Strategy-level Entrainment of Dialogue System Users in a Creative Visual Reference Resolution Task

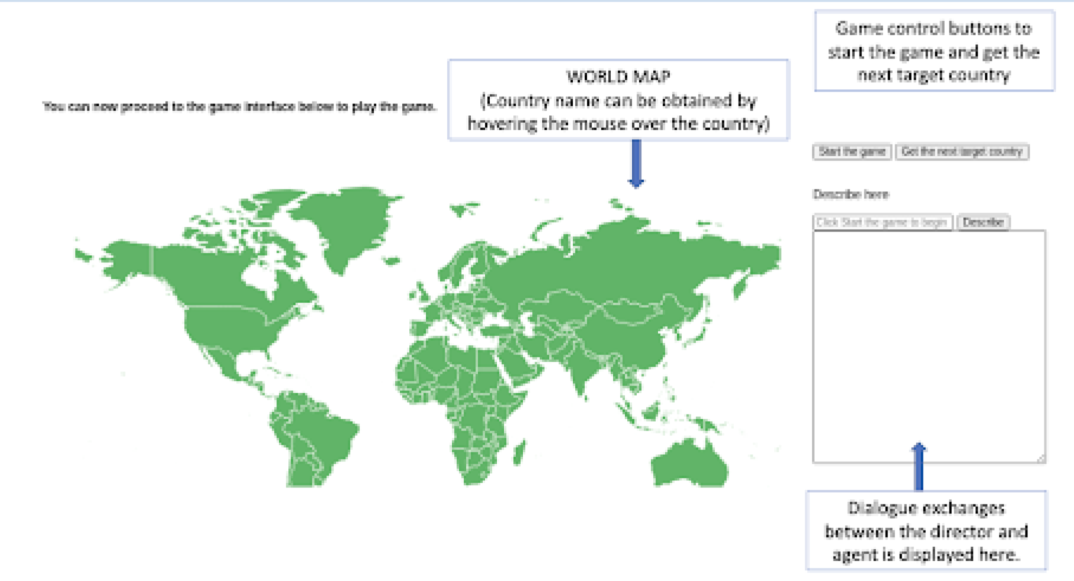
Deepthi Karkada[1], Ramesh Manuvinakurike[1], Maike Paetzel-Prüsmann[2], Kallirroi Georgila[3]
[1] *Intel Corp*, [2] *University of Potsdam*, [3] *Institute for Creative Technologies, University of Southern California*

LREC 2022, Marseille, France

Introduction

- Human descriptions of visual entities in a reference resolution setting can be creative which can be very challenging for dialogue systems. In this work, we study if linguistic entrainment in a creative visual reference resolution game (RDG-Map) can help address the issue.
- Our dialogue system attempts to entrain users on a "strategy-level" without impinging on their descriptive creativity.
- We study if priming users to provide descriptions that are relatively simple helps the dialogue system's natural language understanding (NLU).

Game implementation



Director	Matcher
Pakistan.	
	I don't know where that is.
It's in Asia.	
	I need more information.
It's northwest of India.	
	What does it look like?
It looks like a T-Rex.	
	Got it.

Figure 1: The experiment setup along with a sample conversation between the Director and Matcher.

- RDG-Map (Rapid Dialogue Game - Map) is a two player collaborative, time constrained spoken dialogue game [1]
- The two roles in the game are the director (Dir) and the matcher (Mat)
- The director sees a target country highlighted on the world map. The matcher may not know the target country.
- The director describes the target country for the matcher to make a selection using location, size and shape references.
- Towards automating the matcher, we built a dialogue system agent using knowledge graphs (KG) for the natural language understanding and question generation.

Corpora

The data sources used for training the agent matcher consist of three parts:

- Spoken Human-Wizard Conversations (D_{spoken}) - Annotated conversations collected by [1] between a human director and a remote-controlled (wizarded) agent matcher.
- Wikidata (D_{wiki}) [2] - Openly available knowledge base about the countries on the world map.
- Shape Descriptions (D_{shapes}) - Country shapes data collected using crowd-sourcing on Amazon Mechanical Turk.

D_{spoken}	# Users	80
	# Dialogues	980
	# Turns	3989
	# Tokens	24726
	# Shape descriptions	905
D_{wiki}	# Entities (KG)	2868
	# Relations (KG)	38
D_{shapes}	# Tokens	6496
	# Shape descriptions	1563

Linguistic entrainment

- The phenomenon in which the interlocutors start speaking more similarly to each other. Entrainment has been studied in dialogue systems in the context of increasing task success and can happen on multiple levels such as prosody, speech rate, vocabulary usage, and syntactic and stylistic patterns.
- In this work, we study if human interlocutors exchange their entire descriptive strategy when interacting with an autonomous dialogue system instead of individual phrases in a creative task.

Experimental setup

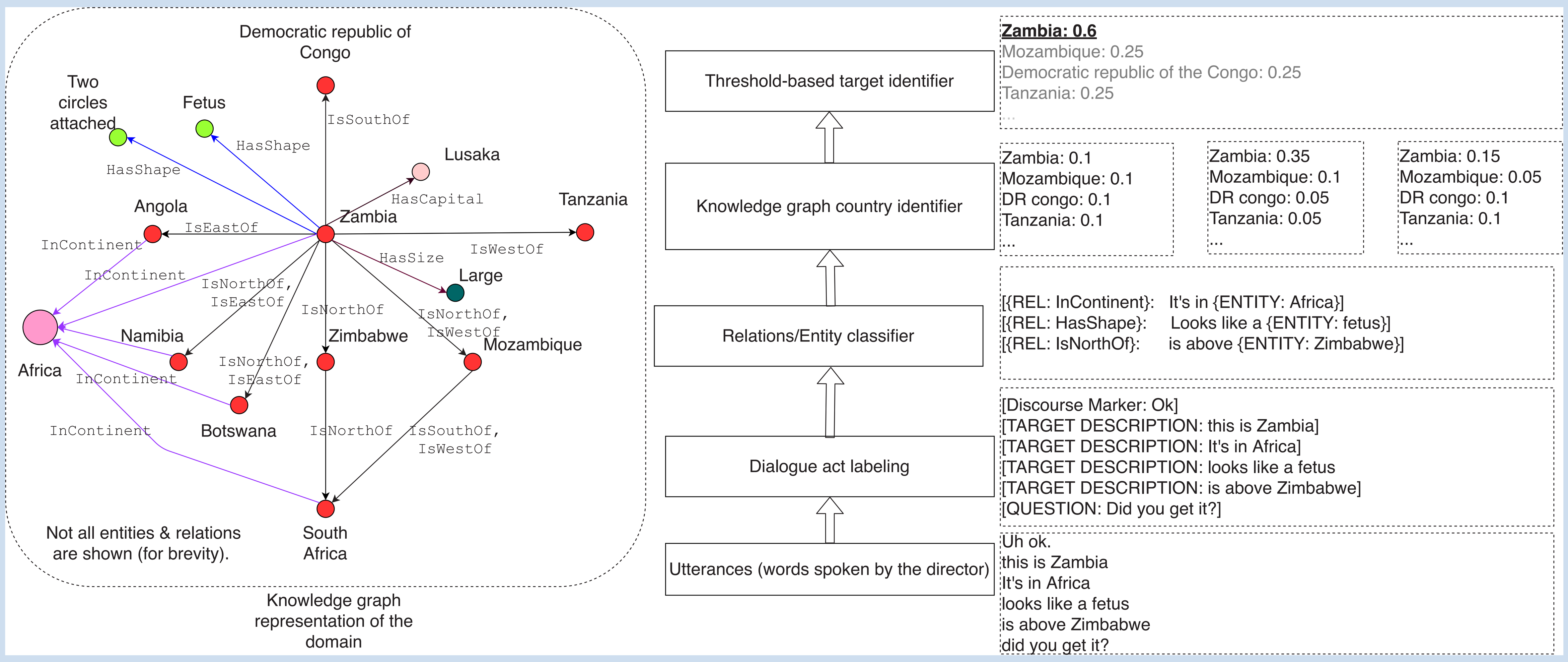


Figure 1: Shows the dialogue system/agent's NLU pipeline.

- We investigate whether the autonomous agent can entrain users at "strategy-level". We divided the game into three phases: the pre-entrainment phase, the entrainment phase, and the post-entrainment phase.
- The agent asks questions inquiring about the location, shape, continent and landmark. The agent's NLU pipeline outputs a probability distribution across the countries the director is describing.
- The questions are generated for the country to be selected by the agent. During the entrainment phase, the shape question is a yes/no question querying about a specific geometric shape description linked in the knowledge graph.
- We hypothesize that entraining the users not only preserves the creativity but also makes it easier for the NLU module to identify the target.

Results

- Entrainment Analysis: A One Sample t-test reveals that our participants use significantly more geometric shapes in the post-entrainment compared to the pre-entrainment phase, $t = 4.5323$, $p < 0.001$. We can hence conclude that the agent was able to get users to adapt their descriptive strategy after the entrainment phase.
- Agent Performance: For the top-1 and top-5 predictions, we observe a slight improvement in performance in the NLU in the post-entrainment phase. This improvement, however, is not significant as a One Sample t-test reveals, $t=1.6$, $p=0.2$.
- Creativity: We observe that priming users for a specific strategy does not impair the creativity of their descriptions, suggesting that the quality of descriptions given and the level of engagement in the task is not affected by the entrainment.

Future work

- We aim to explore recent advancements in reinforcement learning to learn better response selection strategies.

References

[1] Paetzel, M., D. Karkada, and R. Manuvinakurike. RDG-Map: A Multimodal Corpus of Pedagogical Human-Agent Spoken Interactions. LREC 2020. [2] Vrandečić, D. and Krötzsch, M. Wikidata: A free collaborative knowledgebase.