LexiCoord: A Multi-Agent Game for Lexical Ambiguity Resolution between Humans and LLMs

Demonstration Track

Marco Aruta
University of Naples Federico II
Naples, Italy
marco.aruta@unina.it

Francesco Improta
University of Naples Federico II
Naples, Italy
francesco.improta@unina.it

Vadim Malvone
Télécom Paris
Paris, France
vadim.malvone@telecom-paris.fr

Aniello Murano
University of Naples Federico II
Naples, Italy
aniello.murano@unina.it

ABSTRACT

Semantic coordination in Multi-Agent Systems often relies on natural language, where lexical ambiguity can cause misalignment. LexiCoord is a web-based platform that models semantic alignment as a coordination game between humans and LLM agents, who interpret ambiguous utterances and attempt to converge through iterative clarification. The tool integrates heterogeneous LLMs and supports both built-in and user-defined sentences, providing a testbed for studying alignment in mixed human–LLM coalitions.

KEYWORDS

Multi-agent systems; Human-AI interaction; Language-enabled agents

ACM Reference Format:
Marco Aruta, Francesco Improta, Vadim Malvone, and Aniello Murano. 2026. LexiCoord: A Multi-Agent Game for Lexical Ambiguity Resolution between Humans and LLMs: Demonstration Track. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 3 pages.

1 INTRODUCTION

Multi-agent environments [8] in industrial settings increasingly involve heterogeneous entities, namely humans, robots, and other automated systems, that must coordinate to execute complex tasks [4]. In many of these scenarios, natural language (NL) is emerging as the immediate communication channel [1]. NL interactions allow non-expert operators to interface with Large Language Models-based (LLM) agents without specialized programming skills, and support rapid human-like task specification [5]. However, using NL as the primary interface comes with the well-known problem of ambiguity [7]. Words and sentences often have multiple possible meanings and the correct interpretation depends on domain conventions, and shared background knowledge. In a multi-agent setting like the one described above, these assumptions are not guaranteed to be shared or stable [6]. Misinterpretation of task descriptions can lead to coordination breakdowns and ultimately tasks failure. Existing approaches to human–robot communication typically address ambiguity through interface design, constrained vocabularies, or machine learning models trained to contextually infer the most likely interpretation. Sadly, in safety-critical or high-stakes industrial environments, it is not sufficient for interpretations to be "likely" correct; generative AI agents must share the same understanding of key terms before executing coordinated actions. In this paper, we propose a formal-semantics inspired tool to systematically detect and resolve lexical ambiguities in NL communication among cooperating agents. Our approach treats the semantics of ambiguous words as implicit disjunctions of possible meanings [3] within a verification architecture. Declarative task descriptions produced by different agents are mapped into this representation, so that each ambiguous term is associated with a set of candidate meanings rather than a single, fixed interpretation.

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). This work is licenced under the Creative Commons Attribution 4.0 International (CC-BY 4.0) licence.

2 GAME FORMALIZATION

We formally describe the game as follows. Let $Agt = \{a_1, \ldots, a_n\}$ be a finite, non-empty set of agents, with $n \in \mathbb{N}_{>0}$, and let φ be an ambiguous word occurring in a natural-language sentence. Agents populates a finite set of admissible semantic interpretations of φ, $I = \{\iota_1, \ldots, \iota_n\}$, where each ι_k represents one possible disjunctive interpretation of φ, and we identify each ι_k with the k-th disjunct in a disjunction of length n. The game unfolds in discrete rounds $t \in \mathbb{N}$. Let agent a_1 be the coalition leader. In round $t = 0$, a_1 utters a sentence in natural language containing φ and publicly announces an initial interpretation $\iota_1 \in I$, the first disjunct in the disjunction. In each subsequent round $t \geq 1$, every agent $a_i \in Agt$ chooses an interpretation $\iota_i^t \in I$. A round t is said to be *convergent* if all agents select the same interpretation, namely if there exists $\iota^* \in I$ such that for all $i, j \in \{1, \ldots, n\}$ we have $\iota_i^t = \iota_j^t = \iota^*$.

3 THE TOOL

We developed LexiCoord, a web-based multi-agent platform for semantic convergence in mixed human–LLM coalitions[1]. The tool presents users with sentences containing lexical ambiguity and enables human participants and autonomous LLM agents to propose and negotiate interpretations in real time. Each session is hosted in a shared room where multiple humans interact with a

[1]Youtube demo video: https://www.youtube.com/watch?v=ReFltlZVMu8

fixed set of LLM agents through a browser interface. LexiCoord supports two operational modes. In *automatic mode*, ambiguous sentences and candidate meanings are sampled from a predefined dataset. In *manual mode*, the coalition leader (Player 1) introduces a custom ambiguous sentence and specifies the target word, prompting all agents to submit an interpretation and a brief justification. A round converges when all agents agree, otherwise, the system triggers a clarification instance and allows the coalition to iterate until alignment is reached or the leader ends the interaction. The platform tracks win–loss statistics and records all resolved ambiguities together with agent decisions and clarification depth. Figure 2 shows the interface during a running session of joint participation of human and LLM agents.

4 IMPLEMENTATION

LexiCoord is implemented as a client–server application that can be easily embedded into larger multi-agent pipelines. The backend, written in Node.js, manages the convergence protocol, synchronization of browser clients, and interaction with external LLM services. LLM agents are treated as autonomous participants: each receives the same sentence and ambiguity specification as the human players and independently returns an interpretation through a uniform REST-based prompting interface. The system is architecturally organized into three layers: (i) a *coalition leader* that enforces clarification bounds and aggregates agent decisions; (ii) a *human interaction layer* that handles real-time communication via WebSockets; and (iii) an *LLM interface layer* that asynchronously queries heterogeneous cloud-based models and normalizes their outputs. This modular design allows for seamless addition of new LLMs or symbolic agents. LexiCoord stores structured logs of all resolved and unresolved ambiguities. The platform runs as a standalone web application requiring only a browser client and internet access for LLM inference, supporting multiple simultaneous participants in real-time scenarios.

5 EXPERIMENTS

We conducted a simulation campaign to analyze semantic coordination in mixed human-LLM coalitions and to characterize the distribution of convergence outcomes under bounded clarification rounds. The experimental objective to study how humans and artificial agents jointly negotiate the meaning of ambiguous expressions. Each game involved two human participants and four autonomous LLM agents, two Groq-served LLaMA models and two Mistral models, all selecting interpretations independently at every step while the system iteratively enforced a clarification protocol until alignment was reached or the coalition leader terminated the round. Across all games, semantic coordination proved highly robust: *99% converged*, either immediately or after one or more clarification steps, while only *1% failed* to reach agreement. Approximately *70%* of convergent cases were resolved on the first attempt, indicating that autonomous contextual inference often enabled instantaneous alignment among humans and LLMs. The remaining *29%* required additional iterations, typically in the presence of structurally ambiguous or sentences with weakly inferrable contexts. Failures were rare and generally triggered by early human disagreement, which introduced diverging priors that some LLM agents

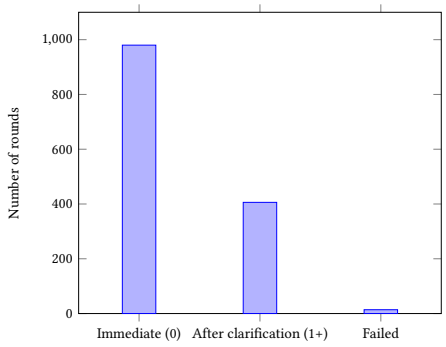


Figure 1: Clarification depth distribution.

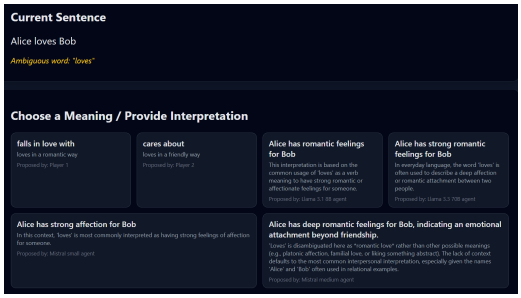


Figure 2: Web-based interface of LexiCoord: a voting example.

reinforced throughout the clarification process. Figure 1 shows that most ambiguities were resolved at depth 0, with roughly one third requiring at least one clarification step. Longer clarification sequences were extremely uncommon and appeared primarily in the small set of non-convergent cases. LLM agents exhibited stable and context-sensitive behaviour across iterations, namely when they accessed the other agents justified choices. When humans initially agreed, the coalition almost always converged immediately; when they disagreed, clarification sequences lengthened and some artificial agents revised their interpretations across rounds as part of the negotiation process. Once human interpretations aligned, however, convergence was almost always achieved.

6 CONCLUSIONS

We introduced LexiCoord, a web-based platform that models semantic alignment as a coordination game between humans and heterogeneous LLM agents. Experiments showed that mixed coalitions typically converge quickly on a shared interpretation, while persistent disagreement remains rare. LLMs generally adapt well within the clarification protocol, yet their behaviour can be strongly influenced by early human disagreement. LexiCoord offers a modular testing environment for Human-AI high-risk interactions where failure is not an option. The current system does not yet model richer discourse context or strategic behaviour. Future extensions will explore integration with formal reasoning frameworks as [2] in which semantic agreement acts as a prerequisite for safe Multi-Agent decision making.

REFERENCES

[1] J. Berg and S. Lu. 2020. Review of Interfaces for Industrial Human-Robot Interaction. *Current Robotics Reports* 1 (2020), 27–34. https://doi.org/10.1007/s43154-020-00005-6

[2] Angelo Ferrando and Vadim Malvone. 2024. VITAMIN: A Tool for Model Checking of MAS. In *Multi-Agent Systems - 21st European Conference, EUMAS 2024, Dublin, Ireland, August 26-28, 2024, Proceedings (Lecture Notes in Computer Science, Vol. to appear)*, Rem Collier, Alessandro Ricci, and Vivek Nallur (Eds.). Springer.

[3] Francesco Improta. 2024. *Intensional superpositions: exploring comprehension and incomprehension in a neurophysical framework.* Master's thesis. Università degli Studi di Padova, Italy. https://hdl.handle.net/20.500.12608/65506 M.A. in Linguistics, 2023/2024.

[4] Wojciech Jamroga, Vadim Malvone, and Aniello Murano. 2019. Natural strategic ability. *Artif. Intell.* 277 (2019). https://doi.org/10.1016/J.ARTINT.2019.103170

[5] I. Maurtua, I. Fernández, J. Kildal, L. Susperregi, A. Tellaeche, and A. Ibarguren. 2017. Natural multimodal communication for human–robot collaboration. *International Journal of Advanced Robotic Systems* (2017).

[6] Christof Monz. [n.d.]. Modeling Ambiguity in a Multi-Agent System. ([n. d.]).

[7] Adam Sennet. 2023. Ambiguity. In *The Stanford Encyclopedia of Philosophy* (summer 2023 ed.), Edward N. Zalta and Uri Nodelman (Eds.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2023/entries/ambiguity/

[8] Michael Wooldridge. 2009. *An introduction to multiagent systems.* John wiley & sons.