Gaze Behaviour & Conversation Unfolding in the HCRC Map Task Corpus

Anaïs Claire Murat and Carl Vogel
Trinity Centre for Computing and Language Studies
Trinity College Dublin, the University of Dublin
{murata,vogel}@tcd.ie

Abstract

Dialogue interactions have varied internal structure, with flow varying, *inter alia*, in face of both difficulty and agreement. This study investigates eye-gaze in the linguistic progression of interactions. We observe the relation of gaze to illocutionary functions of turns through dialogue acts, and to how turns present "new" or "old" content, through lexical entropy and repetition. Results on the HCRC Map Task corpus, enabled by an event alignment annotation method described, show how gaze is related to linguistic progression. A gaze towards the conversation partner at the end of a turn tends to align with complexity and difficulties being expressed in the turn, while keeping gaze down at the map is more typical of obstacle- and disagreement-free interactions. Addressees who look up or off at the start of a turn show evidence of lexicon adaptation to gaze values.

Keywords: gaze, entropy, repetition, dialogue acts, cross-modal event alignment, interaction analysis

1. Introduction

Most Human-Human interactions can rely on at least two modalities: the auditory and the visual ones. These modalities complement each other, and within the visual modality, eye gaze has been shown to be a catalyst for smoother interactions. Gaze has been explored as a tool for turn-taking and for signalling attention or emotional states – multiple functions are evident. This paper examines eye-gaze behaviour in dyadic conversation in relation to the overall purpose of turns through dialogue acts, and abstractions of their content; through computed entropy and repetition.To do so, we explore the HCRC Map Task corpus and, more specifically, the gaze behaviour of both participants at the start and end of turns, to statistically assess the co-presence of gaze and turn qualities. We are interested in whether the choice of dialogue act and two measures of linguistic unfolding – turn by turn repetition and categories entropy behaviour (i.e. increase, decrease, or stable states) interact with gaze behaviour. Fundamental to this analysis is the temporal alignment of the distinct modalities within a representation framework that supports counting and other forms of measurement of events within each modality and their overlap; in support of this we use an annotation method called Beginning-Middle-End (BME), for the facets of events noted. We use observational methods to develop generalisations over statistically significant interactions in categorial representations of gaze types and in sec. 5 discuss their coherence, both in this research work and with respect to the current literature.

2. Background Literature

Gaze has been studied as a paralinguistic tool. It has been noted in turn taking management (Bavelas et al., 2002). Initially, it can signal communicative intent (Cary, 1978; Pfeiffer et al., 2013) and resolve cases of simultaneous starts (Zima et al., 2019). Although Kendon (1967) and Ho et al. (2015) reported that speakers avert their gaze briefly after turn-taking and look back at the listener at the turn end, Beattie (1978) reported finding more averted gaze at turn ends, and that turns containing gaze up at their end were not correlated with prompt answers. In fact, Degutyte and Astell (2021) reviewed 29 papers from the past 50 years in the gaze literature and found no single pattern, as had been believed since Kendon (1967). Additionally, there seem to be influential factors beyond the mere difference in dialogue nature (free flowing conversations or question/answers (Rossano, 2010; Beattie, 1978)) or study design (Kendon, 1978).

Gaze may also provide insights into the emotional state of participants. A direct gaze at the conversational partner may signal attention, engagement (Bavelas et al., 2002; Kendon, 1967; Goodwin, 1980; Argyle and Cook, 1976). Gaze favours success in task-based conversations by supposedly preventing and repairing conversation breakdowns (Boyle et al., 1994; Streeck, 2014; Nicholson et al., 2005). In map task corpora, Boyle et al. (1994) notably showed that direct gaze were more frequent in instances of difficulty (i.e., mismatch between two maps) than otherwise, Doherty-Sneddon (1995) that one of the gaze function was to access feedback, and Nicholson et al. (2005) that instruction givers were more likely to look up at the follower

Proceedings of the 22nd Joint ACL - ISO Workshop on Interoperable Semantic Annotation and Representation (ISA-22), pages 99–110
12 May 2026.

after the latter expressed difficulties. On the other hand, an averted gaze could help identify problematic interruptions by the listener (Brône et al., 2017; Spaniol et al., 2023) and signal unwillingness to pursue an interaction in the current terms (Jokinen et al., 2010).

Additionally, gaze has been scrutinised in relation to the content of the conversation. The deictic function of referential gaze has been used to disambiguate and ground text (Mehlmann et al., 2014), and has been shown to be more efficient than speech for solving referring expressions or searching tasks (Brennan et al., 2008; Neider et al., 2010; Hanna et al., 2020). Directed gaze has been shown to mark questions in Italian (Rossano, 2010), and to indicate preferred answers in polar questions (Kendrick and Holler, 2017). Information may also be encoded in gaze. Mutual gaze windows were shown to accompany minimal responses (such as back-channels or continuers) (Bavelas et al., 2002), and Edlund et al. (2012) showed that, indeed, 3^{rd} party's gaze of a conversation is likely to reveal which one of the speakers is providing the most information.

In relation to entropy, it has been proposed by Genzel and Charniak (2002) that entropy measures ignoring context should increase for sentences in relation to their ordinal position in a text, as a consequence of an entropy constancy rate principle. Evidence for this effect has been provided using gaze during reading as an indicator of cognitive effort associated with higher entropy but not the position of a sentence within a text (reading of sentences does not ignore contexts) (Keller, 2004). Using gaze events as tokens and gaze categories as types, the entropy rate constancy principle has also been demonstrated for listener gaze behaviours in interactive dialogue (Wang et al., 2024). Our work relates specific gaze behaviours of speakers and listeners to categories of lexical entropy change, local repetition, and dialogue acts that interact with cognitive effort (e.g. Doubt vs Explanation).

3. Method

3.1. Description of the HCRC Corpus

This research uses the HCRC Map Task Corpus (Thompson et al., 1993). It is made of dyadic conversations whose task involves closing an information gap between the interlocutors: one designated party is to linguistically indicate a path to follow (the Giver "G") and the dialogue partner is to reproduce that path (the Follower "F"), both with reference to imperfectly matching maps. The entire corpus includes 128 dialogues; half of the corpus had participants having access to mutual eye contact, in the other half, eye contact was obscured with a screen. 46 dialogues had gaze annotations but we focus in this paper on the 31 annotated conversations which allowed for direct eye-contact.

We reused the timestamped annotations as provided in the XML version of the dataset. [1] The gaze tier is continuously annotated and annotations can take one of the three following values: "gaze down" when looking at the map, "gaze up" when looking at the other participant, "gaze off" when looking elsewhere. We examine tuples of co-occurring G and F gazes, also referred as "paired gaze" in what follows.

Similarly, we reused the turns annotations as available in the dataset. Due to the hierarchical organisation of the HCRC annotations, the timestamps for the turn transcripts were retrieved from the timed-units (i.e. words): matching the onset of the first word of a turn, and the offset of the last.

The corpus also includes dialogue act (DA) augmentation for every turn.

As a matter of indication, the HCRC corpus includes 64 talkers (half male, half female), mostly from Glasgow, Scotland; half of the conversations included acquaintances, and each participant participated in 4 conversations, alternating between the role of Giver and Follower.

3.2. Data selection

We used the 31 conversations which were annotated with gaze values and which allowed for direct eye-contact between the participants (in total, 10894 paired gazes, 5943 turns). Some gaze annotation anomalies were spotted. For instance, we noticed several gaze values for a single millisecond, and some starting times occurring after the endtimes. For privacy reasons, the video recordings of the corpus are not available outside the HCRC community, we therefore could not resolve such anomalies and had to remove these gazes from the dataset, along with any turn and other gaze that could have been impacted. The final dataset comprised 10761 paired gaze and 5889 turns.

3.3. BME Modality Alignment

The BME method is an alignment technique which focuses on events (Beginnings, Middles, and Endings of annotations) to reveal the structure of multimodal datasets (Murat et al., 2026). The method is similar to that used by other researchers investigating multi-modal dialogue (Magnusson, 2000; Hunyadi et al., 2018; Magnusson, 2020). Assuming each modality is captured on a different tier, such as what annotation software like Elan (Wittenburg

[1] Data available https://groups.inf.ed.ac.uk/MapTask/index.html, last accessed in March 2026.

Turns	Gaze
	B - Gaze 1
B - Turn 1	M - Gaze 1
M - Turn 1	E - Gaze 1
M - Turn 1	B - Gaze 2
E - Turn 1	M - Gaze 2
	E- Gaze 2

Table 1: Example of alignment of 1 Turn and 2 gaze annotations using the BME method

et al., 2006) produces, it arranges multi-modal data based on the relative onsets and offsets of the tiers' annotations. It retrieves the beginning time (B) and ending time (E) of each annotation and organises them relatively to each other. Middles (M) mark, for a given annotation, the B and E lines of another tier occurring in-between its onset and offset. Ms can then be enumerated as a measure of duration to account for how many events happened during a given annotation: that is, event durations and overlaps may still be reckoned in terms of absolute time elapsed, or by the count of intervening events. Table 1 provides an example of how three overlapping annotations could be organised. This fictive example reads "Gaze 1 started outside a turn, Turn 1 started inside Gaze 1, Gaze 1 ended inside Turn 1, Gaze 2 started inside Turn 1, Turn 1 ended inside Gaze 2, Gaze 2 ended outside a turn." It can also be read as "1 turn event (Middle) occurred in Gaze 1, 2 gaze events occurred in Turn 1, and 1 Turn event occurred in Gaze 2".

The annotation alignment technique supports analysis event type interactions via the counts of the sorts of things that happen in one tier at the same time as starts or stops in another tier, such as used in contingency tables.

Timestamps of interest are those of turns and gaze annotations, as described in section 3.1.

3.4. Illocutionary function through Dialogue Acts

To generalise dialogue act (DA) behaviour during conversations and to have sufficient observations to conduct meaningful statistical tests (i.e., using χ^2-tests, limiting the number of expected values lower than 5), we classify basic DAs into supra-categories INSTRUCTION, CONFIRMATION, EXPLANATION, DOUBT, and REJECTION (see Table 2). INSTRUCTION is only representative of the Giver's utterances as it represent instances where the Giver commands the Follower to carry out an action (e.g., take a certain route, move to a certain landmark, etc.), CONFIRMATION corresponds to instances where one validates the position of the other, EXPLANATION where one develops informative answers, DOUBT where one expresses the need to confront their understanding or position, and REJECTION when one answers the other by the negative. Further description of the individual dialogue acts can be found in the manual Carletta et al. (1996).

3.5. Characterizing the Lexical Development of a Conversation

In order to confront the gaze behaviour to the unfolding of the conversation, we need to find ways to measure lexical choice over time. Similarly to the method (Murat and Vogel, 2025) used to track engagement throughout conversations via linguistic cues, we study word use on two scales: firstly, Lexical entropy from the start to the end of a conversation, measured at turn boundaries; secondly, Linguistic repetitions from one turn to another.

3.5.1. Transcript Processing

In a pre-processing step the HCRC transcripts were normalised: we wrote a Python script to remove punctuation, lowercase letters, split "pronoun + verb" contractions (e.g "you're" $\Rightarrow$ "you", "are"), and split contracted negated auxiliaries (e.g "can't" $\Rightarrow$ "can", "not"). Additionally, using the Spacy small model for English (Montani et al., 2023), we noted for each lexical token its part of speech category and, from it, determined whether the token belonged to open-class (OC) or closed-class (CC) words (as exhaustive and mutually exclusive possibilities; i.e., if their part-of-speech tag was either "NOUN", "VERB", "ADJ" (adjective), or "PROPN" (proper noun) they were considered as open-class; and closed-class otherwise). We distinguish thusly because, by their frequency, closed-class items are more likely to be repeated than open-class items.

3.5.2. Entropy

Initially adapted from Gnjatović et al. (2018), we reuse the method developed in Murat and Vogel (2025) to measure entropy evolution in sequential lexical choice. For a dialogue D and lexical types (v), $V = \langle v_1, v_2, ..., v_k \rangle$ is the *sequence* of lexical items used. D is defined as a sequence of turns T; $D = T_1, T_2, ..., T_n$ (abbreviated T_1^n) which are *sets* T of lexical types produced in the turn. The (cumulative) entropy H of a dialogue from its opening (1) to its nth turn (T_1^n) is based on Shannon Entropy:

$$H(T_1^m) = -\sum_{1}^{k} P(v_i) \log_e P(v_i) \quad (1)$$

The probability $P(v_i)$ is estimated from the relative frequency of the type v_i in the dialogue sequence from the start to the point of measurement.

CONFIRMATION (10888)	DOUBT (6437)	EXPLANATION (4267)	INSTRUCTION (4267)	REJECTION (884)
Acknowledge (5599) Reply_y (3229) Ready (2060)	Check (2133) Align (1778) Query_yn (1758) Query_w (768)	Clarify (1192) Explain (2160) Reply_w (915)	Instruct (4267)	Reply_n (884)

Table 2: Summarised DA into the supra-categories CONFIRMATION, DOUBT, EXPLANATION, INSTRUCTION, and REJECTION. In brackets are the numbers of turns of each kind.

These calculations can be evaluated on a sequential basis and assessed either word after word, or turn after turn. It is relevant to examine the difference between the entropy at the end of any turn h_n and at the end of the immediately prior turn h_{n-1}:

$$\Delta H = \Delta(h_n - h_{n-1}) \quad (2)$$

Eq. 2 thus allows to tell that a certain turn led to an increase or a decrease in entropy. An increase is typical of scenarios adding up new or relatively rare words, while a decrease in entropy suggests that the turn reuses well-known and frequent words. In this work, we look into entropy calculated for the overall conversation vocabulary (H_C), but also at the speaker's level by only accounting for the vocabulary *they* have produced (H_S), and investigate the impact of a turn on both (respectively ΔH_C , and ΔH_S). For this study, we categorise turns (exhaustively and mutually exclusively) thusly :

- $\Delta H_C < 0$, $\Delta H_S < 0$: the turn leads to a decrease in both conversational (ΔH_C) and speaker entropy (ΔH_S).
- $\Delta H_S < 0 < \Delta H_C$: the turn only leads to a decrease in speaker entropy.
- $\Delta H_C < 0 < \Delta H_S$: the turn only leads to a decrease in conversational entropy.
- $\Delta H_C = \Delta H_S$: the turn leads to a similar increase in conversational and speaker entropy.
- $\Delta H_S > \Delta H_C > 0$: the turn leads to an increase in both conversational and speaker entropy, greater increase for the speaker.
- $\Delta H_C > \Delta H_S > 0$: the turn leads to an increase in both conversational and speaker entropy, greater increase for the conversation.

In a nutshell, this measure of entropy allows us to track the overall reuse of tokens in a dialogue. Closed-class words tend to be often used and have a negative impact on entropy; they add little information. In contrast, novel content words lead to entropy increase, and their impact on entropy diminishes as novelty fades. To address conversation dynamics that related to closer events, such as short-term planning and spontaneous reactions, we add another dimension, whose window is smaller: direct repetitions from one turn to another.

3.5.3. Repetitions

There is not one way to quantify repetition in the literature. Distinctions occur in the nature of what is being counted (lexical vs syntactic repetition (Reitter and Moore, 2007; Healey et al., 2014), bag of words vs set of words (Murat et al., 2022), lemma vs tokens (Reverdy and Vogel, 2017), different size of n-grams, etc.), but also in terms of calculations where there can be strict counts as well as length- or frequency-based normalisations (Mekhaldi, 2006). Furthermore, assessing the overall similarities of all interlocutors' productions is not enough: it does not allow one to draw any conclusion on the impact of what one previously said on the current production (Mehler et al., 2011). Temporal cues are therefore crucial and we do so by comparing strictly ordered turns (Reverdy et al., 2020; Murat et al., 2022).

Given the nature of this work, where repetition is an object of study in its relation to gaze, we qualify repetition by the strict presence or absence of a repeated token between the current turn and the preceding one, while also relying on sub-classes of repetition. We make a distinction between, on the one hand, self-repetitions (SR), the speaker is repeating content from their last turn, which might signal grounding or involvement; and on the other hand, other-repetitions (OR), when the speaker is repeating what the addressee said in their last turn, which can be an example of discourse planning or floor-holding (Koutsombogera and Vogel, 2019). We also track as a separate category turns accounting for both self- and other-repetitions (OR-SR) as they might signal even further efforts to co-build through the conversation. As closed-class words are more likely to be repeated than content words, we also made a distinction between repetitions which only counted closed-class words (CC) and turns which also accounted for open-class repetition (OC).

4. Results

The data mining was conducted through identification of statistically significant co-occuring behaviours through the use of χ^2-tests. As such, 12 χ^2-tests were performed (F or G's turns $\times$ DA, entropy or repetitions $\times$ at the start or at the end of a turn). We test cross-categorial interactions only

Table 3: Abbreviation key

Short form	Concept
CC	tokens in closed-class categories
OC	tokens in open-class categories
OR	other-repetition
SR	self-repetition
G	information giver
F	information follower
H_C	entropy for conversation
H_S	entropy for speaker
Δ	change

where less than 20% of expected values were less than 5 (the maximum we have is 13% of cells < 5 for the entropy test in relation to F's turns), and none was less than 1. To palliate this exigence, we had to merge all pairs of gaze including at least one gaze off into one single category (may they be from the Follower or the Giver).[2] Furthermore, considering the number of tests and the relatively high number of categories, we adjusted the p-value of 0.05 by the number of dimensions d ($\alpha = 0.05/d$) in each table for determining significance. Similarly, the critical value z for residual was calculated $P(Z > z) = \frac{\alpha}{2}$. For the sake of space, we do not report non-significant tests. The 8 tables across fig. 4, 5 & 6 report the 8 χ^2-tests showing significant interactions in these conservative conditions.

4.1. Dialogue Acts & Gaze

Exploring dialogue acts enables assessment of whether the evident purpose of an utterance is reflected in the gaze behaviour of interlocutors. The INSTRUCTION category was not included in tests about Follower's turns as, by definition, Followers had little chance to utter such turns.

All 4 χ^2-tests (DA $\times$ Gaze-pairings for $\{F, G\} \times$ {Start, End}; Table 4) showed significant interactions: Giver and Follower's paired gazes at the start and at the end of each turn interact with the DA being performed. When G is speaking, INSTRUCTIONS are very likely to start while both F and G are looking down ($r = 7.82$) and unlikely in any other context ($r < -3.48$). The end of an INSTRUCTION turn, however, is not typical of one gaze category, but has significant dearth where G is looking down and F, up ($r = -4.94$). An expression of DOUBT by G is rather expected when G is looking up but F is looking down ($r = 8.80$) and rather not expected if G is looking down ($r < -3.46$). F's DOUBT turns, however, can also appear when both are looking down ($r = 3.75$), on top of not being likely to occur when F is looking down but G is looking up ($r = -6.49$). For both F and G, the end of a DOUBT is likely occurring in the situation where the non-speaking participant is now looking down and the other looking up ($r > 10.47$). EXPLANATION turns are mostly significantly characterised by their ending behaviour: the speaker tends to look up ($r = 1.24$ for G when both look up, but $r > 3.11$ in all other cases). For CONFIRMATION turns – which intervene in response to the other participant's content, they tend to start in a context where the addressee is looking up but the speaker is already looking down (particularly significant for G's case $r = 6.03$), and they end in a context where the speaker is looking down and the other participant is looking either down or up ($r > 3.86$). On the other hand, REJECTION turns have similar behaviour, to the exception that in G's case, they might both be looking up (hence, making eye contact) at the start ($r = 4.42$).

These results suggest that both participants looking down at the object of interest (the map) is a sign of conversation progressing without obstacles: it is the context in which further INSTRUCTIONS are likely provided ($r = 7.82$) or CONFIRMATION turns can end ($r > 3.86$). F's DOUBT turns are also likely to start in that category ($r = 3.75$). This distinction might come from the need for F to interrupt and express doubts as they lose track of G's elaborations.[3]

Gazes up seem to sometimes signal conversation flow disruption. DOUBT turns end with their speaker gazing up at the partner, and the potential responses, a CONFIRMATION or a REJECTION still start during a look up, while the responder looks down at the map. At the end of such turns, CONFIRMATIONS can lead the non-speaking person to look back down at the map ($r > 3.8$) or not ($r > 5.94$) while REJECTIONS, which are more divisive answers, only favour the non-speaking participant to look up ($r > 6.09$). EXPLANATIONS are also characterised by the speaker looking up to the other participant at the end, possibly checking for feedback ($r > 3.11$).

This section has focused on dialogue acts, i.e, the apparent intentions with every turn start or stop, in relation to the gaze behaviours of the interlocutors. Shared gaze behaviours (both up or both down) distribute differently across DAs than asymmetric behaviours, varying by whether the turn owner is the information Follower or Giver. The next two sections examine gaze in relation to lexical novelty.

4.2. Entropy & Gaze

Entropy allows quantification of the novelty of vocabulary of one turn over the conversation: are the words that have just been uttered relatively new to

[2] In practice, 332 of the gaze-offs at the start of a turn were due to F looking off; against only 73 due to G. At turn ends, these numbers are respectively 327 and 71.

[3] A Pearson's χ^2 test revealed that, indeed, F's DOUBT turns are more likely than G's to occur in the middle of the other participant's turns ($p = 3.775e^{-12}, r = 6.98$).

Start	G's Turns					F's Turns				
G's Gaze	Down		Up		Off	Down		Up		Off
F's Gaze	Down	Up	Down	Up		Down	Up	Down	Up	
Instruction	**7.82**	**-5.52**	**-4.93**	**-3.48**	0.74					
Doubt	**-3.46**	**-4.07**	**8.80**	1.15	0.90	**3.75**	3.05	**-6.49**	1.31	0.68
Explanation	-0.50	1.04	-0.87	-0.50	0.84	-0.49	-1.51	2.08	-1.30	-0.56
Confirmation	-2.52	**6.03**	-2.10	1.42	-1.76	-1.70	-1.69	2.75	-0.63	0.29
Rejection	**-5.38**	**6.90**	-0.22	**4.42**	-1.68	**-3.67**	0.03	**4.45**	1.04	-1.35

(a) **(b)**

End	G's Turns					F's Turns				
G's Gaze	Down		Up		Off	Down		Up		Off
F's Gaze	Down	Up	Down	Up		Down	Up	Down	Up	
Instruction	2.96	**-4.94**	0.03	-0.58	0.31					
Doubt	**-7.08**	**-4.75**	**10.47**	0.99	1.04	-2.27	**10.64**	**-7.19**	2.71	0.46
Explanation	**-3.39**	-1.41	**3.70**	1.24	1.08	-0.98	**3.11**	-2.24	**3.21**	-0.89
Confirmation	**7.45**	**7.67**	**-12.23**	-1.61	-1.69	**3.86**	**-11.35**	**5.94**	**-4.30**	0.55
Rejection	-2.53	**8.77**	-2.31	0.45	-1.63	-2.95	-1.62	**6.09**	-1.38	-0.88

(c) **(d)**

Table 4: Standard residuals of 4 significant χ^2 testing the likelihood of a certain dialogue act given the gaze values of the Follower (F) and the Giver (G) at the start of G's (sub-table **a**, $p < 2.2e^{-16}$) or F's turns (sub-table **b**, $p = 1.396e^{-09}$) and the end of G's (sub-table **c**, $p =< 2.2e^{-16}$) and F's turns (sub-table **d**, $p =< 2.2e^{-16}$). In **bold** are significant residuals.

End	G's Turns					F's Turns				
G's Gaze	Down		Up		Off	Down		Up		Off
F's Gaze	Down	Up	Down	Up		Down	Up	Down	Up	
$\Delta H_S > \Delta H_C > 0$	**-5.13**	2.42	**4.03**	0.64	0.01	**-3.81**	**4.96**	-0.61	**3.01**	-0.16
$\Delta H_C < 0 < \Delta H_S$	0.16	0.67	-1.46	0.75	0.74	-0.68	1.83	-1.53	0.12	1.37
$\Delta H_C = \Delta H_S$	-0.51	-1.58	2.42	1.57	-2.42	1.07	-0.65	-0.68	-0.70	0.36
$\Delta H_S < 0 < \Delta H_C$	1.60	-0.57	-1.77	1.60	-0.74	1.20	-2.22	2.33	-2.66	-1.18
$\Delta H_C > \Delta H_S > 0$	2.88	0.32	-2.64	-1.39	-0.50	1.02	-1.81	0.21	-0.85	0.60
$\Delta H_C < 0, \Delta H_S < 0$	**4.27**	-1.50	**-4.11**	-2.45	2.12	**3.16**	**-4.33**	0.30	-1.26	-0.22

(a) **(b)**

Table 5: Standard residuals of 2 significant χ^2-tests testing the likelihood of a certain entropy ratios given the gaze values of the Giver (G) and the Follower (F) at the end of G's (sub-table **a**, $p = 5.62e^{-08}$) or F's turns (sub-table **b**, $p = 2.48e^{-06}$). In **bold** are significant residuals.

the context (increase in entropy), or have they already been extensively used (decrease in entropy)? Similarly to the DA results, we conducted four (turn edge $\times$ turn owner) χ^2-tests. Unlike the DA results, only the ones matching the turns with the gaze value at their end were shown significant (Table 5).

Turns which had the speaker's entropy increase greater than the conversation's entropy ($\Delta H_S > \Delta H_C > 0$) tend to end in a context where the speaker looks up and the addressee looks down ($r > 4.03$). In the case of F's turns, it can even result in a mutual gaze ($r = 3.01$). In any case, it is very unlikely to result in both gazes converging down to the map ($r \leq -3.81$).

On the opposite, both entropy decreasing ($\Delta H_C < 0$, $\Delta H_S < 0$) are rather positively correlated to turns ending in a context where both participants look down ($r \geq 3.16$), and negatively correlated to the speaker looking up while the other participant looks down ($r \leq -4.11$).

These results show only partial alignment with observations made about DA. Only gaze values at turn ends revealed significant interactions with entropy categories: entropy behaviour types are more equally distributed across gaze types at turn starts. However, turns with specific entropy leading to significantly different gaze behaviour at the end of a turn is a good indicator that lexical novelty can

be assessed through gaze. $\Delta H_S > \Delta H_C > 0$ corresponds to turns where the vocabulary is relatively more novel to the speaker than to the conversation (e.g. the speaker reusing words introduced by their interlocutor). It is, similarly to DOUBT turns, a potential marker of hesitation and need for feedback on the present utterance. The opposite $\Delta H_C < 0$, $\Delta H_S < 0$ turns, where the utterance does not increase entropy for the speaker nor the conversation, behave similar to CONFIRMATION turns: both look down at the map, likely not inviting a break in flow.

4.3. Repetitions & Gaze

The next results address the most tangible aspect of lexical spread: direct repetitions (other and/or self) as described in section 3.5.3. We again realised a set of four χ^2-tests, and Table 6 shows the standard residuals of the two tests which were significant. Amongst these tests are the gaze values at the start of the G's turns (sub-table **a**) and the gaze values at the end of F's turns (sub-table **b**).

At the Start of the Giver's Turns: G is more likely to repeat F when G is looking at the map while F is looking at G ($r = 3.04$) but also if either one of them is looking off ($r = 3.05$) – in practice, 75% of the gazes away in the "off" bin are produced by F, it might hence be a strategy for G to get F's attention back when they are not looking at the map. However, even if this tendency is also significant when running a χ^2-test over a contingency table differentiating between when open-class tokens are being repeated or only closed-class ones (Table 7), none of the residuals returned an amplitude greater than the significance threshold ($|r| \geq 3.19$). Thus, it cannot be established whether OC or CC repeated content played a greater role than the other.

At the End of the Follower's Turns: On the other hand, F's turns are significantly correlated with the gaze values present at the end of their turns (Table 6). When F does not repeat anything, they are more likely to end on a look down at the map, while G looks up ($r = 4.65$). G looking down while F looks up, on the contrary, is unlikely for such turns ($r = -4.21$). Yet, the perfect opposite is likely for F's turns which include both self and other repetitions: the Follower then looks up and the Giver down ($r = 3.62$), and the reversed behaviour is unlikely ($r = -3.45$). Differentiating between OC and CC content (Table 7b), it seems that such a behaviour holds mostly for OC repetitions ($r = 3.22$), and is also visible for self-repetition only ($r = 3.69$).

These results align with previous results. F's repeating both G and themselves results in F gazing up while G looks down at the map ($r > 3.22$), similarly to EXPLAIN or DOUBT turns, as if they were looking for feedback. When F does not repeat any, however, and likely moves forwards, they tend to look down only while the Giver is likely looking at them ($r = 4.65$), similarly to CONFIRMATION or REJECTION turns. The gaze behaviour correlated with the start of the Giver's turns is however singular enough. It is the only one which significantly relates with off-gazes ($r = 3.05$) along with F's gaze up when G is focusing on the map ($r = 3.04$). Other-repetition might then be used to shape the new turn in a way that addresses the needs of the Follower: either to solve their hesitation, as looks up suggest, or maybe return attention to the map.

5. Discussion

Our findings suggest that paired gaze behaviours meaningfully interact with distinct categorisations of dialogue turn contents. If we consider the types of paired gazes that were significantly correlated with the turn qualities, we observe a disparity between symmetrical and asymmetrical gaze. Similarly to a finding of Sbranna et al. (2025), our results suggest that a gaze does not have to be mutual to impact the conversation. In fact, mutual gaze was rarely significantly correlated with a certain type of turn, as opposed to one-sided gazes to the other participants (5 out of 40 vs 28 out of 80; $\chi^2 = 6.7$, $p = 0.009$). As they did, we note that this distinction may be an artefact of the conversation being *task*-based and with a visual competitor (the map).

We investigated the communicative intent of gaze by matching it with dialogue acts.[4] Foundational work (Kendon, 1967; Ho et al., 2015) suggested that, at the end of turns, speakers look up at the addressees to check their availability to speak. While our research does not oppose this claim, and arguably will, in this discussion, agree that looks up are inviting a response, not all turns end with a look up in our dataset. Rather, we have to call for a more qualified answer where gaze behaviour at the end of a turn is dependent on the intent of the participants. Bavelas et al. (2002) found most backchannelling to be happening during "gaze windows" (i.e. mutual gaze instances). While our data does not suggest that CONFIRMATION turns – nor $\Delta H_C < 0$, $\Delta H_S < 0$ which tend to host such turns – start in such windows, the Giver's CONFIRMATION turns were indeed more likely when the Follower was looking at them. The corresponding tendency for the Follower's turns was, however, not found significant. Kendrick and Holler (2017) found that dis-preferred answers were associated with averted gaze, while preferred answers were associated with looks at the conversational partner. Only based on the contrast between CONFIRMATION and REJECTION turns, our observation does not match either as both DAs seemed to display similar behaviours: a tendency for the answerer to look down while

[4] In work under review we study sequential measurements of entropy and repetition in interaction with DAs.

	Start					End				
	G's Turns					F's Turns				
G's Gaze	Down		Up		Off	Down		Up		Off
F's Gaze	Down	Up	Down	Up		Down	Up	Down	Up	
OR	**-3.65**	**3.04**	-0.42	0.05	**3.05**	0.69	0.85	-1.69	1.26	-0.69
SR	2.44	-2.65	1.49	-2.41	-1.17	-0.60	1.76	-1.52	-0.55	1.81
OR-SR	-1.26	1.69	-0.73	0.44	0.62	-0.82	**3.62**	**-3.45**	1.01	1.78
ø	1.54	-1.20	-0.43	1.70	-1.72	0.34	**-4.21**	**4.65**	-1.32	-1.79
	(a)					**(b)**				

Table 6: Standard residuals for 2 significant χ^2-tests testing the likelihood of a certain repetition type given the gaze values of the Follower (F) and the Giver (G) at the start of G's turns (sub-table **a**, $p = 1.37e^{-04}$) and end of F's turns (sub-table **b**, $p = 8.51e^{-07}$). In **bold** are significant residuals.

		Start					End				
		G's Turns					F's Turns				
G's Gaze		Down		Up		Off	Down		Up		Off
F's Gaze		Down	Up	Down	Up		Down	Up	Down	Up	
OR	CC	-2.80	1.60	1.22	-1.52	2.48	-0.25	0.07	-0.76	1.75	0.25
	OC	-2.08	2.66	-2.18	1.98	1.57	1.28	1.13	-1.53	-0.22	-1.27
SR	CC	2.04	-3.13	2.20	-1.71	-1.07	0.52	0.02	-1.21	-0.69	1.42
	OC	1.09	-0.04	-0.54	-1.48	-0.42	-2.28	**3.69**	-0.89	0.17	1.09
OR-SR	CC	-0.80	-0.85	2.02	0.49	-0.03	-0.54	1.64	-1.71	0.71	1.17
	OC	-0.91	2.76	-2.51	0.15	0.78	-0.58	**3.22**	-2.95	0.68	1.27
ø	ø	1.54	-1.20	-0.43	1.70	-1.72	0.34	**-4.21**	**4.65**	-1.32	-1.79
		(a)					**(b)**				

Table 7: Standard residuals for 2 significant χ^2-tests testing the likelihood of a certain repetition type, differentiating between open-class (OC) and closed-class (CC) tokens, given the gaze values of the Follower (F) and the Giver (G) at the start of G's turns (sub-table **a**, $p = 3.82e^{-06}$) and end of G's turns (sub-table **b**, $p = 4.71e^{-06}$). In **bold** are significant residuals.

the addressee looks up. However, our findings match studies which take into account interaction difficulty (Boyle et al., 1994; Nicholson et al., 2005). For instance, Beattie (1978) reported that "Questions terminating with gaze were judged to be more difficult than questions without gaze" (abstract). Although we do not directly work with questions, this notion of difficulty seems to extend to our analyses. DOUBT and EXPLANATION, corresponding to instances in which the participants had to discuss and negotiate meaning, as opposed to the conversation flowing with no obstacle, but also F's turns containing self-repetitions of open-class content – suggesting more planning, are indeed correlated with looks at the addressee at their end.

Beyond abstract intentions observable through dialogue acts, we also find a correlation between the way turns are shaped and gaze behaviour. Sentences which lead to a decrease of both the speaker and conversation entropy tend not to divert gazes from the object of attention – the map – while those with particularly greater speaker's than conversation entropy – that is, those where the speaker uses vocabulary that is newer to them than to the conversation – are correlated with the speaker gazing up at the end. Furthermore, the study of repetition and gaze, shows that the Giver tends to spontaneously – understand from one turn to another – re-use lexical items from the Follower when starting while the latter is looking up. This might be evidence that the shape of a turn is also adapted depending on the gaze of the addressee. Differences in gaze behaviours based on the participant's role were also noted. For instance, DOUBT turns uttered by the Follower can start while both are looking down at the map, while the Giver's DOUBT turns are unlikely to occur in such a context. This can also be explained by the higher likelihood for F's DOUBT to be expressed during G's turns ($p = 3.775e^{-12}$).

Interpreting these results provides further insights into mutual understanding. CONFIRMATIONS seem to lead the speaker to keep looking down at turn ends. REJECTION turns only seem to end with the speaker looking down and the other participant looking up – a finding in favour of analysing referential looks down as a sign that the understanding is mutual. The alignment theory (Pickering and Garrod, 2004) states that participants will align their

beliefs as communicative intent unfolds, and this is visible through different means, linguistic and paralinguistic (Garrod and Pickering, 2007). Such a result adds to past work on repetitions and mutual gaze (Murat et al., 2022) which hypothesised mutual gaze to convey evidence of mutual understanding. In fact, our results are further evidence that mutual gaze is relatively rare in a conversation which includes a gaze competitor (here, the map) and that it occurs in situations of struggle: at the start of a REJECTION turn, at the end of an EXPLANATION or at the end of a turn in which the Follower tries to make the conversation's words theirs ($\Delta H_S > \Delta H_C > 0$). This matches previous literature which sees gaze up as a way to check the availability of the other participant to speak – and hence contribute (Kendon, 1967; Ho et al., 2015). It is also similar to the early gaze studies reported on the HCRC Map Task Corpus (Boyle et al., 1994), which highlighted that gazes up tended to mark moments of difficulty. Interestingly, the method they had then used differed considerably from ours. They focused their analysis on the mismatch between the maps and found significantly more gazes up when non-matching references were discussed than when matching references were discussed. On our end, the methods were purely quantitative. We did not inspect the maps nor the details of the lexicon being used beyond their frequency and their part-of-speech. Yet, we were able to highlight moments of difficulty and draw similar conclusions (presumably because the language used was *about* the maps).

We have reported on a reduction of temporal information to categories and counts based on events and their overlap, without direct attention to durations associated with the relevant boundaries. These have been shown to be important (e.g., Oertel et al. (2012)). Our approach does not destroy information about duration beyond counts of events; rather, the method may complement such analyses.

In sum, exploring gaze behaviour at the start and end of each turn allowed us to study anteriority of events (is the gaze or the turn coming first) and, thus, contributes insights to the turn-taking literature (Degutyte and Astell, 2021) (see §2). The relationship between gaze and turn qualities shows different coherent patterns: as the property of a turn display a need for feedback, the speaker's gaze ends up, and as the gazes are up, the content gets more reused in response, as a way to clarify the situation.

6. Conclusion

This paper investigates the relationship between eye-gaze behaviour in the dyadic conversations and the linguistic progression of the interaction. To do so, we studied the correlation between paired gaze and dialogue acts to reveal the intention behind turns, along with their lexical properties, as measured by entropy and repetition, using a cross-modality event annotation scheme that enables count-based interaction analysis. Our results on the HCRC Map Task corpus show that gaze at the turn level (their start or end) is correlated with the progression of the interaction. A gaze towards the conversation partner at the end of a turn tends to align with complexity and difficulties expressed in the turn, while keeping a look down at the map is more typical of situations without obstacles, and neither participant shows signs of disagreement. There is also evidence that the Guide may adapt the content of their turns to the Follower when the latter is looking up at the start of the turn, as such a case is correlated with the presence of other-repetition.

7. Ethics Statement & Limitations

This observational study uses part of only one – albeit large – corpus: the HCRC corpus (Thompson et al., 1993). Its strict set-up – a task having to be performed by strictly two persons with a visual competitor that is the map, is more natural than some settings, but still far from representative of all face-to-face human interactions. Rather, this paper has to be read as an incremental contribution, and this aspect is accented in the discussion where the results are compared with prior findings.

While the HCRC Map Task corpus is widely studied and this observational study relied on its ethical provenance, the results here have some bias. Even if the corpus is gender-balanced, its tightly controlled design records only native English speakers, most of them Scots. We have not yet analysed Map Task data constructed in other languages. The raw HCRC recordings are not available for privacy reasons, so all the gaze annotations have to be trusted – in fact, due to contradictions in some gaze annotations and absence of material to reconcile them, we had to omit excerpts (see sec. 3.2). We have found prior descriptions of gaze annotation in this corpus (e.g. Boyle et al. (1994); Anderson et al. (1997); Doherty-Sneddon (1995), but they imply a different label inventory and temporal precision than is in the corpus. We have not spotted a clear discussion of the released annotations.

8. Acknowledgements

This publication has emanated from research conducted with the financial support of Taighde Éireann – Research Ireland under Grant No. 18/CRT/6223. We are grateful to the anonymous reviewers for their constructive feedback. We think this paper is improved by having addressed their comments.

9. Bibliographical References

Anne H Anderson, Ellen Gurman Bard, Cathy Sotillo, Alison Newlands, and Gwyneth Doherty-Sneddon. 1997. Limited visual control of the intelligibility of speech in face-to-face dialogue. *Perception & Psychophysics*, 59(4):580–592.

Michael Argyle and Mark Cook. 1976. *Gaze and mutual gaze*. Cambridge Univ. Press, Cambridge.

Janet Beavin Bavelas, Linda Coates, and Trudy Johnson. 2002. Listener Responses as a Collaborative Process: The Role of Gaze. *Journal of Communication*, 52(3):566–580.

Geoffrey W. Beattie. 1978. Floor apportionment and gaze in conversational dyads. *British Journal of Social and Clinical Psychology*, 17(1):7–15.

Elizabeth A. Boyle, Anne H. Anderson, and Alison Newlands. 1994. The Effects of Visibility on Dialogue and Performance in a Cooperative Problem Solving Task. *Language and Speech*, 37(1):1–20.

Susan E. Brennan, Xin Chen, Christopher A. Dickinson, Mark B. Neider, and Gregory J. Zelinsky. 2008. Coordinating cognition: the costs and benefits of shared gaze during collaborative search. *Cognition*, 106(3):1465–1477.

Geert Brône, Bert Oben, Annelies Jehoul, Jelena Vranjes, and Kurt Feyaerts. 2017. Eye gaze and viewpoint in multimodal interaction management. *Cognitive Linguistics*, 28(3):449–483. Publisher: De Gruyter Mouton.

Jean Carletta, Amy Isard, Stephen Isard, Jacqueline Kowtko, Alison Newlands, Gwyneth Doherty-Sneddon, and Anne Anderson. 1996. HCRC dialogue structure coding manual. Technical report, University of Edinburgh. HCRC Technical Report 82.

Mark S. Cary. 1978. The Role of Gaze in the Initiation of Conversation. *Social Psychology*, 41(3):269.

Ziedune Degutyte and Arlene Astell. 2021. The Role of Eye Gaze in Regulating Turn Taking in Conversations: A Systematized Review of Methods and Findings. *Frontiers in Psychology*, 12.

Gwyneth Doherty-Sneddon. 1995. *The development of conversational and communication skills*. Ph.D. thesis, University of Glasgow.

Jens Edlund, Simon Alexandersson, Jonas Beskow, Lisa Gustavsson, Mattias Heldner, Anna Hjalmarsson, Petter Kallionen, and Ellen Marklund. 2012. 3rd party observer gaze as a continuous measure of dialogue flow. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, pages 1354–1358, Istanbul, Turkey. European Language Resources Association (ELRA).

Simon Garrod and Martin J. Pickering. 2007. Alignment in dialogue. In M. Gareth Gaskell, editor, *The Oxford Handbook of Psycholinguistics*, pages 442–452. Oxford University Press.

Dmitriy Genzel and Eugene Charniak. 2002. Entropy rate constancy in text. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 199–206.

Milan Gnjatović, Jovica Tasevski, Branislav Borovac, and Nemanja Maček. 2018. An Entropy-Based Approach to Automatic Detection of Critical Changes in Human-Machine Interaction. In *2018 9th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, pages 000175–000178. ISSN: 2380-7350.

Charles Goodwin. 1980. Restarts, Pauses, and the Achievement of a State of Mutual Gaze at Turn-Beginning. *Sociological Inquiry*, 50(3-4):272–302.

Joy E. Hanna, Susan E. Brennan, and Kelly J. Savietta. 2020. Eye Gaze and Head Orientation Cues in Face-to-Face Referential Communication. *Discourse Processes*, 57(3):201–223. _eprint: https://doi.org/10.1080/0163853X.2019.1675467.

Patrick G. T. Healey, Matthew Purver, and Christine Howes. 2014. Divergence in dialogue. *PloS one*, 9(6):e98598. Https://doi.org/10.1371/journal.pone.0098598.

Simon Ho, Tom Foulsham, and Alan Kingstone. 2015. Speaking and Listening with the Eyes: Gaze Signaling during Dyadic Interactions. *PLOS ONE*, 10(8):e0136905. Publisher: Public Library of Science.

Laszlo Hunyadi, Tamás Váradi, Gy Kovács, István Szekrényes, Hermina Kiss, and Karolina Takács. 2018. Human-human, human-machine communication: on the HuComTech multimodal corpus. *Selected papers from the CLARIN Annual Conference 2018. Linköping Electronic Conference Proceedings*, 159:56–65.

Kristiina Jokinen, Kazuaki Harada, Masafumi Nishida, and Seiichi Yamamoto. 2010. Turn-Alignment Using Eye-Gaze and Speech in Conversational Interaction. *Interspeech*, pages 2018–2021.

Frank Keller. 2004. The entropy rate principle as a predictor of processing effort: An evaluation against eye-tracking data. In *Proceedings of the 2004 conference on empirical methods in natural language processing*, pages 317–324.

Adam Kendon. 1967. Some functions of gaze-direction in social interaction. *Acta Psychologica*, 26:22–63.

Adam Kendon. 1978. Looking in conversation and the regulation of turns at talk: A comment on the papers of G Beattie and D. R. Rutter *et al* . *British Journal of Social and Clinical Psychology*, 17(1):23–24.

Kobin H. Kendrick and Judith Holler. 2017. Gaze Direction Signals Response Preference in Conversation. *Research on Language and Social Interaction*, 50(1):12–32.

Maria Koutsombogera and Carl Vogel. 2019. Observing Collaboration in Small-Group Interaction. *Multimodal Technologies and Interaction*, 3(3):45.

Magnus S. Magnusson. 2000. Discovering hidden time patterns in behavior: T-patterns and their detection. *Behavior Research Methods, Instruments, & Computers*, 32(1):93–110.

Magnus S. Magnusson. 2020. T-Pattern Detection and Analysis (TPA) With THEMETM: A Mixed Methods Approach. *Frontiers in Psychology*, 10.

Alexander Mehler, Andy Lücking, and Peter Menke. 2011. Assessing Lexical Alignment in Spontaneous Direction Dialogue Data by Means of a Lexicon Network Model. In Alexander F. Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing*, volume 6608, pages 368–379. Springer Berlin Heidelberg, Berlin, Heidelberg. Series Title: Lecture Notes in Computer Science.

Gregor Mehlmann, Markus Häring, Kathrin Janowski, Tobias Baur, Patrick Gebhard, and Elisabeth André. 2014. Exploring a Model of Gaze for Grounding in Multimodal HRI. In *Proceedings of the 16th International Conference on Multimodal Interaction*, pages 247–254, Istanbul Turkey. ACM.

Dalila Mekhaldi. 2006. *A study on multimodal document alignment: bridging the gap between textual documents and spoken language*. Ph.D. thesis, Faculty of Science, University of Fribourg (Switzerland).

Ines Montani, Matthew Honnibal, Adriane Boyd, Sofie Van Landeghem, and Henning Peters. 2023. explosion/spaCy: v3.7.2: Fixes for APIs and requirements.

Anaïs Claire Murat, Maria Koutsombogera, and Carl Vogel. 2026. Event chronography in multimodal data: the BME method for quantitative analyses. In *Proceedings of the 2026 Language Resources and Evaluation Conference*. To appear.

Anaïs Claire Murat, Maria Koutsombogera, and Carl Vogel. 2022. Mutual Gaze and Linguistic Repetition in a Multimodal Corpus. In *Proceedings of the Language Resources and Evaluation Conference*, pages 2771–2780, Marseille, France. European Language Resources Association (ELRA).

Anaïs Claire Murat and Carl Vogel. 2025. Online meetings: When repetition signals engagement. In *2025 IEEE 16th International Conference on Cognitive Infocommunications (CogInfoCom)*, pages 37–42, Vienna, Austria. IEEE.

Mark B. Neider, Xin Chen, Christopher A. Dickinson, Susan E. Brennan, and Gregory J. Zelinsky. 2010. Coordinating spatial referencing using shared gaze. *Psychonomic Bulletin & Review*, 17(5):718–724.

Hannele Nicholson, Ellen Gurman Bard, Robin Lickley, Anne H. Anderson, and Catriona Havard (3) Yiya Chen. 2005. Disfluency and behaviour in dialogue: evidence from eye-gaze. In *Disfluency in Spontaneous Speech (DiSS 2005)*, pages 133–138.

Catharine Oertel, Marcin Włodarczak, Jens Edlund, Petra Wagner, and Joakim Gustafson. 2012. Gaze patterns in turn-taking. In *13th Annual Conference of the International Speech Communication Association 2012 (INTERSPEECH 2012)*, pages 2246–2249.

Ulrich J. Pfeiffer, Kai Vogeley, and Leonhard Schilbach. 2013. From gaze cueing to dual eye-tracking: Novel approaches to investigate the neural correlates of gaze in social interaction. *Neuroscience & Biobehavioral Reviews*, 37(10):2516–2528.

Martin J. Pickering and Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(02).

David Reitter and Johanna D Moore. 2007. Predicting success in dialogue. In *Proceedings of the 45th Annual Meeting of the Association of Comp utational Linguistics*, pages 808–815. Association for Computational Linguistics.

Justine Reverdy, Maria Koutsombogera, and Carl Vogel. 2020. Linguistic Repetition in Three-Party Conversations. In Anna Esposito, Marcos Faundez-Zanuy, Francesco Carlo Morabito, and

Eros Pasero, editors, *Neural Approaches to Dynamics of Signal Exchanges*, volume 151, pages 359–370. Springer Singapore.

Justine Reverdy and Carl Vogel. 2017. Measuring Synchrony in Task-Based Dialogues. In *Interspeech 2017*, pages 1701–1705, Stockholm, Sweden. ISCA.

Federico Rossano. 2010. Questioning and responding in Italian. *Journal of Pragmatics*, 42(10):2756–2771.

Simona Sbranna, Michelina Savino, Florence Baills, and Martine Grice. 2025. Gaze behaviour and vocal feedback in task-based dyadic conversations with and without eye contact. *Frontiers in Communication*, 10. Publisher: Frontiers.

Malin Spaniol, Alicia Janz, Simon Wehrle, Kai Vogeley, and Martine Grice. 2023. Multimodal signalling: The interplay of oral and visual feedback in conversation. In *Proceedings of the 20th International Congress of Phonetic Sciences. ICPhS*.

Jürgen Streeck. 2014. Mutual gaze and recognition: Revisiting Kendon's "Gaze direction in two-person conversation". In Mandana Seyfeddinipur and Marianne Gullberg, editors, *From Gesture in Conversation to Visible Action as Utterance: Essays in honor of Adam Kendon*, pages 35–56. John Benjamins Publishing Company.

Yu Wang, Yang Xu, Gabriel Skantze, and Hendrik Buschmeier. 2024. How much does nonverbal communication conform to entropy rate constancy?: A case study on listener gaze in interaction. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 3533–3545.

Peter Wittenburg, Hennie Brugman, Albert Russel, Alex Klassmann, and Han Sloetjes. 2006. ELAN: a professional framework for multimodality research. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy. European Language Resources Association (ELRA).

Elisabeth Zima, Clarissa Weiß, and Geert Brône. 2019. Gaze and overlap resolution in triadic interactions. *Journal of Pragmatics*, 140:49–69.

10. Language Resource References

Thompson, Henry S. and Anderson, Anne and Bard, Ellen Gurman and Doherty-Sneddon, Gwyneth and Newlands, Alison and Sotillo, Cathy. 1993. *The HCRC Map Task corpus: natural dialogue for speech recognition*. Association for Computational Linguistics, ISLRN 777-455-577-608-1.