satisfied. So not only does the agent learn more about how to recognise golden delicious apples, but also how to recognise granny smiths—i.e., those same visual features, via the truth conditions of corrections, will be added to the negative support set for granny smith. In other words, recognising the teacher’s embodied conversation message as corrective feedback carries entailments that go beyond those of simply asserting T2 without considering how that assertion semantically relates to its context. We aim to show in this chapter that exploiting such entailments enhances learning efficiency for embodied lifelong agents.

[Appelgren and Lascarides \(2020\)](#) have studied how embodied corrective feedback can be used for ITL. They focussed on solving an instance of tasks under unawareness (block stacking) with the constraints for the domain model estimation dictated by coherence relations between the user’s embodied corrective feedback and the action ex-

ecuted in the environment. However, in their work, the agent could not engage in embodied conversation more generally: they used *only* the embodied corrective feedback for interactive symbol grounding. In this chapter, we aim to *integrate* the insights from FSA of embodied corrective feedback with the algorithms developed so far in this thesis, which support other types of speech acts such as asking and answering questions. We aim to show that processing *both* embodied corrective feedback and learning and deploying a policy for engaging in embodied conversation yields compounding benefits when solving tasks under unawareness. This leads to our first experimental hypothesis in this chapter:

Experiment hypothesis III: An ITL agent that is both contributing to embodied conversations and processing corrective feedback using FSA masters the task of rearrangement under unawareness more efficiently than an agent that uses just one of these learning signals.

Another facet to explore is the overall *robustness* of the proposed architecture in Chapter 3 and how it can cope with false beliefs that may be initially postulated by the agent (i.e., unknown knowns). In particular, this chapter takes a closer look at what happens when the base weights in the belief state take on values other than the default weight (of 0.5, corresponding to complete ignorance) and how those various weights (in particular, weights that contribute to false belief) influence grounding and decision-making. This leads to our second empirical hypothesis:

Experiment hypothesis IV: An ITL agent processing embodied conversation that includes corrective feedback using FSA is robust to prior (and possibly false) beliefs and can revise them when solving rearrangement under unawareness more efficiently than an agent that does use FSA.

The rest of this chapter formulates the task of rearrangement under unawareness, outlines how corrections on execution actions can be integrated into the embodied life-long agent architecture from Chapter 3, and how a FSA of corrective feedback can contribute to data efficiency and the above two empirical hypotheses.

6.2 Rearrangement under Unawareness

Rearrangement (Batra et al., 2020) is a manipulation task in which, given a goal specification, the agent manipulates rigid bodies in the environment from the initial state to a state that satisfies the goal, using its sensory observations to inform its plan. This

chapter considers a special case of rearrangement, in which, when observing the goal specification in the form of a task instruction $t \in \mathbb{N}_{\text{task}}$, the agent is unaware of at least one concept that features in the task instruction. We call this the task of *rearrangement under unawareness*.

Two kinds of rearrangements in simulated and real-world environments are considered:

- *move rearrangement*: the teacher requests the learner to move objects in the simulated blocks tabletop environment to other locations relative to some other objects in the environment, with a task instruction such as:

Teacher: move $\underbrace{\text{every red cube}}_{r_{\text{direct}}}$ to the left of $\underbrace{\text{the one blue cylinder}}_{r_{\text{indirect}}}$.

where $r_{\text{direct}}, r_{\text{indirect}} \in \mathbb{N}_{\text{ref}}$ are referential expressions in the task instruction t , denoting direct objects and indirect objects of the move rearrangement, respectively.

- *put rearrangement*: the teacher requests the learner to place some fruits inside the basket in the real-world tabletop environment (more details are given in Section 6.4.1), with a task instruction such as:

Teacher: put $\underbrace{\text{the two granny smiths}}_{r_{\text{direct}}}$ inside the basket.

where $r_{\text{direct}} \in \mathbb{N}_{\text{ref}}$ is the referential expression in the task instruction denoting the direct object of the put manipulation (the indirect object is the basket, which for this kind of rearrangement is fixed).

To solve both kinds of rearrangement under unawareness, the agent has to identify denotations of the referential expressions that are part of the task instruction by estimating the domain model $\hat{\mathcal{M}}$, given its current belief state $b \in \mathbb{B}$. The agent can reduce the uncertainty of the belief state by asking the teacher questions in the embodied conversation, as studied in Chapter 5. In this chapter, we consider questions $\mathbb{N}_{\text{quest}}$, for which embodied conversation messages have the following form:

Learner: Before this, show me $\underbrace{\text{every red cube}}_{r_{\text{show}}}$.

where $r_{\text{show}} \in \mathbb{N}_{\text{ref}}$ is a referential expression like r_{direct} or r_{indirect} , save that there is an option of replacing the quantifier in the referential expression with an existential

(“a/an”) or a universal (“every/all”) quantifier. These questions are similar to the exploration questions in the task of interactive reference resolution under unawareness, which we studied in Section 5.2.

Upon receiving the agent’s question, the teacher responds with an embodied conversation message that designates a referent:

Teacher: Here it is. (points to $\Phi(r_{\text{show}})^{\mathcal{M}}$)

Compared to tasks of reference resolution from Chapters 4 and 5, the mistakes in the domain model estimation are not observed by the teacher immediately but rather get exposed as the learner acts in the environment in an attempt to solve the task. In other words, the learner chooses a sequence of actions $act \in \mathbb{A}$ of pick and place moves to achieve a state that satisfies the goal, and before executing that plan learner utters the following embodied conversation message:

Learner: Okay. Let’s try to execute the task.

As rearrangement under unawareness is a sequential decision-making problem, requiring potentially multiple execution actions to reach a goal state, the learner performs Task and Motion Planning (TAMP), as depicted in the agent’s architecture presented in Chapter 3. Using the TAMP module, the agent computes the sequence of manipulation moves in the environment that will, according to its beliefs, transform the current state to a goal state. For the move and put rearrangements, the agent uses two types of manipulation moves:

- *pick*: picking an object in the environment;
- *place*: placing the picked object in the desired location in the environment.

These manipulation moves are atomic and deterministic.¹ The sequence of these manipulation moves defines the task plan. When this plan is executed, after every manipulation move, the teacher evaluates whether that latest move is part of an optimal plan or not. If it is not, the teacher immediately responds with an embodied conversation message—corrective feedback $c \in \mathbb{N}_{\text{corr}}$ —of the following form (see, for instance,

¹In practice, the actions are executed continuously using motion planners and controllers. Treating manipulation moves as continuous leads to various timing considerations, partially studied by [Appelgren and Lascarides \(2021\)](#), but they are out of the scope of this thesis.

Figure 6.1):

Teacher: No. This is $\underbrace{\text{a blue cube}}_{r_{\text{corr}}}$. (points to o_{corr})

where $r_{\text{corr}} \in \mathbb{N}_{\text{ref}}$ is a referential expression and $o_{\text{corr}} \in \mathbb{O}$ is an object designated by the teacher in the embodied conversation message of the corrective feedback c .

Additionally, when the learner finishes executing the task plan, agent performs the *complete* action, which constitutes uttering the following:

Learner: I have finished the task.

The teacher's response to this message is one of the two embodied conversation messages:

$$\text{Teacher: } \begin{cases} \text{You have executed the task correctly.} & \text{if goal state is reached.} \\ \text{No. This is } \underbrace{\text{a red cube}}_{r_{\text{corr}}}. \text{ (points to } o_{\text{corr}} \text{)} & \text{otherwise.} \end{cases}$$

The second option is corrective feedback on a premature completion: that is, the learner believes it has finished the task, but the teacher indicates this is not the case by pointing to an item that still needs to be moved.

Given the outlined setup for the task of rearrangement under unawareness, the agent's action space increases when compared to the action space of interactive reference resolution under unawareness (Section 5.2) by three additional execution actions $\mathbb{P} = \{\text{pick}, \text{place}, \text{complete}\}$, leading to the overall action space $\mathbb{A} = \mathbb{N}_{\text{quest}} \cup \{\text{act}\} \cup \mathbb{P}$. Because of the way the interaction for solving the task under unawareness is set, *act* denotes a decision to act in the environment given the agent's current beliefs, and by choosing this action, the actions in \mathbb{P} become valid for an agent in the current belief state while $\mathbb{N}_{\text{quest}}$ are not (so the agent does not ask the user mid-execution about the domain); at the same time, in case of the corrective feedback, the agent not only revises their belief but as well is again faced with the ability to engage in the embodied conversation ($\mathbb{N}_{\text{quest}}$ are valid actions) but no longer executed actions in the environment (\mathbb{P} are not valid actions).

In this chapter, the ITL agent will perform FSA of the teacher's corrections on its executed actions. We assume corrections are explicitly signalled with "No." and always correct the agent's latest action. Therefore, the teacher and the learner know which move is being corrected (and that this move was not a part of an optimal plan). FSA

in this scenario is used to construct a well-formed formula, which in turn updates the belief state (in particular Δ). Therefore, this, in turn, updates interaction planning (e.g., on updating beliefs, the agent may now decide to ask a question). We study how these formulae are constructed in the next Section.

When solving rearrangement, we use the reward function in Eq. 5.2 from Chapter 5. However, we limit the agent's attempts to solve each task to four attempts. If this limit is reached, then the agent is given a new task to solve.

6.3 Corrections on Execution Actions

This section presents FSA of corrective feedback on executed actions within our task domain of rearrangement under unawareness. We draw on coherence-based semantic theories of discourse, and in particular the coherence relation *Correction* (Asher and Lascarides, 2003; Hunter et al., 2018).

6.3.1 Discourse Coherence

Coherence-based theories of discourse interpretation are built upon the assumption that when a participant in the conversation makes a dialogue move, then in coherent conversation, that move is semantically related to a prior and salient part of the discourse context. Hunter et al. (2018) go further and argue that non-linguistic events (which in our case would be pick and place moves that the agent performs) can also be appropriated by a speaker to become a part of the content of the conversation, even though the agent of those moves did not mean for it to be communicative. The teacher's corrective feedback to a pick or place move is a specific example of this phenomenon: when the agent performs an execution action $a \in \mathbb{P}$ that is not a part of the optimal plan, then the teacher utters (embodied) corrective feedback $c \in \mathbb{N}_{\text{corr}}$, such that *Correction*(a, c) holds, and is a part of the content of the conversation (where *Correction* is a coherence relation.) In this thesis, we are concerned with how FSA of coherence relations, particularly the relation *Correction*, enhances the agent's capacity to estimate the domain model.

According to Lascarides and Asher (2009), the content of *Correction*(a, c) can be true only if:

1. corrective feedback $c \in \mathbb{N}_{\text{corr}}$ is true (or to put this another way, correction is a

type of assertion, and so carries the assertion’s illocutionary force). In our case, that means that the object $o \in \mathbb{O}$ that the teacher designates with their pointing gesture is truly denoted by the referential expression $r_{\text{corr}} \in \mathbb{N}_{\text{ref}}$ that features in c . So this can be used to build a well-formed formula using semantic analysis ξ studied before;

2. The corrected action $a \in \mathbb{P}$ is not a part of an optimal plan. Given the task setup, this leads to further consequences, which are detailed in the next section. But informally, if a is *pick*, then the object picked should not have been picked (it cannot be picking *per se* that is wrong, given the tasks call for pick and place to be performed alternately); if a is *place*, then the picked object was correctly picked (because otherwise, the prior pick action would have been corrected), but this object should not have been placed where it was placed, and if a is *complete* then the task is not completed and the designated object has to be manipulated.

Taken together, these inferences about corrections, in particular what follows from the executed action a not being a part of an optimal plan, make some domain models unsatisfiable.

The second effect is essential when processing embodied conversation in ITL because it reveals to the agent their false beliefs.

6.3.2 Formal Semantic Analysis

Given the logical consequences of corrective feedback in our specific domain, we can assume that the following well-formed formulae are constructed.

As mentioned earlier, for $pick \in \mathbb{P}$ followed by the corrective feedback c , FSA dictates that the picked object should have *not* been picked. Hence the picked object is not denoted by the direct object referential expression r_{direct} in the task instruction t . To illustrate, consider the task instruction t = “move the one cube in front of a cylinder.” and a correction c = “No. This is a sphere. (points to $o \in \mathbb{O}$)”. In this scenario, the following well-formed formulae can be constructed: $\text{sphere}(o)$ stems from the first effect of the coherent corrective feedback, and $\neg \text{cube}(o)$ from the second effect (because the only objects in an optimal plan that should be picked are cubes). More generally, FSA for a corrective move c of a $pick \in \mathbb{P}$ is as follows:

$$\zeta(\text{pick}, t, c) = \xi(r_{\text{corr}}, \{\{o_{\text{corr}}\}\}) \wedge \neg \xi(r'_{\text{direct}}, \{\{o_{\text{corr}}\}\}) \quad (6.1)$$

where r'_{direct} is a referential expression just like r_{direct} but with an existential quantifier to make the $\xi(r'_{\text{direct}}, \{\{o_{\text{corr}}\}\})$ construction valid (e.g. if $r_{\text{direct}} = \text{“the one cube”}$, then $r'_{\text{direct}} = \text{“a cube”}$).

For $place \in \mathbb{P}$ followed by the corrective feedback c , FSA is similar to the $pick \in \mathbb{P}$ case, except this time the object designated in the embodied conversation message indicates that it not part of the referent of the indirect object referential expression r_{indirect} in the task instruction t . Following the same illustrative example above, if the corrective feedback $c = \text{“No. This is a sphere. (points to } o \in \mathbb{O}\text{)”}$ is produced in response to $place \in \mathbb{P}$, then the agent would (again) be able to infer that $sphere(o)$ (by the first effect of corrective feedback, that the correction is true). But this time, the agent would also infer that $\neg cylinder(o)$, because the critical choice with object placement is to ensure that, as a result of the action, the placed object is in the intended spatial relation to the indirect object.² More generally, FSA for $place \in \mathbb{P}$ is as follows:

$$\zeta(place, t, c) = \xi(r_{\text{corr}}, \{\{o_{\text{corr}}\}\}) \wedge \neg \xi(r'_{\text{indirect}}, \{\{o_{\text{corr}}\}\}) \quad (6.2)$$

where r'_{indirect} is a referential expression just like r_{indirect} but with existential quantifier to make $\xi(r'_{\text{direct}}, \{\{o_{\text{corr}}\}\})$ construction valid similar to r'_{direct} from FSA of $pick$.

For $complete \in \mathbb{P}$ followed by the corrective feedback c , the correction entails that the task is *not* complete (for a correction always makes an utterance that it corrects false Asher and Lascarides (2003)). What is unique for $complete$ is that both well-formed formulae inferred from the effects of the corrective feedback using FSA are the same. To illustrate, consider correction $c = \text{“No. This is a cube (points to } o \in \mathbb{O}\text{)”}$ in response to the $complete$ execution action, using the first effect of the corrective feedback, the agent can infer that $cube(o)$ is true. As predicted by WMC, the agent must, therefore, reject any domain model for which $\neg cube(o)$ is true. This will inevitably have a material effect on the agent’s beliefs because if the agent thought the plan was complete before, then it must have believed, with sufficient confidence, that the object o was not a cube. So FSA for $complete \in \mathbb{P}$ leads to the following well-formed formula:

$$\zeta(complete, t, c) = \xi(r_{\text{corr}}, \{\{o_{\text{corr}}\}\}) \quad (6.3)$$

²There could have been different embodied conversation messages to correct $place$, like not designating the object, leaving it to the agent to infer which object in the scene the teacher is referring to via its knowledge of coherent corrections. However, this makes reasoning much more complex because it introduces ambiguity of reference, which must be resolved in the context of uncertainty about the domain model. We avoid this within the scope of this thesis, but, in principle, the reasoning methods presented here, particularly WMC, can handle such cases as well.

Furthermore, the second effect of the corrective move c entails that the current state does not satisfy the goal: in other words, there is a cube that is *not* in front of a cylinder. Treating the utterance “this is a cube (points to o)” as an *Explanation* of the corrective cue “no” entails that it is o that is the renegade cube, not in front of a cylinder. With the ground truth to spatial relations, the agent could infer from this that every object in front of o is not a cylinder. But we forego exploiting this entailment in our experiments and simply show that the logical consequences on the symbols that denote the designated object o are enough to improve data efficiency.

Using the outlined FSA, the above well-formed formulae can be constructed and used to update the belief state following Algorithm 1. In this semantic analysis, we have both the notion of the base analysis (i.e., the first effect of the corrective feedback that simply carries the illocutionary force of an assertion) and the full semantic analysis exploiting logical consequences (i.e., the second effect of corrective feedback). What is important compared to FSA studied before is that exploiting logical consequences is *necessary*, as otherwise, the corrective feedback would not have an effect on the learner’s behaviour. For example, consider the scenario in which for the task with an instruction t = “move the one cube in front of a cylinder” *pick* is followed by the correction c = “No. This is a sphere (points to $o \in \mathbb{O}$)”. If the agent ignores the logical consequences of this interaction (i.e., the second effect of corrective feedback), then the belief state update may not lead to the desired change in the domain model, and the agent will still believe what it picked up was a cube (as well as a sphere). Thus, the learner would continue to execute the plan without considering the teacher’s corrective feedback.

6.4 Experiments

This section describes the experiments to evaluate the two hypotheses of this chapter.

6.4.1 Experimental Setup

In the experiments, we consider two domains: the simulated blocks domain from the previous chapters and the real-world domain of tabletop fruit manipulation. In both cases, the agent has to solve rearrangement under unawareness, but in the latter case, we focus on showing how our architecture can fix the problem when a large vision model is not (pre-)trained to identify a particular symbol.



(a) Environment 1



(b) Environment 2



(c) Environment 3



(d) Environment 4

Table 6.1: Visualisation of apple manipulation environments, their ground-truth labels, and initial grounding DINO predictions *before* interaction with the environment and the user via embodied conversation.

For generating task instructions in the simulated blocks domain, we use template-based natural language generation to produce two referential expressions $r_{\text{direct}} \in \mathbb{N}_{\text{ref}}$ and $r_{\text{indirect}} \in \mathbb{N}_{\text{sub}}$, ensuring that the objects denoted by r_{direct} and r_{indirect} are not overlapping and they are not yet in the target spatial relationship (that is the current state is not a goal state). For reasoning about spatial relationships, we use rudimentary spatial reasoning functions that take the locations of two objects and evaluate them depending on their relative position. In these tasks, we only consider simple 2D spatial relationships signalled by the words “to the left of”, “to the right of”, “in front of”, and “behind”.³ The possible questions that the agent can use to engage in the embodied conversation are also generated via template-based natural language generation, as in Chapter 5. When the agent decides in the interaction planning to $act \in \mathbb{A}$, TAMP module is used to attempt to solve the task. In particular, the estimated domain model $\hat{\mathcal{M}}$ is used to estimate $\Phi(r_{\text{direct}})^{\hat{\mathcal{M}}}$ and $\Phi(r_{\text{indirect}})^{\hat{\mathcal{M}}}$. Because of the way the task has been set up, task planning is simple and can be achieved by a sequence of pick moves of objects in $\Phi(r_{\text{direct}})^{\hat{\mathcal{M}}}$ and placing them in the target spatial relationship to $\Phi(r_{\text{indirect}})^{\hat{\mathcal{M}}}$, with the location determined by sampling a collision-free placement in the tabletop environment.

The manipulation moves (motion planning) are executed by using an operational space controller (Khatib, 1987) and the rudimentary logic for setting control targets based on the pose of the object of interest. After each manipulation move is executed, the oracle is queried to check if the move that was just performed in the simulated environment is part of an optimal plan (that is, if the object from $\Phi(r_{\text{direct}})^{\mathcal{M}}$ is picked and is placed in the target spatial relationship to $\Phi(r_{\text{direct}})^{\mathcal{M}}$) and if not, producing the corrective feedback $c \in \mathbb{N}_{\text{corr}}$ that we described earlier. In response to this feedback, the robot undoes the manipulation move (returns the object to its location before the move) and uses the updated belief state to plan subsequent interactions. In these experiments, the agent is given one task per environment (as opposed to two tasks per environment from the previous chapters) as ITL agents have longer interactions with the same environment due to more complex tasks than in Chapters 4 and 5. In these tasks, the agent can attempt to solve the task up to 5 times (that is, execute 5 task plans and receive 5 corrections). If that is not enough, the episode terminates and a new task is presented to the agent in a new environment instance.

³Reasoning about arbitrary spatial relationships between objects depends not only on object position but the identities of the objects themselves (Levit and Roy, 2007). This thesis avoids the need for such complex reasoning by only considering block-like objects and simple 2D spatial relationships.

We also conduct real-world experiments in a fruit domain (see Table 6.1). In this domain, the Franka Emika Panda robot is tasked to solve a rearrangement with the task instruction “put the two granny smith apples inside the basket.” in 4 different environments. To detect fruits of interest and assign base weights for `grannysmith`, we use grounding DINO (Liu et al., 2023): an open-vocabulary object detection and grounding model, which when given the prompt “granny smith apple”⁴ localizes all the apples in each environment and outputs the similarity score between the prompt and localised area in the environment. This score is used as a base weight for `grannysmith`, and in all environments, this ranges between 0.4 to 0.7 before interaction.

To evaluate ITL agents, we conduct runs with all possible sequences of environment permutations (6 in total) in which the robot starts unaware of `grannysmith` and through embodied conversation with the human teacher (both questions and corrections) gathers evidence. As in the simulation experiments, the robot can attempt to solve the task 5 times before termination and addressing the next task in the new environment. For real robot experiments, we have modified unit pointing $c_{\text{point}} = 0.2$ and reference costs $c_{\text{ref}} = 0.6$ to make querying more expensive and encourage the robot to query only on situations of high uncertainty—these tend to arise at the beginning of interaction when a neologism is encountered. When the robot decides to execute the plan, it performs manipulation moves using MoveIt (Chitta, 2016) to plan and execute collision-free trajectories with pick-and-place locations, determined using transporter networks (Zeng et al., 2020) from a ZED 2i camera feed. The human teacher monitors the execution and, after each pick-and-place, corrects if necessary or stays silent.

To evaluate the experimental hypotheses III and IV of this chapter, we consider the following ITL agents:

- π_{sem} : the agent can engage in conversation before execution and utilises FSA for interaction planning (see Chapter 5), for grounding (see Chapter 4), and for processing corrective feedback (see Section 6.3). For the dialogue strategy, the agent uses Q function parameterised by $\theta_{\text{sem}} = [0.2, 1.5]^{\top}$, as in Chapter 5;
- π_{base} : the agent can engage in conversation before execution but uses just the base semantic analysis for both interaction planning and grounding; it process corrective feedback using FSA outlined in Section 6.3.⁵ For the dialogue strat-

⁴We tried just “granny smith” as a prompt, but grounding DINO fails to detect any object for this prompt in all environments, suggesting a limited awareness about such a concept.

⁵As explained earlier, if the semantic effects of being a correction as opposed to simply an assertion

weights\agents	π_{sem} VS. π_{base}		π_{sem} VS. π_{corr}		π_{base} VS. π_{corr}	
	cR	mF1	cR	mF1	cR	mF1
w_b	$1 \cdot 10^{-22}$	$1 \cdot 10^{-30}$	$4 \cdot 10^{-26}$	$3 \cdot 10^{-42}$	$1 \cdot 10^{-21}$	$4 \cdot 10^{-8}$
w_b^{opt}	$7 \cdot 10^{-1*}$	$5 \cdot 10^{-37}$	$4 \cdot 10^{-28}$	$3 \cdot 10^{-32}$	$1 \cdot 10^{-26}$	$9 \cdot 10^{-1*}$
w_b^{pes}	$7 \cdot 10^{-1*}$	$3 \cdot 10^{-17}$	$3 \cdot 10^{-28}$	$2 \cdot 10^{-6}$	$7 \cdot 10^{-25}$	$9 \cdot 10^{-14}$

Table 6.2: P-values for comparing the agent’s learning curves for default, optimistic, and pessimistic base weights with * indicating no statistically significant difference.

egy, the agent uses Q function parameterised by $\theta_{\text{base}} = [0.2, 1.0]^\top$, as in Chapter 5;

- π_{corr} : the agent lacks all querying capabilities but processes corrections using FSA as outlined in Section 6.3.

As in the other experiments in this thesis, the ITL agents performance is evaluated on cR and mF1 learning curves. This leads to the following evaluation protocols:

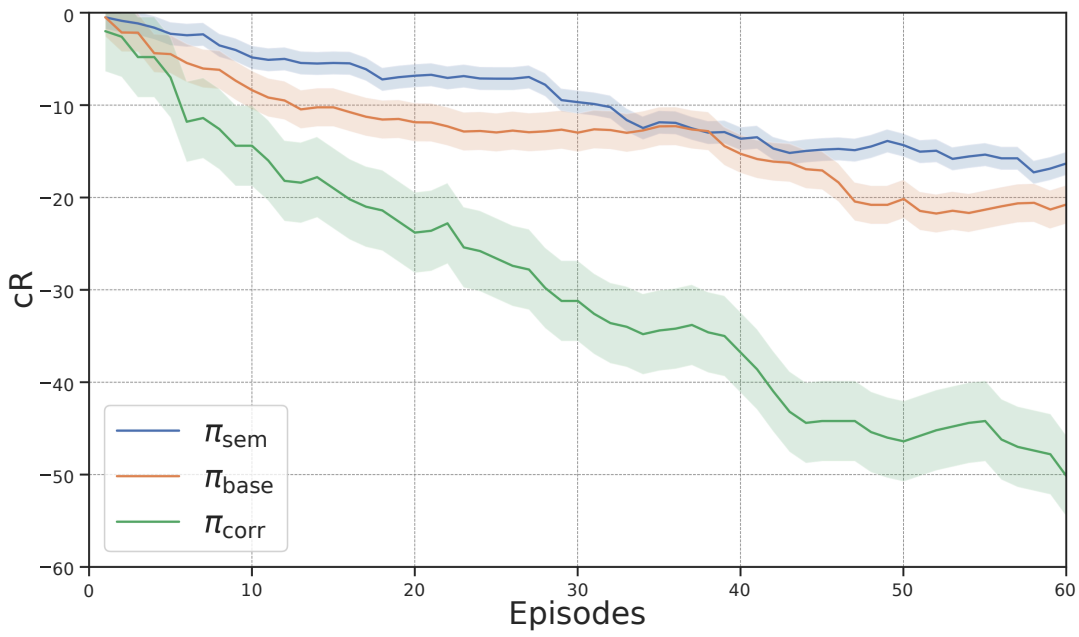
Experiment hypothesis III evaluation protocol: If both contributing to the embodied conversation and processing corrective feedback using FSA is beneficial, then π_{sem} learning curves for cR and mF1 should be above and statistically significantly different (paired t-test p-value on learning curves < 0.05) to π_{base} and π_{corr} .

Experiment hypothesis IV evaluation protocol: If FSA for processing embodied conversation that includes corrective feedback is robust to possibly false prior beliefs, then the agent comparison conclusions from evaluating hypothesis III, should not change even if we change the base weights to be optimistic or pessimistic.

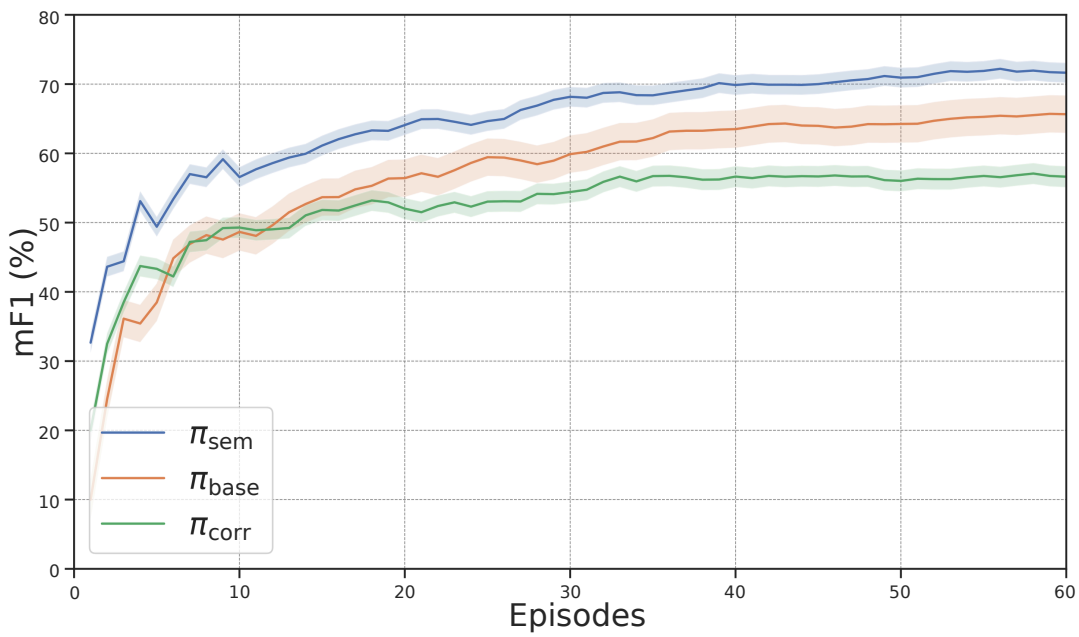
6.4.2 Results and Discussion

The results of the simulation experiments with the initial default setup are given in Figure 6.2. As before, cR taking negative values indicates both robot querying and corrections in execution occurring during interaction before succeeding (or not) to execute the task. From the results, we observe that ITL agents that engage in conversation before execution (π_{sem} and π_{base}) perform statistically significantly better than π_{corr} (see Table 6.2). This signifies the value of engaging in conversation and using it for more

get ignored, then there will be no revision of a false belief, even in principle.



(a) Commulative reward



(b) Mean F1 score

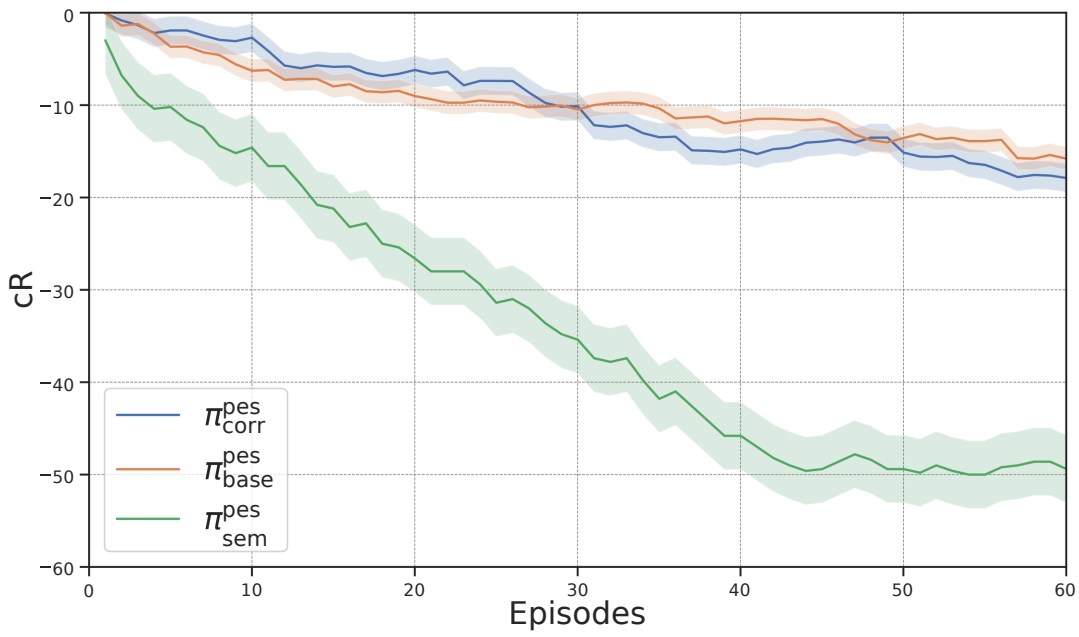
Figure 6.2: Learning curves for solving move rearrangement under unawareness for π_{sem} , π_{base} , and π_{corr} given the default setup. Each curve is plotted with a 95% confidence interval.

sample-efficient learning than corrective feedback alone, which is qualitatively similar to RL from the sparse reward signal. The difference in performance between π_{sem} and π_{base} is also statistically significant (see Table 6.2). Comparing cR for π_{sem} and π_{base} in more detail, for the majority of the tasks π_{sem} performs better than π_{base} . The mF1 score for π_{sem} is also (on average) 5% better than that for π_{base} and 10% better than π_{corr} 's. This signifies the benefits of using an ITL agent whose strategic decisions factor in the additional knowledge one gains from the semantic analysis: e.g., designating “every cube” entails the non-designated objects are not cubes.

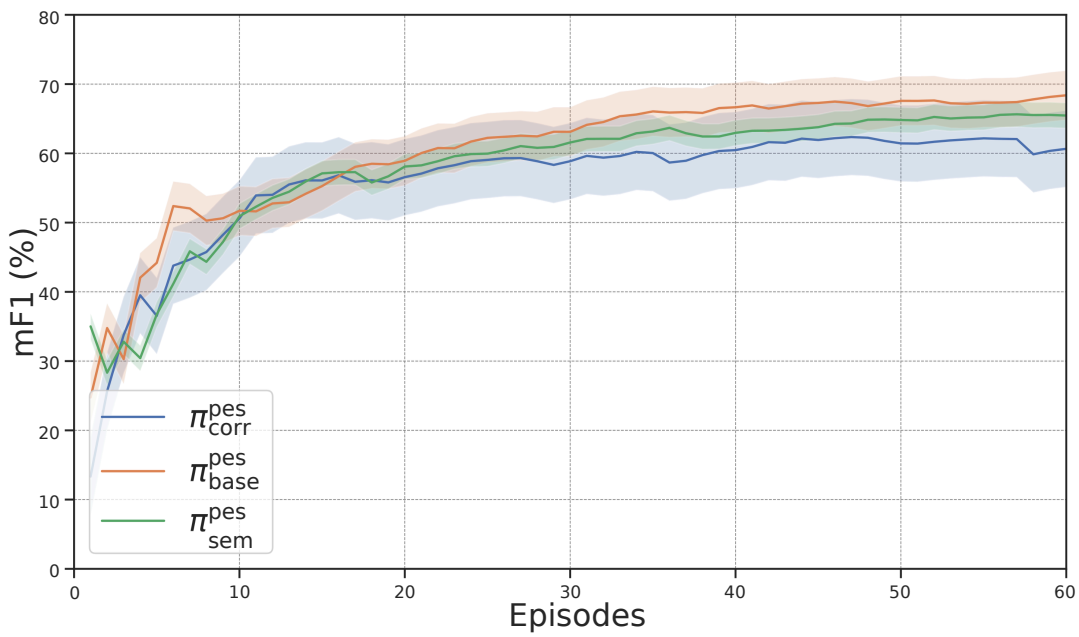
We additionally test whether our proposed ITL agent can cope with (initial) biased beliefs that might be false through its reasoning about observed evidence. To evaluate this, we consider a modified version of the above experiments, in which for all textures (plain, dotted, starry), we assigned optimistic ($w_b^{\text{opt}} = 0.7$) and pessimistic ($w_b^{\text{pes}} = 0.3$) base weights: these respectively impose an initial bias that all objects are (and respectively are not) plain, dotted and starry. Thus they encourage (respectively discourage) symbol prediction. Figure 6.4 and Figure 6.3 show the results. When comparing cR values for both pessimistic and optimistic, (π_{corr} performs statistically significantly worse than ITL agents that engage in embodied conversation (see Table 6.2). For π_{base} and π_{sem} , for the period of the first 20 tasks, π_{sem} performs better for both optimistic and pessimistic cases while on later trials the performance difference is not conclusive, fluctuating between π_{base} or π_{sem} . Their mF1 scores show that having optimistic beliefs hurts π_{base} , making its grounding performance similar to π_{corr} , while π_{sem} has similar grounding performance as with default base weights. On the other hand, pessimistic base weights do not yield any statistically significant difference in cR when comparing π_{sem} with π_{base} and π_{corr} . This is not surprising: the effects of FSA diminish because the logical consequences they yield that predict atoms are false, which matches the pessimistic bias.

Table 6.4 shows the summary results of the real-world experiments. π_{sem} has highest and statistically significant cR values after interaction in all 4 environments while it is not statistically different for mF1. We observe the following behaviours of the ITL agents on learning to the new concept `grannysmith`:

- π_{sem} : is highly uncertain about the denotations of `grannysmith` and keeps asking for evidence using the question “show me the two granny smiths”, after which no reward is obtained for the task (-1 for the question and +1 for task success). After several such questions, the agent can correctly identify denotations

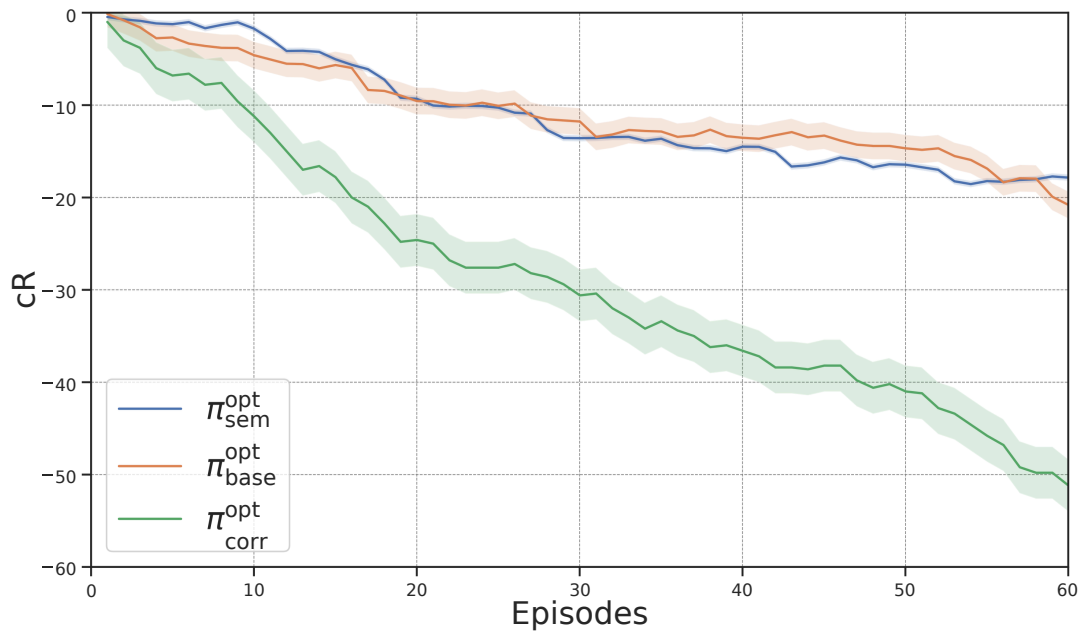


(a) Commulative reward

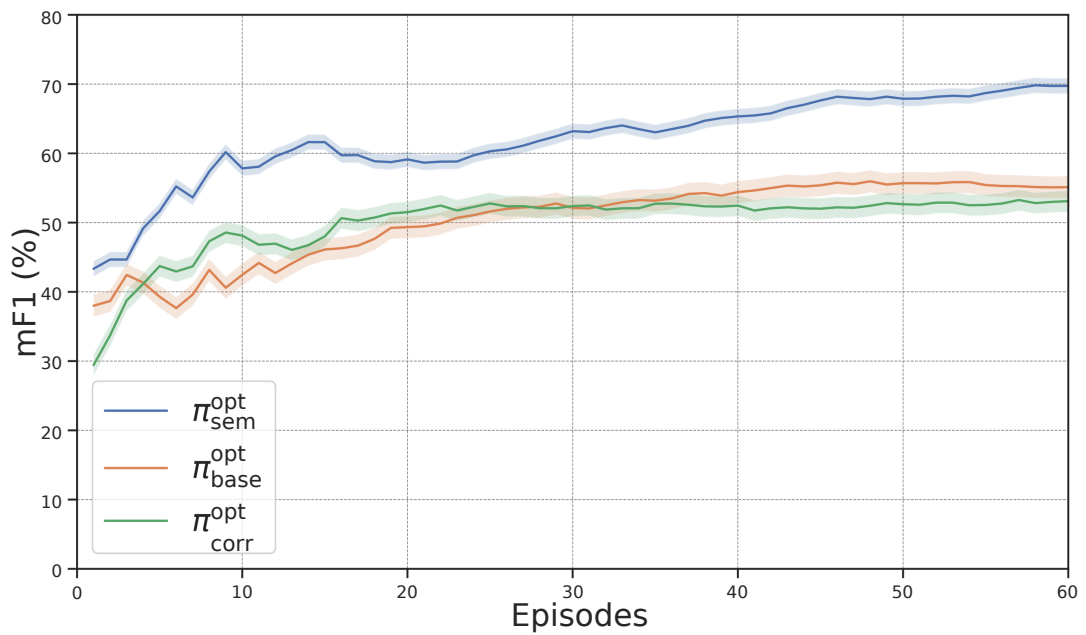


(b) Mean F1 score

Figure 6.3: Learning curves for solving move rearrangement under unawareness for pessimistic agents: $\pi_{\text{sem}}^{\text{pes}}$, $\pi_{\text{base}}^{\text{pes}}$, and $\pi_{\text{corr}}^{\text{pes}}$. Each curve is plotted with a 95% confidence interval.



(a) Commulative reward



(b) Mean F1 score

Figure 6.4: Learning curves for solving move rearrangement under unawareness for optimistic agents: π_{sem}^{opt} , π_{base}^{opt} , and π_{corr}^{opt} . Each curve is plotted with a 95% confidence interval.

and is confident enough to decide to attempt to solve the task;

- π_{base} : is highly uncertain about the domain and often first opts to ask a question “show me a granny smith”, which always results in several corrections on falsely picked fruits afterwards. Importantly, from the evidence gathered, π_{base} is not able to significantly change its grounding from the evidence gathered from its previous tasks and always requires some kind of interaction with the user to solve the task.
- π_{corr} : is attempting to solve the task from corrections, leading to lowest cR, on average requiring 3 attempts to solve the task. At the same time, this gathers a lot of negative evidence, so after attempting to solve several tasks, this policy succeeds later in identifying referents as indicated on average higher mF1 than for π_{base} .

For additional intuitions about agent behaviours, see Table 6.4 for interaction traces for different agents.

Given the experimental results, we conclude that the ITL agent that can *both* engage in embodied conversation and processes corrective feedback using embodied conversation can effectively solve a task of rearrangement under unawareness. Additionally, conducted experiments to evaluate the overall robustness of the proposed architecture have shown that FSA is particularly valuable in scenarios when an agent may have strong optimistic beliefs about the domain (that is, the agent strongly believes that a certain symbol denotes something while it is not.).

Initial belief state

Embodied conversation

Updated belief state

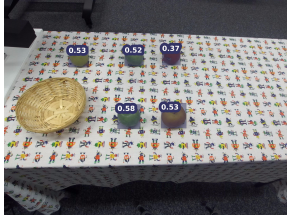

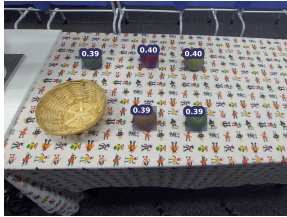
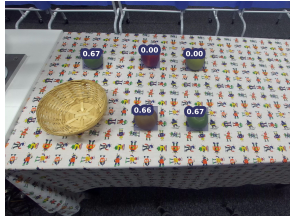
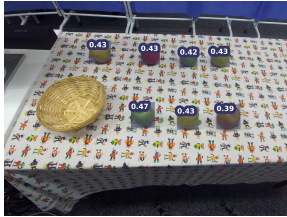
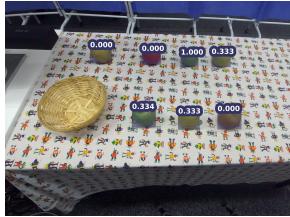


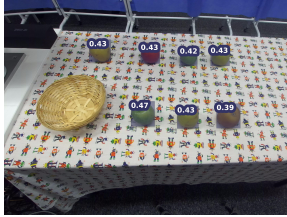



Initial belief state	Embodied conversation	Updated belief state
π_{corr}		
	<p>Teacher: Put the two granny smiths inside the basket. Learner: Okay. Let's try to achieve the task Teacher: No. This is a golden delicious. [-1] Learner: Sorry. Let's try again. Teacher: No. This is a russet. [-1] Learner: Sorry. Let's try again. Teacher: No. This is a red delicious. [-1] Learner: Sorry. Let's try again. Learner: I have finished the task. Teacher: Correct. [1]</p>	
	<p>Teacher: Put the two granny smiths inside the basket. Learner: Okay. Let's try to achieve the task Teacher: No. This is a red delicious. [-1] Learner: Sorry. Let's try again. Teacher: No. This is a golden delicious. [-1] Learner: Sorry. Let's try again. Learner: I have finished the task. Teacher: Correct. [1]</p>	
π_{base}		
	<p>Teacher: Put the two granny smiths inside the basket. Learner: Before that, show me a granny smith. Teacher: (designates one of the granny smiths) [-0.8] Learner: Okay. Let's try to achieve the task Teacher: No. This is a golden delicious. [-1] Learner: Sorry. Let's try again. Teacher: No. This is a golden delicious. [-1] Learner: Sorry. Let's try again. Learner: I have finished the task. Teacher: Correct. [1]</p>	
	<p>Teacher: Put the two granny smiths inside the basket. Learner: Okay. Let's try to achieve the task Teacher: No. This is a russet. [-1] Learner: Sorry. Let's try again. Learner: I have finished the task. Teacher: Correct. [1]</p>	
π_{sem}		
	<p>Teacher: Put the two granny smiths inside the basket. Learner: Before that, show me the two granny smiths. [-1] Teacher: (designates the two granny smiths) Learner: Okay. Let's try to achieve the task Learner: I have finished the task. Teacher: Correct. [1]</p>	
	<p>Teacher: Put the two granny smiths inside the basket. Learner: Okay. Let's try to achieve the task Learner: I have finished the task. Teacher: Correct. [1]</p>	

Table 6.3: Learner behaviour traces for in two consequent environments showing initial belief state, embodied conversation with rewards given in angle brackets and the updated belief state.

	cR	mF1
π_{corr}	-3.70 ± 0.53	50.54 ± 3.72
π_{base}	-3.45 ± 0.72	45.11 ± 3.08
π_{sem}	$0.47 \pm 0.44^*$	51.63 ± 3.99

Table 6.4: Real-world experiment results with 95% confidence interval with * indicates a statistically significant difference between other agents.

6.5 Summary

In this chapter, we have studied how corrective feedback, prompted by an example of false belief (unknown knowns) when trying to solve the task of rearrangement under unawareness, can make learning more efficient. We conducted experiments to demonstrate this in both a simulated blocks domain and the real-world fruit domain. We formalised how the FSA of corrections of pick and place moves and of falsely believing the current state is a goal state, drawing on coherent-based semantic theories of discourse. We showed how this semantics can be used in these scenarios by conducting experiments to validate the experimental hypothesis following accompanying evaluation protocols that an ITL agent with the capacity to both engage in embodied conversation before attempting to solve the task as well as process corrections of its executed actions, can learn to solve task under unawareness more sample efficiently than an agent that lacks such capabilities. We also evaluate the overall robustness of using FSA with possibly false beliefs in which we observe that FSA is robust and does not degenerate in case unfavourable pessimistic scenarios whilst having additional benefits in case of optimistic scenarios. The result of the experiments conducted are nevertheless limited as we concentrated on rearrangement with pre-defined FSA for our execution actions, simplifying a lot of analysis from the general theory of interpreting corrective dialogue moves.

Chapter 7

Conclusions

7.1 Findings

In this thesis, we have developed a framework for an embodied lifelong agent that processes embodied conversation in service to ITL, focusing on learning tasks where the agent is unaware of domain-level concepts critical to solving it. The development of such a framework is motivated by the contingencies the generalist agent, even with contemporary models of visions, language, and action, faces in real-world scenarios when they encounter new concepts and have to learn through natural interaction with the user and not through extensive retraining or fine-tuning. Embodied conversation is studied as a mode of natural interaction, and its Formal Semantic Analysis (FSA) is explored. The unique feature of our agents is that their decisions and learning are influenced by the logical consequences of the teacher’s utterances, including determiners, negation, and the satisfaction conditions of corrective feedback on actions. We empirically demonstrated the benefits of FSA via simulations and real-world experiments.

We probed the value of FSA in three ways. Firstly, we looked into how utilising the FSA of referential expressions—in particular, the truth conditions of generalized quantifiers—contributes to learning grounding models and, therefore, also helps the learner to identify referents of expressions in the embodied conversation within the current visual scene. We showed that an agent who knows and uses the truth conditions of quantifiers learns to ground and perform reference resolution under unawareness more efficiently than a learner who lacks these capabilities.

Secondly, we studied how FSA can be used to learn to make strategic decisions about

what to do during the learning process, especially when the learner is unaware of concepts that matter. Should the learner ask the teacher a question about the domain (which comes at a small cost) or risk attempting to solve the task using their current beliefs (which comes at a big cost if the learner gets it wrong)? Our experiments revealed that FSA helps with decision-making and grounding: performing FSA for both is better than exploiting it only for grounding or when considering what actions to take.

Finally, we considered using FSA when processing the teacher's corrective feedback to a suboptimal action done by the learner. We showed how FSA has benefits in overcoming and revising false beliefs about the domain model. We conducted experiments by considering agents that can both engage in the conversation with and without the FSA and can process corrections in execution. The results of these experiments have shown once again the value of FSA when processing embodied conversation for ITL.

In creating this framework for processing embodied conversation for ITL, we made several unique contributions to the field. Firstly, we developed an interactive neuro-symbolic grounding approach that uses symbolic evidence from the FSA and dense sensory observations to update the agent's belief state. Secondly, we designed a dialogue strategy that can cope with a non-stationary action space when solving tasks under unawareness and integrates FSA in the decision-making. Finally, we created and tested the overall framework on top of the robotic system for solving rearrangements under unawareness (a type of sequential decision task) by engaging in embodied conversation and processing corrective feedback, again utilising the insights of FSA.

The findings of this thesis have a broad value for ML and NLP communities. For ML practitioners, this thesis offers a look at how to develop methods that go beyond the standard supervised, unsupervised, and reinforcement learning, by showing a way to learn using incidental supervision (Roth, 2017) that comes from the user choosing to utter linguistic expressions in a seemingly idiosyncratic, but inferentially valuable way. Such learning methods have merits for future development of embodied lifelong agents, since the features of ITL scenarios that we have considered here—referential expressions featuring quantifiers and neologisms, and corrections—may well occur in domains in which ML practitioners aim to deploy robots interacting with users in the real world. For NLP practitioners, this thesis offers a return to the old ideas developed by computational linguists in the 20th century in the formal and computational semantics of the natural language. This thesis integrates these insights with current ML methods, including neural approaches to interpreting visual sensory data.

7.2 Future Work

Several future directions can be pursued using the work conducted in this thesis to enhance further the capabilities of ITL agents that process the embodied conversation.

Firstly, this thesis explored a limited range of linguistic phenomena that can be modelled using FSA. Phenomena like conditionals or modals (Gamut, 1991b) can be embedded in the current framework without significant modification, as they broadly can be reduced to well-formed formulas in first-order logic. However, many linguistic phenomena are best tackled with non-monotonic rather than classical first-order logic (Gabbay, 1985; Stachniak, 1995), such as generic rules (e.g., “birds fly”) or mass terms (e.g., “coffee in the mug”). Handling these linguistic constructions requires non-trivial adaptation to the proposed framework. For processing generics, we would need to remove the assumption that the logical form constructed by processing embodied conversation yields only monotonic inferences was used to build a logic theory. For mass terms, we would need to extend our theory of reference to take into account the non-discrete nature of mass terms and their interaction with other linguistic expressions (Nicolás, 2024).

Secondly, to utilize FSA for reasoning about the domain, we took an object-centric view of the environment in which computational resources to perform various probabilistic queries depend on the number of objects, predicates, and the arity of predicates as discussed in Section 4.4.2. To cope with this combinational complexity in real-world scenarios, approximate inference methods should be explored and evaluated in their suitability to still observe the benefits of FSA without exact computation over possible domain models.

Thirdly, while we conducted experiments on real robots (see Chapter 6), the ITL agent developed in this thesis works in relatively artificial settings: the range of teacher’s utterances that we allowed was quite limited, there was no occlusion of objects on the table, and the lighting conditions were kept constant. The adoption of the insights offered by this thesis in the real-world system would require an extension of generalist robotic agents that use contemporary models for language, vision, and action to make them capable of learning from the learner-teacher style interaction with embodied conversation. This thesis does not deal with the full gamut of phenomena that arise in such conversations but lays the foundations for developing a robot-human interaction interface for the next generation of robots.

Finally, this thesis only considered embodied conversation in which strategic thinking about what to say is only given to the learner. In reality, teachers consider carefully how the learner will react to what they say (Csibra and Gergely, 2009), and while there is a lot of research on this within computational paradigms, e.g. using the optimal transport (Shafto et al., 2021), we have not addressed teacher strategies here. Nevertheless, the notion of unforeseen possibilities would have an effect on teaching strategy. Upon becoming aware of the learner’s knowledge gap, the teacher would need to adapt their behaviour to include signals for teaching information they may have assumed the learner has already mastered (Shafto et al., 2014). As alluded to in Section 2.2, the teacher following a pre-defined and hand-crafted dialogue policy is not realistic, and it is a matter of future work to integrate the models we have developed here to include proactive teachers with bi-directional embodied conversation initiatives.

Appendix A

Prompts

Prompts are used with CodeLlama 7B (Rozière et al., 2023) to perform NLP routines using in-context learning (Brown et al., 2020). For generating referential expressions, the following prompt is used to order words to follow the conventional order of nouns and adjectives for English descriptions:

[INST] Given a list of words for an English description, return them in the correct order as they would appear in an English description. The correct order should follow the typical adjective order: quantity, quality, size, age, shape, colour, proper adjective (often nationality, another place of origin, or material), purpose or qualifier. You return a single string and nothing else. For example:

(red, sphere, dotted) = “dotted red sphere”

(blue, cone, plain) = “plain blue cone”

(cone, red) = “red cone”

(green, sphere, small) = “small green sphere”

(stary, purple, block) = “stary purple block”

(plain, green, cube) = “plain green cube”

(plain, cube, green) = “plain green cube”

(cube, plain, green) = “plain green cube”

(red, sphere, big) = “big red sphere”

(blue, rectangle, dotted) = “dotted blue rectangle”

(green, sphere, small) = “small green sphere”

Now do yourself: (words) = [/INST]

where **words** are words in the prompt used to generate a referential expression.

To obtain logical forms of referential expressions from their surface forms, the following prompt is used:

[INST] You are a helper in a virtual environment. You are asked to translate referential expressions to their logical forms. Referential expressions like “the one red block” have a logical form $\langle _the_1_q\ x.red(x) \wedge block(x) \rangle$. following the pattern $\langle Q\ x.\phi \rangle$. Q is a quantifier. We consider the following quantifiers:

- *existential with surface form: “a/an” and denoted as symbol: $_a_q$*
- *universal with surface form: “every/all” and denoted as a symbol: $_every_q$*
- *uniqueness with surface form: “the n” and denoted as symbol: $_the_n_q$, where n is a natural number like “one”, “two”, etc.*

x is a variable e.g. $x, x1, x12$. Note that x is the only free variable in ϕ and is not bound by a quantifier. If the formula has only one variable, it should be named x . ϕ is the formula of predicate logic. Each formula is constructed recursively:

- *predicates like $red(x)$, $above(x1, x2)$, $left(x1, x3)$ are a well-formed formulas*
- *negation of ϕ like $neg(red(x1))$ is a well-formed formula*
- *conjunction of ϕ s like $red(x) \wedge block(x)$ is a well-formed formula*
- *disjunction of ϕ s like $red(x) \vee block(x)$ is a well-formed formula*
- *implication of ϕ s like $red(x) \rightarrow block(x)$ is a well-formed formula*
- *bicondition of ϕ s like $red(x) \leftrightarrow block(x)$ is a well-formed formula*
- *structure $Q\ x.(\phi, \psi)$ like $_the_1_q\ x.(red(x), block(x))$ is a well-formed formula*

Here are some examples of referential expressions and their logical forms:

“a block.” = $\langle _a_q\ x.block(x) \rangle$

“the one block.” = $\langle _the_1_q\ x.block(x) \rangle$

“the two plain objects.” = $\langle _the_2_q\ x.plain(x) \wedge object(x) \rangle$

“every magenta sphere.” = $\langle _every_q\ x.magenta(x) \wedge sphere(x) \rangle$

“not a block above a sphere.” = $\langle _a_q\ x. neg(_a_q\ x1.(sphere(x1), block(x) \wedge above(x, x1))) \rangle$

“a sphere to the left of every green cone.” = $\langle _a_q\ x. _every_q\ x1.(green(x1) \wedge cone(x1), sphere(x) \wedge left(x, x1)) \rangle$

“every sphere to the left of every green object.” = $\langle _every_q\ x. _every_q\ x1.(green(x1) \wedge$

$\text{object}(x1), \text{sphere}(x) \wedge \text{left}(x, x1))\rangle$

“a sphere to the right of the two green cones.” = $\langle _a_q\ x.\ _the_2_q\ x1.(\text{green}(x1) \wedge \text{cone}(x1), \text{sphere}(x) \wedge \text{right}(x, x1))\rangle$

“the one sphere in front of every green cone.” = $\langle _the_1_q\ x.\ _every_q\ x1.(\text{green}(x1) \wedge \text{cone}(x1), \text{sphere}(x) \wedge \text{front}(x, x1))\rangle$

Now, please translate the following referential expression to its logical form. Just give the logical form—no extra information. Referential expression: **refexp** [/INST]

where **refexp** is a referential expression in a prompt that is parsed to its logical form.

Bibliography

- Ahn, M., Brohan, A., Brown, N., Chebotar, Y., Cortes, O., David, B., Finn, C., Fu, C., Gopalakrishnan, K., Hausman, K., Herzog, A., Ho, D., Hsu, J., Ibarz, J., Ichter, B., Irpan, A., Jang, E., Ruano, R. J., Jeffrey, K., Jesmonth, S., Joshi, N., Julian, R., Kalashnikov, D., Kuang, Y., Lee, K.-H., Levine, S., Lu, Y., Luu, L., Parada, C., Pastor, P., Quiambao, J., Rao, K., Rettinghouse, J., Reyes, D., Sermanet, P., Sievers, N., Tan, C., Toshev, A., Vanhoucke, V., Xia, F., Xiao, T., Xu, P., Xu, S., Yan, M., and Zeng, A. (2022). Do as i can and not as i say: Grounding language in robotic affordances. In *arXiv preprint arXiv:2204.01691*.
- Alomari, M., Duckworth, P., Hawasly, M., Hogg, D. C., and Cohn, A. G. (2017a). Natural language grounding and grammar induction for robotic manipulation commands. In Bansal, M., Matuszek, C., Andreas, J., Artzi, Y., and Bisk, Y., editors, *Proceedings of the First Workshop on Language Grounding for Robotics*, pages 35–43, Vancouver, Canada. Association for Computational Linguistics.
- Alomari, M., Duckworth, P., Hogg, D. C., and Cohn, A. G. (2017b). Natural language acquisition and grounding for embodied robotic systems. In *Thirty-First AAAI Conference on Artificial Intelligence*, pages 4349–4356. AAAI Press.
- Alomari, M., Li, F., Hogg, D. C., and Cohn, A. G. (2022). Online perceptual learning and natural language acquisition for autonomous robots. *Artificial Intelligence*, 303:103637.
- Andreas, J. and Klein, D. (2016). Reasoning about pragmatics with neural listeners and speakers. In Su, J., Duh, K., and Carreras, X., editors, *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1173–1182, Austin, Texas. Association for Computational Linguistics.
- Antol, S., Agrawal, A., Lu, J., Mitchell, M., Batra, D., Zitnick, C. L., and Parikh, D.

- (2015). VQA: Visual Question Answering. In *International Conference on Computer Vision (ICCV)*.
- Appelgren, M. and Lascarides, A. (2020). Interactive task learning via embodied corrective feedback. *Journal of Autonomous Agents and Multi-Agent Systems*, 34(54).
- Appelgren, M. and Lascarides, A. (2021). Symbol grounding and task learning from imperfect corrections. In Alikhani, M., Blukis, V., Kordjamshidi, P., Padmakumar, A., and Tan, H., editors, *Proceedings of Second International Combined Workshop on Spatial Language Understanding and Grounded Communication for Robotics*, pages 1–10, Online. Association for Computational Linguistics.
- Asher, N. and Lascarides, A. (2003). *Logics of Conversation*. Cambridge University Press.
- Asher, N. and Lascarides, A. (2013). Strategic conversation. *Semantics and Pragmatics*, 6(2):2:1–:62.
- Asher, N. and Morreau, M. (1991). Commonsense entailment: A modal theory of non-monotonic reasoning. In Mylopoulos, J. and Reiter, R., editors, *Proceedings of the 12th International Joint Conference on Artificial Intelligence. Sydney, Australia, August 24-30, 1991*, pages 387–392. Morgan Kaufmann.
- Baral, C., Gelfond, M., and Rushton, N. (2009). Probabilistic reasoning with answer sets. *Theory and Practice of Logic Programming*, 9(1):57–144.
- Bartneck, C., Belpaeme, T., Eyssel, F., Kanda, T., Keijsers, M., and Sabanovic, S. (2020). *Human-Robot Interaction: An Introduction*. Cambridge University Press.
- Barwise, J. and Cooper, R. (1981). Generalized quantifiers and natural language. *Linguistics and Philosophy*, 4:159–219.
- Batra, D., Chang, A. X., Chernova, S., Davison, A. J., Deng, J., Koltun, V., Levine, S., Malik, J., Mordatch, I., Mottaghi, R., Savva, M., and Su, H. (2020). Rearrangement: A challenge for embodied AI. *CoRR*, abs/2011.01975.
- Beame, P., den Broeck, G. V., Gribkoff, E., and Suciu, D. (2015). Symmetric weighted first-order model counting. In Milo, T. and Calvanese, D., editors, *Proceedings of the 34th ACM Symposium on Principles of Database Systems, PODS 2015, Melbourne, Victoria, Australia, May 31 - June 4, 2015*, pages 313–328. ACM.

- Bengio, Y., Courville, A. C., and Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(8):1798–1828.
- Berger-Tal, O., Nathan, J., Meron, E., and Saltz, D. (2014). The exploration-exploitation dilemma: a multidisciplinary framework. *PloS one*, 9(4):e95693.
- Blackburn, P. and Bos, J. (2005). *Representation and Inference for Natural Language - a First Course in Computational Semantics*. CSLI Studies in Computational Linguistics. CSLI Publications.
- Blackburn, P., van Benthem, J., and Wolter, F. (2007). Handbook of modal logic. In *Studies in logic and practical reasoning*.
- Bochman, A. (2007). Nonmonotonic reasoning. In Gabbay, D. M. and Woods, J., editors, *The Many Valued and Nonmonotonic Turn in Logic*, volume 8 of *Handbook of the History of Logic*, pages 557–632. North-Holland.
- Brohan, A., Brown, N., Carbajal, J., Chebotar, Y., Chen, X., Choromanski, K., Ding, T., Driess, D., Dubey, A., Finn, C., Florence, P., Fu, C., Arenas, M. G., Gopalakrishnan, K., Han, K., Hausman, K., Herzog, A., Hsu, J., Ichter, B., Irpan, A., Joshi, N., Julian, R., Kalashnikov, D., Kuang, Y., Leal, I., Lee, L., Lee, T.-W. E., Levine, S., Lu, Y., Michalewski, H., Mordatch, I., Pertsch, K., Rao, K., Reymann, K., Ryoo, M., Salazar, G., Sanketi, P., Sermanet, P., Singh, J., Singh, A., Soricut, R., Tran, H., Vanhoucke, V., Vuong, Q., Wahid, A., Welker, S., Wohlhart, P., Wu, J., Xia, F., Xiao, T., Xu, P., Xu, S., Yu, T., and Zitkovich, B. (2023). Rt-2: Vision-language-action models transfer web knowledge to robotic control. In *arXiv preprint arXiv:2307.15818*.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., and Amodei, D. (2020). Language models are few-shot learners. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H., editors, *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*.
- Bundy, A. and Wallen, L. (1984). *Skolemization*, pages 123–123. Springer Berlin Heidelberg, Berlin, Heidelberg.

- Caelen, J. and Xuereb, A. (2011). Dialogue and game theory. In *2011 6th Conference on Speech Technology and Human-Computer Dialogue (SpeD)*, pages 1–10.
- Cano Santín, J. M., Dobnik, S., and Ghanimifard, M. (2020). Fast visual grounding in interaction: bringing few-shot learning with neural networks to an interactive robot. In Howes, C., Chatzikyriakidis, S., Ek, A., and Somashekarappa, V., editors, *Proceedings of the Probability and Meaning Conference (PaM 2020)*, pages 53–61, Gothenburg. Association for Computational Linguistics.
- Caruana, R. (1997). Multitask learning. *Mach. Learn.*, 28(1):41–75.
- Cassell, J. (2001). Embodied conversational agents: Representation and intelligence in user interfaces. *AI Mag.*, 22(4):67–84.
- Chai, J. Y., Gao, Q., She, L., Yang, S., Saba-Sadiya, S., and Xu, G. (2018). Language to action: Towards interactive task learning with physical agents. In Lang, J., editor, *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden*, pages 2–9. ijcai.org.
- Chandu, K. R., Bisk, Y., and Black, A. W. (2021). Grounding ‘grounding’ in NLP. In Zong, C., Xia, F., Li, W., and Navigli, R., editors, *Findings of the Association for Computational Linguistics: ACL/IJCNLP 2021, Online Event, August 1-6, 2021*, volume ACL/IJCNLP 2021 of *Findings of ACL*, pages 4283–4305. Association for Computational Linguistics.
- Channell, J. (1994). *Vague Language*. Oxford University Press.
- Chavira, M. and Darwiche, A. (2008). On probabilistic inference by weighted model counting. *Artif. Intell.*, 172(6-7):772–799.
- Chen, Z. and Liu, B. (2018). *Lifelong Machine Learning, Second Edition*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool Publishers.
- Chitta, S. (2016). Moveit!: an introduction. *Robot Operating System (ROS) The Complete Reference (Volume 1)*, pages 3–27.
- Choi, Y., Vergari, A., and Van den Broeck, G. (2020). Probabilistic circuits: A unifying framework for tractable probabilistic models. (*Unpublished manuscript*).
- Clark, H. H. (1996). *Using Language*. Cambridge University Press.

- Cooper, R. (2023). *From perception to communication: a theory of types for action and meaning*. Oxford University Press.
- Cooper, R., Dobnik, S., Lappin, S., and Larsson, S. (2015). Probabilistic type theory and natural language semantics. *Linguistic Issues in Language Technology*, 10.
- Copetake, A., Emerson, G., Goodman, M. W., Horvat, M., Kuhnle, A., and Muszyńska, E. (2016). Resources for building applications with dependency Minimal Recursion Semantics. In Calzolari, N., Choukri, K., Declerck, T., Goggi, S., Grobelnik, M., Maegaard, B., Mariani, J., Mazo, H., Moreno, A., Odijk, J., and Piperidis, S., editors, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 1240–1247, Portorož, Slovenia. European Language Resources Association (ELRA).
- Copetake, A. and Flickinger, D. (2000). An open source grammar development environment and broad-coverage English grammar using HPSG. In Gavriliidou, M., Carayannis, G., Markantonatou, S., Piperidis, S., and Stainhauer, G., editors, *Proceedings of the Second International Conference on Language Resources and Evaluation (LREC'00)*, Athens, Greece. European Language Resources Association (ELRA).
- Cover, T. M. and Thomas, J. A. (2005). *Elements of Information Theory*. Wiley-Interscience.
- Csibra, G. and Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, 13:148–153.
- Dale, R. and Reiter, E. (1995). Computational interpretations of the gricean maxims in the generation of referring expressions. *Cognitive science*, 19(2):233–263.
- Das, A., Kottur, S., Gupta, K., Singh, A., Yadav, D., Moura, J. M., Parikh, D., and Batra, D. (2017). Visual Dialog. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Datta, S., Sikka, K., Roy, A., Ahuja, K., Parikh, D., and Divakaran, A. (2019). Align2ground: Weakly supervised phrase grounding guided by image-caption alignment. In *2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019*, pages 2601–2610. IEEE.

- den Broeck, G. V. (2011). On the completeness of first-order knowledge compilation for lifted probabilistic inference. In Shawe-Taylor, J., Zemel, R. S., Bartlett, P. L., Pereira, F. C. N., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems 2011. Proceedings of a meeting held 12-14 December 2011, Granada, Spain*, pages 1386–1394.
- DeVault, D., Kariaeva, N., Kothari, A., Oved, I., and Stone, M. (2005). An information-state approach to collaborative reference. In Nagata, M. and Pedersen, T., editors, *Proceedings of the ACL Interactive Poster and Demonstration Sessions*, pages 1–4, Ann Arbor, Michigan. Association for Computational Linguistics.
- Dobnik, S., Cooper, R., Ek, A., Noble, B., Larsson, S., Ilinykh, N., Maraev, V., and Somashekarappa, V. (2022). In search of meaning and its representations for computational linguistics. In Dobnik, S., Grove, J., and Sayeed, A., editors, *Proceedings of the 2022 CLASP Conference on (Dis)embodiment*, pages 30–44, Gothenburg, Sweden. Association for Computational Linguistics.
- Du, Y., Fu, Z., Liu, Q., and Wang, Y. (2021). Visual grounding with transformers. *2022 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6.
- Dubray, A., Schaus, P., and Nijssen, S. (2024). Anytime weighted model counting with approximation guarantees for probabilistic inference. In Shaw, P., editor, *30th International Conference on Principles and Practice of Constraint Programming, CP 2024, September 2-6, 2024, Girona, Spain*, volume 307 of *LIPICs*, pages 10:1–10:16. Schloss Dagstuhl - Leibniz-Zentrum für Informatik.
- Frank, M. C. and Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, 336(6084):998–998.
- Fried, D., Chiu, J., and Klein, D. (2021). Reference-centric models for grounded collaborative dialogue. In Moens, M.-F., Huang, X., Specia, L., and Yih, S. W.-t., editors, *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 2130–2147, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Friedman, N., Getoor, L., Koller, D., and Pfeffer, A. (1999). Learning probabilistic relational models. In Dean, T., editor, *Proceedings of the Sixteenth International*

- Joint Conference on Artificial Intelligence, IJCAI 99, Stockholm, Sweden, July 31 - August 6, 1999. 2 Volumes, 1450 pages, pages 1300–1309.* Morgan Kaufmann.
- Fudenberg, D. and Levine, D. K. (1998). *The Theory of Learning in Games*. MIT Press.
- Fürnkranz, J. and Hüllermeier, E. (2005). Preference learning. *Künstliche Intell.*, 19(1):60.
- Gabbay, D. M. (1985). Theoretical foundations for non-monotonic reasoning in expert systems. In Apt, K. R., editor, *Logics and Models of Concurrent Systems*, pages 439–457, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Gamut, L. T. F. (1991a). *Logic, language, and meaning. Vol.1, Introduction to logic ; L.T.F. Gamut*. University of Chicago Press.
- Gamut, L. T. F. (1991b). *Logic, language, and meaning. Vol.2, Intensional logic and logical grammar; L.T.F. Gamut*. University of Chicago Press.
- Garnelo, M. and Shanahan, M. (2019). Reconciling deep learning with symbolic artificial intelligence: representing objects and relations. *Current Opinion in Behavioral Sciences*, 29:17–23. Artificial Intelligence.
- Garrett, C. R., Chitnis, R., Holladay, R., Kim, B., Silver, T., Kaelbling, L. P., and Lozano-Perez, T. (2021). Integrated task and motion planning. *Annual review of control, robotics, and autonomous systems*, 4.
- Getoor, L. and Taskar, B. (2007). *Introduction to Statistical Relational Learning*. MIT Press, 1 edition.
- Giulianelli, M. and Fernández, R. (2021). Analysing human strategies of information transmission as a function of discourse context. In Bisazza, A. and Abend, O., editors, *Proceedings of the 25th Conference on Computational Natural Language Learning*, pages 647–660, Online. Association for Computational Linguistics.
- Giulianelli, M., Sinclair, A., and Fernández, R. (2021). Is information density uniform in task-oriented dialogues? In Moens, M.-F., Huang, X., Specia, L., and Yih, S. W.-t., editors, *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 8271–8283, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

- Goodman, N. D. and Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, 20(11):818–829.
- Goodman, N. D. and Stuhlmüller, A. (2013). Knowledge and implicature: Modeling language understanding as social cognition. *Top. Cogn. Sci.*, 5(1):173–184.
- Grice, H. P. (1975). Logic and conversation. In Cole, P. and Morgan, J. L., editors, *Syntax and Semantics Volume 3: Speech Acts*, pages 41–58. Academic Press.
- Gu, W., Sah, A., and Gopalan, N. (2024). Interactive visual task learning for robots. In Wooldridge, M. J., Dy, J. G., and Natarajan, S., editors, *Thirty-Eighth AAAI Conference on Artificial Intelligence, AAAI 2024, Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence, IAAI 2024, Fourteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2024, February 20-27, 2024, Vancouver, Canada*, pages 10297–10305. AAAI Press.
- Haber, J., Baumgärtner, T., Takmaz, E., Gelderloos, L., Bruni, E., and Fernández, R. (2019). The PhotoBook dataset: Building common ground through visually-grounded dialogue. In Korhonen, A., Traum, D., and Màrquez, L., editors, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1895–1910, Florence, Italy. Association for Computational Linguistics.
- Hansson, S. O. (2022). Logic of Belief Revision. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Spring 2022 edition.
- Hao, X., Jhaveri, Y., and Shafto, P. (2023). Common ground in cooperative communication. In Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., and Levine, S., editors, *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*.
- Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, 42(1):335–346.
- Haslum, P., Lipovetzky, N., Magazzeni, D., and Muise, C. (2019). *An Introduction to the Planning Domain Definition Language*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool Publishers.
- Hawkins, R., Kwon, M., Sadigh, D., and Goodman, N. (2020). Continual adaptation

- for efficient machine communication. In Fernández, R. and Linzen, T., editors, *Proceedings of the 24th Conference on Computational Natural Language Learning*, pages 408–419, Online. Association for Computational Linguistics.
- Holyoak, K. J. and Morrison, B. (2005). *The Cambridge handbook of thinking and reasoning*. Cambridge University Press.
- Hristov, Y. (2021). *Interactive robot learning with human alignment*. PhD thesis, The University of Edinburgh, UK.
- Hu, R., Rohrbach, M., Andreas, J., Darrell, T., and Saenko, K. (2016a). Modeling relationships in referential expressions with compositional modular networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4418–4427.
- Hu, R., Xu, H., Rohrbach, M., Feng, J., Saenko, K., and Darrell, T. (2016b). Natural language object retrieval. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- Hunter, J., Asher, N., and Lascarides, A. (2018). A formal semantics for situated conversation. *Semantics and Pragmatics*.
- Hupkes, D., Dankers, V., Mul, M., and Bruni, E. (2020). Compositionality decomposed: how do neural networks generalise?
- Innes, C. and Ramamoorthy, S. (2020). Elaborating on learned demonstrations with temporal logic specifications. In Toussaint, M., Bicchi, A., and Hermans, T., editors, *Robotics: Science and Systems XVI, Virtual Event / Corvallis, Oregon, USA, July 12-16, 2020*.
- Jara-Ettinger, J., Gweon, H., Schulz, L. E., and Tenenbaum, J. B. (2016). The naïve utility calculus: Computational principles underlying commonsense psychology. *Trends in Cognitive Sciences*, 20:589–604.
- Kamp, H. and Reyle, U. (1993). *From Discourse to Logic: Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Kluwer Academic Publishers, Dordrecht.
- Karbstein, K., Kösters, L., Hodač, L., Hofmann, M., Hörandl, E., Tomasello, S., Wagner, N. D., Emerson, B. C., Albach, D. C., Scheu, S., Bradler, S., de Vries, J., Irisarri, I., Li, H., Soltis, P., Mäder, P., and Wäldchen, J. (2024). Species delimitation 4.0:

- integrative taxonomy meets artificial intelligence. *Trends in Ecology Evolution*, 39(8):771–784.
- Khatib, O. (1987). A unified approach for motion and force control of robot manipulators: The operational space formulation. *IEEE J. Robotics Autom.*, 3(1):43–53.
- Kisa, D., den Broeck, G. V., Choi, A., and Darwiche, A. (2014). Probabilistic sentential decision diagrams. In Baral, C., Giacomo, G. D., and Eiter, T., editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Fourteenth International Conference, KR 2014, Vienna, Austria, July 20-24, 2014*. AAAI Press.
- Koller, D. and Friedman, N. (2009). *Probabilistic Graphical Models - Principles and Techniques*. MIT Press.
- Kottur, S., Moura, J. M. F., Parikh, D., Batra, D., and Rohrbach, M. (2018). Visual coreference resolution in visual dialog using neural module networks. In *The European Conference on Computer Vision (ECCV)*.
- Kottur, S., Moura, J. M. F., Parikh, D., Batra, D., and Rohrbach, M. (2019). CLEVR-dialog: A diagnostic dataset for multi-round reasoning in visual dialog. In Burstein, J., Doran, C., and Solorio, T., editors, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 582–595, Minneapolis, Minnesota. Association for Computational Linguistics.
- Krishna, R., Zhu, Y., Groth, O., Johnson, J., Hata, K., Kravitz, J., Chen, S., Kalantidis, Y., Li, L., Shamma, D. A., Bernstein, M. S., and Fei-Fei, L. (2016). Visual genome: Connecting language and vision using crowdsourced dense image annotations. *CoRR*, abs/1602.07332.
- Kuhnle, A. and Copestake, A. (2018). Deep learning evaluation using deep linguistic processing. In Bisk, Y., Levy, O., and Yatskar, M., editors, *Proceedings of the Workshop on Generalization in the Age of Deep Learning*, pages 17–23, New Orleans, Louisiana. Association for Computational Linguistics.
- Kuhnle, A. and Copestake, A. A. (2017). Shapeworld - A new test methodology for multimodal language understanding. *CoRR*, abs/1704.04517.
- Kuzelka, O. (2021). Weighted first-order model counting in the two-variable fragment with counting quantifiers. *J. Artif. Intell. Res.*, 70:1281–1307.

- Laird, J. E., Gluck, K. A., Anderson, J. R., Forbus, K. D., Jenkins, O. C., Lebiere, C., Salvucci, D. D., Scheutz, M., Thomaz, A., Trafton, J. G., Wray, R. E., Mohan, S., and Kirk, J. R. (2017). Interactive task learning. *IEEE Intelligent Systems*, 32(4):6–21.
- Larsson, S. (2013). Formal semantics for perceptual classification. *Journal of Logic and Computation*, 25(2):335–369.
- Larsson, S. (2021). The role of definitions in coordinating on perceptual meanings. In *Proceedings of the 25th Workshop on the Semantics and Pragmatics of Dialogue - Full Papers*, Potsdam, Germany. SEMDIAL.
- Larsson, S., Bernardy, J.-P., and Cooper, R. (2021). Semantic learning in a probabilistic type theory with records. In Chatzikyriakidis, S. and Osswald, R., editors, *Proceedings of the ESSLLI 2021 Workshop on Computing Semantics with Types, Frames and Related Structures*, pages 35–44, Utrecht, The Netherlands (online). Association for Computational Linguistics.
- Lascarides, A. and Asher, N. (2009). Agreement, disputes and commitments in dialogue. *Journal of Semantics*, 26(2):109–158.
- LaValle, S. M. (2006). *Planning Algorithms*. Cambridge University Press.
- Levine, S., Kumar, A., Tucker, G., and Fu, J. (2020). Offline reinforcement learning: Tutorial, review, and perspectives on open problems. *CoRR*, abs/2005.01643.
- Levit, M. and Roy, D. (2007). Interpretation of spatial language in a map navigation task. *IEEE Trans. Syst. Man Cybern. Part B*, 37(3):667–679.
- Lifschitz, V. (2008). What is answer set programming? In Fox, D. and Gomes, C. P., editors, *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence, AAAI 2008, Chicago, Illinois, USA, July 13-17, 2008*, pages 1594–1597. AAAI Press.
- Liu, C., Chai, J. Y., Shukla, N., and Zhu, S. (2016). Task learning through visual demonstration and situated dialogue. In Kephart, J. O., Rosenthal, S., Veloso, M. M., and Rudnický, A., editors, *Symbiotic Cognitive Systems, Papers from the 2016 AAAI Workshop, Phoenix, Arizona, USA, February 13, 2016*, volume WS-16-14 of AAAI Technical Report. AAAI Press.
- Liu, S., Zeng, Z., Ren, T., Li, F., Zhang, H., Yang, J., Li, C., Yang, J., Su, H., Zhu,

- J., and Zhang, L. (2023). Grounding DINO: marrying DINO with grounded pre-training for open-set object detection. *CoRR*, abs/2303.05499.
- Loáiciga, S., Dobnik, S., and Schlangen, D. (2021). Reference and coreference in situated dialogue. In Xin, Hu, R., Hudson, D., Fu, T.-J., Rohrbach, M., and Fried, D., editors, *Proceedings of the Second Workshop on Advances in Language and Vision Research*, pages 39–44, Online. Association for Computational Linguistics.
- MacKay, D. J. C. (2003). *Information theory, inference, and learning algorithms*. Cambridge University Press.
- Malhotra, S. and Serafini, L. (2022). Weighted model counting in FO2 with cardinality constraints and counting quantifiers: A closed form formula. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022*, pages 5817–5824. AAAI Press.
- Manhaeve, R., Dumancic, S., Kimmig, A., Demeester, T., and Raedt, L. D. (2018). Deepprolog: Neural probabilistic logic programming. In Bengio, S., Wallach, H. M., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R., editors, *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*, pages 3753–3763.
- Manhaeve, R., Marra, G., Demeester, T., Dumancic, S., Kimmig, A., and Raedt, L. D. (2021). Neuro-symbolic AI = neural + logical + probabilistic AI. In Hitzler, P. and Sarker, M. K., editors, *Neuro-Symbolic Artificial Intelligence: The State of the Art*, volume 342 of *Frontiers in Artificial Intelligence and Applications*, pages 173–191. IOS Press.
- Mao, J., Gan, C., Kohli, P., Tenenbaum, J. B., and Wu, J. (2019). The neuro-symbolic concept learner: Interpreting scenes, words, and sentences from natural supervision. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net.
- Mao, J., Huang, J., Toshev, A., Camburu, O., Yuille, A. L., and Murphy, K. (2016). Generation and comprehension of unambiguous object descriptions. In *2016 IEEE*

- Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 11–20. IEEE Computer Society.
- Mao, J., Shi, H., Wu, J., Levy, R. P., and Tenenbaum, J. B. (2021). Grammar-Based Grounded Lexicon Learning. In *Advances in Neural Information Processing Systems (NeurIPS)*.
- Marra, G., Dumancic, S., Manhaeve, R., and Raedt, L. D. (2021). From statistical relational to neural symbolic artificial intelligence: a survey. *CoRR*, abs/2108.11451.
- Matuszek, C. (2018). Grounded language learning: Where robotics and NLP meet. In Lang, J., editor, *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden*, pages 5687–5691. ijcai.org.
- McCloskey, M. and Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In Bower, G. H., editor, *Psychology of Learning and Motivation*, volume 24, pages 109–165. Academic Press.
- Mendez, J. A. and Eaton, E. (2022). How to reuse and compose knowledge for a lifetime of tasks: A survey on continual learning and functional composition. *CoRR*, abs/2207.07730.
- Metropolis, N. and Ulam, S. (1949). The monte carlo method. *Journal of the American Statistical Association*, 44(247):335–341.
- Mohri, M., Rostamizadeh, A., and Talwalkar, A. (2012). Foundations of machine learning. In *Adaptive computation and machine learning*.
- Monroe, W., Hawkins, R. X. D., Goodman, N. D., and Potts, C. (2017). Colors in context: A pragmatic neural model for grounded language understanding. *Trans. Assoc. Comput. Linguistics*, 5:325–338.
- Mooney, R. J. (2007). Learning for semantic parsing. In Gelbukh, A. F., editor, *Computational Linguistics and Intelligent Text Processing, 8th International Conference, CICLing 2007, Mexico City, Mexico, February 18-24, 2007*, volume 4394 of *Lecture Notes in Computer Science*, pages 311–324. Springer.
- Mykel J. Kochenderfer, T. A. W. and Wray, K. H. (2022). *Algorithms for Decision Making*. The MIT Press, Boston Massachusetts.
- Nicolas, D. (2024). The Logic of Mass Expressions. In Zalta, E. N. and Nodelman,

- U., editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Spring 2024 edition.
- Noble, B. and Ilinykh, N. (2023). Describe me an auklet: Generating grounded perceptual category descriptions. In Bouamor, H., Pino, J., and Bali, K., editors, *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 9330–9347, Singapore. Association for Computational Linguistics.
- Noble, B., Larsson, S., and Cooper, R. (2022). Classification systems: Combining taxonomical and perceptual lexical meaning. In Kalouli, A.-L. and Chatzikyriakidis, S., editors, *Proceedings of the 3rd Natural Logic Meets Machine Learning Workshop (NALOMA III)*, pages 11–16, Galway, Ireland. Association for Computational Linguistics.
- Octo Model Team, Ghosh, D., Walke, H., Pertsch, K., Black, K., Mees, O., Dasari, S., Hejna, J., Xu, C., Luo, J., Kreiman, T., Tan, Y., Chen, L. Y., Sanketi, P., Vuong, Q., Xiao, T., Sadigh, D., Finn, C., and Levine, S. (2024). Octo: An open-source generalist robot policy. In *Proceedings of Robotics: Science and Systems*, Delft, Netherlands.
- Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., Assran, M., Ballas, N., Galuba, W., Howes, R., Huang, P., Li, S., Misra, I., Rabbat, M. G., Sharma, V., Synnaeve, G., Xu, H., Jégou, H., Mairal, J., Labatut, P., Joulin, A., and Bojanowski, P. (2023). DINOv2: Learning robust visual features without supervision. *CoRR*, abs/2304.07193.
- Pan, S. J. and Yang, Q. (2010). A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.*, 22(10):1345–1359.
- Park, Y., Casey, D., Joshi, I., Zhu, J., and Cheng, F. (2020). Emergence of new disease: How can artificial intelligence help? *Trends in Molecular Medicine*, 26(7):627–629.
- Post, E. L. (1921). Introduction to a general theory of elementary propositions. *American Journal of Mathematics*, 43:163.
- Raedt, L. D., Kimmig, A., and Toivonen, H. (2007). Problog: A probabilistic prolog and its application in link discovery. In Veloso, M. M., editor, *IJCAI 2007, Proceedings of the 20th International Joint Conference on Artificial Intelligence, Hyderabad, India, January 6-12, 2007*, pages 2462–2467.

- Rafferty, A. N., Brunskill, E., Griffiths, T. L., and Shafto, P. (2016). Faster teaching via POMDP planning. *Cogn. Sci.*, 40(6):1290–1332.
- Reed, S. E., Zolna, K., Parisotto, E., Colmenarejo, S. G., Novikov, A., Barth-Maron, G., Gimenez, M., Sulsky, Y., Kay, J., Springenberg, J. T., Eccles, T., Bruce, J., Razavi, A., Edwards, A., Heess, N., Chen, Y., Hadsell, R., Vinyals, O., Bordbar, M., and de Freitas, N. (2022). A generalist agent. *Trans. Mach. Learn. Res.*, 2022.
- Reiter, R. (1980). A logic for default reasoning. *Artif. Intell.*, 13(1-2):81–132.
- Richardson, M. and Domingos, P. M. (2006). Markov logic networks. *Mach. Learn.*, 62(1-2):107–136.
- Roth, D. (2017). Incidental supervision: Moving beyond supervised learning. In Singh, S. and Markovitch, S., editors, *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA*, pages 4885–4890. AAAI Press.
- Rozière, B., Gehring, J., Gloeckle, F., Sootla, S., Gat, I., Tan, X. E., Adi, Y., Liu, J., Remez, T., Rapin, J., Kozhevnikov, A., Evtimov, I., Bitton, J., Bhatt, M., Canton-Ferrer, C., Grattafiori, A., Xiong, W., Défossez, A., Copet, J., Azhar, F., Touvron, H., Martin, L., Usunier, N., Scialom, T., and Synnaeve, G. (2023). Code llama: Open foundation models for code. *CoRR*, abs/2308.12950.
- Rubavicius, R., Fagan, P. D., Lascarides, A., and Ramamoorthy, S. (2025). SE-CURE: Semantics-aware embodied conversation under unawareness for lifelong robot learning. In Eaton, E., Liu, B., Mahmood, R., Rannen-Triki, A., Chandar, S., and Pascanu, R., editors, *Proceedings of The 4th Conference on Lifelong Learning Agents*, Proceedings of Machine Learning Research (PMLR).
- Rubavicius, R. and Lascarides, A. (2022). Interactive symbol grounding with complex referential expressions. In Carpuat, M., de Marneffe, M.-C., and Meza Ruiz, I. V., editors, *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4863–4874, Seattle, United States. Association for Computational Linguistics.
- Rubavicius, R., Lascarides, A., and Ramamoorthy, S. (in press). Learning neuro-symbolic dialogue strategies for interactive symbol grounding. *Linguistic Issues in Language Technology*.

- Rummery, G. A. and Niranjan, M. (1994). On-line Q-learning using connectionist systems. Technical Report TR 166, Cambridge University Engineering Department, Cambridge, England.
- Russell, B. (1917). Knowledge by acquaintance and knowledge by description. In *Mysticism and Logic*, pages 152–167. London: Longmans Green.
- Russell, S. and Norvig, P. (2020). *Artificial Intelligence: A Modern Approach (4th Edition)*. Pearson.
- Saltelli, A., Chan, K., and Scott, E. (2000). *Sensitivity Analysis: Gauging the Worth of Scientific Models*. John Wiley and Sons.
- Saravanan, R. (2021). *The Rumsfeld Matrix: Degrees of Knowledge*, page 193–211. Cambridge University Press.
- Sarker, M. K., Zhou, L., Eberhart, A., and Hitzler, P. (2021). Neuro-symbolic artificial intelligence. *AI Commun.*, 34(3):197–209.
- Settles, B. (2012). *Active Learning*. Morgan & Claypool Publishers.
- Shafto, P., Goodman, N. D., and Griffiths, T. L. (2014). A rational account of pedagogical reasoning: Teaching by, and learning from, examples. *Cognitive Psychology*, 71:55–89.
- Shafto, P., Wang, J., and Wang, P. (2021). Cooperative communication as belief transport. *Trends in Cognitive Sciences*, 25(10):826–828.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.*, 27(4):623–656.
- Silver, D. L., Yang, Q., and Li, L. (2013). Lifelong machine learning systems: Beyond learning algorithms. In *Lifelong Machine Learning, Papers from the 2013 AAAI Spring Symposium, Palo Alto, California, USA, March 25-27, 2013*, volume SS-13-05 of AAAI Technical Report. AAAI.
- Sperber, D. and Wilson, D. (1986). *Relevance: Communication and Cognition*. Wiley-Blackwell.
- Stachniak, Z. (1995). Nonmonotonic theories and their axiomatic varieties. *Journal of Logic, Language and Information*, 4(4):317–334.

- Sutton, R. S. and Barto, A. G. (1998). *Reinforcement learning - an introduction*. Adaptive computation and machine learning. MIT Press.
- Takmaz, E., Brandizzi, N., Giulianelli, M., Pezzelle, S., and Fernandez, R. (2023). Speaking the language of your listener: Audience-aware adaptation via plug-and-play theory of mind. In Rogers, A., Boyd-Graber, J., and Okazaki, N., editors, *Findings of the Association for Computational Linguistics: ACL 2023*, pages 4198–4217, Toronto, Canada. Association for Computational Linguistics.
- Takmaz, E., Giulianelli, M., Pezzelle, S., Sinclair, A., and Fernández, R. (2020). Refer, Reuse, Reduce: Generating Subsequent References in Visual and Conversational Contexts. In Webber, B., Cohn, T., He, Y., and Liu, Y., editors, *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4350–4368, Online. Association for Computational Linguistics.
- Takmaz, E., Pezzelle, S., and Fernández, R. (2022). Less descriptive yet discriminative: Quantifying the properties of multimodal referring utterances via CLIP. In Chersoni, E., Hollenstein, N., Jacobs, C., Oseki, Y., Prévot, L., and Santus, E., editors, *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*, pages 36–42, Dublin, Ireland. Association for Computational Linguistics.
- Tanevska, A., Rea, F., Sandini, G., Cañamero, L., and Sciutti, A. (2020). A socially adaptable framework for human-robot interaction. *Frontiers Robotics AI*, 7:121.
- Tarski, A. (1931). Sur les ensembles définissables de nombres réels. *Fundamenta Mathematicae*, 17:210–239.
- Thompson, W. and Kaufmann, S. (2010). Signaling games with partially observable actions as a model of conversational grounding. In *Interactive Decision Theory and Game Theory, Papers from the 2010 AAI Workshop, Atlanta, Georgia, USA, July 12, 2010*, volume WS-10-03 of *AAAI Technical Report*. AAAI.
- Thrun, S. and Mitchell, T. M. (1995). Lifelong robot learning. *Robotics and Autonomous Systems*, 15(1):25–46. The Biology and Technology of Intelligent Autonomous Agents.
- Todorov, E., Erez, T., and Tassa, Y. (2012). Mujoco: A physics engine for model-based control. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2012, Vilamoura, Algarve, Portugal, October 7-12, 2012*, pages 5026–5033. IEEE.

- Udandarao, V., Prabhu, A., Ghosh, A., Sharma, Y., Torr, P. H. S., Bibi, A., Albanie, S., and Bethge, M. (2024). No "zero-shot" without exponential data: Pretraining concept frequency determines multimodal model performance. *CoRR*, abs/2404.04125.
- Valiant, L. G. (1979). The complexity of computing the permanent. *Theor. Comput. Sci.*, 8:189–201.
- van Bremen, T. and Kuzelka, O. (2020). Approximate weighted first-order model counting: Exploiting fast approximate model counters and symmetry. In Bessiere, C., editor, *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*, pages 4252–4258. ijcai.org.
- van de Ven, G. M., Tuytelaars, T., and Toliás, A. S. (2022). Three types of incremental learning. *Nat. Mac. Intell.*, 4(12):1185–1197.
- van Harmelen, F., Lifschitz, V., and Porter, B. W., editors (2008). *Handbook of Knowledge Representation*, volume 3 of *Foundations of Artificial Intelligence*. Elsevier.
- Vedantam, R., Bengio, S., Murphy, K., Parikh, D., and Chechik, G. (2017). Context-aware captions from context-agnostic supervision. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 1070–1079. IEEE Computer Society.
- Vilalta, R. and Drissi, Y. (2002). A perspective view and survey of meta-learning. *Artif. Intell. Rev.*, 18(2):77–95.
- Vinyals, O., Blundell, C., Lillicrap, T., Kavukcuoglu, K., and Wierstra, D. (2016). Matching networks for one shot learning. In Lee, D. D., Sugiyama, M., von Luxburg, U., Guyon, I., and Garnett, R., editors, *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, pages 3630–3638.
- Wang, P., Wang, J., Paranamana, P., and Shafto, P. (2020a). A mathematical theory of cooperative communication. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H., editors, *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*.
- Wang, R., Mao, J., Hsu, J., Zhao, H., Wu, J., and Gao, Y. (2023). Programmatically grounded, compositionally generalizable robotic manipulation. In *The Eleventh In-*

- ternational Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net.
- Wang, S. I. (2017). *Learning adaptive language interfaces through interaction*. PhD thesis, Stanford University, USA.
- Wang, S. I., Liang, P., and Manning, C. D. (2016). Learning language games through interaction. In Erk, K. and Smith, N. A., editors, *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2368–2378, Berlin, Germany. Association for Computational Linguistics.
- Wang, Y., Yao, Q., Kwok, J. T., and Ni, L. M. (2020b). Generalizing from a few examples: A survey on few-shot learning. *ACM Comput. Surv.*, 53(3):63:1–63:34.
- Wei, X., Song, Y., Aodha, O. M., Wu, J., Peng, Y., Tang, J., Yang, J., and Belongie, S. J. (2022). Fine-grained image analysis with deep learning: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(12):8927–8948.
- White, J., Mu, J., and Goodman, N. D. (2020). Learning to refer informatively by amortizing pragmatic reasoning. In Denison, S., Mack, M., Xu, Y., and Armstrong, B. C., editors, *Proceedings of the 42th Annual Meeting of the Cognitive Science Society - Developing a Mind: Learning in Humans, Animals, and Machines, CogSci 2020, virtual, July 29 - August 1, 2020*. cognitivesciencesociety.org.
- Wray, R. E., Kirk, J. R., and Laird, J. E. (2021). Language models as a knowledge source for cognitive agents. *CoRR*, abs/2109.08270.
- Xu, J., Zhang, Z., Friedman, T., Liang, Y., and den Broeck, G. V. (2018). A semantic loss function for deep learning with symbolic knowledge. In Dy, J. G. and Krause, A., editors, *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, volume 80 of *Proceedings of Machine Learning Research*, pages 5498–5507. PMLR.
- Yang, Z., Han, Y., Yu, G., and Zhang, X. (2019). Prototypical networks for multi-label learning. *CoRR*, abs/1911.07203.
- Ye, L., Rochan, M., Liu, Z., and Wang, Y. (2019). Cross-modal self-attention network for referring image segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pages 10502–10511. Computer Vision Foundation / IEEE.

- Yosinski, J., Clune, J., Bengio, Y., and Lipson, H. (2014). How transferable are features in deep neural networks? In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. D., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pages 3320–3328.
- Yu, L., Tan, H., Bansal, M., and Berg, T. L. (2017). A joint speaker-listener-reinforcer model for referring expressions. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 3521–3529. IEEE Computer Society.
- Zarri , S. and Schlangen, D. (2019). Know what you don’t know: Modeling a pragmatic speaker that refers to objects of unknown categories. In Korhonen, A., Traum, D., and M rquez, L., editors, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 654–659, Florence, Italy. Association for Computational Linguistics.
- Zeng, A., Florence, P., Tompson, J., Welker, S., Chien, J., Attarian, M., Armstrong, T., Krasin, I., Duong, D., Sindhwani, V., and Lee, J. (2020). Transporter networks: Rearranging the visual world for robotic manipulation. In Kober, J., Ramos, F., and Tomlin, C. J., editors, *4th Conference on Robot Learning, CoRL 2020, 16-18 November 2020, Virtual Event / Cambridge, MA, USA*, volume 155 of *Proceedings of Machine Learning Research*, pages 726–747. PMLR.
- Zhao, W., Queralt, J. P., and Westerlund, T. (2020). Sim-to-real transfer in deep reinforcement learning for robotics: a survey. In *2020 IEEE Symposium Series on Computational Intelligence, SSCI 2020, Canberra, Australia, December 1-4, 2020*, pages 737–744. IEEE.
- Zitkovich, B., Yu, T., Xu, S., Xu, P., Xiao, T., Xia, F., Wu, J., Wohllhart, P., Welker, S., Wahid, A., Vuong, Q., Vanhoucke, V., Tran, H. T., Soricut, R., Singh, A., Singh, J., Sermanet, P., Sanketi, P. R., Salazar, G., Ryoo, M. S., Reymann, K., Rao, K., Pertsch, K., Mordatch, I., Michalewski, H., Lu, Y., Levine, S., Lee, L., Lee, T. E., Leal, I., Kuang, Y., Kalashnikov, D., Julian, R., Joshi, N. J., Irpan, A., Ichter, B., Hsu, J., Herzog, A., Hausman, K., Gopalakrishnan, K., Fu, C., Florence, P., Finn, C., Dubey, K. A., Driess, D., Ding, T., Choromanski, K. M., Chen, X., Chebotar, Y., Carbajal, J., Brown, N., Brohan, A., Arenas, M. G., and Han, K. (2023). RT-2: vision-language-action models transfer web knowledge to robotic control. In

Tan, J., Toussaint, M., and Darvish, K., editors, *Conference on Robot Learning, CoRL 2023, 6-9 November 2023, Atlanta, GA, USA*, volume 229 of *Proceedings of Machine Learning Research*, pages 2165–2183. PMLR.