Utrecht University

MINT WORKSHOP · EMNLP 2026

Gaze as Evidence for Common Grounding

A cross-corpus analysis of MapTask and MUNDEX

Nan Li[1] · Albert Gatt[1] · Massimo Poesio[1 2]

[1]Utrecht University · [2]Queen Mary University of London

Grounding runs on evidence; gaze may supply some of it

THEORY · EVIDENCE OF UNDERSTANDING

Clark & Schaefer (1989)

1. Continued attention
2. Initiation of the relevant next contribution
3. Acknowledgement
4. Demonstration
5. Display

OUR HYPOTHESES · WHAT GAZE MAY SHOW

TASK GAZE — continued attention to the joint task

PARTNER GAZE — attention to the partner; monitoring and inviting feedback (Bavelas et al. 2002; Clark & Krych 2004) ; can also signal trouble

AWAY GAZE — processing load or a topic shift (Doherty-Sneddon & Phelps 2005; Lazarov & Grimminger 2026)

Gaze labels here record **where** people look; one look can serve several of these functions.

RESEARCH QUESTIONS

Three questions about gaze and grounding

Q1 ASSOCIATIONS

Which gaze patterns go with grounding in both tasks?

Overall; then by who produced the reference, visibility, and whose label

MAPTASK MUNDEX

Q2 PROCESS

How does gaze change when a referent becomes aligned?

Same-speaker pairs along reference chains

MAPTASK

Q3 PREDICTION

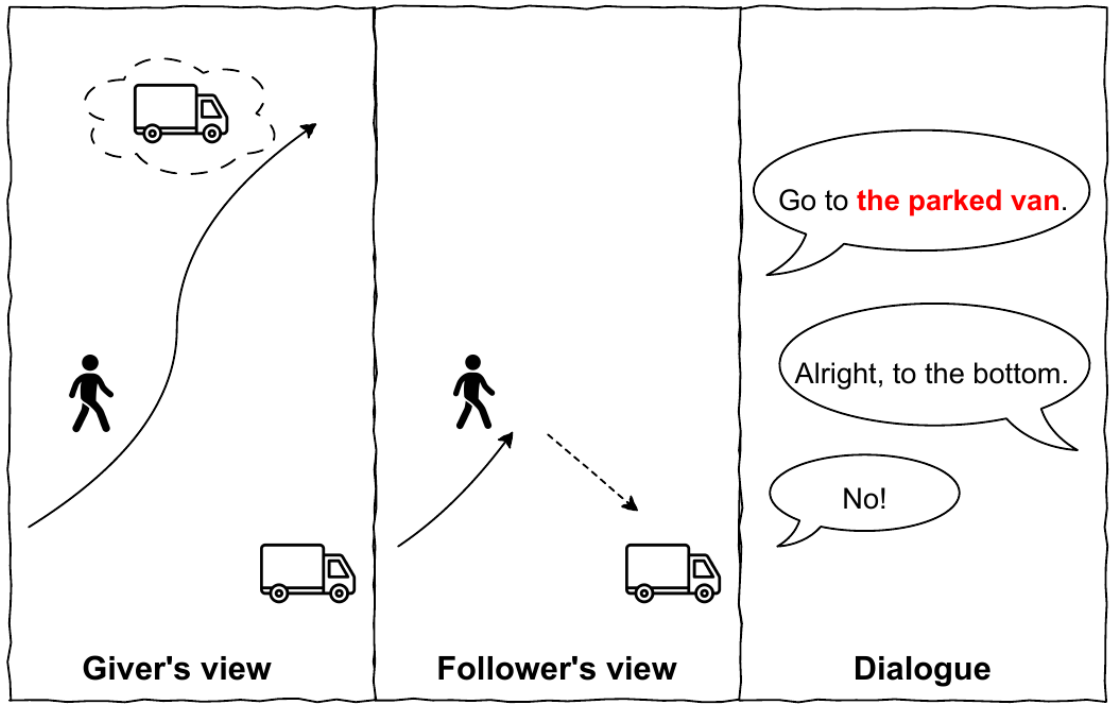
How well do gaze features predict grounding labels, and which feature groups help?

MAPTASK MUNDEX

Earlier gaze studies work within one corpus (Boyle et al. 1994; Nakano et al. 2003; Murat & Vogel 2026; Wang et al. 2026). We use one shared vocabulary across two corpora.

TASKS · MAPTASK

“Go to the parked van”: one expression, two interpretations



THE TASK · ENGLISH

A **giver** guides a **follower** along a route. Their maps differ by design, and neither sees the other’s map.

PERSPECTIVIST LABELS

Our LREC 2026 annotation (Li et al. 2026): for every reference, the speaker’s **intended** landmark and the addressee’s **interpreted** landmark, recorded separately.

ALIGNED same landmark

PENDING not yet resolved

MISUNDERSTOOD different referents

Anderson, A. H., Bader, M., Bard, E. G., et al. (1991). The HCRC Map Task Corpus. *Language and Speech* , 34(4), 351–366. Li, N., Gatt, A., & Poesio, M. (2026). Grounded misunderstandings in asymmetric dialogue: A perspectivist annotation scheme for MapTask. *Proceedings of LREC 2026* , 4988–5001.

Two perspectives on one understanding event

An explainer (EX) teaches an explainee (EE) a board game; afterwards, each labels EE's understanding from the video.

ONE REAL EXCHANGE · TRANSLATED AND SHORTENED

EX At some point you have to say: okay, from now on I'm turning back. And then you can only go back.

EE Okay. Do I lose oxygen doing that too?

EX … If you have treasures, yes. If you have no treasures, no.

EE Yeah. … Not then, okay.

MUNDEX EX_13–EE_c, 03:25–03:37.

ACROSS THE CORPUS

44 of 135 linked annotation pairs disagree on UND vs non-UND.



91 agree · 44 disagree

We keep both labels as separate observations.

Explainer · judges the EE — **PARTIALLY UNDERSTOOD**

"She understood the rule that she uses oxygen, but not quite when it applies."

Explainee · self-report — **UNDERSTOOD**

"I understood that well too, because it was phrased precisely."

Türk, O., Wagner, P., Buschmeier, H., Grimminger, A., Wang, Y., & Lazarov, S. (2023). MUNDEX: A multimodal corpus for the study of the understanding of explanations. *Proceedings of the 1st International Multimodal Communication Symposium*.

Two asymmetric tasks, two kinds of grounding label

HCRC MAPTASK · ENGLISH

Route giving with maps that differ

46
dialogues
31 eye contact · 15 without

5,144
windows
reference + 1.5 s

74%
aligned



aligned 3,807 **pending** 1,261 **misund.** 76

MUNDEX · GERMAN

Explaining a board game to a newcomer

26
interactions
9 explainers

807
windows
annotation ± 2 s

458 · 349
labelled
by explainer · by explainee



UND 360 **partial** 199 **not** 151 **mis.** 97

METHOD · GAZE FEATURES

One gaze vocabulary: partner · task · away

MAPTASK	MUNDEX		SHARED
up*	EX / EE	→	partner
down	TABLE	→	task
off	AWAY	→	away

* literally partner-directed only with eye contact

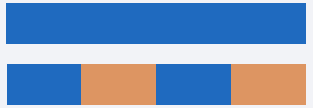
RAW PROPORTIONS · PER WINDOW

Share of time on each target, per person · mutual gaze

STRUCTURED = RAW + …

Entropy: how spread gaze time is

Transitions: how often gaze switches



A entropy 0 · 0 switches

B entropy 1 bit · 3 switches

Two 2-s windows: **task** · **partner**

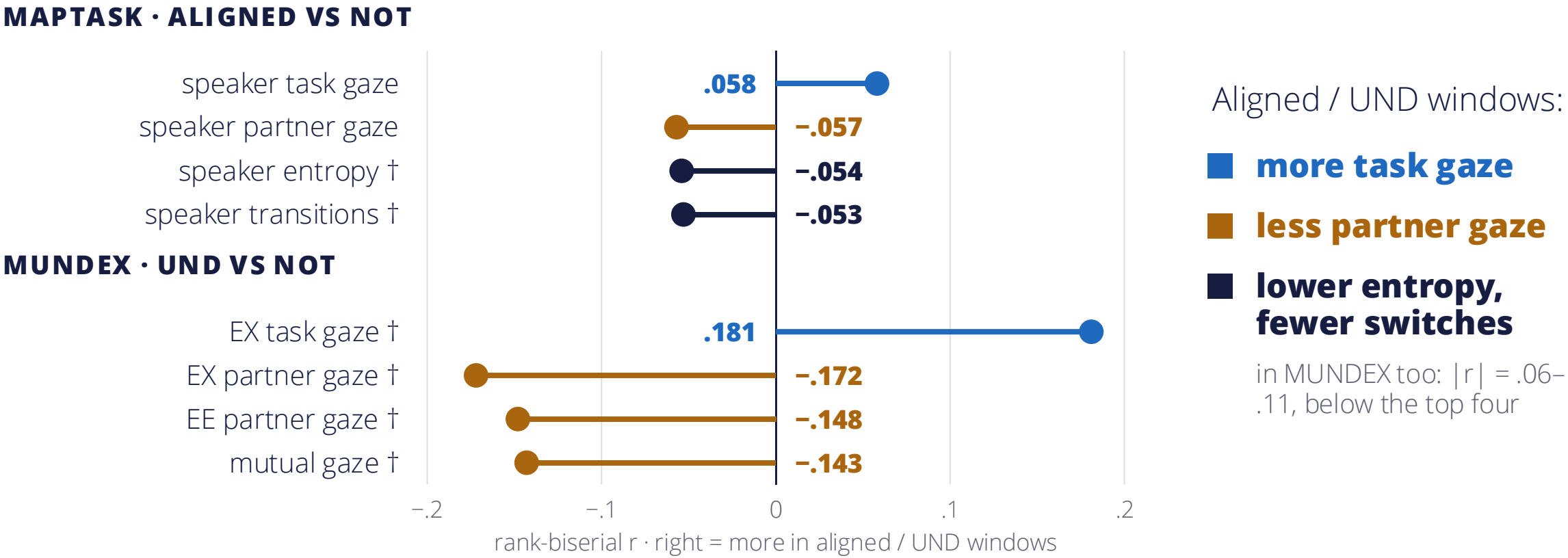
EXTENDED GROUPS · prediction only

temporal (timing) · bigrams (switch direction) · coordination (joint gaze) · ratios

Coded from video, not eye-tracked: the three-target vocabulary can map onto corpora with similar discrete gaze coding.

Q1 ASSOCIATIONS · OVERALL

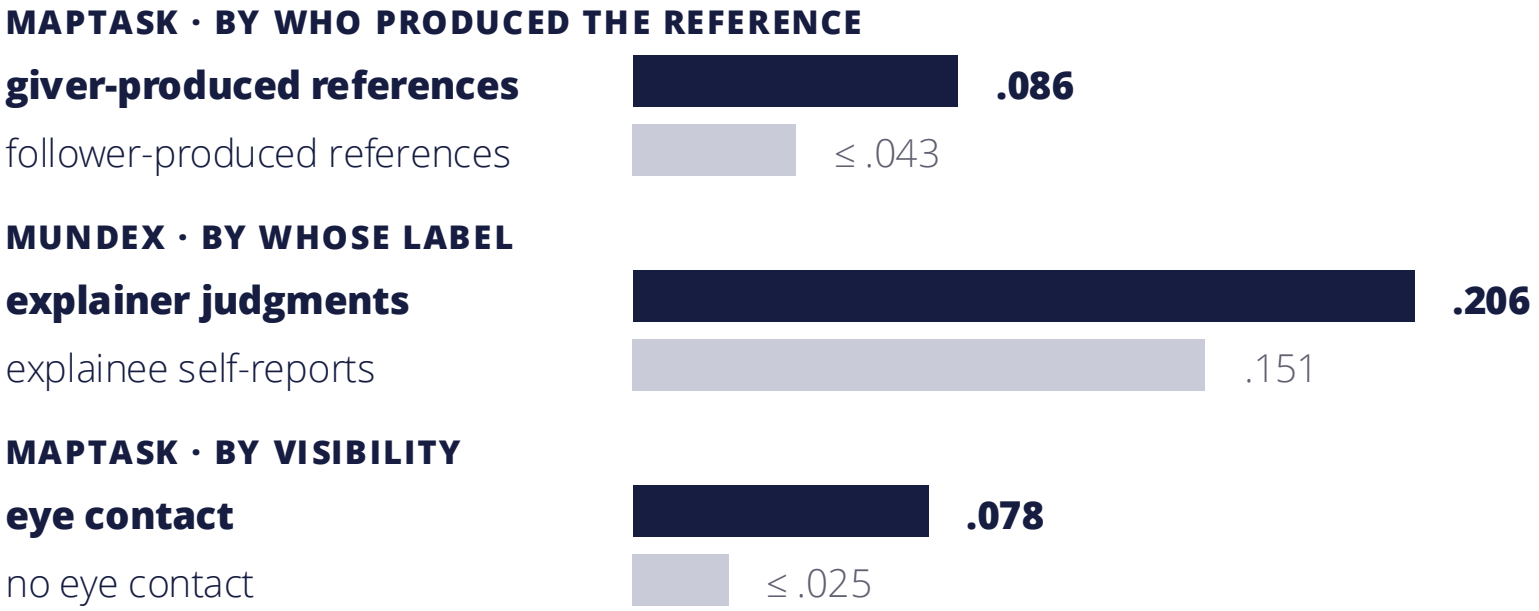
Same direction in both corpora, small effects



Four largest |r| per corpus. All Mann–Whitney p < .001; † also significant under cluster-robust GEE (FDR q < .05).

Q1 ASSOCIATIONS · BY ROLE AND VISIBILITY

Clearest in giver-produced references and explainer judgments



Largest |r| within each stratum.

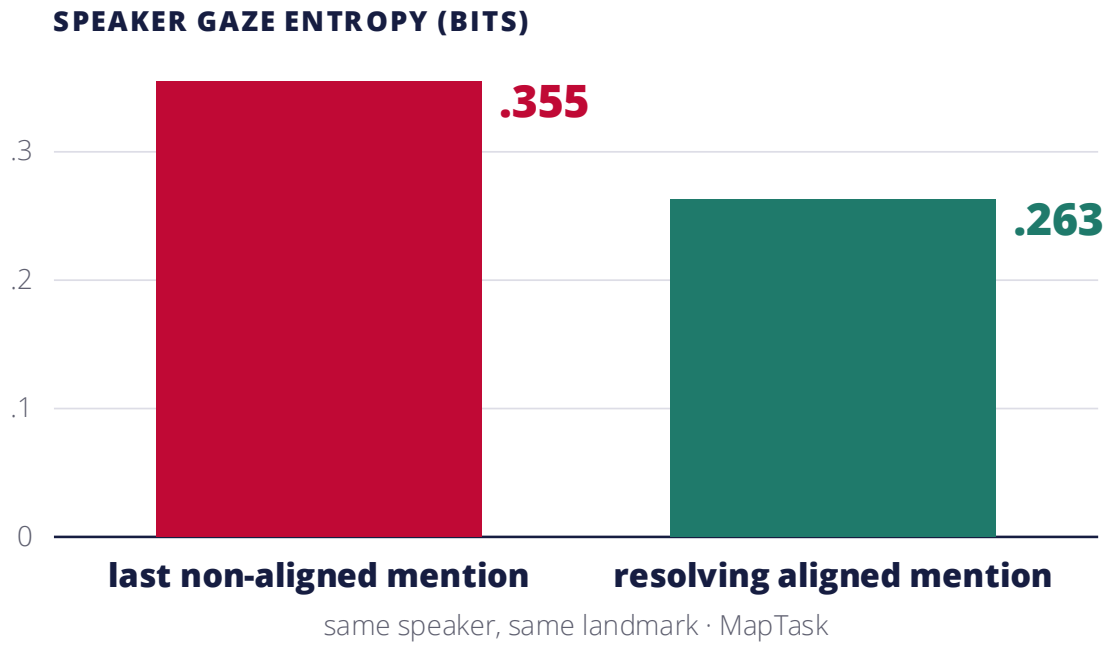
ONE POSSIBLE READING

The participant leading the task : givers produce the instructions being grounded; explainers monitor the explainee.

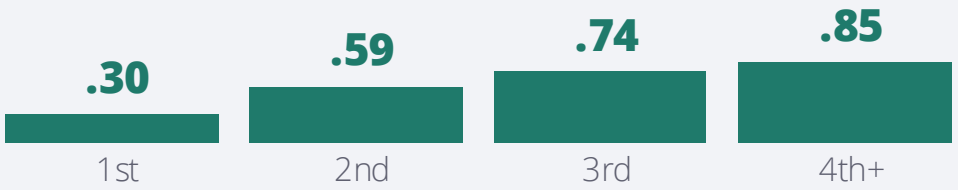
Speakers monitor addressees for understanding (Clark & Krych 2004).

Q2 PROCESS · ALONG REFERENCE CHAINS

When a referent becomes aligned, speaker entropy drops



same speaker, same landmark · MapTask



Repetition and alignment rise together.
We compare the same **speaker and landmark** at the mention that resolves it.

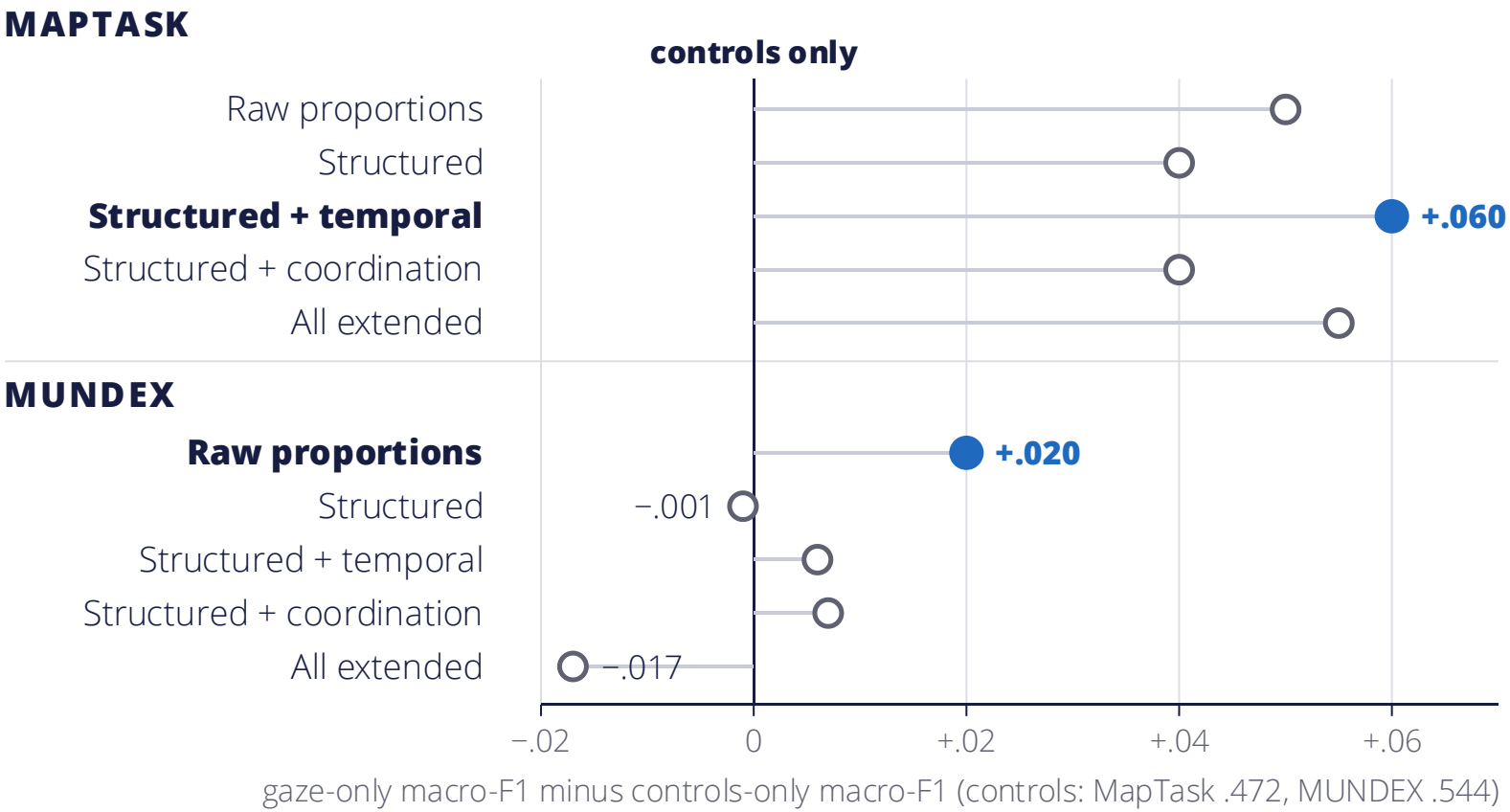
SAME DIRECTION

task gaze ↑ **partner gaze ↓** **transitions ↓**

Once the referent is aligned, the speaker's gaze is less spread out.

Q3 PREDICTION · VS CONTROLS

Gaze-only beats controls-only, modestly; the best features differ by task



- **Best gaze-only model vs controls-only:** +.060 MapTask, +.020 MUNDEX.
- **MapTask:** every gaze set beats controls; timing helps most.
- **MUNDEX:** raw proportions score highest; structured and all extended do not exceed controls.

Logistic regression, macro-F1 on held-out dialogues (MapTask) and explainers (MUNDEX). Controls: task condition and speaker role (MapTask); annotator role (MUNDEX).

CONCLUSIONS

Where people look echoes how they ground

ATTENTION · Q1

Grounded: more task gaze, less partner gaze, lower entropy

→ **continued attention to the joint task**

MONITORING · Q1

Clearest for givers and in explainers' judgments

→ **the leading partner watches for evidence of understanding**

PROCESS · Q2

When a referent becomes aligned, speaker gaze settles

→ **gaze may change as grounding is reached**

FEATURE · Q3

Gaze alone beats controls on held-out dialogues and explainers

→ **a signal to combine with other cues**

Gaze is a small but directionally consistent signal of grounding: the shared pattern spans referential alignment and reported understanding; a feature to model together with language and task context.

PAPER

aclanthology.org/2026.mint-1.18

CODE + DATA

github.com/chnln/gaze-as-grounding-evidence

Thank you

n.li@uu.nl

CONCLUSIONS

Where people look echoes how they ground

ATTENTION · Q1

Grounded: more task gaze, less partner gaze, lower entropy

→ **continued attention to the joint task**

MONITORING · Q1

Clearest for givers and in explainers' judgments

→ **the leading partner watches for evidence of understanding**

PROCESS · Q2

When a referent becomes aligned, speaker gaze settles

→ **gaze may change as grounding is reached**

FEATURE · Q3

Gaze alone beats controls on held-out dialogues and explainers

→ **a signal to combine with other cues**

Gaze is a small but directionally consistent signal of grounding: the shared pattern spans referential alignment and reported understanding; a feature to model together with language and task context.

PAPER

aclanthology.org/2026.mint-1.18

CODE + DATA

github.com/chnln/gaze-as-grounding-evidence

Thank you

n.li@uu.nl