Automatic Analysis of Collaboration Through Human Conversational Data Resources: A Review

Yi Yu[1,2], Maria Boritchev[2], Chloé Clavel[1,2]
[1]INRIA Paris, ALMAnaCH, [2]LTCI, Télécom Paris, Institut Polytechnique de Paris, France
48, rue Barrault 75013 Paris, 19 place Marguerite Perey, 91120 Palaiseau
{yi.yu, chloe.clavel}@inria.fr, maria.boritchev@telecom-paris.fr

Abstract

Collaboration is a task-oriented, high-level human behavior. In most cases, conversation serves as the primary medium for information exchange and coordination, making conversational data a valuable resource for the automatic analysis of collaborative processes. In this paper, we focus on verbal aspects of collaboration and conduct a review of collaboration analysis using task-oriented conversation resources, encompassing related theories, coding schemes, tasks, and modeling approaches. We aim to address the question of how to utilize task-oriented human-human conversational data for collaboration analysis. We hope our review will serve as a practical resource and illuminate unexplored areas for future collaboration analysis.

Keywords: Multimodal Task-Oriented Conversation Resources, Human Collaboration Analysis, Literature Review

1. Introduction

Collaboration analysis (CollA) seeks to use computational methods to model how people coordinate, think, and learn in shared tasks in order to gain insights that improve both collaboration processes and outcomes (Martinez-Maldonado et al., 2021). As collaboration is a fundamental human behavior, CollA has broad applications, including education (Jaques et al., 2023), management (Casey-Campbell and Martens, 2009), interface design (Prati et al., 2021), and AI agent development (Enayet and Sukthankar, 2023a; Zhang et al., 2024).

Computational methods for CollA require data to understand the phenomena in play. Human conversational resources are irreplaceable for two main reasons. First, conversations are instances of joint action (Clark, 1996), and human conversational data is a sequential record of multimodal communicative behavior that reflects both individual contributions and interpersonal dynamics. Additionally, linguistic research on human interpersonal phenomena supplies features for CollA, such as referring expressions (Heeman and Hirst, 1995; Clark and Wilkes-Gibbs, 1986) and multilevel entrainment (Lubold and Pon-Barry, 2014), as discussed in Section 4. We illustrate in Figure 1 how informative human task-oriented conversation can be for CollA. Second, although we have relatively mature metrics for evaluating task-related dimensions of CollA in task-oriented conversations (*e.g.*, task decomposition, task completion, etc. Guan et al. (2025)), interpersonal dynamics, which directly affect the collaboration process and quality, remain largely unexplored. For collaborative learning scenarios, collaboration itself can be a way of learning. Collaboration directly fosters both the internalization and the sharing of knowledge, which provide new dimensions for CollA. In addition, the quality of interpersonal dynamics can significantly influence learning outcomes (Yang, 2023). The variety of interaction patterns in human task-oriented dialogue also makes it possible to investigate how people come to know each other (*e.g.*, through language use, personality traits, or educational background) during collaboration (Guo et al., 2025). CollA using human conversational data provides valuable insights for designing human-machine collaboration systems that balance individual benefits with overall task performance of the group, and provide a better understanding of humanity.

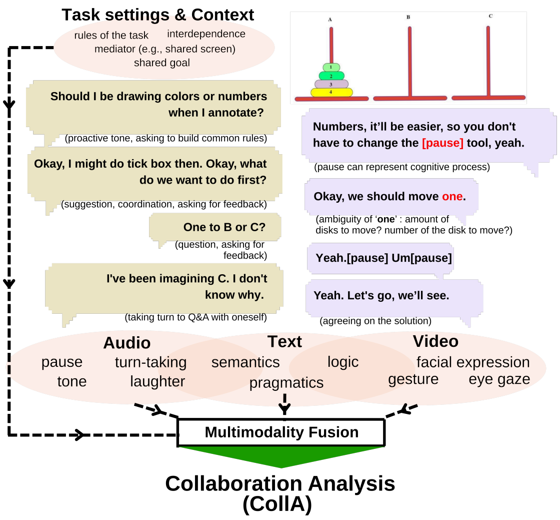

Figure 1: Two players playing the Hanoi tower game and conversing with each other to solve the puzzle. Their conversation provides many useful elements for CollA.

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 7248–7261
11–16 May 2026. ©ELRA Language Resources Association (ELRA), 2026

Several surveys have covered related but distinct topics: Praharaj et al. (2021) and Schürmann et al. (2024) examine CollA multimodal features and existing collaboration measurements, but focus on the educational context. Zou et al. (2025) centers on LLM-based human-agent system building and examines how human feedback and control contribute to performance improvement. They review works on human-LLM systems and the conversation corpora for these systems, focusing on tasks where human feedback can be crucial for defining and assessing evaluation metrics. Vaccaro et al. (2024) surveys recent work comparing the performance of humans working alone, humans collaborating with machines, and machines operating independently, and highlights the challenges involved in integrating human intelligence with computational systems. While this study investigates the conditions under which human–machine collaboration surpasses human-only or machine-only performance, our work instead reviews computational models of multimodal discourse in collaborative contexts.

A holistic review of human task-oriented conversation resources and their usage for CollA remains unexplored. By task-oriented conversation, we mean conversation directed with clear intention toward the completion of tasks Grosz and Sidner (1986). For collaboration, building on the definitions provided by Wood and Gray (1991); Randrup et al. (2016), we conceptualize collaboration as an interactive process comprising four core elements: **a shared goal**, **a shared understanding of the task** (including rules, norms, and structures), **positive interdependency**, and **joint individual commitments** reflected in participants' actions and decisions. Then, we apply this definition to select task-oriented corpora from peer-reviewed published papers that provide collaboration annotations and evaluations, focusing particularly on settings that create positive interdependencies among participants (*e.g.*, tasks that cannot be completed by a single participant under the defined settings). The step-by-step procedure and criteria for paper selection are explained in Appendix A, Figure 2. We conduct the review by systematically analyzing the coding schemes for capturing collaboration in conversation, extracting salient multimodal features, and examining recent collaboration modeling approaches applied to the selected corpora.

After discussing coding schemes in Section 2, we present the criteria used to select CollA corpora and review their task settings in Section 3. We then examine CollA studies based on at least one of the selected corpora, discussing salient features and modeling approaches in Sections 4 and 5. We conclude by discussing recent advances and future directions of CollA in Section 6.

2. Coding Schemes for CollA

Coding schemes serve as a lens to showcase important elements for CollA (Chen et al., 2020) and are used to annotate collaboration in conversations for computational collaboration model building. This section reviews coding schemes employed in the corpora discussed in Section 3, highlighting how different coding approaches capture both individual and group (including dyad) aspects of collaboration. We also examine applied questionnaires, given their flexibility (*e.g.*, self-reports, annotation instruments, external evaluation) and wide application in CollA.

The full details of all the employed coding schemes are given in Table 1. We discuss the main categories of coding schemes and questionnaires, their theoretical foundations, and how they capture different aspects of collaboration to illuminate current trends and challenges.

2.1. Individual Perspective

From the individual perspective, applied coding schemes and questionnaires focus on the single-participant collaborative aspect, such as *individual collaborative/cooperative behaviors*, *engagement*, and individual variables, such as *collective orientation* that refers to the propensity to work in a collective manner in a team setting (Driskell et al., 2010). They are applied to individual audio and visual data to model participants' collaborative behaviors and their impacts on the collaboration process.

Using **linguistic** aspects, Cavicchio and Poesio (2012) use a *cooperation* coding scheme in Rovereto corpus, developped by Davies (1997, 2007). This scheme analyzes individual collaborativeness using Grice's cooperative principle (Grice, 1975) for conversation analysis. It is evaluative, *i.e.*, not only used to label what the speakers do but also to assess it in terms of appropriateness, which can be subjective and hard to agree on for the annotators. Findings in classroom discourse research reveal the important role of individual argument in the collaborative learning process (Engle and Conant, 2002). Based on that, Olshefski et al. (2020) develop a coding scheme to capture the function of collaborative argument moves for students' discussion in the Discussion Tracker corpus to explore the collaboration dimension of argument in large-group collaborative learning tasks.

Using **behavior** studies, Richey et al. (2016) applied a modified version of the collaborative behavior coding scheme from Johnson and Johnson (2013) to the SRI corpus. These coding schemes are rooted in social interdependence theory (Johnson and Johnson, 2009) and Vygotsky's cognitive developmental theory (Tudge and Rogoff, 2014), which highlight the interactive aspects of individual behavior and collaborative indicators of learning.

	Corpus, Description	Scheme/Questionnaire	Details	Reference
Individual	**SRI** speech-based collaborative learning corpus	I Code (individual collaboration indicators)	regulative/logistical, interaction, and cognitive indicators of teamwork behavior	Richey et al. (2016)
	MISC information-seeking conversations	adapted User Engagement Scale (UES)	scaled ratings of partner's collaborativeness	McDuff et al. (2017)
	RoomReader multimodal, multiparty conversational interactions corpus	online engagement	continuous scaled engagement in groups based on collaborators' behaviors and perceived intentions	Reverdy et al. (2022)
	Discussion Tracker multiparty discussions	collaborative argumentation functions	classification of arguments as: new ideas, agreements, extensions, probes/challenge	Olshefski et al. (2020)
	Rovereto emotion and cooperation corpus	cooperative dialogue effort	evaluation of each turn along three dimensions: knowledge sharing, non-cooperative behavior, and cooperation level (scaled)	Cavicchio and Poesio (2012)
	MULTICOLLAB multimodal dialogues	extreme emotion (frustration) in collaboration	participant self-assessed frustration level	Peechatt et al. (2024)
	Teams multiparty dialogues for entrainment	collective orientation	participant self-assessed preference for teamwork	Litman et al. (2016)
Dyad & Group	**SRI**	Q Code (team collaboration quality; triads)	team-level quality states (*e.g.*, good collaboration, follow-the-leader), based on number of engaged participants	Richey et al. (2016)
	MISC	adapted User Engagement Scale (UES)	scaled ratings of collaboration process	McDuff et al. (2017)
	Teams	team cohesion, satisfaction, potency/efficacy	between and post-game questionnaires elicit perceptions of teams processes	Litman et al. (2016)
	GAME-ON group analysis of multimodal expression of cohesion corpus	modified Group Environment Questionnaire (GEQ) for group cohesion	highlights of instrumental function (social vs task facets); affective function optional depending on study	Maman et al. (2020)
	GAP group affect and performance corpus	teamwork experience (self-report)	ratings of teamwork performance (time management, efficiency, overall work quality)	Braley and Murray (2018)
	AMI augmented multiparty interaction corpus), **PCC** patient consultation corpus	group cohesion	ratings of task cohesion, social cohesion, and leadership	Hung and Gatica-Perez (2010) Kantharaju and Pelachaud (2021)
	MULTISIMO multimodal group interaction corpus	collaboration quality (overall)	scaled rating of overall collaboration quality	Koutsombogera and Vogel (2018)
	PhotoBook visually-grounded dialogues	collaboration performance	scaled ratings of overall collaboration performance and perceived mutual understanding	Haber et al. (2019)

Table 1: Applied coding schemes that capture different aspects of collaboration in conversational data.

Reverdy et al. (2022) adapts a multifacet classroom engagement behavior coding scheme (Goldberg et al., 2021) for online interaction in the RoomReader corpus, where head and hand position and eye gaze/focus play an important role in annotating individual engagement levels. Peechatt et al. (2024) design a coding scheme for annotating the frustration level in MULTICOLLAB to predict critical moments in collaboration process.

From a **human factors** perspective, Litman et al. (2016) collect collective orientation ("the propensity to work in a collective manner", Driskell et al. (2010)) from each participant via a self-assessment questionnaire in the Teams corpus, enabling further collaboration analysis alongside individual variables.

2.2. Dyad and Group Perspective

The dyad and group interaction dynamics of the collaboration process have attracted lots of attention from the research community. These coding schemes or questionnaires focus on the interpersonal dynamics and group behaviors, such as *group cohesion*, self-assessed *collaboration performance*, and perceived *collaboration quality*.

For collaboration quality evaluation, Haber et al. (2019) use a questionnaire in the dyadic conversation corpus PhotoBook to collect self-assessed, scaled overall collaboration performance, while McDuff et al. (2017) select items from the `User Engagement Scale` questionnaire (O'Brien and Toms, 2010) in MISC, for collaboration process evaluation. For three-person groups, Richey et al. (2016) develop a coding scheme, `Q codes` (as shown in Table. 4 of Richey et al. (2016)), in which the collaboration quality is defined as "nb of the group members' actively contribute to the task completion" with a special focus on balanced involvement of each member.

Group cohesion is a group phenomenon defined as "the group members' inclinations to forge social bonds" Casey-Campbell and Martens (2009), which impacts the collaboration process (Hung and Gatica-Perez, 2010; Kantharaju and Pelachaud, 2021). Hung and Gatica-Perez (2010) study both the social and task aspects of cohesion in a computational way. Their 27-item questionnaire for perceived group cohesion is based on group research (Carron and Brawley, 2000) and psychology literature (Siebold, 1999), and it has been applied to the AMI corpus (Kraaij et al., 2005) and the PCC corpus (Kantharaju and Pelachaud, 2021).

Severt and Estrada (2015) provides another coding scheme for a cohesion study that includes functional and structural dimensions of cohesion. This scheme has two psychological functions: *affective* and *instrumental*, but only the last one, with *social* and *task* facets, has been used in the GAME-ON corpus (Maman et al., 2020), to align with the dominant approach (Braun et al., 2020).

2.3. Main Trends for CollA Coding Schemes

There is no universal coding scheme for CollA. For segment level, earlier studies tend to provide annotations for individual-level collaborativeness, while recent coding schemes are mostly applied to study group-level collaborative phenomena. One possible explanation could be that a group is not simply the sum of its dyads. Group-level collaboration involves emergent dynamics that individual or dyadic models cannot capture. However, group-level emergent collaborative interaction is context-dependent and can be hard to capture with individual-level cues. Human manual annotation remains the most reliable approach for these studies, given that large language models (LLMs) are increasingly involved in recent annotation processes (Wang et al., 2024).

We observe that task-level annotation has been widely chosen to evaluate both individual-level (McDuff et al., 2017; Haber et al., 2019) and group-level (Koutsombogera and Vogel, 2018; Braley and Murray, 2018; Litman et al., 2016) collaboration. Most annotations are based on task-adapted questionnaires, such as perceived collaboration quality using `User Engagement Scale` (O'Brien and Toms, 2010) in MISC and self-assessed overall collaboration quality and satisfaction in MULTISIMO, GAP, and Teams. Both external and self-assessed annotations are valuable for CollA, from modeling collaborative conversation to detecting individual variables for dialogue system adaptation. However, task-level granularity can be insufficient for analyzing emergent phenomena.

3. Task-Oriented Corpora for CollA

In this section, we examine **open-source**, **human-human**, **task-oriented** conversation corpora created **in the last 20 years (from 2005)** that **require collaboration** in their task-settings and have **direct annotations of collaboration** (*e.g.*, collaboration quality/skills, group cohesion, conflicts, etc) to understand the recent trends in task setting design for CollA. Corpora with only measurable collaboration task results will not be included here, *e.g.* ELEA (Sanchez-Cortes et al., 2011), since task results alone cannot reflect collaboration quality.

The task settings are vital for building CollA corpora, as they determine the level of interdependence among participants during the collaboration process. We first discuss CollA task settings, including group size and how they promote collaboration between participants, and provide details on their collaboration annotations in Table 2. We then compare different choices and synthesize the main trends.

Game Game scenarios constitute the dominant task setting category in our selection of corpora. We categorize a corpus task as a game when the keyword "game" is used to describe the corpus' scenario.

Social game settings have been intensively used for group cohesion analysis. The Teams corpus (Litman et al., 2016) is built on a role-playing social game for CollA. A group of 3 or 4 players, each with a different adventurer role, must discuss strategies to collect enough treasures to complete the game. The GAME-ON corpus (Maman et al., 2020) is based on a multitask social game in which 3 players must cooperatively discover clues and solve several puzzles within a limited time frame. Perceived cohesion is evaluated through self-assessment after each puzzle. They argue that cohesion can take considerable time to emerge in groups of strangers, so they recruit only real-world friends to play the game.

Game settings can be easily adapted to a particular aspect of CollA. The MISC corpus (McDuff et al., 2017) employs a role-play, information-seeking setting, assigning the "seeker" role to one participant in each pair. This setting helps understand human interests in this collaborative information-seeking process to design and evaluate human-machine interfaces. The PhotoBook corpus (Haber et al., 2019) uses a remote multi-round image identification game setting. Each group of two players has access to different sets of images, and participants must mark each image as either common or different by discussing it with their partner. This setting enables referring analysis[1] in a collaborative task.

Education Education is a domain where collaboration has been extensively studied in the context of collaborative learning. Due to privacy concerns, a significant number of corpora under this task setting are not public (Schneider and Bryant, 2022; Lämsä et al., 2021; Olsen et al., 2020; Spikol et al., 2017; Salinas et al., 2021). Recent corpora in collaborative learning settings often involve large-group interactions, making the capture of multimodal data more difficult.

Among the accessible resources, the Discussion Tracker corpus (Olshefski et al., 2020) captures classroom teacher-student interactions, focusing on collaborative argumentation in literature discussions. The SRI corpus (Richey et al., 2016) exemplifies collaborative problem-solving with triadic groups solving math problems, capturing both social and cognitive dimensions mainly from audio.

The only corpus we found with video data and collaborative learning task setting is the Room-

[1] "John didn't come to class because **he** was sick." Referring analysis studies what the word "he" refers to, tracking who or what is being talked about, a key part of collaborative conversation.

Corpus	Lang.	Hours	Task Type	Group Size	Audio	Video	Transc.	Sensors	Collaboration Annotation				Dataset Link
									Object	Obj. Level	Annotator	Granularity	
AMI 2005	EN	100	meeting	4	✓	✓	✓	✗	cohesion	group	external	segment	Dataset link
Rovereto 2012	IT	4.67	game	4	✓	✓	✓	✓	cooperativeness	individual	external	segment	Dataset on request
SRI 2016	EN	26.6	education	3	✓	✗	✓	✗	I Code, Q Code	both	external	segment	Dataset link
☆Teams 2016	EN	47	game	3-4	✓	✓	✓	✗	participant's collective orientation; group cohesion, satisfaction, potency	both	self	task	Dataset link
❀MISC 2017	EN	42	game	2	✓	✓	✓auto	✗	perceived collaboration (help/ understanding/communication)	individual	external	task	Dataset link
MULTISIMO 2018	EN	4	game	3	✓	✓	✓	✗	collaboration quality	group	external	task	Dataset link
GAP 2018	EN	4	meeting	2,3,4	✓	✗	✓	✗	teamwork experience	group	self	task	Dataset link
PhotoBook 2019	EN	-	game	2	✗	✗	✓	✗	collaboration performance	group	self	task	Dataset link
GAME-ON 2020	IT	11	game	3	✓	✓	✗	✓	cohesion	group	self	task	Dataset link
Discussion Tracker 2020	EN	-	education	15	✗	✗	✓	✗	collaborative argumentation	individual	external	segment	Dataset link
PCC 2021	EN	2	other	3,4	✓	✓	✗	✗	cohesion	group	external	segment	Dataset on request
RoomReader 2022	EN	38	education	4,5	✓	✓	✓	✗	engagement, cohesion	both	both	segment	Dataset link
☆MULTICOLLAB 2023	EN	3	other	2	✓	✓	✓	✓	extreme emotion (frustration)	individual	self	both	Dataset will go public

❀ The transcripts in MISC are auto-generated without mentioning any manual verification process.

☆MULTICOLLAB is currently not public but is to be made available in the future. The video, transcription, and questionnaire data of Teams will be available in a future release.

Table 2: Overview of collaboration corpora arranged chronologically, to highlight the evolution of research focus in CollA. Comparative assessment across dimensions: language, recording size, task type, group size, multimodal data availability, and collaboration annotation (annotation object, object level, annotator type, and temporal granularity).

Reader corpus (Reverdy et al., 2022). It is based on online computer-supported student-tutor conversations and can be used to analyze engagement in collaboration and conversational dynamics.

Meetings and Others Scenarios adapted from real-world tasks are frequently used for CollA.

A part of the AMI corpus (Kraaij et al., 2005) elicits collaboration using a role-playing functional meeting within a 4-person design team for new product protocol development. The GAP corpus (Braley and Murray, 2018) also applies a meeting setting for a small group to make a decision on the rank of the most important items for a plane crash. The PCC corpus (Kantharaju and Pelachaud, 2021) simulates health consultations between patients and healthcare professionals by equipping professionals with detailed background information to study cohesion.

The MULTICOLLAB corpus (Peechatt et al., 2024) adopts a role-playing setting, *i.e.*, instructor and builder, for a block building task in which some builders are instructed to deliberately disobey to stimulate critical moments in the collaboration process with strong interdependency. Their task settings yield half of their data from non-collaborative builders.

Main Trends of CollA Task Settings Common task settings for CollA include role-play and information asymmetry, which promote interdependence among participants and foster the observation of active joint contributions during the collaboration process. Co-locating collaboration corpora with video recordings, especially with both individual- and room-level recordings, is interesting for studying interpersonal and group dynamic aspects of CollA (Kraaij et al., 2005; Litman et al., 2016; Koutsombogera and Vogel, 2018; Maman et al., 2020) and have attracted more attention in recent CollA.

Remote meetings can also provide a view of group synchrony (*e.g.*, RoomReader (Reverdy et al., 2022)), while it is recognized that remote settings may inhibit natural interaction (Poel et al., 2008). We also observe a trend toward analyzing group collaboration phenomena using segment-level annotations, rather than individual collaborativeness as in earlier CollA studies.

4. CollA Features

This section synthesizes experimentally supported features extracted from text, audio, video, sensor signals[2], and cross-modal, for CollA. We narrow the discussion to features extracted from our selection of conversation corpora presented in Section 3, arguing that task-oriented corpora built with particular settings (*e.g.*, interdependency, information asymmetry between participants) that foster collaboration are better suited for identifying CollA features. Both features that have shown significant associations with collaboration quality and those used in individual- and group-level collaboration modeling are included.

We also cover different feature-generation methods (*e.g.*, natural language processing, signal processing, questionnaires) and how these features are exploited (*e.g.*, used directly in modeling or incorporated into high-level construct building) to make the discussion more practical. An overview of features per modality is available in Table 3.

[2]We find the following sensor signals applied for CollA: electrocardiogram, electrodermal signals, galvanic skin response, photoplethysmography, and body motion.

	Group-level Collaboration		Individual Collaborativeness		
	Group Cohesion and Entrainment	Teamwork Process and Performance	Engagement	Collaboration Behaviors	Perceived Personality and Role
Text 4.1	pronoun usage (Enayet and Sukthankar, 2021a) lexical entrainment (Rahimi and Litman, 2020) paralinguistic mimicry (Nanninga et al., 2017)	syntactic entrainment embedding (Enayet and Sukthankar, 2021a) DACTs sequence embedding, sentiment embedding (Enayet and Sukthankar, 2023a, 2021a) SUBTL score (Murray and Oertel, 2018) lexical cohesion (Rahimi and Litman, 2020) dependency parse feature (Murray and Oertel, 2018) word psycholinguistic score (Murray and Oertel, 2018)	word embedding (Rahimi and Litman, 2020) dialogue act Be-Positive (Kantharaju et al., 2020)	underexplored	BERT embedding (Fenech et al., 2022)
Audio 4.2	intensity, frequency, shimmer, jitter (Peechatt et al., 2024; Litman et al., 2016) turn-taking (Sabry et al., 2021; Sassier-Roublin et al., 2025) laughter,backchannels (Kantharaju and Pelachaud, 2021)	total overlapping, pause time (Hung and Gatica-Perez, 2010)	audio embedding (Li et al., 2024)	underexplored	total speaking time, pitch, jitter, loudness (Peechatt et al., 2024; Litman et al., 2016; Sabry et al., 2021) shimmer, harmonics-to-noise (Sabry et al., 2021) eGeMAPS features (Fenech et al., 2022)
Video 4.3	mutual gaze (Kantharaju and Pelachaud, 2021) automatic extracted facial expression, head nods duration (Kantharaju et al., 2020)	underexplored	eye gaze (Hung and Gatica-Perez, 2010)	saccade peak velocity (Peechatt et al., 2024)	focus of attention without mutual engagement (Sabry et al., 2021) facial action units from OpenFace (Fenech et al., 2022)
Sensor 4.3	group and individual proxemics and kinesics features (Sabry et al., 2021)	bodily motion energy synchrony (Kantharaju and Pelachaud, 2021)	underexplored	galvanic skin response (GSR) (Peechatt et al., 2024)	traveled distance, kinetic energy, posture expansion, amount of walking, and hand gesture (Sabry et al., 2021)
Cross-modality 4.4	representation based leadership (Sabry et al., 2021) interpersonal synchrony (Sassier-Roublin et al., 2025)	mutual gaze instance during interruption (Kantharaju et al., 2020)	MUMIN coding on turn management (Murray and Oertel, 2018)	underexplored	ratio between successful interruptions and speaking turns (Murray and Oertel, 2018)

Table 3: Features are categorized by modality and their application in group-level and individual-level collaboration analysis, highlighting current trends and potential underexplored areas in CollA research. Detailed discussions of each modality are referenced in their respective sections.

4.1. Text-Based Features

Text-based features coming from lexical, syntactic, and semantic properties have been studied for CollA. *Pronoun usage* (*e.g.*, comparison of singular/plural pronoun usage, 1st/2nd/3rd-person pronoun usage) can reflect group cohesion level (Enayet and Sukthankar, 2021a). *Discourse markers* (*e.g.*, "okay", "but", "because") signal the communicative function of a phrase (*e.g.*, agreement, disagreement) and they can be used as individual-level collaborative behavior features (Koutsombogera and Vogel, 2018),

Both lexical and syntactic *entrainment*, describing how team members adopt similar speaking styles during conversation, have been studied: syntactic entrainment, calculated using automatic part-of-speech tagging, has been shown to be an effective predictor of team performance, but is expected to be effective only in the late stages of collaboration (Enayet and Sukthankar, 2021a). Lexical entrainment of function words based on LIWC-derived categories of function words (Pennebaker et al., 2001) has been used to identify influencers, connectors, and passive members (Rahimi and Litman, 2020) for multiparty collaboration.

BERT-based pretrained models (Devlin et al., 2019; Liu et al., 2019; He et al., 2020) can be used for text embedding generation in conversation, mapping high-dimensional spaces to low dimensions while retaining only the most effective representations as sparse vectors. This approach has been employed for feature generation in engagement modeling (Li et al., 2024) and conflict modeling (Enayet and Sukthankar, 2023a).

4.2. Audio

Audio of the collaborator's speech contains many useful features for CollA, such as *intensity*, *frequency*, and their variations (*e.g.*, shimmer, jitter). They can be used to measure speaking energy, pitch, voice quality and excitement, which have been found to be positively correlated with both extreme emotion like frustration (Peechatt et al., 2024) and group cohesion (Litman et al., 2016), and often require cross-modality verification. For frustration identification, *voice features* (*e.g.*, F0, intensity) can be more salient than *visual features*, such as chin raises and brow furrows (Peechatt et al., 2024). Audio features can also automatically be extracted using tools such as OpenSMILE (Eyben et al., 2010) and pretrained wav2vec (Schneider et al., 2019). This has been applied in student engagement prediction (Li et al., 2024).

4.3. Video and Sensor Signal

Eye Gaze Eye gaze refers to the direction and movement of a person's eyes, often used in communication to signal attention, engagement, and turn-taking (Kraaij et al., 2005). It helps regulate conversational flow; for example, it has been found that speakers avert their gaze to signal turn initiation and re-establish eye contact to yield the floor (Hung and Gatica-Perez, 2010). *Mutual gaze*, indicating shared attention, is found to be related to social cohesion (Kantharaju and Pelachaud, 2021). Additionally, metrics such as *saccade peak velocity*, which can be measured via eye-tracking sensors, have been shown to indicate frustration during collaboration (Peechatt et al., 2024).

Facial Expression Perceived facial expressions provide insights into participants' emotional states and can contribute to the functional meanings of human interactions. For example, *lip corner puller*, automatically extracted using OpenFace (Amos et al., 2016), is observed as a more frequent and longer action in high-cohesive segments than in low-cohesive segments for small group meetings, while there are no significant differences for *outer brow raiser* and *brow lowerer* (Kantharaju et al., 2020). Kantharaju and Pelachaud (2021) conduct a comprehensive facial unit study to examine its correlation with high- and low-cohesive segments in the collaboration process.

Head, Hand, Body Motion and Sensor Signals Interlocutors synchronize in overall body movement. Bodily motion energy synchrony, a simplified version of interpersonal synchrony, is frequently observed in highly cohesive teams (Kantharaju and Pelachaud, 2021). Sabry et al. (2021) compute *proxemics* (*e.g.*, interpersonal distance) and *kinesics* (*e.g.*, amount of walking, energy synchrony) features of motion caption to model the emergent leadership and group cohesion. Biophysical signals, such as *Galvanic Skin Response* (GSR), provide physiological indicators of states, such as frustration, during collaboration (Peechatt et al., 2024).

4.4. Dialogic and Cross-Modality

Dialogue act (DACT) annotation codes the speaker's intention. Kantharaju et al. (2020) study 15 DACTs following the coding scheme used in the corpus AMI[3] and find the "Be-Positive" DACT as highly related to group cohesion. However, as far as we know, no other DACTs have experimentally been proven to be salient in CollA.

Turn-taking management can reveal both participant dominance and disengagement (Koutsombogera and Vogel, 2018), but requires further validation from other modalities to determine specific functions of turns. The *total pause time* during the collaborative conversation can reflect participants' attentiveness and is consistently high in highly cohesive meetings (Hung and Gatica-Perez, 2010).

Cross-modality inter-speaker features are often more grounded than unimodal individual features for group analysis. Hung and Gatica-Perez (2010) state that *silent motion* (*i.e.*, gestures when not talking), *audiovisual synchrony* (*e.g.*, mirroring another's gesture, finishing another's turn), and *backchannels* (*e.g.*, "um-huh") accompanied by head nodding can indicate engagement and support.

Audiovisual synchrony, within an individual (*e.g.*, body gestures aligned with speech) or across group members (*e.g.*, interlocutors align with one another in both motion and prosody) is a strong indicator of task cohesion among the group (Hung and Gatica-Perez, 2010). Both the *number of interruptions* and the *number of mutual gaze instances occurring during interruption* are positively correlated with group cohesion (Kantharaju et al., 2020). *Overlapping speech*, combined with *visual expressions* and *prosodic energy*, helps identify dominance and task cohesion (Hung and Gatica-Perez, 2010). Shared laughter can be observed more frequently and can also last longer in high-cohesion situations (Kantharaju and Pelachaud, 2021).

4.5. Main Trends of CollA Features

Many features have been shown to be statistically significant and are used in collaboration modeling, from collected individual low-level cues to assembled group-level constructs (*e.g.*, entrainment, convergence). To better capture group-level collaborative phenomena, some studies explored feature-level fusion across modalities, or used cross-modality validation to obtain better-grounded features (Hung and Gatica-Perez, 2010; Kantharaju et al., 2020; Kantharaju and Pelachaud, 2021).

We also want to highlight the usage of pretrained models for feature embedding in CollA studies. This approach has been seen a lot in recent CollA, as it enables cross-corpora generalization in the results. However, its effectiveness depends heavily on the robustness and stability of the underlying pretrained models. When applying such models for feature extraction, careful consideration must be given to the alignment between the pretraining data and the intended modeling objectives. For instance, studies have shown that automatically extracted facial action units from pretrained models performed poorly in classifying self-assessed frustration during collaboration (Peechatt et al., 2024), highlighting the importance of validating pretrained feature extractors for specific CollA tasks.

5. Models

As discussed in Section 3, existing CollA corpora can include both task-level collaboration annotations as well as segment-level annotations. In this section, we discuss existing collaboration models based on segment-level and task-level annotations, comparing them across their modeling objects, modeling approaches, feature designs, and modality fusion.

The objective is to understand the possible reasons behind different choices of analysis granularity in the CollA research community and identify valuable future directions. We discuss CollA studies that use at least one of the corpora listed in Section 3.

[3] https://groups.inf.ed.ac.uk/ami/corpus/Guidelines/dialogue_acts_manual_1.0.pdf

5.1. Predicting Task-Level Annotations

Task-level collaboration annotations have previously been used to model team collaboration quality (Litman et al., 2016; Haber et al., 2019), individual collaboration effort (McDuff et al., 2017), and group cohesion (Maman et al., 2020).

These evaluations are suitable for correlation studies between collaboration and **task-level, temporal-difference** CollA features, such as entrainment (Rahimi and Litman, 2020; Paletz et al., 2023), convergence (Rahimi and Litman, 2018), dominance (Vogel et al., 2023) that change over time during the collaboration process. We also observe that when modeling with task-level annotations, features are typically aggregated across the entire session (*e.g.*, average, standard deviation, min max value, distribution (Walocha et al., 2020)).

Multimodal embeddings of conversations generated by pre-trained models have also been tested in the prediction of task-level collaboration (Enayet and Sukthankar, 2021a, 2023a; Rahimi and Litman, 2020). For perceived collaboration quality, scaling data for supervised learning is relatively easier, since task-level evaluations of perceived dimensions can be added to existing task-oriented corpora, enabling model training on combined corpora. We observe that there are more deep-learning approaches (*e.g.*, LSTM (Enayet and Sukthankar, 2023b,a) for conflict prediction, multimodal transformer (Fenech et al., 2022) for personality modeling) applied to combined, relatively large corpora with task-level evaluation (Enayet and Sukthankar, 2023b,a). We observe that deep-learning approaches (*e.g.*, LSTM (Enayet and Sukthankar, 2023b,a) for conflict prediction and multimodal transformers (Fenech et al., 2022) for personality modeling) are more common in studies using combined, relatively large corpora with task-level evaluations.

5.2. Predicting Segment-Level Annotations

Most segment-level annotations are not directly for group-level collaboration quality modeling, but rather for group-level emergent collaboration phenomena modeling (Kantharaju et al., 2020; Peechatt et al., 2024). Due to the context-dependent nature of collaboration, existing segment-level direct collaboration annotations can be too scenario-specific to be utilized in further studies of different collaborative situations (Richey et al., 2016). Collaboration-related aspects with broader applicability to human interaction (*e.g.*, engagement (Reverdy et al., 2022), cohesion (Kantharaju et al., 2020), and frustration (Peechatt et al., 2024)) have attracted attention from the research community.

Segment-level annotations enable research on emergent group behavior and phenomena by leveraging low-level cues and sequential temporal features that capture their context-dependent nature. However, collaboration modeling based on these segment-level annotations still faces several challenges. First, the relatively small, often imbalanced datasets limit the choice of supervised models to classical classifiers such as support vector machines and logistic regression (Enayet and Sukthankar, 2021a; Kantharaju and Pelachaud, 2021). Second, the automatically extracted multimodal features can have alignment issues, and it is challenging to create a common representation space that preserves cross-modal relationships without losing modality-specific nuances.

Given data scarcity, solutions have been explored at different levels: at the data level, corpus combination (Enayet and Sukthankar, 2023b,a), and synonym replacement in conversational text (Enayet and Sukthankar, 2023a); at the feature level, features oversampling (Corbellini et al., 2023). These methods require a careful adaptation to the modeling object to avoid introducing bias: for example, data augmentation with synonym replacement can delete lexical entrainment in collaboration dynamics. At the model level, graph-based neural network (GNN) methods have achieved state-of-the-art performance on the RoomReader corpus, which contains only 8 hours of recordings (Li et al., 2024), and have remained effective for social interaction classification across several corpora with different task settings (Corbellini et al., 2023). We therefore expect more studies on the potential of GNNs and other graph-learning methods in CollA.

5.3. State-Of-The-Art Performance

Direct comparison of results remains challenging due to variations in collaboration tasks, group sizes, contextual settings, analytical approaches, and evaluation metrics.

Model performance in collaboration analysis depends strongly on task formulation and label structure. Model performance in collaboration analysis depends strongly on task formulation and label structure. For example, binary classification tasks, such as high/low group cohesion or high/low arousal, tend to yield higher accuracy, reaching up to 78% on AMI and PCC (Kantharaju and Pelachaud, 2021). Imbalanced multiclass problems are more challenging: in four-class group-level Q-code classification, average unweighted accuracy drops to around 49% on the relatively small SRI corpus (Bassiou et al., 2016). Similarly, tasks with well-defined and balanced classes, such as binary engagement detection, can exceed 90% accuracy with multimodal features (Li et al., 2024).

Overall, deep-learning approaches outperform classical methods in both performance and scalability for multimodal modeling, especially on larger or combined datasets.

For conflict prediction, Rahimi and Litman (2018) achieves 67.74% accuracy using SVM, while (Enayet and Sukthankar, 2021b) use an LSTM for feature engineering and achieve 73.33% accuracy on GitHub Issue Dataset[4], with training performed on the Teams corpus.

Late fusion strategies, which combine modality-specific embeddings at the decision level, lose modality interactions but have been adopted in recent studies as a trade-off to enable training on several corpora. These results underscore the importance of both task formulation and model architecture in advancing automatic collaboration analysis.

6. Conclusion and Future Direction

This survey provides an overview of recent advances in the analysis of collaboration through human-human conversational corpora. We discuss coding schemes for individual- and group-level collaboration annotation, different task settings for building the CollA corpora, salient multimodal features, and modeling approaches with different granularity. We highlight the evolution from classical statistical methods to deep learning and large language models, as well as the growing integration of multimodality to provide a more nuanced understanding of collaborative processes.

Several directions for future research on collaboration analysis emerge. First, deeper analysis and modeling of individual collaboration strategies using linguistic frameworks would enhance our understanding of collaboration dynamics (Haber et al., 2019). Then, more public multimodal corpora for CollA are needed, especially with recording settings that enable group-level multimodal feature capture, such as room-level video (Koutsombogera and Vogel, 2018) or wearable sociometric badges used in TeamSense corpus[5] (Zhang et al., 2018). Zero-few shot and in-context learning approaches have been applied in many recent studies for the CollA of human-machine collaboration systems. These studies often choose an aspect of human collaboration evaluation and aim to align human-machine or machine-only collaborative conversations towards human level, while human conversations are inevitably downgraded in their modality diversity to be comparable. To the best of our knowledge, the use of these methods for CollA within multimodal human conversational data is understudied.

A key insight is that collaboration analysis is inherently task-oriented and interpersonal, which is driving its growing application in human-centered human-machine collaboration systems. Automatic analysis of collaboration through conversational data is a rapidly evolving field. Continued interdisciplinary efforts, combining linguistics, computer science, psychology, and education, will be essential to address challenges and unlock the full potential of collaborative technologies in both research and real-world applications.

[4] https://github.com/ayeshaEnayet/DAC-USE

[5] TeamSense is not a public corpus and is not included in the range of this review.

7. Acknowledgements

Comments from the reviewers are greatly appreciated. Special thanks also go to Aina Garí Soler and Elodie Etienne for their comments. This work was funded by the ANR-23-CE23-0033-01 SINNet project and additional support was provided by the ANR under the France 2030 program PRAIRIE (ANR-23-IACL-0008).

8. Bibliographical References

Brandon Amos, Bartosz Ludwiczuk, Mahadev Satyanarayanan, et al. 2016. Openface: A General-Purpose Face Recognition Library with Mobile Applications. *CMU School of Computer Science*, 6(2):20.

Nikoletta Bassiou, Andreas Tsiartas, Jennifer Smith, Harry Bratt, Colleen Richey, Elizabeth Shriberg, Cynthia M D'Angelo, and Nonye Alozie. 2016. Privacy-Preserving Speech Analytics for Automatic Assessment of Student Collaboration. In *Interspeech*, pages 888–892.

McKenzie Braley and Gabriel Murray. 2018. The Group Affect and Performance (GAP) Corpus. In *Proceedings of the Group Interaction Frontiers in Technology*, GIFT'18. Association for Computing Machinery.

Michael T Braun, Steve WJ Kozlowski, Tara A Brown, and Richard P DeShon. 2020. Exploring the Dynamic Team Cohesion–Performance and Coordination–Performance Relationships of Newly Formed Teams. *Small Group Research*, 51(5):551–580.

Albert V Carron and Lawrence R Brawley. 2000. Cohesion: Conceptual and Measurement Issues. *Small group research*, 31(1):89–106.

Milly Casey-Campbell and Martin L Martens. 2009. Sticking It All Together: A Critical Assessment of the Group Cohesion–Performance Literature. *International Journal of Management Reviews*, 11(2):223–246.

Federica Cavicchio and Massimo Poesio. 2012. The Rovereto Emotion and Cooperation Corpus: A New Resource to Investigate Cooperation and

Emotions. *Language Resources and Evaluation*, 46(1):117–130.

Yuxin Chen, Christopher D Andrews, Cindy E Hmelo-Silver, and Cynthia D'Angelo. 2020. Coding Schemes as Lenses on Collaborative Learning. *Information and Learning Sciences*, 121(1-2):1–18.

Herbert H Clark. 1996. *Using Language*. Cambridge University Press.

Herbert H Clark and Deanna Wilkes-Gibbs. 1986. Referring as a Collaborative Process. *Cognition*, 22(1):1–39.

Nicola Corbellini, Jhony H Giraldo, Giovanna Varni, and Gualtiero Volpe. 2023. Few Labels are Enough! Semi-Supervised Graph Learning for Social Interaction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3060–3068.

Bethan L Davies. 1997. Empirical Examination of Cooperation, Effort and Risk in Task-Oriented Dialogue. *Annexe Thesis Digitisation Project 2016 Block 5*.

Bethan L Davies. 2007. Grice's Cooperative Principle: Meaning and Rationality. *Journal of pragmatics*, 39(12):2308–2331.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *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 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

James E Driskell, Eduardo Salas, and Sandra Hughes. 2010. Collective Orientation and Team Performance: Development of an Individual Differences Measure. *Human factors*, 52(2):316–328.

Ayesha Enayet and Gita Sukthankar. 2021a. Analyzing Team Performance with Embeddings from Multiparty Dialogues. In *2021 IEEE 15th International Conference on Semantic Computing (ICSC)*, pages 33–39. IEEE.

Ayesha Enayet and Gita Sukthankar. 2021b. Learning a Generalizable Model of Team Conflict from Multiparty Dialogues. *International Journal of Semantic Computing*, 15(04):441–460.

Ayesha Enayet and Gita Sukthankar. 2023a. A Proactive and Generalizable Conflict Prediction Model. In *2023 IEEE 17th International Conference on Semantic Computing (ICSC)*, pages 216–220. IEEE.

Ayesha Enayet and Gita Sukthankar. 2023b. Improving the Generalizability of Collaborative Dialogue Analysis With Multi-Feature Embeddings. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 3551–3565, Dubrovnik, Croatia. Association for Computational Linguistics.

Randi A Engle and Faith R Conant. 2002. Guiding Principles for Fostering Productive Disciplinary Engagement: Explaining an Emergent Argument in a Community of Learners Classroom. *Cognition and instruction*, 20(4):399–483.

Florian Eyben, Martin Wöllmer, and Björn Schuller. 2010. Opensmile: The Munich Versatile and Fast Open-Source Audio Feature Extractor. In *Proceedings of the 18th ACM international conference on Multimedia*, pages 1459–1462.

Kristian Fenech, Ádám Fodor, Sean P Bergeron, Rachid R Saboundji, Catharine Oertel, and András Lőrincz. 2022. Perceived Personality State Estimation in Dyadic and Small Group Interaction with Deep Learning Methods. *arXiv preprint arXiv:2211.04979*.

Patricia Goldberg, Ömer Sümer, Kathleen Stürmer, Wolfgang Wagner, Richard Göllner, Peter Gerjets, Enkelejda Kasneci, and Ulrich Trautwein. 2021. Attentive or Not? Toward a Machine Learning Approach to Assessing Students' Visible Engagement in Classroom Instruction. *Educational Psychology Review*, 33(1):27–49.

Herbert Paul Grice. 1975. Logic and Conversation. *Syntax and semantics*, 3:43–58.

Barbara J Grosz and Candace L Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational linguistics*, 12(3):175–204.

Shengyue Guan, Haoyi Xiong, Jindong Wang, Jiang Bian, Bin Zhu, and Jian-guang Lou. 2025. Evaluating LLM-Based Agents for Multi-Turn Conversations: A Survey. *arXiv preprint arXiv:2503.22458*.

Ao Guo, Atsumoto Ohashi, Ryu Hirai, Yuya Chiba, Yuiko Tsunomori, and Ryuichiro Higashinaka. 2025. User Personality and Its Influence on the Performance of Pipeline and End-to-End Task-Oriented Dialogue Systems. *Scientific Reports*, 15(1):20745.

Janosch Haber, Tim Baumgärtner, Ece Takmaz, Lieke Gelderloos, Elia Bruni, and Raquel Fernández. 2019. The PhotoBook Dataset: Building Common Ground through Visually-Grounded Dialogue. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1895–1910, Florence, Italy. Association for Computational Linguistics.

Pengcheng He, Xiaodong Liu, Jianfeng Gao, and Weizhu Chen. 2020. Deberta: Decoding-enhanced bert with disentangled attention. *arXiv preprint arXiv:2006.03654*.

Peter A Heeman and Graeme Hirst. 1995. Collaborating on Referring Expressions. *arXiv preprint cmp-lg/9504003*.

Hayley Hung and Daniel Gatica-Perez. 2010. Estimating Cohesion in Small Groups Using Audio-visual Nonverbal Behavior. *IEEE Transactions on Multimedia*, 12(6):563–575.

Patrícia Jaques, Adja Andrade, João Jung, Rafael Bordini, and Rosa Vicari. 2023. Using Pedagogical Agents to Support Collaborative Distance Learning. In *Computer support for collaborative learning*, pages 546–547. Routledge.

David W Johnson and Roger T Johnson. 2009. An Educational Psychology Success Story: Social Interdependence Theory and Cooperative Learning. *Educational researcher*, 38(5):365–379.

David W Johnson and Roger T Johnson. 2013. Cooperation and the Use of Technology. In *Handbook of research on educational communications and technology*, pages 777–803. Routledge.

Reshmashree Bangalore Kantharaju, Caroline Langlet, Mukesh Barange, Chloé Clavel, and Catherine Pelachaud. 2020. Multimodal Analysis of Cohesion in Multi-Party Interactions. In *Proceedings of the twelfth language resources and evaluation conference*, pages 498–507.

Reshmashree Bangalore Kantharaju and Catherine Pelachaud. 2021. Social Signals of Cohesion in Multi-Party Interactions. In *Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents*, pages 9–16.

Maria Koutsombogera and Carl Vogel. 2018. Modeling Collaborative Multimodal Behavior in Group Dialogues: The MULTISIMO Corpus. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

Wessel Kraaij, Thomas Hain, Mike Lincoln, and Wilfried Post. 2005. The AMI Meeting Corpus. In *Proc. International Conference on Methods and Techniques in Behavioral Research*, pages 1–4.

Ming Li, Xiaosheng Zhuang, Lu Bai, and Weiping Ding. 2024. Multimodal Graph Learning Based on 3D Haar Semi-Tight Framelet for Student Engagement Prediction. *Information Fusion*, 105:102224.

Diane Litman, Susannah Paletz, Zahra Rahimi, Stefani Allegretti, and Caitlin Rice. 2016. The Teams Corpus and Entrainment in Multi-party Spoken Dialogues. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1421–1431.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A Robustly Optimized Bert Pretraining Approach. *arXiv preprint arXiv:1907.11692*.

Nichola Lubold and Heather Pon-Barry. 2014. Acoustic-prosodic Entrainment and Rapport in Collaborative Learning Dialogues. In *Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge*, pages 5–12.

Joni Lämsä, Pablo Uribe, Abelino Jiménez, Daniela Caballero, Raija H. Hämäläinen, and R. Araya. 2021. Deep Networks for Collaboration Analytics: Promoting Automatic Analysis of Face-to-Face Interaction in the Context of Inquiry-Based Learning.

Lucien Maman, Eleonora Ceccaldi, Nale Lehmann-Willenbrock, Laurence Likforman-Sulem, Mohamed Chetouani, Gualtiero Volpe, and Giovanna Varni. 2020. Game-On: A Multimodal Dataset for Cohesion and Group Analysis. *IEEE Access*, 8:124185–124203.

Roberto Martinez-Maldonado, Dragan Gašević, Vanessa Echeverria, Gloria Fernandez Nieto, Zachari Swiecki, and Simon Buckingham Shum. 2021. What Do You Mean by Collaboration Analytics? A Conceptual Model. *Journal of Learning Analytics*, 8(1):126–153.

Daniel McDuff, Paul Thomas, Mary Czerwinski, and Nick Craswell. 2017. Multimodal Analysis of Vocal Collaborative Search: A Public Corpus and Results. In *Proceedings of the 19th acm international conference on multimodal interaction*, pages 456–463.

Gabriel Murray and Catharine Oertel. 2018. Predicting Group Performance in Task-based Interaction. In *Proceedings of the 20th ACM Inter-*

national Conference on Multimodal Interaction, pages 14–20.

Marjolein C Nanninga, Yanxia Zhang, Nale Lehmann-Willenbrock, Zoltán Szlávik, and Hayley Hung. 2017. Estimating Verbal Expressions of Task and Social Cohesion in Meetings by Quantifying Paralinguistic Mimicry. In *Proceedings of the 19th ACM international conference on multimodal interaction*, pages 206–215.

Heather L O'Brien and Elaine G Toms. 2010. The Development and Evaluation of a Survey to Measure User Engagement. *Journal of the American Society for Information Science and Technology*, 61(1):50–69.

Jennifer K Olsen, Kshitij Sharma, Nikol Rummel, and Vincent Aleven. 2020. Temporal Analysis of Multimodal Data to Predict Collaborative Learning Outcomes. *British Journal of Educational Technology*, 51(5):1527–1547.

Christopher Olshefski, Luca Lugini, Ravneet Singh, Diane Litman, and Amanda Godley. 2020. The Discussion Tracker Corpus of Collaborative Argumentation. *Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC'20)*.

Susannah BF Paletz, Diane Litman, Valerie Karuzis, Kelly M Jones, and Zahra Rahimi. 2023. Speaking Similarly: Team Personality Composition and Acoustic-Prosodic Entrainment. *Small Group Research*, 54(6):860–898.

Michael Peechatt, Cecilia Ovesdotter Alm, and Reynold Bailey. 2024. MULTICOLLAB: A Multimodal Corpus of Dialogues for Analyzing Collaboration and Frustration in Language. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 11713–11722.

James W Pennebaker, Martha E Francis, Roger J Booth, et al. 2001. Linguistic Inquiry and Word Count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates*, 71(2001):2001.

Mannes Poel, Ronald Poppe, and Anton Nijholt. 2008. Meeting Behavior Detection in Smart Environments: Nonverbal Cues that Help to Obtain Natural Interaction. In *2008 8th IEEE International Conference on Automatic Face & Gesture Recognition*, pages 1–6. IEEE.

Sambit Praharaj, Maren Scheffel, Hendrik Drachsler, and Marcus Specht. 2021. Literature Review on Co-located Collaboration Modeling Using Multimodal Learning Analytics—Can We Go the Whole Nine Yards? *IEEE Transactions on Learning Technologies*, 14(3):367–385.

Elisa Prati, Valeria Villani, Fabio Grandi, Margherita Peruzzini, and Lorenzo Sabattini. 2021. Use of Interaction Design Methodologies for Human-Robot Collaboration in Industrial Scenarios. *IEEE Transactions on Automation Science and Engineering*, 19(4):3126–3138.

Zahra Rahimi and Diane Litman. 2018. Weighting Model Based on Group Dynamics to Measure Convergence in Multi-party Dialogue. In *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*, pages 385–390, Melbourne, Australia. Association for Computational Linguistics.

Zahra Rahimi and Diane Litman. 2020. Entrainment2vec: Embedding Entrainment for Multi-Party Dialogues. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 8681–8688.

Nils Randrup, Douglas Druckenmiller, and Robert O Briggs. 2016. Philosophy of Collaboration. In *2016 49th Hawaii international conference on system sciences (HICSS)*, pages 898–907. IEEE.

Justine Reverdy, Sam O'Connor Russell, Louise Duquenne, Diego Garaialde, Benjamin R Cowan, and Naomi Harte. 2022. Roomreader: A Multimodal Corpus of Online Multiparty Conversational Interactions. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 2517–2527.

Colleen Richey, Cynthia M D'Angelo, Nonye Alozie, Harry Bratt, and Elizabeth Shriberg. 2016. The SRI Speech-Based Collaborative Learning Corpus. In *INTERSPEECH*, pages 1550–1554.

Soumaya Sabry, Lucien Maman, and Giovanna Varni. 2021. An Exploratory Computational Study on the Effect of Emergent Leadership on Social and Task Cohesion. In *Companion Publication of the 2021 International Conference on Multimodal Interaction*, pages 263–272.

Omar Salinas, Fabian Riquelme, Roberto Muñoz, Cristian Cechinel, Roberto Martinez, and Diego Monsalves. 2021. Can Analytics of Speaking Time Serve as Indicators of Effective Team Communication and Collaboration? In *Proceedings of the X Latin American Conference on Human Computer Interaction*, pages 1–4.

Dairazalia Sanchez-Cortes, Oya Aran, and Daniel Gatica-Perez. 2011. An Audio Visual Corpus for Emergent Leader Analysis. In *Workshop on* *multimodal corpora for machine learning: taking stock and road mapping the future, ICMI-MLMI*.

Mathilde Sassier-Roublin, Julien Saunier, and Alexandre Pauchet. 2025. Toward Real-Time Cohesion Estimation in Hybrid Groups: A Multimodal Social Signal Processing Approach. In *Proceedings of the 25th ACM International Conference on Intelligent Virtual Agents*, pages 1–9.

Bertrand Schneider and Tonya Bryant. 2022. Using Mobile Dual Eye-Tracking to Capture Cycles of Collaboration and Cooperation in Co-located Dyads.

Steffen Schneider, Alexei Baevski, Ronan Collobert, and Michael Auli. 2019. Wav2vec: Unsupervised Pre-training for Speech Recognition. *arXiv preprint arXiv:1904.05862*.

Verena Schürmann, Nicki Marquardt, and Daniel Bodemer. 2024. Conceptualization and Measurement of Peer Collaboration in Higher Education: A Systematic Review. *Small Group Research*, 55(1):89–138.

Jamie B Severt and Armando X Estrada. 2015. On the Function and Structure of Group Cohesion. In *Team cohesion: Advances in psychological theory, methods and practice*, pages 3–24. Emerald Group publishing limited.

Guy L Siebold. 1999. The Evolution of the Measurement of Cohesion. *Military psychology*, 11(1):5–26.

Daniel Spikol, E Ruffaldi, and M Cukurova. 2017. Using Multimodal Learning Analytics to Identify Aspects of Collaboration in Project-Based Learning.

Jonathan Tudge and Barbara Rogoff. 2014. Peer Influences on Cognitive Development: Piagetian and Vygotskian Perspectives. In *Interaction in human development*, pages 17–40. Psychology Press.

Michelle Vaccaro, Abdullah Almaatouq, and Thomas Malone. 2024. When Combinations of Humans and AI are Useful: A Systematic Review and Meta-Analysis. *Nature Human Behaviour*, 8(12):2293–2303.

Carl Vogel, Maria Koutsombogera, and Justine Reverdy. 2023. Aspects of Dynamics in Dialogue Collaboration. *Electronics*, 12(10):2210.

Fabian Walocha, Lucien Maman, Mohamed Chetouani, and Giovanna Varni. 2020. Modeling Dynamics of Task and Social Cohesion from the Group Perspective Using Nonverbal Motion Capture-Based Features. In *Companion publication of the 2020 international conference on multimodal interaction*, pages 182–190.

Xinru Wang, Hannah Kim, Sajjadur Rahman, Kushan Mitra, and Zhengjie Miao. 2024. Human-LLM Collaborative Annotation Through Effective Verification of LLM Labels. In *Proceedings of the 2024 CHI conference on human factors in computing systems*, pages 1–21.

Donna J Wood and Barbara Gray. 1991. Toward a Comprehensive Theory of Collaboration. *The Journal of Applied Behavioral Science*, 27(2):139–162.

Xigui Yang. 2023. A Historical Review of Collaborative Learning and Cooperative Learning. *TechTrends*, 67(4):718–728.

Jintian Zhang, Xin Xu, Ningyu Zhang, Ruibo Liu, Bryan Hooi, and Shumin Deng. 2024. Exploring Collaboration Mechanisms for LLM Agents: A Social Psychology View. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 14544–14607, Bangkok, Thailand. Association for Computational Linguistics.

Yanxia Zhang, Jeffrey Olenick, Chu-Hsiang Chang, Steve WJ Kozlowski, and Hayley Hung. 2018. TeamSense: Assessing Personal Affect and Group Cohesion in Small Teams Through Dyadic Interaction and Behavior Analysis with Wearable Sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(3):1–22.

Henry Peng Zou, Wei-Chieh Huang, Yaozu Wu, Yankai Chen, Chunyu Miao, Hoang Nguyen, Yue Zhou, Weizhi Zhang, Liancheng Fang, Langzhou He, et al. 2025. A Survey on Large Language Model Based Human-Agent Systems. *Authorea Preprints*.

9. Language Resource References

Braley, McKenzie and Murray, Gabriel. 2018. *The Group Affect and Performance (GAP) Corpus*. Association for Computing Machinery, GIFT'18.

Cavicchio, Federica and Poesio, Massimo. 2012. *The Rovereto Emotion and Cooperation Corpus: A New Resource to Investigate Cooperation and Emotions*. Springer.

Haber, Janosch and Baumgärtner, Tim and Takmaz, Ece and Gelderloos, Lieke and Bruni, Elia and Fernández, Raquel. 2019. *The PhotoBook Dataset: Building Common Ground through*

Visually-Grounded Dialogue. Association for Computational Linguistics.

Kantharaju, Reshmashree B and Pelachaud, Catherine. 2021. *Social Signals of Cohesion in Multi-Party Interactions*.

Koutsombogera, Maria and Vogel, Carl. 2018. *Modeling Collaborative Multimodal Behavior in Group Dialogues: The MULTISIMO Corpus*.

Kraaij, Wessel and Hain, Thomas and Lincoln, Mike and Post, Wilfried. 2005. *The AMI Meeting Corpus*.

Litman, Diane and Paletz, Susannah and Rahimi, Zahra and Allegretti, Stefani and Rice, Caitlin. 2016. *The Teams Corpus and Entrainment in Multi-party Spoken Dialogues*.

Maman, Lucien and Ceccaldi, Eleonora and Lehmann-Willenbrock, Nale and Likforman-Sulem, Laurence and Chetouani, Mohamed and Volpe, Gualtiero and Varni, Giovanna. 2020. *Game-on: A Multimodal Dataset for Cohesion and Group Analysis*. IEEE.

McDuff, Daniel and Thomas, Paul and Czerwinski, Mary and Craswell, Nick. 2017. *Multimodal Analysis of Vocal Collaborative Search: A Public Corpus and Results*.

Olshefski, Christopher and Lugini, Luca and Singh, Ravneet and Litman, Diane and Godley, Amanda. 2020. *The Discussion Tracker Corpus of Collaborative Argumentation*.

Peechatt, Michael and Alm, Cecilia Ovesdotter and Bailey, Reynold. 2024. *MULTICOLLAB: A Multimodal Corpus of Dialogues for Analyzing Collaboration and Frustration in Language*.

Reverdy, Justine and Russell, Sam O'Connor and Duquenne, Louise and Garaialde, Diego and Cowan, Benjamin R and Harte, Naomi. 2022. *Roomreader: A Multimodal Corpus of Online Multiparty Conversational Interactions*.

Richey, Colleen and D'Angelo, Cynthia M and Alozie, Nonye and Bratt, Harry and Shriberg, Elizabeth. 2016. *The SRI Speech-Based Collaborative Learning Corpus.* ISLRN 199-041-455-836-2.

Appendix A - Corpora Selection Process for CollA: A Relatively Less-Resourced Domain

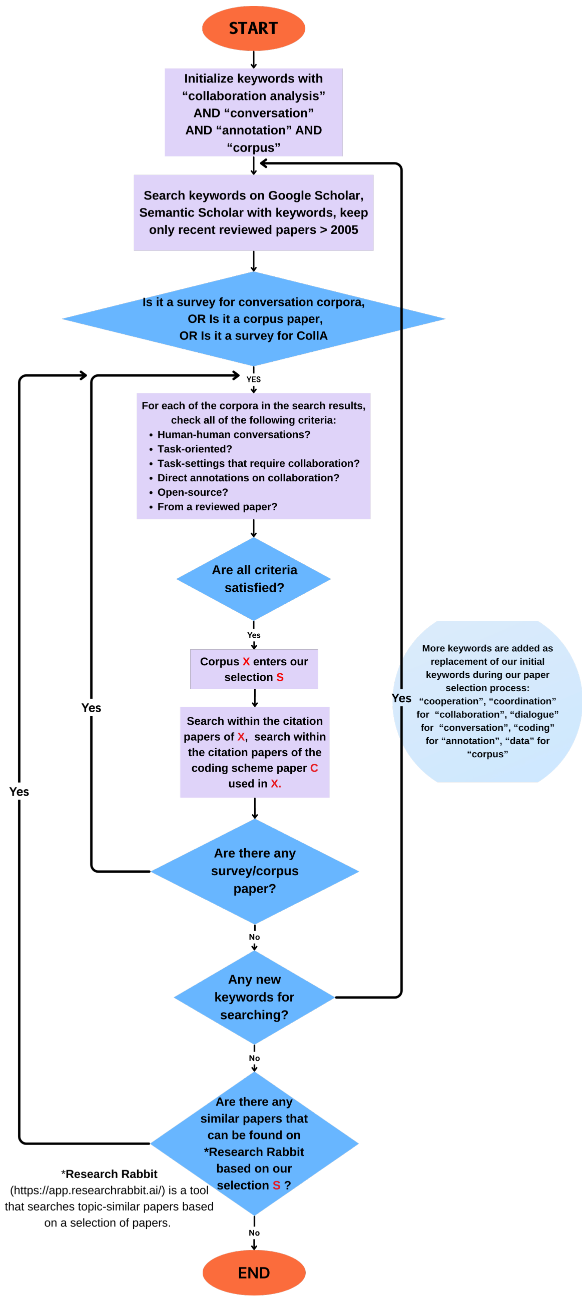

Figure 2: Human-human conversation corpus building with manual annotations is a costly process. It was particularly difficult to find corpora with direct annotations on collaboration. Our recursive approach is time-consuming, but the boundary is clearly defined, which results in a relatively grounded and complete selection for an overview of CollA corpora.