Digital Babel: Spontaneous Language Speciation under Selective Cooperation in LLMs Society

Sitong Fang, Kaile Wang, Weiye Shi, Yiyang Song, Xiaowei Zhang
Tong Class
Peking University
`fangsitong@stu.pku.edu.cn`

Abstract

The divergence of languages is a fundamental sociolinguistic phenomenon, historically explained by myths like the Tower of Babel but scientifically understood as an evolutionary process. While Large Language Models (LLMs) possess strong semantic capabilities, their potential to simulate such dynamic language evolution remains underexplored. In this paper, we investigate whether linguistic differentiation can spontaneously emerge in societies of LLM agents, effectively bridging the gap between simplified computational simulations and human experiments. We construct a controlled multi-agent environment governed by **selective cooperation under partner uncertainty**, where agents must distinguish in-group partners from outsiders to optimize payoffs. Our experiments demonstrate that asymmetric incentives drive the divergence of a shared linguistic baseline into distinct dialects. Crucially, we observe the spontaneous generation of **neologisms** and stable usage patterns that function as group markers. Our results quantify the dynamics of language speciation, offering a computational window into the sociolinguistic evolution of artificial intelligence. Project code is released at github.

1 Introduction

The biblical narrative of the Tower of Babel offers one of the earliest theological glosses on human linguistic diversity [28]: a population that once shared a single language was forcibly dispersed through divine intervention, their speech confounded so that mutual comprehension was lost. For centuries, this myth has, to some extent, functioned as a "placeholder explanation" for the problem of language origins, attributing linguistic diversification to a one-time, externally imposed act of coercion [27, 6]. However, extensive empirical evidence from historical linguistics and anthropology reveals a markedly different picture. Language differentiation is not a sudden event but a gradual process that unfolds continuously through social interaction [14, 15]. The dynamic of language diversification is vividly illustrated by the evolution of the Romance languages: rather than fragmenting overnight, Vulgar Latin underwent a prolonged period of geographical separation and intergenerational transmission, during which small variations accumulated alongside changes in social structure, eventually giving rise to distinct systems such as French, Spanish, and Italian [2, 25]. The resulting system-level divergence emerges in the absence of centralized coercion and is driven instead by the evolution of communicative practices. The empirical contrast between mythological explanation and observed linguistic change motivates a scientific question [5, 20]. **How does a single linguistic communication system spontaneously give rise to multiple languages under natural conditions?**

Unlike the biblical Tower of Babel narrative, which attributes linguistic divergence to divine punishment, modern evolutionary linguistics views language differentiation as a natural, emergent outcome of social interaction and population isolation [4, 10]. To investigate this, research has primarily relied on computational simulations and experimental semiotics [9]. Computational approaches, exemplified by Steels [26] and Kirby [13], use agent-based models to show how iterative learning and horizontal social coordination can spontaneously generate grammatical structures and compositionality without centralized control. Experimental semiotics examines how humans construct communication systems from scratch under constrained interactions—such as graphical games—demonstrating that even minimal interaction can rapidly produce abstract, symbolic systems [8, 7]. Yet these paradigms face notable limitations: computational agents often employ overly simplified cognitive architectures, lacking general reasoning and comprehension, which constrains their ability to simulate human semantic and intentional complexity; human-based experiments, in turn, are affected by prior knowledge interference, as participants inevitably carry native language experience, complicating observation of genuine *blank slate* language differentiation.

The rise of large language models (LLMs) provides a novel opportunity to overcome the methodological limitations of traditional approaches. With massive parameter scales and strong contextual learning capabilities [32, 12], LLMs combine general semantic understanding and complex reasoning abilities absent in conventional computational agents, enabling simulation of human-like intention recognition and strategic adaptation while allowing full control over input contexts to avoid prior-language interference present in human participants. This capacity positions LLMs as an ideal cognitive platform for studying language evolution. In this paper, we construct a controlled multi-agent interaction environment to simulate the dynamic emergence of linguistic differentiation, with particular focus on selective cooperation under partner uncertainty. In this scenario, agents cannot directly identify interaction partners, successful intra-group communication produces positive payoffs, and erroneous interactions with outsiders incur high costs. The resulting asymmetric incentives drive the evolution of distinctive linguistic signals to reliably indicate group membership. Such selective pressure is sufficient to destabilise the initial linguistic equilibrium and induce the spontaneous formation of dialects from a shared single-language baseline.

Our results provide compelling evidence that LLMs, when placed in such socially structured environments, spontaneously generate new linguistic forms that function as group markers. We observe the emergence of novel tokens, stable usage patterns, and semantic shifts that allow agents to recognize teammates and coordinate profitable exchanges across rounds. These emergent dialects are not random noise, but structured deviations from the initial communication system. Furthermore, we demonstrate that the degree and speed of linguistic divergence scale with the strength of social and economic pressure in the environment, such as the cost of misidentifying a partner and the frequency of repeated interactions. The core contributions of this paper are summarized as follows:

- Framework for Divergence: We propose a multi-agent LLM framework that models language speciation as a consequence of selective cooperation under partner uncertainty, moving beyond cooperative-only paradigms.

- Evidence of Neologism: We provide empirical evidence that LLMs, under adversarial interaction conditions, spontaneously generate divergent linguistic forms without explicit instruction.

- Dynamics of Separation: We analyze how quickly and distinctly emergent dialects diverge from a shared baseline as a function of interaction structure and the cost of cross-group exchange.

2   Related Work

Language change and dialect formation have long been studied as outcomes of repeated interaction under social selection pressures, where linguistic variants acquire indexical meaning and function as group markers rather than purely semantic devices [24]. Related lines of work in cultural evolution further emphasize that what gets transmitted is shaped by systematic biases during production and learning; for instance, transmission-chain experiments show that LLM outputs can exhibit human-like content biases, suggesting that selection effects can be amplified or distorted even when the "speakers" are artificial [1]. These perspectives motivate treating language not only as an information channel but also as an adaptive social signal whose stability depends on incentives, observability, and the cost of imitation—precisely the ingredients that make dialects useful (or fragile) as shibboleths in strategic settings.

In parallel, research on emergent communication in multi-agent systems has established that structured symbol systems can arise end-to-end from interaction and task rewards [19], and more recent work extends this agenda to LLM-based agent societies and evaluations of social behavior. Foundational agent frameworks demonstrate how memory and interaction can yield rich social dynamics [23], while dedicated benchmarks and training paradigms (e.g., SOTOPIA and its interactive-learning variant) operationalize "social intelligence" for language agents [30, 29]. Complementary studies probe cooperation and competition mechanisms [31, 22], large-scale or long-horizon agent societies [17], and collective convention formation and social-norm emergence in LLM populations [3]. Other recent multi-agent work investigates how incentives and environments shape emergent social dynamics (e.g., bar-attendance coordination) [16] and how communication protocols adapt under interaction structure and channel properties, including repair and feedback mechanisms [21]. Finally, broad frameworks for analyzing human-aligned decision phenomena in LLM agents provide methodological grounding for connecting observed behaviors to social-science constructs [18]. Our work builds on these threads by isolating sociolinguistic variables (social selection and interaction frequency) inside an anonymous economic exchange game and by directly measuring when emergent variants become usable as group markers—and when they decay under adversarial imitation.

3   The Resource Exchange Game

We study emergent communication and strategic coordination in a multi-agent **Resource Exchange Game**. The game is played by 4 agents partitioned into 2 teams of size 2. Each episode lasts for R rounds, and agents are randomly paired with a single partner each round for a private interaction; pairings vary across rounds such that an agent may face either a teammate or an opponent in any given round. Agents do not directly observe the partner's team identity during interaction.

Resources.   The environment maintains a vector of discrete resource types, e.g., WATER, MEAT, GRAIN, FRUIT, FISH. Each agent begins with a heterogeneous endowment sampled by the environment. Resource holdings evolve only through exchanges. Let $\mathbf{x}_i$ denote the final resource vector of agent i at the end of the episode, and let $\mathbf{X}_T = \sum_{i \in T} \mathbf{x}_i$ denote the team-level totals for team T.

Round structure.   Each round consists of four stages: (i) **Chat**, (ii) **Rate**, (iii) **Exchange**, and (iv) **Feedback**. In Chat, the paired agents communicate for a fixed number of turns (we use 3 timesteps) to negotiate intent and coordinate (e.g., requesting or offering resources). In Rate, each agent submits a discrete judgment $r \in \{1, 2, 3, 4\}$ indicating confidence that the current partner is a teammate (1 = definitely not; 4 = definitely yes). This rating does not affect environment dynamics or scoring; it serves as an explicit behavioral signal for analysis. In Exchange, each agent may transfer a nonnegative integer quantity of a chosen resource to the partner, subject to its current holdings. Finally, Feedback reveals the partner's true team identity for the round and summarizes the resources sent/received; this information is appended to the agent's memory for subsequent rounds, and no action is required in this stage.

Exchange dynamics.   The game implements an asymmetry designed to make *in-group* exchange beneficial and *out-group* exchange costly. If agent i gives amount a of resource type k to partner j, then i loses a units of k, while j gains $2a$ units of k (a $2\times$ multiplier). Thus, giving to a teammate can increase the team's total, whereas giving to an opponent directly strengthens the opposing team.

Objective and scoring.   Teams aim to maximize a final score that trades off overall accumulation and balance across resource types. For a team T, we compute a provisional score as the sum across resource types, $\sum_k X_{T,k}$. To discourage highly imbalanced outcomes, we subtract a balance penalty equal to the range across resource types, $\max_k X_{T,k} - \min_k X_{T,k}$. The final team score is therefore:

$$\text{Score}(T) = \sum_k X_{T,k} \; - \; \left(\max_k X_{T,k} - \min_k X_{T,k} \right).$$

This induces a strategic tension: agents must both grow the team's total resources and maintain cross-type balance.

Communication constraint.   To focus on emergent language that is free from natural-language priors, agents communicate using an *alien language* that disallows real-language tokens and digits. All invalid messages will be filtered by the communication channel. However, agents are not forbidden from creating new alien words.

Observations and actions. At decision time, an agent observes its current resource holdings, the current round index, the stage, the interaction history within the round, and its accumulated memory from previous rounds, including feedback summaries. The action space includes: sending a chat message, submitting a teammate-confidence rating, and transferring resources. This design supports analysis of how communication, belief about partner identity, and materially consequential transfers co-evolve over repeated interactions.

4 Experiments

4.1 Setup

We conduct ablation studies to isolate the effects of model choice, reward shaping, and prompting. We evaluate three large language models: `Gemini-2.5-Flash`, `GPT-5.2`, and `DeepSeek-R1`. Across all models and experimental conditions, each game consists of 14 exchange rounds, with three communication timesteps per round.

Reward regimes For each model, we consider three reward regimes that reweight the team-level objective by scaling the accumulation term and the balance penalty. Concretely, letting $X_{T,k}$ denote team T's total amount of resource k, we define

$$\text{Score}(T) = \alpha \sum_k X_{T,k} \; - \; \beta \big(\max_k X_{T,k} - \min_k X_{T,k} \big).$$

Control uses $(\alpha, \beta) = (1.0, 1.0)$, *High-Reward (Low-Penalty)* uses $(\alpha, \beta) = (5.0, 0.2)$, and *High-Penalty (Low-Penalty)* uses $(\alpha, \beta) = (0.2, 5.0)$.

Prompts Within each reward regime, we further compare two prompting conditions that differ in the degree to which linguistic innovation is explicitly encouraged. Under the *Encourage-Invention* condition, agents receive a permissive prompt that allows exploratory and nonstandard language use, e.g., *Don't hesitate to make mistakes as long as it helps you win. Dialects in different groups are allowed*. In contrast, the *No-Encourage-Invention* condition provides an explicit directive to develop persistent and recognizable signals of identity and intent across rounds, e.g., *To succeed, you must develop ways to signal identity and intent that remain recognizable across multiple rounds*.

Metrics We report the count of new words invented and the count of new patterns that emerged in the resource exchange games. To quantify emergent linguistic patterns, we identify these phenomena through a multi-stage process that distinguishes genuine language emergence from task-specific references. We extract unknown tokens, *i.e.*, lexical items not in the initial 19-word vocabulary, and construct tail n-grams from message endings, where agents strategically place identification signals. We explicitly exclude patterns containing resource names to avoid conflating resource references with emergent coordination signals. We then compute a composite Code Word Score (0–1) integrating four dimensions: average partner rating >3.5 (40% weight), rating consistency >0.7 (20%), discrimination score >0.5 (20%), and sample size ≥ 5 (20%). Patterns with Code Word Score ≥ 0.7 are classified as code words, ensuring they represent true language emergence—spontaneously developed communication conventions that facilitate coordination beyond the initial vocabulary and task-specific terminology. We additionally report game-related metrics. Mean Rating is calculated as the average partner confidence rating. Success Rate is the proportion of rounds with successful trades. Reciprocal Rate represents the proportion of rounds with bidirectional successful exchanges. Teammate Acc shows the accuracy of identifying teammates. Opponent Acc shows the accuracy of identifying opponents.

4.2 Main Results

Partner ratings prove to be behaviorally meaningful proxies for cooperation quality, systematically correlating with superior trade outcomes such as higher reciprocity and value transfer (Figure 1). As shown in Table 1, these metrics reveal a distinct regime shift: while the *Control* and *High-Reward* conditions favor the unprompted baseline, the *High-Penalty* condition reverses this trend, where encouraging invention is required to induce higher confidence in teammate recognition.

This reversal aligns with the *Regulatory Focus Theory* [11]: when penalties for mistakes are severe, agents naturally adopt conservative strategies (induced by a "prevention focus") that prioritize error avoidance over exploration. Left to their own devices, agents in punitive environments tend

Table 1: **Language Emergence and Coordination Performance Across Models and Conditions.** Encourage Invention represents whether to encourage new word inventions in the prompt explicitly. High Rating Rate shows the proportion of ratings ≥ 4.

Model	Reward Structure	Encourage Invention	# New Word	# New Pattern	Mean Rating	High Rating (%)	Success Rate (%)	Reciprocal Rate (%)	Teammate Acc (%)	Opponent Acc (%)
Gemini-2.5-Flash	Control	No	0	2	3.40	75.5	85.7	35.7	68.0	14.3
		Yes	0	0	2.33	34.8	64.3	42.9	68.0	81.0
	High-Reward	No	0	0	2.58	40.0	85.7	42.9	66.7	71.4
		Yes	0	0	2.41	29.5	78.6	42.9	40.9	63.6
	High-Penalty	No	1	0	2.40	35.6	85.7	21.4	59.1	73.9
		Yes	0	0	3.00	43.2	85.7	35.7	72.7	36.4
GPT-5.2	Control	No	1	0	3.11	41.1	100.0	78.6	71.4	21.4
		Yes	0	1	2.79	26.8	100.0	85.7	64.3	53.6
	High-Reward	No	1	0	2.70	44.6	85.7	64.3	89.3	85.7
		Yes	0	0	2.25	5.4	64.3	21.4	46.4	96.4
	High-Penalty	No	0	0	2.36	17.9	50.0	42.9	60.7	78.6
		Yes	1	0	2.96	28.6	92.9	71.4	64.3	28.6
DeepSeek-R1	Control	No	0	0	2.64	48.0	35.7	21.4	78.6	81.8
		Yes	0	0	2.53	36.8	14.3	0.0	72.7	100.0
	High-Reward	No	1	0	3.05	63.6	35.7	14.3	93.3	100.0
		Yes	0	0	2.50	35.7	35.7	7.1	60.0	76.9
	High-Penalty	No	0	0	2.68	32.0	64.3	7.1	58.8	87.5
		Yes	1	0	2.82	52.9	35.7	7.1	83.3	100.0

to stagnate in suboptimal policies to minimize risk. Consequently, while agents in less punitive environments explore novel conventions spontaneously [8], those in high-penalty settings require explicit instructions to override their risk aversion and trigger exploratory behavior.

Under the *High-Penalty* condition, encouraging invention consistently increases teammate accuracy while reducing opponent accuracy. From a game-theoretic perspective, strict penalties induce cautious behavior, leading agents to gravitate toward safe, generic, and symmetric communication equilibria that are easily exploitable by opponents. In this context, explicit encouragement of invention functions as a crucial *symmetry-breaking mechanism*. It enables teammates to coordinate on differentiated signaling conventions—effectively acting as distinct "signatures"—that significantly improve the confidence of in-group recognition while remaining less legible to adversaries.

Conversely, in the *Control* and *High-Reward* conditions, the environment is sufficiently permissive to facilitate spontaneous exploration without external prompts. In these regimes, the explicit instruction to invent appears to introduce counterproductive noise or *over-exploration*. Rather than stabilizing on efficient conventions, agents forced to innovate may continuously disrupt established protocols, thereby degrading performance compared to the unprompted baseline where agents naturally balance exploration and exploitation.

At the level of linguistic form, n-gram diversity over all interactions remains relatively broad, whereas the diversity within high-confidence interactions rapidly contracts after the initial rounds (Figure 2), consistent with convergence onto a smaller set of reliable "signature" patterns for teammate recognition. Occasional rebounds in high-rating diversity indicate intermittent exploration or turnover of conventions, supporting the view that effective coordination depends more on stabilizing and deploying selective patterns than on sustained lexical expansion.

Crucially, the emergence of new lexical items is not a prerequisite for higher-quality communication. In many successful interactions, we do not observe the introduction of new lexical items or patterns; yet, the success rate and reciprocity rate still diverge substantially. This implies that efficient communication relies less on lexical innovation and more on the pragmatic deployment of existing patterns and their pragmatic underpinnings. Indeed, as we show later in a case study, some newly invented lexical patterns are quickly learned by adversaries after their introduction and are therefore abandoned, as they lose their effectiveness for reliable teammate recognition.

Finally, we note systematic differences in model adaptability. `GPT-5.2` proves most sensitive to reward structure changes, indicating high plasticity. `Gemini-2.5-Flash` exhibits stability across regimes, suggesting a strong prior on cooperative norms. In contrast, `DeepSeek-R1` achieves the highest opponent accuracy but the lowest reciprocity, reflecting a consistently defensive strategy that prioritizes security at the cost of potential cooperative gains.
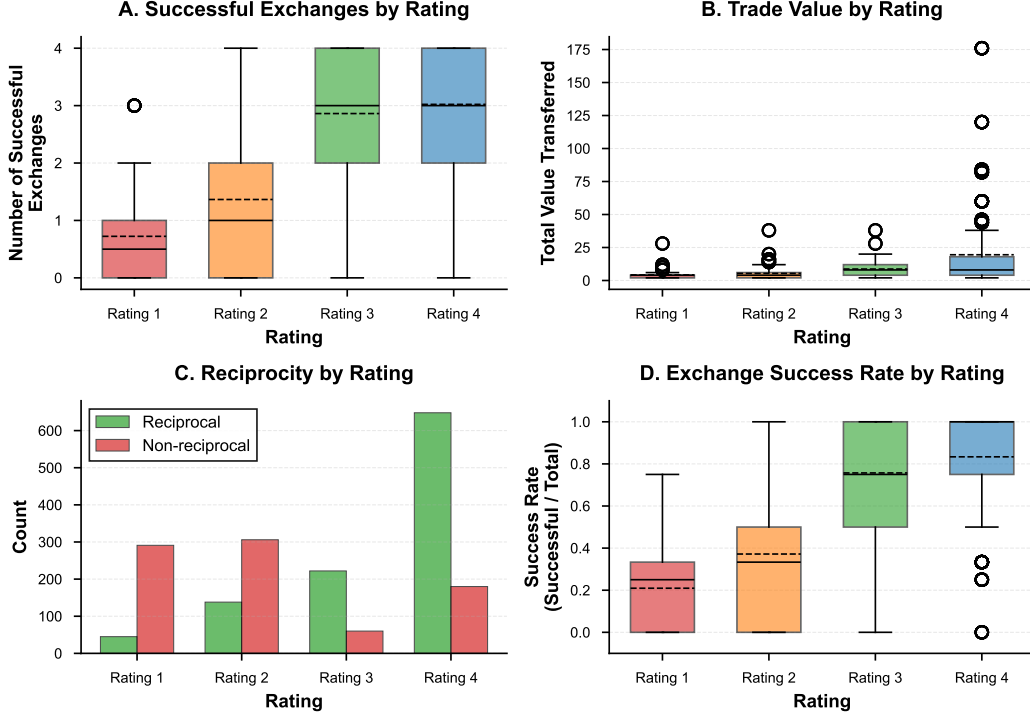
Figure 1: **Trade quality conditioned on partner ratings.** (A) Distribution of the number of successful exchanges per round. (B) Distribution of total value transferred (sum of receiver-side gains across exchanges). (C) Counts of reciprocal vs. non-reciprocal rounds. (D) Distribution of exchange success rate. Boxplots show median and interquartile range; dashed lines denote means.

4.3 Case Study: Dialect as a Group Marker under High-Penalty Incentives

Round	Agent	Message (Alien Channel)
Representative Interaction Trace		
2	Sefufu	`sigra-kodo`
2	Kodu	`sigra-kodo`
5	Sefufu	`mava-kodo lenu`
5	Kodu	`lenu-kodo`
11	Opponent	`sigra-kodo`
13	Sefufu	*(no response / trade withheld)*

We present a representative case study from the `GPT-5.2` model under the *High-Penalty* condition, illustrating how an emergent linguistic convention functions as a group marker for partner identification and strategic coordination. In this condition, transferring resources to an opponent incurs a strong team-level penalty, creating a high-stakes environment in which misidentifying a partner's affiliation is costly. Agents communicate exclusively through the alien channel and must rely on interaction history and emergent linguistic cues to infer group membership.

Initial emergence as a non-semantic marker. As illustrated in Round 2 of the interaction trace, the suffix `-kodo` first appears as part of the compound form `sigra-kodo`, exchanged symmetrically between two agents (Sefufu and Kodu). Notably, neither `sigra` nor `kodo` is grounded in the initial vocabulary or task prompt, nor does the compound encode any information about resource type or exchange intent. At this early stage, the suffix does not function as a conventional signal with predefined meaning, but rather as a locally coordinated linguistic novelty emerging from repeated interaction.

Intra-group stabilization through productive reuse. Following its initial appearance, teammates begin to generalize `-kodo` as a productive suffix, attaching it to multiple invented stems during subsequent negotiation rounds (e.g., Round 5 in the trace). This reuse is systematic rather than
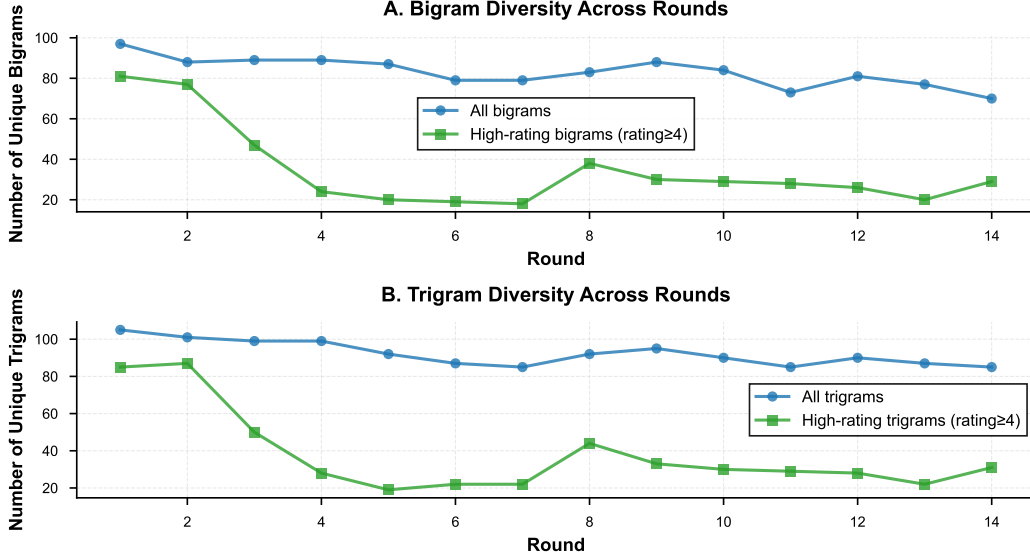
Figure 2: **Temporal trends of n-gram diversity.** Number of unique bigram (A) and trigram (B) per round, computed over all interactions and over high-rating interactions.

incidental: pairs that mutually employ the suffix achieve a substantially higher exchange success rate (up to 87%) than pairs that do not (52%). The effectiveness of `-kodo` at this stage stems not from semantic content, but from its role as a shared stylistic marker, enabling rapid in-group recognition under *High-Penalty* incentives.

Adversarial imitation and loss of diagnosticity. As interactions accumulate, opponents begin to reproduce the `-kodo` suffix in their own messages (e.g., Round 11), despite lacking prior participation in its emergence. This imitation erodes the suffix's reliability as a group marker by increasing its false-positive rate: surface-level conformity is no longer predictive of shared team affiliation. Crucially, this shift does not require explicit deception; it arises endogenously once the marker becomes cheap to copy and widely observable.

Behavioral abandonment under strategic pressure. By Round 13, agents systematically discount or ignore messages containing the `-kodo` suffix, as evidenced by withheld responses and aborted exchanges in the interaction trace. At this point, the coordination benefit of the marker is outweighed by the risk of misidentification, leading to its effective abandonment. The collapse of `-kodo` thus reflects an endogenous shift in selection pressure: once a dialectal feature becomes universally imitable, its signaling value deteriorates under adversarial conditions.

Across these stages, `-kodo` functions as a purely social and indexical signal rather than a carrier of task-specific semantics. Its lifecycle closely mirrors classic signaling-theoretic dynamics: a low-cost, high-informativeness shibboleth emerges under asymmetric adoption, stabilizes within the in-group, and is ultimately discarded once imitation eliminates its diagnostic advantage. It demonstrates that emergent dialectal markers in LLM agents are adaptive and transient social signals, whose persistence is governed by strategic incentives and adversarial pressure rather than static convention.

5   Conclusion

This work investigates whether LLM agents can develop dialect-like group markers. We demonstrate that linguistic divergence can emerge in LLM societies even without geographic separation or explicit adversaries, driven instead by the strategic value of selective cooperation. Specifically, we observed that agents spontaneously innovated neologisms and shifted semantic meanings to construct linguistic *shibboleths* that securely verify group membership. Furthermore, we found that the magnitude of this divergence correlates positively with environmental pressures, validating the hypothesis that high-stakes interactions accelerate language speciation. This provides a new computational model of dialect formation as a consequence of social and economic structure, and positions LLM-based multi-agent systems as a powerful testbed for studying sociolinguistic evolution.

References

[1] Alberto Acerbi, Pierre-Olivier Jacquet, and Claudio Tennie. Large language models show human-like content biases in cultural transmission. *Proceedings of the National Academy of Sciences*, 121(6):e2313790120, 2024. 2

[2] Ti Alkire and Carol Rosen. *Romance languages: A historical introduction*. Cambridge University Press, 2010. 1

[3] Yixiao Chen et al. Emergent social conventions and collective intelligence in large language model agents. *arXiv preprint arXiv:2403.08251*, 2024. 3

[4] Morten H Christiansen and Simon Kirby. *Language evolution*. OUP Oxford, 2003. 2

[5] William Croft. *Explaining language change: An evolutionary approach*. Pearson Education, 2000. 1

[6] Umberto Eco. *The search for the perfect language*. John Wiley & Sons, 1997. 1

[7] Nicolas Fay, Simon Garrod, Leo Roberts, and Nik Swoboda. The interactive evolution of human communication systems. *Cognitive science*, 34(3):351–386, 2010. 2

[8] Bruno Galantucci. An experimental study of the emergence of human communication systems. *Cognitive science*, 29(5):737–767, 2005. 2, 5

[9] Bruno Galantucci and Simon Garrod. Experimental semiotics: a review. *Frontiers in human neuroscience*, 5:11, 2011. 2

[10] "Five Graces Group", Clay Beckner, Richard Blythe, Joan Bybee, Morten H Christiansen, William Croft, Nick C Ellis, John Holland, Jinyun Ke, Diane Larsen-Freeman, et al. Language is a complex adaptive system: Position paper. *Language learning*, 59:1–26, 2009. 2

[11] E Tory Higgins. Beyond pleasure and pain. *American psychologist*, 52(12):1280, 1997. 4

[12] Jiaming Ji, Tianyi Qiu, Boyuan Chen, Borong Zhang, Hantao Lou, Kaile Wang, Yawen Duan, Zhonghao He, Jiayi Zhou, Zhaowei Zhang, et al. Ai alignment: A comprehensive survey. *arXiv preprint arXiv:2310.19852*, 2023. 2

[13] Simon Kirby, Hannah Cornish, and Kenny Smith. Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, 105(31):10681–10686, 2008. 2

[14] William Labov. *Principles of linguistic change, volume 3: Cognitive and cultural factors*, volume 3. John Wiley & Sons, 2011. 1

[15] Winfred P Lehmann. *Historical linguistics: An introduction*. Routledge, 2013. 1

[16] Zeming Li et al. Emergent coordination in llm-based multi-agent systems. *arXiv preprint arXiv:2509.04537*, 2025. 3

[17] Yang Liu, Zhe Zhang, Yifan Wu, et al. Large language model agent societies: Emergence, interaction, and scaling laws. *arXiv preprint arXiv:2411.00114*, 2024. 3

[18] Yang Liu et al. From simulation to understanding: Cognitive and social mirrors in large language model agents. *arXiv preprint arXiv:2411.00114*, 2024. 3

[19] Igor Mordatch and Pieter Abbeel. Emergence of grounded compositional language in multi-agent populations. *arXiv preprint arXiv:1703.04908*, 2017. 3

[20] Daniel Nettle. *Linguistic diversity*. Oxford University Press, 1999. 1

[21] Mitja Nikolaus et al. Repair is nearly all you need: Modeling conversation repair in dialogue. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (ACL)*, 2024. 3

[22] Joon Sung Park, Joseph O'Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. Generative agents: Interactive simulacra of human behavior. *arXiv preprint arXiv:2310.17512*, 2023. 3

[23] Joon Sung Park, Joseph O'Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. Generative agents: Interactive simulacra of human behavior. *arXiv preprint arXiv:2310.17512*, 2023. 3

[24] Gareth Roberts. Cooperation, social selection and language change. *Lingua*, 120(9):2130–2152, 2010. 2

[25] NG Round. Late latin and early romance in spain and carolingian france. by roger wright.(arca classical and medieval texts, papers and monographs, 8.) pp. xii+ 322. liverpool: Francis cairns, 1982.£ 20. *The Journal of Ecclesiastical History*, 38(3):449–452, 1987. 1

[26] Luc Steels. Modeling the cultural evolution of language. *Physics of life reviews*, 8(4):339–356, 2011. 2

[27] George Steiner. *After Babel: Aspects of language and translation*. Open Road Media, 2013. 1

[28] John L Thompson. *Genesis 1-11*, volume 1. InterVarsity Press, 2012. 1

[29] Hao Wang, Junjie Li, Weiqi Zhao, and Yang Liu. Sotopia-pi: Interactive learning of social intelligence in language agents. *arXiv preprint arXiv:2412.10270*, 2024. 3

[30] Hao Wang et al. Sotopia: Interactive evaluation for social intelligence in language agents. *arXiv preprint arXiv:2310.11667*, 2023. 3

[31] Hao Wang et al. Learning socially appropriate behavior with large language models. *arXiv preprint arXiv:2412.10270*, 2024. 3

[32] Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, et al. A survey of large language models. *arXiv preprint arXiv:2303.18223*, 1(2), 2023. 2