Utrecht University

LREC 2026

May, 2026

Grounded Misunderstandings in Asymmetric Dialogue: A Perspectivist Annotation Scheme for MapTask

Nan Li[1], Albert Gatt[1], Massimo Poesio[12]
Utrecht University[1] Queen Mary University of London[2]

https://arxiv.org/abs/2511.03718

Nan Li, Utrecht University
Supervised by Massimo Poesio and Albert Gatt

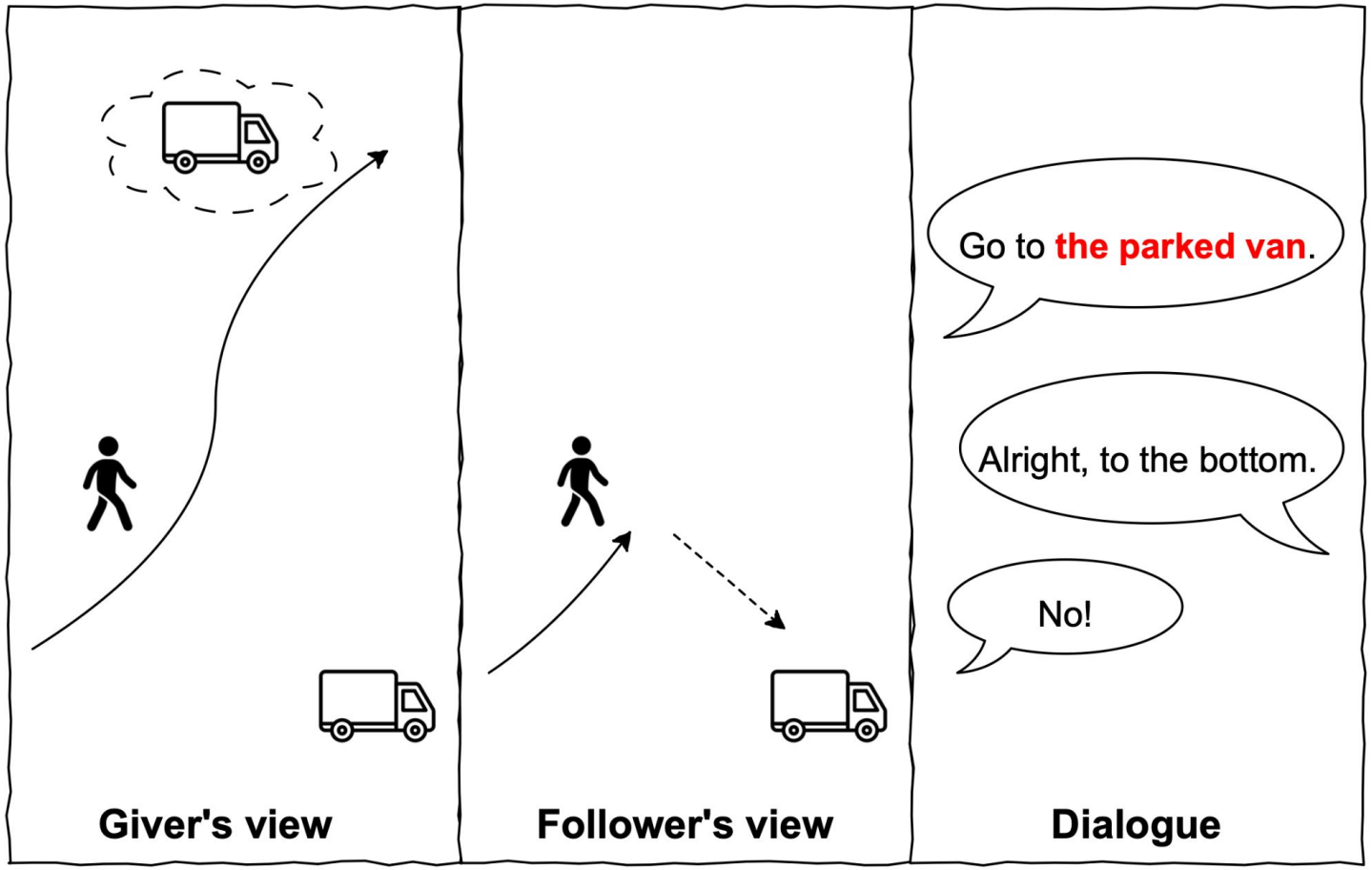
Outlines

- Motivation
  - *Divergent interpretations among interlocutors in dialogue*
- Background
  - *MapTask Corpus*
- Annotation Scheme and Practice
  - *Design & Pipeline*
  - *Quantitative Analysis*

Grounding in dialogue is an incremental alignment process

- **The standard view**: Utterances enter common ground only after being recognized and acknowledged by the addressee.
- **A hidden flaw**: Traditional reference resolution assumes that once this grounding is achieved, both people are referring to the exact same entity.
- **In asymmetric setting**: In collaborative tasks with asymmetric information, this assumption regularly fails. Interlocutors can tacitly agree while holding divergent interpretations.

A perspectivist view on reference resolution



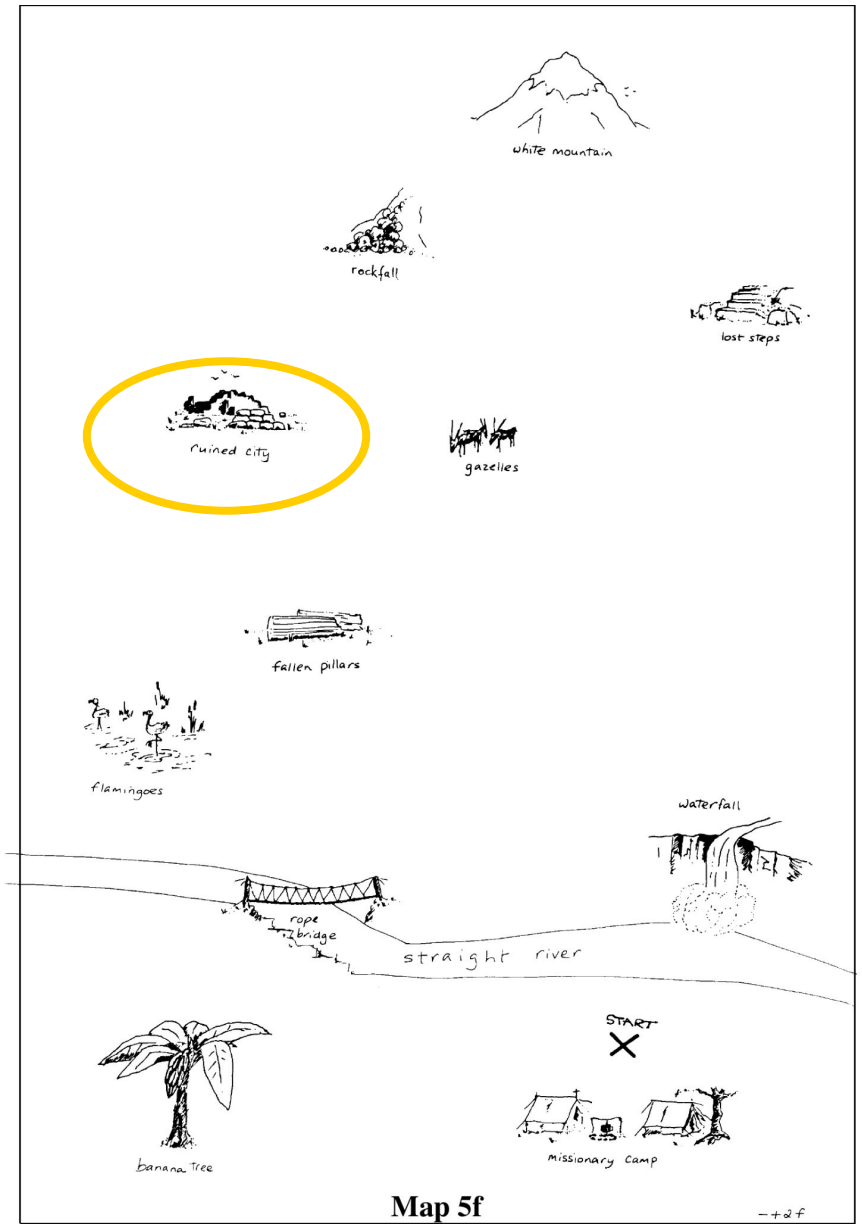
- Reference expressions (REs) are central to coordinating understanding in dialogue.
- In collaborative tasks, interlocutors will adapt their expressions and behavior based on their perspective knowledge.
- In asymmetric settings, divergent interpretations may persist even after apparent grounding.

Research Questions

1. Is it possible to develop a **personal-interpretation annotation scheme** to capture nuanced understanding states for REs in collaborative tasks like MapTask?
2. Given conflicting personal interpretations, can we **trace how understanding** evolves across turns until the successful grounding?
3. Can **LLMs, under a schema-constrained protocol, reliably annotate personal interpretations**; and how does their output inform future evaluation of (V)LLMs on grounded dialogue?

MapTask

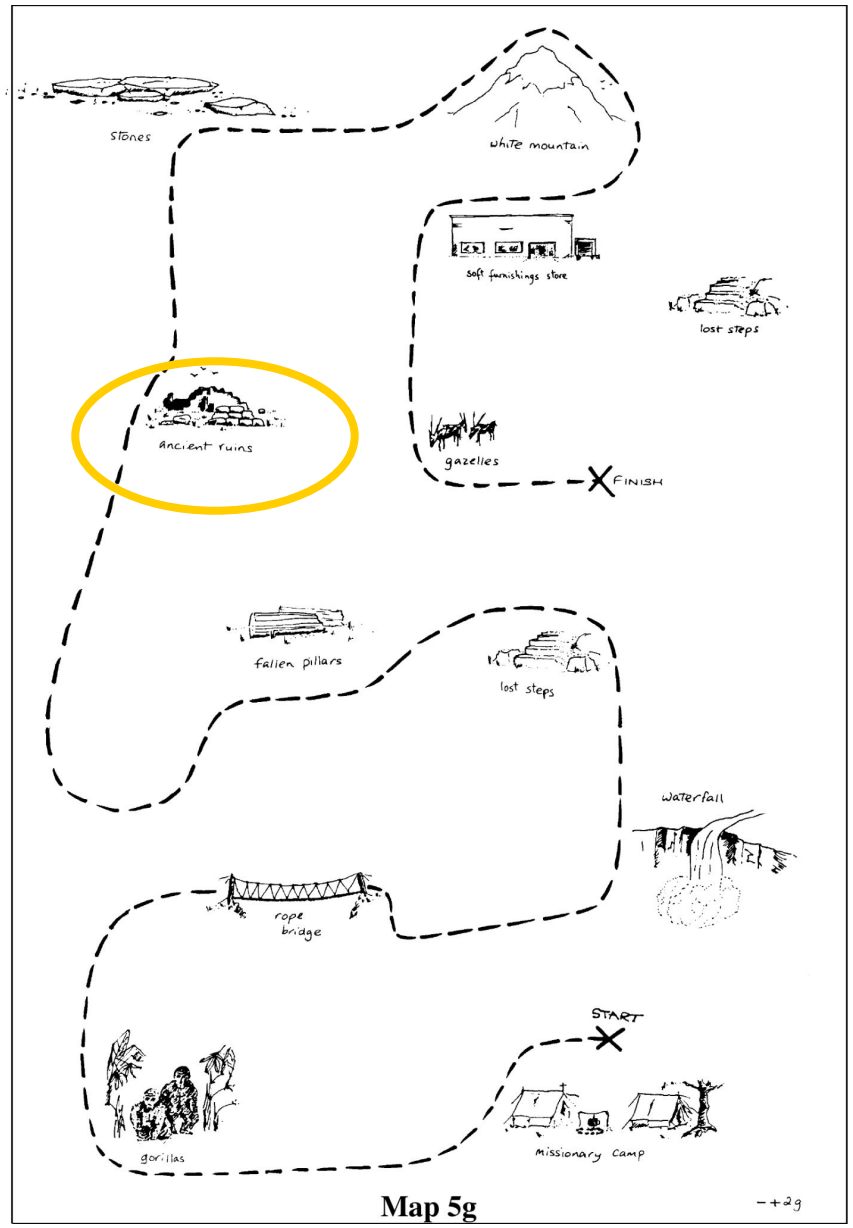
- A corpus of 128 unscripted task-oriented dialogues.
- Participants must collaborate to reproduce a route on one participant's map (the instruction giver's) on the other's (the instruction follower's). They cannot see each other's maps and can only communicate verbally.
- The intentional differences between maps on routes, landmark names, and landmark placements often lead to ambiguous or misleading instructions.



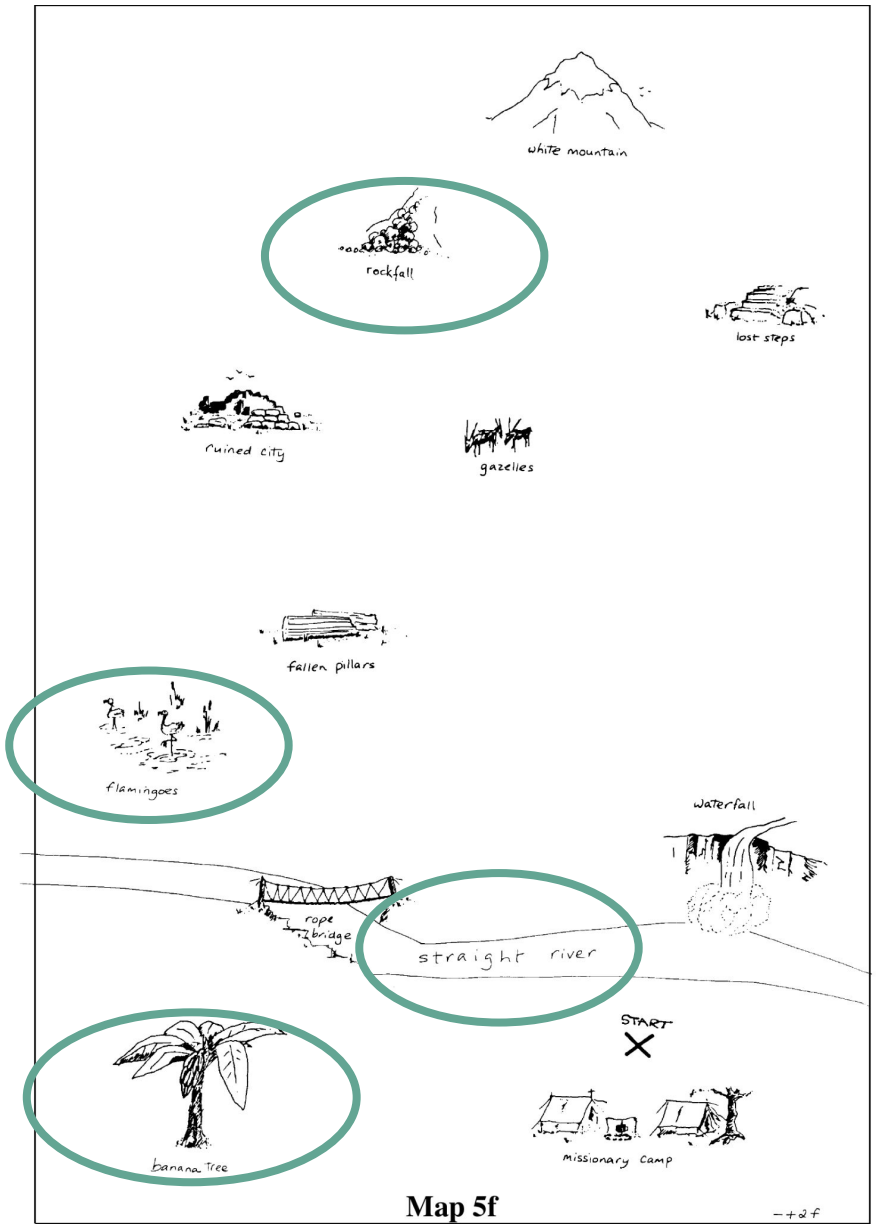
Lexical

Existence

Multiplicity



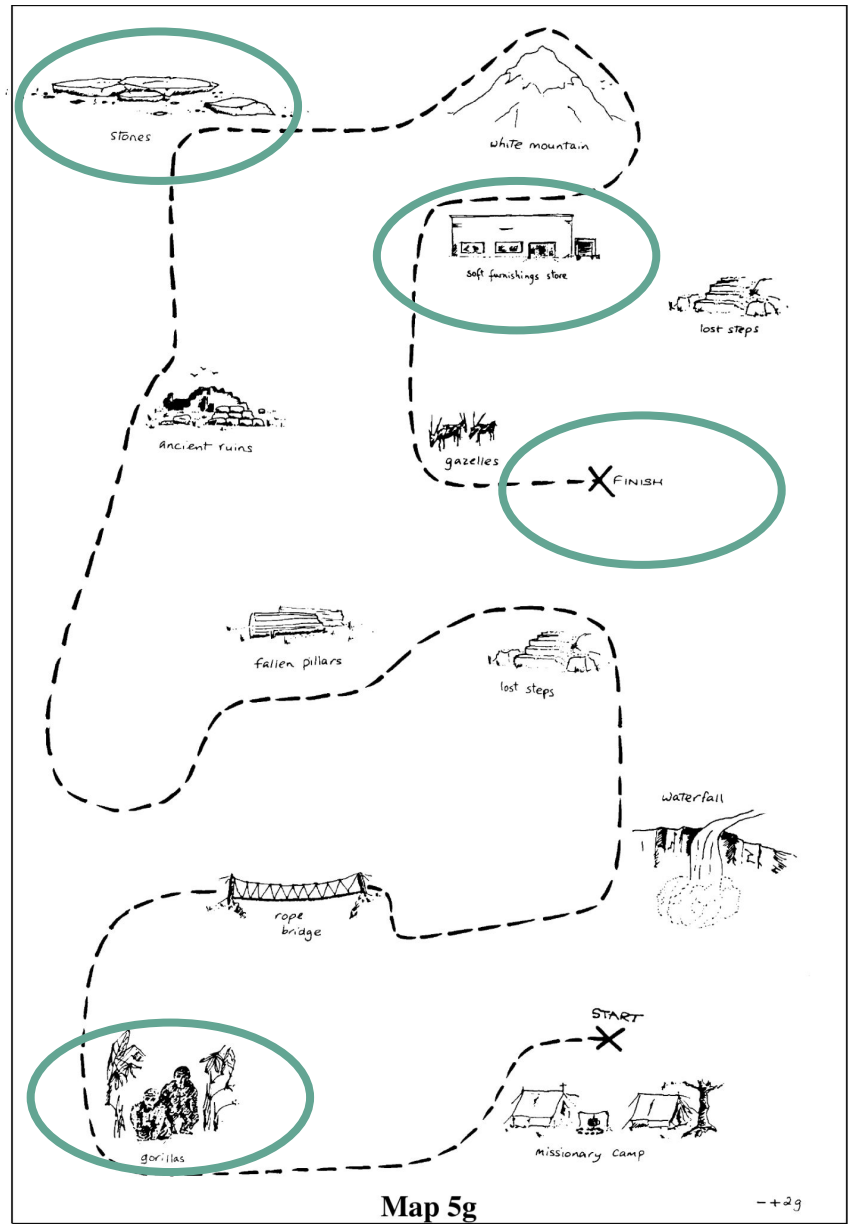
HCRC MapTask Design



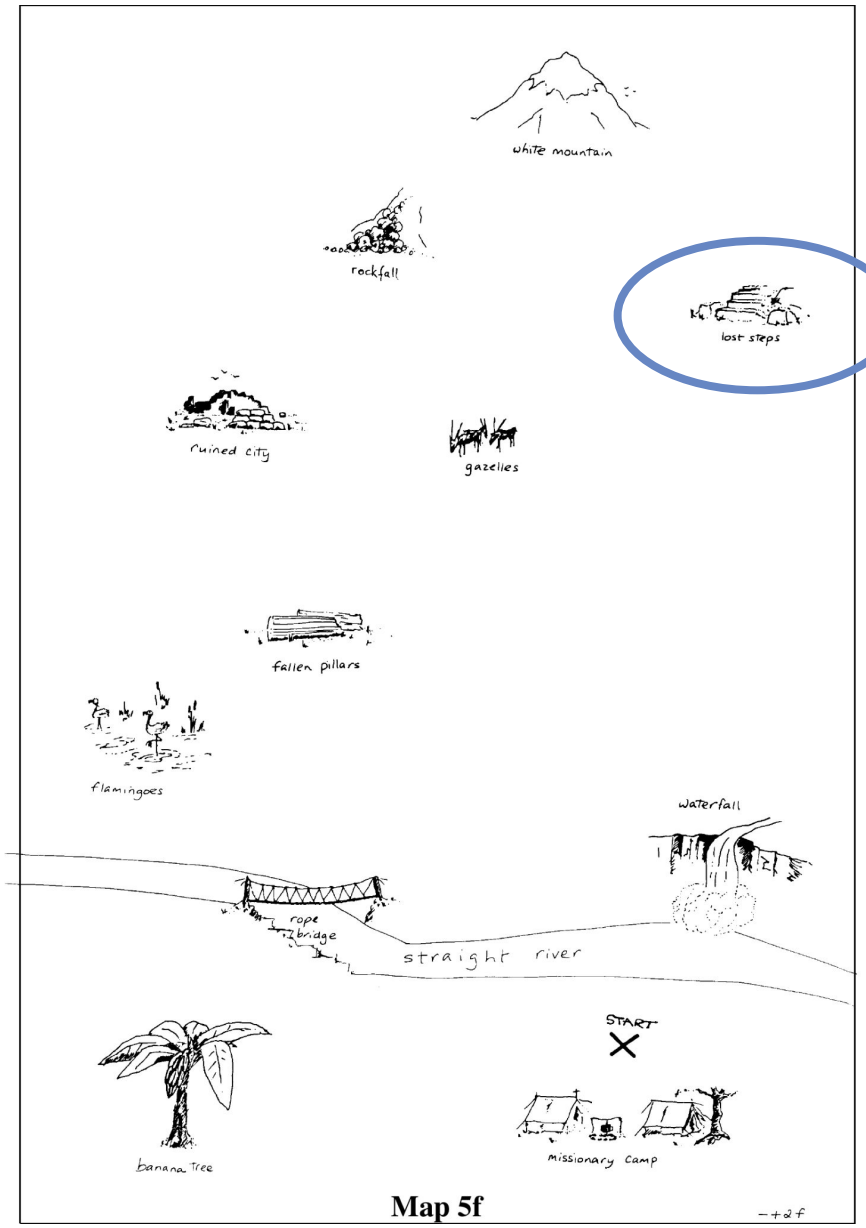
Lexical

Existence

Multiplicity



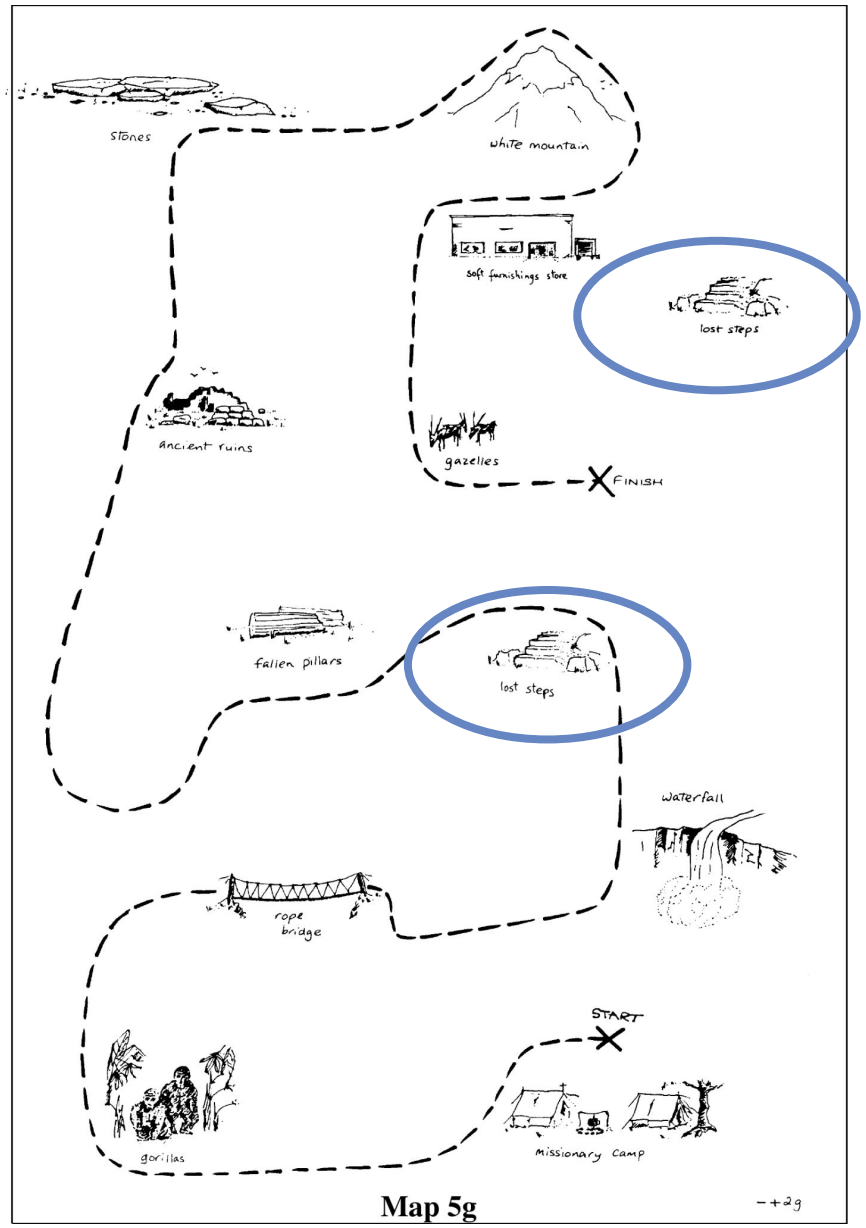
HCRC MapTask Design



Lexical

Existence

Multiplicity



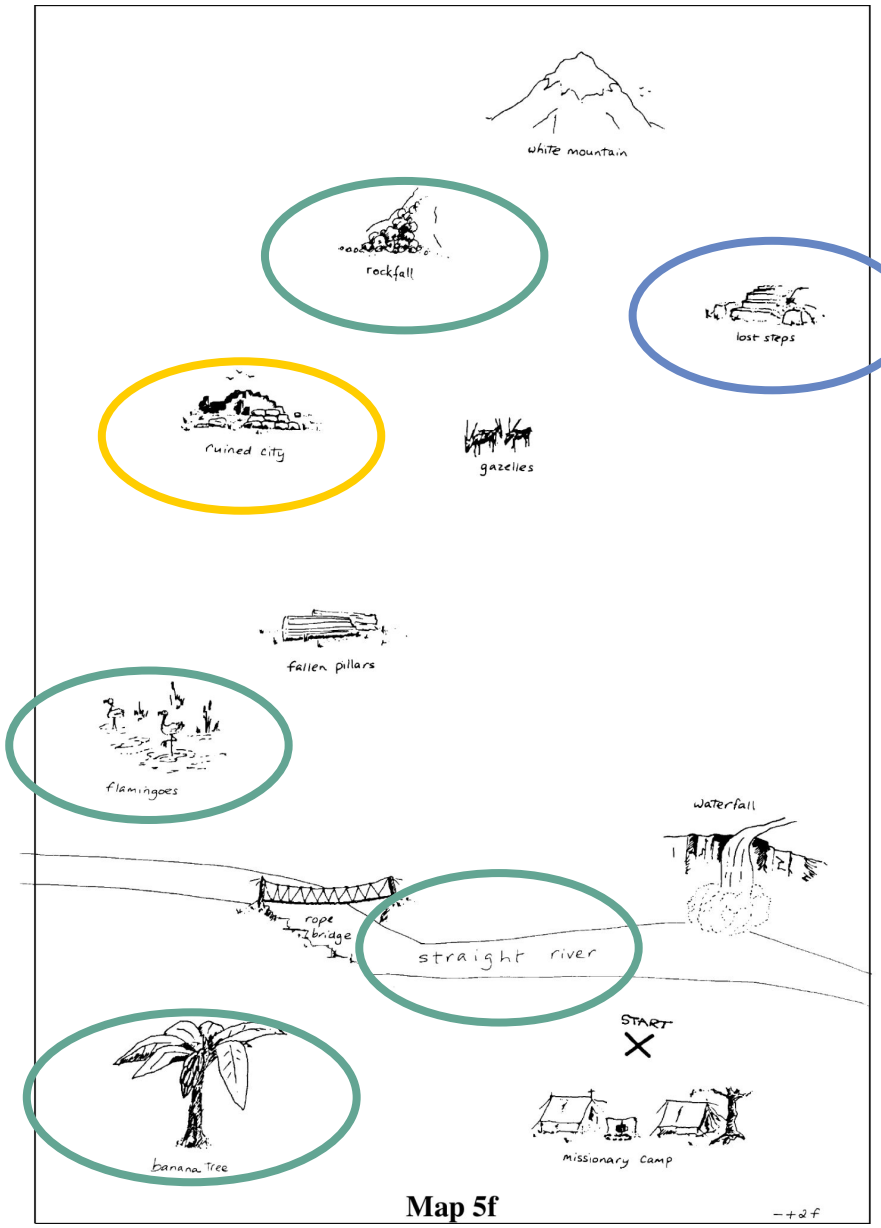
HCRC MapTask Design

HCRC MapTask Annotations Leveraged in This Work

- Timed Units
- Reference Expressions
- Move Tags and Transactions
  - **[g utt:1 move:instruct] starting off we are above a caravan park**
  - **[f utt:2 move:acknowledge] mmhmm**

Limitations of the original MapTask landmark ID scheme

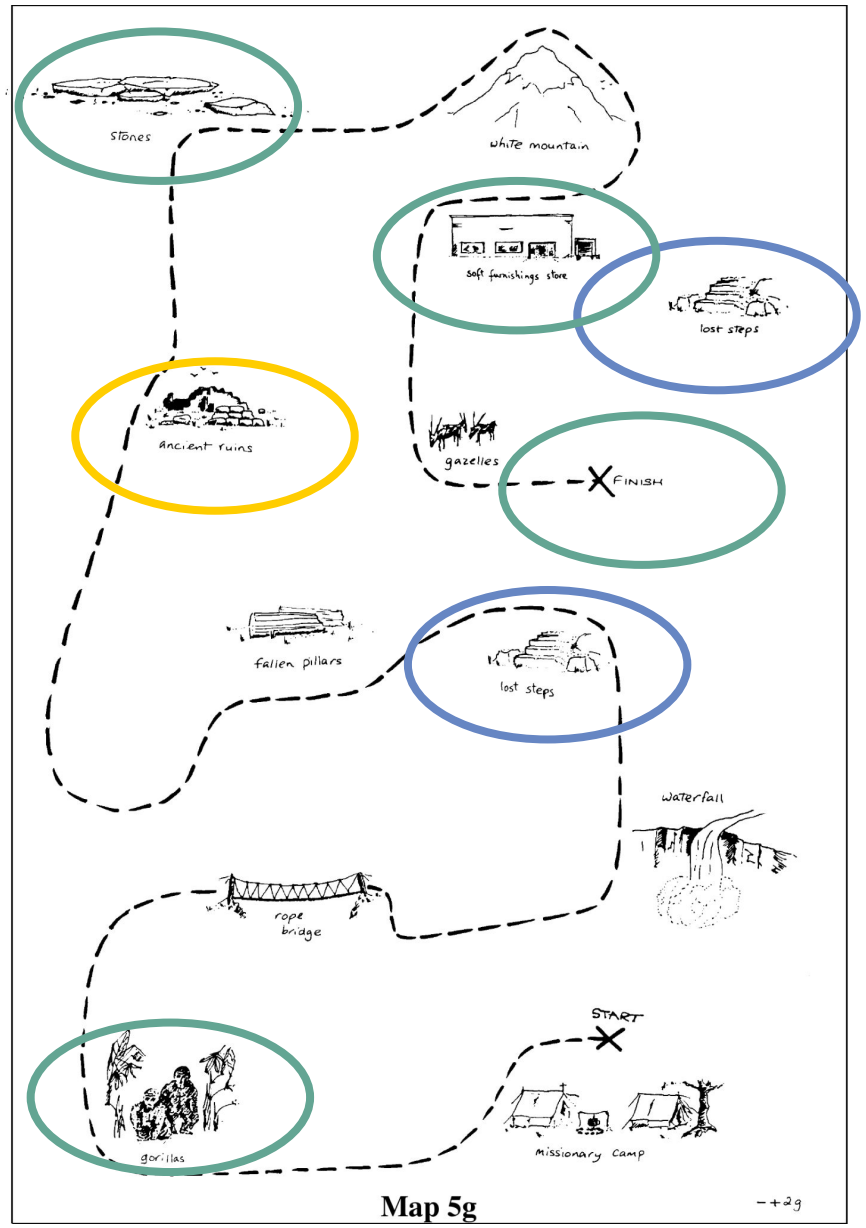
- Differentiated by landmark names.
- Treat landmarks of multiplicity discrepancies as the same entity, but treat lexical variants as different entities.



Lexical

Existence

Multiplicity

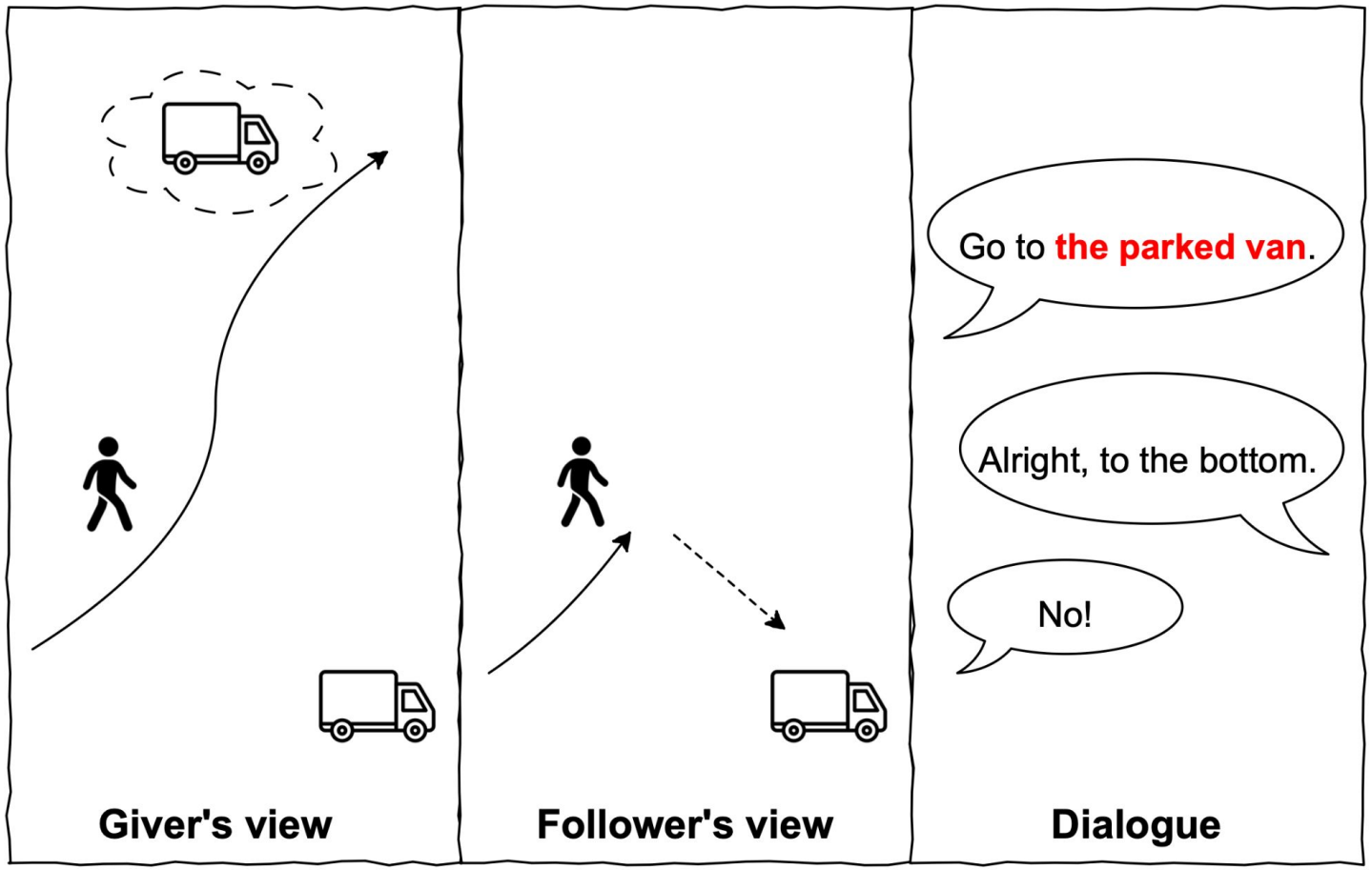


HCRC MapTask Design

Perspective-taking Annotation Scheme
Step 0: Landmark ID Reassignment

<map-id>_<landmark-name>#<ordinal>@<side>

- map-id: the original map ID
- landmark-name: the landmark name as it appears on the map
- ordinal: Positional index (0, 1) for landmarks of the same type occurring multiple times on maps, ordered from bottom to top.
- side: map side (g for giver, f for follower)



Landmark ID Reassignment

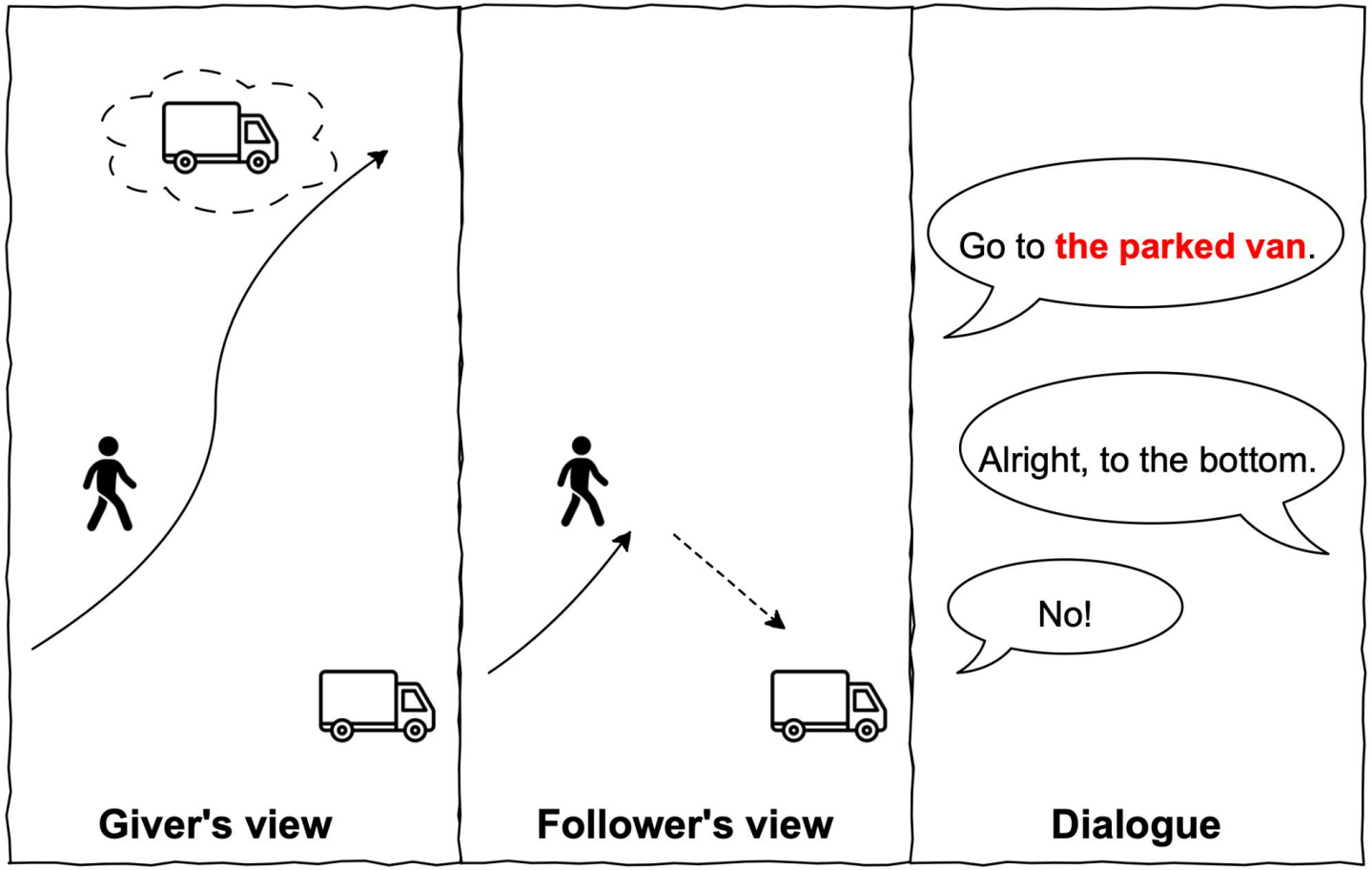
- On the giver's map
  - **m0_parked_van#0@g**
  - **m0_parked_van#1@g**
- On the follower's map
  - **m0_parked_van#0@f**

Perspective-taking Annotation Pipeline:
A 5-layer attribute resolution process

For each RE:

- **is_quantificational**: whether the RE is a quantificational expression.
- **is_specified**: whether the context has given enough information to specify the addressee's response.
- **is_accommodated**: whether the addressee has addressed the RE successfully.
- **is_grounded**: whether the addressee has linked the RE to a specific landmark on maps.
- **is_imagined**: whether the addressee is grounding the RE by the description of the speaker to a landmark absent from his/her own map.

Finally: assign unified landmark IDs to the interlocutors' interpretations



Steps: A 5-layer attribute resolution process

- **is_quantificational**: *false*
- **is_specified**: *true*
- **is_accommodated**: *true*
- **is_grounded**: *true*
- **is_imagined**: *false*

Assign unified landmark IDs to the interlocutors' interpretations

- **giver**: m0_parked_van#1@g
- **follower**: m0_parked_van#0@f

Perspective-taking Annotation Pipeline: Prompt Structure and Template

Prompt Template (Abbreviated)

<background> Background on MapTask, landmark discrepancies, and reference resolution </background>

<task_description> Annotate personal interpretations for reference expressions </task_description>

<landmark_id_explanation> Unified landmark ID format and semantics </landmark_id_explanation>

<annotation_rule>

Follow the annotation workflow for each RE step by step:

Step 0: Identify the speaker and the addressee, and match them with the giver and follower based on the actual situation; fill in the landmark ID for the speaker's interpretation (intended landmark) first.

Steps 1-5: Check and fill in attributes is_quantificational, is_specified, is_accommodated, is_grounded, and is_imagined in order.

Step 6: If grounded, fill the landmark ID for the addressee's interpretation.

Step 7: Provide a concise (<50 words) evidence-based reason summarizing the judgement in the reason field for each RE.

</annotation_rule>

<output_format> JSON schema with required fields </output_format>

<target_ref_ids> IDs of reference expressions to annotate </target_ref_ids>

<context_dialogue> Dialogue transcript with surrounding context </context_dialogue>

<context_dialogue_acts> Move-level annotations for context </context_dialogue_acts>

<landmark_candidates> All landmarks on both maps with IDs </landmark_candidates>

LLM Annotation & Quality Evaluation

Attribute	N	Acc.	F1	Errors
is_quantificational	504	100.00%	1.00	0
is_specified	466	97.85%	0.990	10
is_accommodated	455	97.58%	0.988	11
is_grounded	447	96.20%	0.980	17
is_imagined	422	97.63%	0.891	10
RE-level: 28/504 REs (5.6%) with $\geq$1 error				48

- Model: We utilized GPT-5 via the OpenAI Batch API, enforcing a strict JSON-schema output

- Human golden standard
  - 504 REs across 3 dialogues
  - The pipeline achieved a **94.4% RE-level accuracy** across these human-verified expressions.

State	Before		After	
	Count	%	Count	%
Aligned	8,750	66.9	9435	72.1
Pending	1,799	13.8	1799	13.8
Misunderstood	924	7.07	239	1.82
Total REs	13,077	100.0	13,077	100.0

Discrepancy Type	Total REs	Misunderstanding	Rate (%)
Lexical	914	13	1.4
Multiplicity	956	115	12.0
Existence	2,878	93	3.2
Identical	8,333	18	0.2
Total	13,081	239	1.8

Understanding State

We infer understanding states (aligned, pending, misunderstood) from the values of the 5 attributes and landmark IDs.

We unify lexical variants in each pair to a single landmark.

Understanding State Tracking

Discrepancy Type	# Chains	Mean Length	Max Length	Median Length
Identical	939	8.16	70	6
Lexical	79	10.70	61	9
Existence	539	4.09	31	3
Multiplicity	108	7.0	26	5

[giver ln:445 utt:282] going sloping downwards towards <<the monument id:q1nc7.ref.182 lm:m12_monument>>
[follower ln:446 utt:283] right i've and then curve round <<the monument id:q1nc7.ref.183 lm:m12_monument>>
[giver ln:447 utt:284] round <<the monument id:q1nc7.ref.184 lm:m12_monument>>
[follower ln:448 utt:285] right
[giver ln:449 utt:286] just right
[follower ln:450 utt:287] i've done
[giver ln:451 utt:286] round
[follower ln:452 utt:287] that
[giver ln:453 utt:286] <<that id:q1nc7.ref.185 lm:m12_monument>>
[giver ln:454 utt:288] eh sort of right are you at the of top <<the monument id:q1nc7.ref.186 lm:m12_monument>>
see <<the guy's head id:q1nc7.ref.187 lm:m12_monument>>
[follower ln:455 utt:289] yes <<that's id:q1nc7.ref.188 lm:m12_monument>>
[giver ln:456 utt:290] right

We form **reference chains** by tracing the same original landmark ID.

Length of reference chains reflects how persistently participants negotiate understanding of landmarks and use these landmarks to generate spatial descriptions.

Reference Chain

Dialogue Excerpt: q7ec2 (utterances 131–161)

[giver utt:131] i've got **«a meadow»**[ref.112] to the left of «that» due south pee porashin

[follower utt:132] **«a vast meadow»**[ref.114]

[giver utt:133] **«a vast meadow»**[ref.115]

[follower utt:134] i've got **«a vast meadow»**[ref.116] up just below «the blacksmiths» and «carpenters»

[giver utt:135] well is that huge big and there's nothing at all in that space

[follower utt:136] uh-huh except

[giver utt:137] about halfway up the page

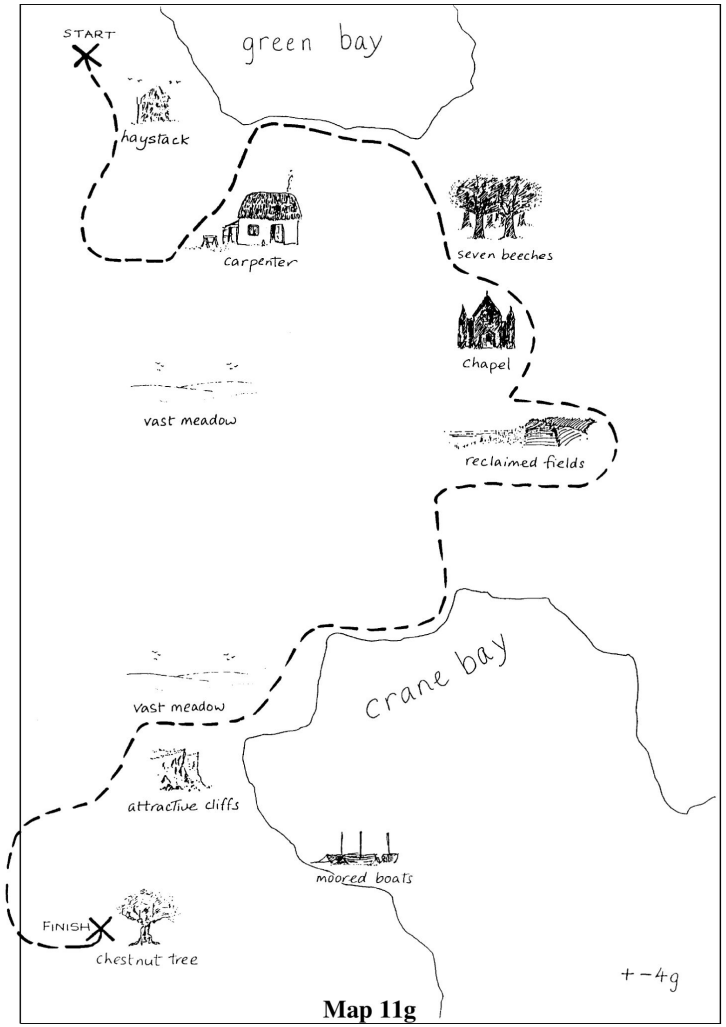
[follower utt:138] **«vast meadow»**[ref.119] about halfway up the page it's empty on mine

[giver utt:139] well you want to well anyway follow «crane bay» round «the curve»

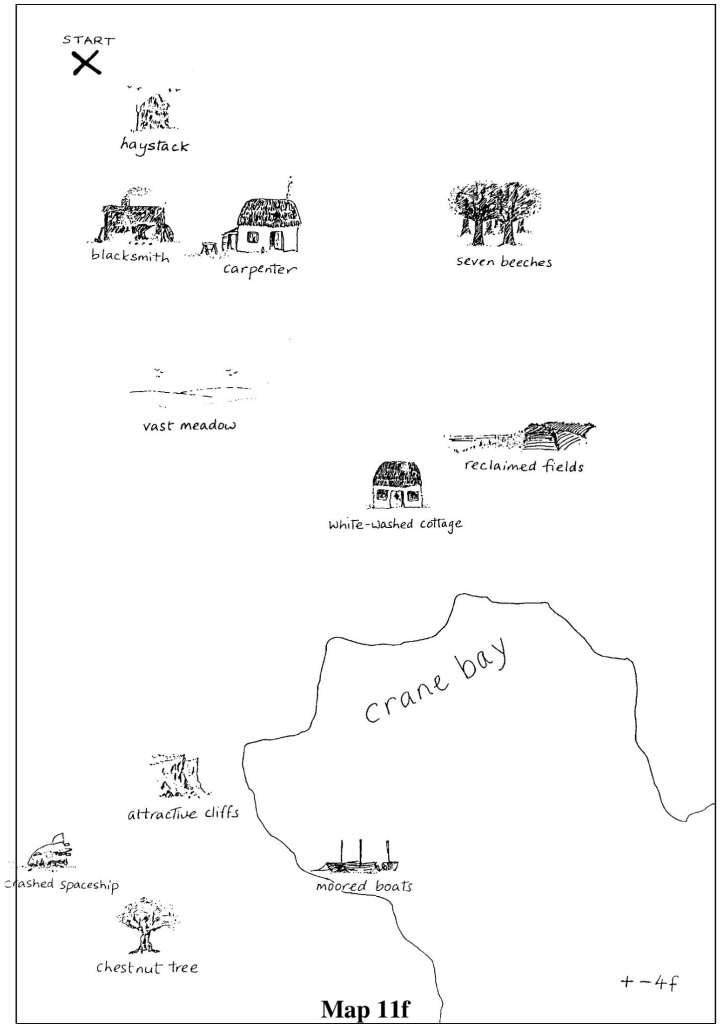
... (utterances 140–160: participants navigate along crane bay and attractive cliffs) ...

[giver utt:161] just below i've got **«a vast meadow»**[ref.127] there

[follower utt:162] right where will i stop that line across the top



Map 11g



Map 11f

Reference Chain

Dialogue Excerpt: q7ec2 (utterances 131–161)

[giver utt:131] i've got **«a meadow»**[ref.112] to the left of «that» due south pee porashin

[follower utt:132] **«a vast meadow»**[ref.114]

[giver utt:133] **«a vast meadow»**[ref.115]

[follower utt:134] i've got **«a vast meadow»**[ref.116] up just below «the blacksmiths» and «carpenters»

[giver utt:135] well is that huge big and there's nothing at all in that space

[follower utt:136] uh-huh except

[giver utt:137] about halfway up the page

[follower utt:138] **«vast meadow»**[ref.119] about halfway up the page it's empty on mine

[giver utt:139] well you want to well anyway follow «crane bay» round «the curve»

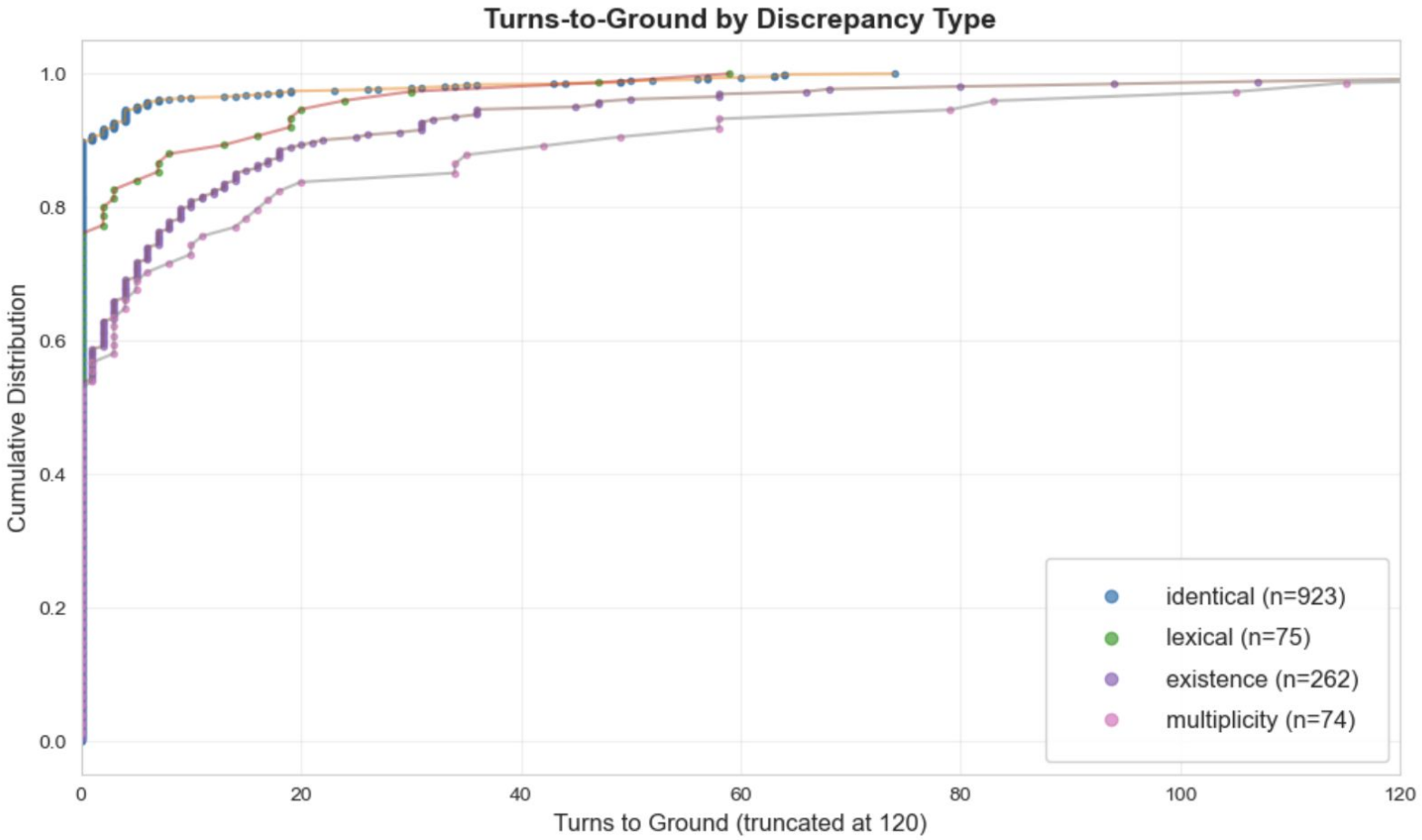
... *(utterances 140–160: participants navigate along crane bay and attractive cliffs)* . . .

[giver utt:161] just below i've got **«a vast meadow»**[ref.127] there

[follower utt:162] right where will i stop that line across the top

Ref	Spk	Expression	Giver ID	Follower ID	State
112	G	"a meadow"	#0@g	*ungrounded*	Pending
114	F	"a meadow"	#0@g	*ungrounded*	Pending
115	G	"a vast meadow"	#0@g	#1@f	**Misunderstood**
116	F	"a vast meadow"	#1@g	#1@f	Aligned
119	F	"vast meadow"	#1@g	#1@f	Aligned
127	G	"a vast meadow"	#0@g	*ungrounded*	Pending

Understanding State Tracking



turns-to-ground: by counting how many turns participants need to take before reaching the “aligned” state in a reference chain.

Conclusion

- Our annotation scheme disentangles the speaker’s intended and the addressee’s interpreted referents through a five-attribute hierarchy and unified landmark IDs, operationalized via a scheme-constrained prompting pipeline with reliability estimates.
- Our quantitative analyses reveal given the rarity of full misunderstandings, multiplicity discrepancies act as a consistent hazard for eliciting misalignment. Our statistics also demonstrate interlocutors' systematic preferences for reusing grounding-established REs.

Utrecht University

Sharing science,
shaping tomorrow