Journal of Linguistic Studies
Theory and Practice
https://jls.uok.ac.ir

Research paper
Year (Vol.), ..-.., Season, 2025

Linguistic Behavior and Deceptive Strategies in Mafia Game in the Iranian Context

Shaghayegh Mojiri[1] | **Seyyed Mohammad Razinejad**[2] | **Afsar Rouhi**[3]

[1] MA in Linguistics, Department of Literature and Humanities, University of Mohaghegh Ardabili, Ardabil, Iran.
(Corresponding Author) E-mail: shaghayegh.mojiri@student.uma.ac.ir

[2] Associate professor of Linguistics, Department of Foreign Language Teaching, Faculty of Literature and Humanities, University of Mohaghegh Ardabili, Ardabil, Iran. E-mail: mrazi@uma.ac.ir

[3] Professor of Linguistics, Department of Foreign Language Teaching, Faculty of Literature and Humanities, University of Mohaghegh Ardabili, Ardabil, Iran.
E-mail: afsarrouhi@uma.ac.ir

Article Info

Article type:
Research article

Article history:
Received: 29 May 2025
Accepted:24 Sep. 2025

Keywords:
linguistic behavior,
deception,
Mafia game,
social interaction,
conversation analysis

ABSTRACT

This study explores linguistic patterns in social deception within the mafia game among the Iranian community. It adopts interpersonal deception theory (IDT) as its theoretical foundation, which posits that language use in deceptive contexts differs from normal language patterns, and investigates the validity of this theory in the context of mafia games. Quantitative and qualitative analyses were done on 70 professional mafia players, homogeneous in age and social group. The result revealed significant correlations between linguistic features and deceptive and non-deceptive roles, such as the number of verbs used by players and different verb forms, such as the subjunctive form, past continuous, and negative commands. Other strategies employed by mafia players are introduced as avoiding leadership roles, shifting the blame, the Serial Position Effect, and gaslighting. These findings contribute to understanding deception and language in social interactions, facilitating further research on the linguistic manifestations of deception in social deduction games.

Cite this article: Mojiri, S., Razinejad, M., Rouhi, A. (2025). "Linguistic Behavior and Deceptive Strategies in Mafia Game in the Iranian Context". *Journal of Linguistic Studies: Theory and Practice*, Year (Vol.), …-…..

Publisher: University of Kurdistan.

DOI: ……

1. Introduction

The mafia game, initially invented by a Russian psychology student, Dmitry Davidoff, as a pedagogical tool to merge psychology research with educational practice, was designed to set an informed minority of players against a larger, uninformed majority (Robertson, 2010). At its core, the game is grounded in deception and relies heavily on participants' linguistic performance to expose hidden roles and influence collective decision-making. Initially gaining popularity among youth in educational and social settings within Russia, the game eventually spread to other countries, including Iran, and gained popularity particularly within university communities. During the COVID-19 pandemic, when face-to-face interaction became limited, television programs began broadcasting the mafia game in the studios for an eager audience.

Given the fundamental scarcity of literature on different profiles of the mafia game in the Iranian context, the present study is conducted to examine the linguistic behavior of the players in the mafia game to shed some light on linguistic patterns found in players' speech. The main question to guide the study is: What are the linguistic behaviors of mafia players in the context of the game?

At the heart of the mafia game lies deception, a delicate skill advanced to help humans control how others perceive them and to help them go through the complexities of social life more effectively. It is important to draw the line between deception and a lie. In Bond Jr and DePaulo's (2006) words, a lie is a statement that is completely false and has no truth; however, deception covers a broader aspect. In deception, the deceiver intentionally and consciously uses different techniques and methods to convince a person to believe an idea, whether true or false. In the context of the game, deception shapes the very essence of the gameplay experience. As participants engage in their roles, they practice the art of persuasion, crafting narratives that mask their intentions while probing the motivations of others, to sway the game to their advantage.

Despite the common belief that dishonesty and manipulation are immoral, people, irrespective of their social standings, participate in lying and deception daily. It is, therefore, unsurprising that the concept of deception has attracted the attention of researchers across multiple fields. Psychologists, criminologists, linguists, and other pursuers of knowledge have tried to unveil the mystery of deception and create a scientific formula for understanding and detecting deception, yet many aspects remain undiscovered. Previous research has consistently highlighted the significance of linguistic factors as potential indicators of deception (e.g., Abouelenien et al., 2014). Studies have demonstrated that certain linguistic markers, such as interruption of speech, negation, and hedging, can offer insights into whether a speaker is truthful or deceptive (Bajaj et al., 2023). While the focus on studying mafia games has mostly centered on the mathematical aspects and structure of the game, for example, the frequency of mafia winning, Zhou and Sung (2008) have discovered that there are linguistic distinctions between the mafia players' and citizen players' speech. The contribution of linguistic factors in identifying mafia players becomes clearer through the findings of O'Gara (2023), which indicates that the discussion time in the game affects the accuracy of finding the mafia players. In games with longer discussion times, players were more successful in correctly identifying the mafia players.

Set in a fictional community, the mafia game places citizens and mafia players in opposition, fostering a dynamic of discussion, accusation, and strategic social interaction. Each role carries distinct goals; mafia players collaborate covertly during night phases to eliminate citizens while blending in during the day; citizens, unaware of others' roles, must identify and vote out mafia players through collective reasoning. The game

alternates between night and day phases, balancing asymmetrical knowledge held by the mafia, with numerical advantage held by the citizens. Victory is achieved when all mafia players are eliminated or their numbers equal those of the citizens.

2. A brief note of previous works

To fully understand the linguistic behaviors in the mafia games, it is helpful to consider two distinct lines of research. The first examines the game, exploring the patterns and strategies players employ in communication from a scientific perspective, such as speech duration and lexical diversity. The second draws from broader linguistic studies on deception, bias, and language, offering a wider scope of the topic through the perspective of linguistics in other contexts, such as detecting deception in daily conversation, biased online reviews, and biased news articles.

Evidence from multiple studies highlights the importance of language use in identifying the mafia players. Ibraheem et al. (2022) worked on creating models capable of finding players suspicious of having mafia roles, showing distinct differences in the language used by mafia players compared to citizens. According to this study, suspicious linguistic patterns can be singled out by training specific classifiers. Another research by O'Gara (2023) on the capabilities of artificial intelligence (AI) in playing social deduction games adds to the former study by pointing out that the discussion time in mafia games provides a reliable index for detecting the deceptive players. The longer the discussions are, the more accurate the identification of mafia members becomes.

In addition to the text-based studies mentioned above, Chittaranjan and Hung (2010) explored the role of audio cues in recognizing mafia players, focusing on the relationship between non-verbal signals and player roles. The findings of this study indicate that a combination of factors, such as pitch variation and total speaking duration, can help predict the mafia players more accurately.

In another intriguing study on linguistic behavior, Niculae et al. (2015) examined the linguistic outputs of players of a strategic computer game right before their betrayal. The nature of the game encouraged the players to build trust and friendship with their comrades, build teams, and collaborate, and in later stages of the game, betray the friendship they had built. The result of this study showed that there are subtle hints in the conversation of the players that indicate their impending betrayal. A lasting friendship showed a form of balance manifested in their language use, while any form of linguistic pattern that appeared to disrupt this balance signaled a betrayal.

Zhou and Sung (2008) focused on some linguistic factors in the language of the mafia players and others. Their research found that mafia players tended to communicate less and use less syntactically complex sentences while using a high level of lexical diversity. A significant finding of this study was the variation in results across different cultures. While Zhou and Sung's study showed a higher diversity of messages in deceivers' output, a similar study in America at the time showed opposite results. In addition, self-reference appeared inconsequential in this study as opposed to American research, emphasizing the crucial impact of culture on language and social behavior.

While the work on linguistic factors regarding deception is vast in various contexts and fields, a few of the more relevant works will be discussed here. With the goal of identifying factors for deception, Abouelenien et al. (2014) used a multimodal approach, combining linguistic factors, physiological responses, and thermal sensing. The result of this study showed that linguistic classifiers programmed based on the lexical output of deceivers and truth tellers, and thermal modalities based on a person's facial temperature, trained together, can potentially be good indicators of deception. Additionally, their experiment showed that the quality of the extracted features is topic-related; things such

as negativity and emotion can confuse the result, and encountering a new topic of conversation requires separate training of models to distinguish deception.

Another study, focused on the proximity of linguistic markers and their relation to deception, allowed for some interesting findings (Bajaj et al., 2023). Among their discoveries, they found that an uncertainty marker following an explainer marker is more likely to be deceptive, while an explainer marker following an uncertainty marker is more likely to be truthful. For example, in a sentence like "She said he is suspicious because she saw him running out of the building or something", the presence of "something" after "because" is considered very suspicious and possibly an indicator of deception. This discovery seems logical since, in an attempt to convince the audience, humans tend to follow an uncertain statement with a personal belief or explanation rather than the other way around.

Another fascinating study focused on the strawman fallacy, a form of argument that distorts an opponent's view to make it appear more extreme and therefore, less acceptable (Schumann et al., 2019), showed that a strawman is more accepted when the speaker's argument is attacked rather than their standing point. To clarify, examples from the experiment are brought here. Consider the sentence, "…it is crucial to better support young parents because having a child means a lot of financial charges." as our informative statement; if a person took this statement and rephrased it by saying: "…Let's raise family allowance since it is only about the money." it is a strawman fallacy that misrepresents the previous sentence's argument by making it more extreme. Further experiments showed the strawman fallacy is more likely to be accepted when the two arguments are simply side by side without a logical connector that would explain how the idea leads to another. For example, misrepresentation in a sentence like "…Let's raise the family allowance. It is only about money." is more likely to go unnoticed than a sentence like "…Let's raise the family allowance since it is only about the money.". Without the connectors, the audience is less likely to think critically about the connection between the two ideas, making the acceptance of the fallacy easier. A strawman is also more acceptable when it echoes the speaker's explicit meaning rather than their implicit meaning, as it targets what is directly stated, making the distortion appear more credible (Schumann et al., 2019).

Moving on to research on online dishonest use of language, Kim et al. (2024) studied the linguistic behavior in online fake reviews against the authentic ones. Findings indicated that fake reviews are more likely to appear informative and positive; however, they often lack detailed product-specific content. They concentrate on general recommendations or broader aspects of the product, rather than addressing specific features or performances. Similar studies by Ansari and Gupta (2021) on people's personal opinions about the authenticity of online reviews revealed that properly contextualized reviews with decent coherence are more likely to be considered honest, while non-verbal flattering expressions, such as excessive punctuations, are considered dishonest.

Spinde et al. (2021) concentrated on biased words in news articles, using a feature-based approach, such as considering biased lexicons extracted from previous biased news; a set of media bias data was created for analyzing and identifying bias in news articles. They have managed to achieve 77% accuracy in detecting biased words by considering the background interaction and context in which the words are set.

Kheirabadi and Kheirabadi (2024) conducted similar studies on the linguistic structure of fake news in advertising posts in Iranian social media. They found that such messages tend to be low in informational content while conveying high levels of certainty. Another

study from Malmir et al. (2023) examined nominalization in Iranian political discourse, showing that such grammatical metaphors were employed not to obscure agency but to emphasize the actor's role in events.

All these explorations raise the question of the possibility of finding patterns of linguistic behavior in mafia game players, to identify players' roles and study deception in language.

3. Theoretical framework

In this study, Interpersonal Deception Theory (IDT) by David Buller and Judee Burgoon (1996) is the guiding framework for analyzing linguistic patterns and behavioral dynamics among participants in the mafia game. This theory suggests that deception manifests through linguistic output, with individuals displaying both conscious and unconscious indicators of deceit. Some conscious strategies that deceivers use could be falsification of a statement, concealing all or partial truths, and fully avoiding the issue. On the other hand, some unconscious indicators of deception are referred to as leakages, such as slick performances, anxiety, forgetting previous information, and nonverbal leakage. Through the lens of this theory, the linguistic output of expert mafia players is examined.

4. Methodology

The sample for this study was collected from the first eight episodes of season 17 of Citizen and Mafia, a reality show aired on Salamat TV (a national Iranian television channel) in July 2022. Episode 7 was excluded from the analysis due to significant omissions in participants' speech caused by media editing. The show featured professional players selected from previously observed games to compete against each other. The study involved 70 male participants, all closely matched in age to minimize potential demographic influences.

While IDT suggests that all deceptions leave a trace in linguistic output, individuals often learn to dwindle these traces to manipulate public opinion successfully. Therefore, professional players were deemed ideal subjects for this study, as some have years of experience perfecting their manipulation skills and masking their deceptive behavior. This study aimed to identify patterns and behaviors that even the most skilled players struggle to conceal, and to examine the extent to which individuals can develop skills to mask deception effectively.

For this study, each participant's linguistic output was transcribed, and a mix of *t*-test and empirical observation of the transcription was used to identify the existing patterns in the participants' speech and connect them to the possibility of deception indicators.

Considering the variety of linguistic factors and their contextual nature (Abouelenien et al., 2014), each participant's linguistic output was transcribed, and the frequency of words occurring during the games was measured. To get a hold of contextual words used in mafia games, words with a frequency of occurrence lower than ten in each game were set aside and deemed insignificant for the current analysis. The remaining words were categorized into relevant groups, such as verbs, adverbs, prepositions, and game-related terminology, as markers for the study. Additionally, factors such as stuttering and emotionally charged words were also taken into account.

To ensure comparability across participants and games, the linguistic markers were normalized. The total number of words and verbs used in each participant's speech was divided by the number of rounds they survived and actively participated in to achieve word and verb means. Other linguistic markers were calculated as percentages relative to

the total number of words each player produced throughout the game. This approach allowed the data to be compared fairly across different participants, regardless of game length or survival duration.

Each linguistic category was compared to the participant's role, their elimination from the game, and the result of each game by comparing the mean of linguistic markers produced by individuals through a *t*-test, with their role and game result. Many markers were cast aside due to their lack of significant value, and the rest were pursued further to show a correlation between the markers and the groups. The results of the analysis are presented below (Table 1). Values that reached statistical significance are shown in bold.

Table 1. ***Key Linguistic Features in the Iranian Mafia Game***

Linguistic Marker	Role			Game result			Individual roles citizen			Individual roles mafia			Eliminated by mafia		
	Mafia	Citizen	Sig.	M. win	C.win	Sig.	C.win	C.lose	Sig.	M.win	M.lose	Sig.	Yes	No	Sig.
Word mean	**187.1**	**213.4**	**.014**	218.4	198.4	.098	206.8	222.3	.206	198.2	178.7	.208	201.5	222.4	.087
Verb mean	**35.67**	**43.90**	**<.001**	42.78	40.42	.307	43.09	44.92	.515	37.78	34.19	.213	**40.77**	**46.25**	**.047**
Subjunctive	**1.60**	**2.05**	**.022**	2.08	1.79	.116	1.91	2.23	.167	1.74	1.52	.405	1.89	2.17	.224
Past Continuous	**0.07**	**0. 15**	**.037**	0.09	0.15	.174	0.19	0.09	.105	0.08	0.07	.843	0.18	0.12	.356
Negetive verbs	2.10	2.13	.928	2.26	2.01	.208	1.96	2.34	.071	2.09	2.13	.938	2.09	2.16	.728
Command	1.23	1.37	.578	1.29	1.35	.817	1.36	1.28	.784	1.32	1.32	.993	**0.91**	**1.71**	**.003**
Negetive command	**0.13**	**0.04**	**.018**	0.04	0.08	.205	0.06	0.01	.080	0.13	0.13	.926	0.04	0.04	.946
'Citizen'	2.63	2.48	.585	2.78	2.33	.061	2.40	2.68	.353	**3.02**	**2.19**	**.032**	2.32	2.61	.327
'Target'	1.09	1.18	.665	**0.85**	**1.38**	**.002**	**1.45**	**0.84**	**.006**	0.87	1.23	.237	1.35	1.04	.178
Time references	0.12	0.08	.633	**0.03**	**0.14**	**.036**	**0.12**	**0.03**	**.008**	0.04	0.20	.378	0.07	0.09	.473
Emotional words	0.77	0.85	.588	0.83	0.82	.921	0.83	0.90	.676	0.69	0.80	.693	**1.06**	**0.69**	**.022**

Explainers	0.29	0.28	.953	0.31	0.27	.582	0.26	0.34	.420	0.24	0.29	.655	0.27	0.29	.836
Pauses	0.83	0.81	.846	0.84	0.80	.769	0.85	0.79	.714	0.95	0.69	.351	0.91	0.72	.223

5. Findings

The findings will be demonstrated in five sections in the same order of their presentation in Table 1: (1) players role, in which the linguistic behaviors of mafia players are compared to citizen players; (2) the game result, which examines the collective behavior of players with the game's outcome; (3) intra-group comparisons of winning and losing players; (4) players eliminated by mafia during night phases; and (5) a general qualitative observation, introducing observed strategies used by mafia players in the game.

5.1 Players Role

The analysis of players' roles revealed that mafia players participate less in group discussions, $t = -2.523$, $p = .014$. While the number of words became significant only after the number of participants increased, the difference in the number of verbs became significant after only two games, $t = -4.208$, $p < .001$, indicating that mafia players use significantly fewer verbs in their speech.

Another feature with significant value is the subjunctive form. As shown in Table 1, citizens used the subjunctive form more frequently than mafia players, $t = -2.351$, $p = .022$. Following up, citizens used past continuous more frequently than mafia players, $t = -2.130$, $p = .037$. The final discovery of this section is the negative command feature. The mafia players used negative commands significantly more than citizens, $t = 2.504$, $p = .018$.

5.2 Game Result

In this section, the role of the players was set aside in favor of analyzing the game outcome based on the linguistic performances of all players. Two markers from words commonly used in the mafia game showed significant value. The first marker is the word "target", which was used significantly more in games where the citizens won, $t = -3.266$, $p = .002$. The second marker was words referring to the past (e.g., "before"), which were used more frequently in games won by citizens, $t = -2.158$, $p = .036$. Both of these markers also showed significant values in the next section.

5.3 Intra-group Comparisons

5.3.1 Winning Citizens vs Losing Citizens

Similar to the previous section, the two significant markers were the word "target" and words denoting the past. The results indicated that citizens who won their games used both features more than citizens who lost their games, $t = 2.857$, $p = .006$, $t = 2.764$, $p = .008$, respectively.

5.3.2 Winning Mafia vs Losing Mafia

The only feature significant in this comparison was the use of the word "citizen". According to the findings, mafia players who have used the word "citizen" more were likely to win their games, $t = 2.318$, $p = .032$.

5.3.3 Eliminated Players by Mafia Team

Furthermore, citizens eliminated by the mafia tend to use fewer verbs compared to other citizens, $t = -2.042$, $p = .047$. In addition, a significant difference was observed in command forms; citizens eliminated by the mafia employed fewer commands than other citizens, $t = -3.175$, $p = .003$. Interestingly, those eliminated by the mafia also used more emotionally charged words, $t = 2.370$, $p = .022$.

5.4 Qualitative Observations

Alongside these quantitative discoveries, some qualitative strategies were observed in the linguistic behaviors of mafia players. Each strategy will be introduced briefly in the following section and then discussed later in the discussion section.

5.4.1 The Leadership Role

One interesting strategy is for mafia players to ally themselves with a well-influenced citizen leader in the game to obscure their role. In such instances, the mafia carefully follows the leader's actions and statements, mirroring them to avoid suspicion.

Two to three players typically take on the leadership role throughout each game. These leaders tend to speak the most, guiding the group through a process of elimination as they attempt to identify the mafia players.

An example of mafia taking the role of leader would be: "…player 4, 6, 7, and 10 are mafia. The rest of you are citizens. Look for mafia only in these four people. Player 7's only target was player 5 and me… but player 7 didn't speak about player 4 and 6, who are the most suspicious in the game."

An example of mafia talking about the flow of the game while not taking the role of leadership would be: "… you have a mafia for sure, don't you? Tell us who you suspect. Who do you want to vote out… I think you are suspicious since you don't have a concrete suspect, yet you want the other citizens to trust you and follow your words".

Successful mafia players often adopt the role of "right-hand man" to a leading citizen. By allying themselves with a leader, they stay under the radar while manipulating the other players. Should the leader be voted out, the mafia player pretends to follow in the footsteps of the eliminated citizen, maintaining their influence.

5.4.2 Shifting the Blame

Mafia players may shift the blame by accusing others of mistakes they themselves had committed. A common example includes statements like, "… you didn't vote for player X, so you must be mafia", even when the accuser had also refrained from voting. Despite its low success rate, this strategy remained frequently observed in the games.

5.4.3 Distancing and the Serial Position Effect

A critical task for mafia players is to subtly distance themselves from their teammates to avoid being linked if one of them is eliminated. One widely observed strategy among mafia players is strategically ordering their accusations, placing their mafia teammate in the middle of a sentence, and ending with targeting a citizen. For instance: "…I think X might be mafia; however, Y is really suspicious, so we should vote him now." This strategic ordering was observed on many occasions of the game, especially amid the chaos of information overload, when the accusations and suspicious behaviors have been pointed out so much that confusion riddles the players' minds.

5.4.4 Gaslighting

Gaslighting happens when a player denies an observable truth to manipulate others' perceptions (e.g., "…you are lying, I did vote for X"). It is occasionally employed by mafia players, albeit sparingly. This strategy was observed primarily during the later stages of the game when the number of remaining players was significantly reduced.

6. Discussion

This section provides an analysis of the key findings from this study, shedding light on their implications and comparing them with existing literature, while exploring how they contribute to a deeper understanding of linguistic behavior in the mafia game.

6.1 Players Role

The first notable finding is the reduced use of words, particularly verbs, by mafia players. This pattern of verb usage aligns with IDT's principle of withdrawal, which suggests that deceivers often avoid engaging directly with the issue at hand. While this disparity in word and verb usage might not be immediately apparent during the game, the data provides undeniable evidence of its significance. Verbs, as the backbone of a sentence and indicators of its core meaning, are particularly noticeable in this context. The substantial gap in verb usage may suggest that mafia players, while attempting to speak just enough to evade suspicion, deliberately limit their use of verbs to avoid disclosing crucial information. It may also reflect a strategy of filling their speech with unnecessary descriptions to waste time without actually contributing to the game's information.

These findings contradict two previous studies. Fay's (2009) finding showed no significant difference between the word-mean and verb-mean of players from different teams, while Bedwell et al.'s (2011) finding revealed that deceivers tend to use more verbs than truth-tellers. The contradictory findings of Bedwell et al.'s study can likely be the outcome of different contexts. In their research, deceivers participated in one-sided conversations, free from the immediate scrutiny or questioning of others. This lack of challenge may have emboldened them to elaborate and craft detailed stories without the fear of being exposed. The difference in Fay's (2009) results could stem from their small number of participants, which may have limited the reliability and scope of their conclusions.

The discoveries regarding citizens' use of more subjunctive forms highlight their lack of concrete information and their tendency to consider all possible scenarios. In contrast, mafia players appear to prioritize sounding confident and persuasive, relying on fewer uncertain statements in favor of influencing others effectively. This finding contradicts Bajaj et al.'s (2023) finding that deceivers use expressions of uncertainty more often than truth-tellers, yet aligns with the findings of Kheirabadi and Kheirabadi (2024), suggesting that in Iranian contexts, deceivers tend to favor expressions of certainty and present themselves with greater confidence.

The citizens' frequent use of the past continuous tense aligns with their efforts to defend themselves and uncover mafia players by providing detailed accounts of the players' past actions and words. By actively recounting past events, citizens built trust and reliability among the players. In addition, describing past events, particularly in a continuous form, often raises suspicion against mafia players. This is likely due to the inconsistencies in mafia players' past actions, which makes their perspective appear less cohesive compared to citizen players.

Finally, in moments of desperation, mafia players are more inclined to use negative command forms as a means of warning others. This strategy is used to create doubt in citizens' minds and prevent them from voting the mafia out. In contrast, citizens tend to rely on logical explanations, trusting that their reasoning will resonate with others and lead to the identification of the mafia players.

6.2 Game Result

A significant difference was observed between games in which the word "target" was frequently used and those in which it was not. Games, where citizens emerged victorious, featured this word more prominently than games won by mafia players. Among the participants in this study, the word "target" was used to clearly express suspicion and even direct accusations toward another player. This finding highlights that the more players actively participate in accusing others and explicitly state their suspicion, the greater the likelihood of a citizen victory.

This suggests that decisive players, unafraid to directly target others, are more effective in identifying mafia players compared to those who hesitate to voice their suspicion. Overall, the assertiveness of citizens plays a crucial role. When they become more vocal and confident in their accusations, their ability to collaborate effectively and eliminate mafia players increases significantly.

Another finding revealed that games resulting in citizen victories featured significantly more words referencing the past. This can be explained by the fact that mafia players' words and actions are more likely to reveal inconsistencies as the game progresses. Naturally, focusing on past events and critically evaluating each player's words and actions from the beginning of the game enhances the ability to identify mafia players more accurately.

This result also aligns with the observed pattern of citizens often voting to eliminate a player by chance; typically, someone who is both supported and targeted by different groups. Once the eliminated player's role is revealed, citizens adjust their strategies according to the new information to identify the mafia players more accurately.

6.3 Intra-group Comparisons

Similar to the results discussed above, citizens were more likely to win their games when they displayed assertiveness and courage in targeting individuals they found suspicious, without fear of retaliation. This approach, combined with a focus on past events and a commitment to connecting each player's words and actions to earlier stages of the game, enables citizens to create a clearer path toward identifying mafia players.

When comparing mafia teams that won to those that lost, it appears that mafia teams using the word 'citizen' more frequently were likely to win. Whether by redirecting focus, creating confusion, or claiming to be citizens themselves, mentioning citizens helps them manipulate the narrative. By actively bringing citizens to the table, mafia players may appear engaged in group reasoning, reducing suspicion about themselves. This tactic likely helps them maintain their cover and influence the group's decisions, ultimately contributing to their success.

6.4 Eliminated Players by Mafia Team

This section's findings show that mafia players eliminate citizens who use fewer verbs in their speech, rely less on the command form, and employ more emotionally charged language. The first two findings can be explained by considering the mafia players' underlying motives. Their primary objective is to sow doubt and suspicion among citizens, ultimately persuading the majority that the citizens are mafia players.

While mafia players often risk revealing their true roles through excessive speech, citizens also face similar risks. They may unintentionally make mistakes, such as falsely accusing an innocent player or contradicting their own actions, offering the mafia opportunities to shift the attention away from themselves, and onto a citizen. Citizens who use fewer verbs (therefore less informative language) and avoid assertive command forms

present fewer opportunities for mistakes to be exploited. As a result, it becomes strategically advantageous for mafia players to eliminate such individuals, as they create fewer opportunities for the mafia to exploit.

Regarding the use of emotionally charged words, while it is tempting to conclude that mafia players eliminate those who express emotions to prevent them from influencing other citizens, a more plausible explanation lies in the dynamics of the game. At the beginning of the game, players often use a more emotional tone, expressing their excitement to be playing with the group, or sharing personal feelings, with statements like "…that was nice." Or "...you shouldn't have said that." In a sympathetic way.

As the game progresses, players adopt a more neutral, logical, and serious tone, focusing on bringing as much logic and reason as possible to the discussion. Those eliminated early in the game by the mafia do not have the opportunity to shift their speech patterns to a more neutral stance. Therefore, their speech is naturally laden with more emotional expressions.

6.5 Insignificant Markers

There are several markers present in Table 1 with no significant value. The reason for their mention is to compare them with the findings of other researchers. According to the discoveries of Kim et al. (2024), fake reviews tend to feature more positive forms than negative forms. However, no significant relationship between negative forms and players' behavior was found in this study. Before dismissing the role of negative forms in deception, it's essential to consider the unique nature of the present study. In this case, participants predominantly used the negative form to report past actions rather than to use it as a tool for manipulation, which could explain the absence of a notable difference between the two groups.

Similarly, Bajaj et al's (2023) research suggested that deceivers use more explanatory language, yet no significant correlation was observed. This result was unexpected, given the assumption that deceivers may either limit their explanations to avoid errors or over-explain to construct a convincing narrative. In the context of the Mafia game, however, time is of the essence; each word must carry weight. Citizens may perceive over-explaining as a sign of suspicion, prompting them to distrust a player who elaborates too much on simple statements.

The final feature in Table 1 is 'pauses', which include disruptions in speech such as stuttering and false starts. Given its relevance to IDT, a significant relationship was anticipated, but none was discovered. One possible explanation for this outcome lies in the expertise of the participants. As professional players, they likely have developed the skill to mask any sign of anxiety, naturally incorporating pauses into their speech without drawing attention to their role, whether as mafia members or citizens.

6.6 Qualitative Observations

This section will discuss the qualitative behavioral linguistic patterns observed in the Mafia game in great detail, diving deep into the reasons for their use in the context of the game and how each behavior affects the success or failure of mafia players.

6.6.1 The Leadership Role

Successful mafia players tend to take the role of second leader or follower rather than the leader. There is a simple explanation for why the success or failure of a mafia player is related to leadership. For citizens, this leadership role poses little risk. Firstly, being the truth-telling leader, they don’t mind being in the spotlight since they have no reason to be

concerned with contradictions between their words and actions. Secondly, if they are targeted and voted out as a player with strong influence, their role is revealed upon elimination, reinforcing the validity of their earlier statements. The only thing remaining would be other citizens picking up where they left off and following their footsteps.

For mafia players, however, becoming a leader carries much higher stakes. If a mafia player takes on the role of a leader, they expose themselves to the risk of contradictory statements or actions more than before. In such a situation, manipulation becomes even more difficult since other players follow their words and actions. For example, a mafia player may strongly accuse another player but hesitate to vote against them since they are also a mafia player. Such inconsistencies draw incredible attention to the leading role, and citizens collectively agree that voting out the leader benefits the whole. Once eliminated, the mafia player's true identity is revealed, and the citizens dismiss any previous statements, effectively neutralizing their influence on the game.

Considering the fact that citizens outnumber mafia players, with careful play, they can afford to vote out a citizen or two without losing the game. In contrast, mafia players must be far more cautious, as losing a teammate costs them much more at the end of the game. Therefore, successful mafia players often adopt the role of 'right-hand man' to a leading citizen. By aligning themselves with a leader, they can stay under the radar while manipulating the other players. Should the leader be voted out, the mafia player can pretend to follow in their footsteps, maintaining their influence and power over the group.

6.6.2 Shifting the Blame

Mafia players may target another player for a crime they also committed themselves, hoping to raise suspicion against a citizen player. While this strategy was heavily relied upon, it would get caught by an alert citizen. Citizens quickly would point out the hypocrisy, noting that the accuser was equally guilty of the same behavior. Despite its low success rate, this strategy remained the most frequently observed in the analyzed games. On rare occasions, when citizens were not paying close attention to the unfolding events or failed to connect the dots, this tactic managed to sow doubt and temporarily shift the narrative in the mafia's favor.

6.6.3 Distancing and the Serial Position Effect

In many instances of the game, mafia players strategically structure their speech to mention a teammate in passing to avoid raising suspicion, yet deliberately enough to prevent being perceived as aligned. Mafia players target their teammates in the middle or beginning of their sentences and finish their speech by targeting a citizen.

This strategy follows the Serial Position Effect introduced by Hermann Ebbinghaus (1913). This psychological theory suggests that people tend to remember certain parts of information more distinctly than others; most notably, the final section (recency effect), followed by the beginning (primacy effect), with the middle portion being the least memorable. By mentioning their teammate in the middle or beginning of the statement and closing with a strong focus on a citizen, mafia players effectively reduce the likelihood that their mention of their teammate would linger in others' minds enough to cause them trouble.

The effectiveness of this strategy was highly evident. Other players would concentrate on the last name mentioned, while the mafia teammate was often overlooked. If the teammate was later revealed as mafia, the mafia accusing them could defend themselves by pointing out that they had mentioned them as a suspect, albeit subtly. This use of the Serial Position Effect proved to be a successful strategy.

6.6.4 Gaslighting

Finally, gaslighting is a strategy where a player denies an observable truth to manipulate others' perceptions (e.g., "…you are lying, I did vote for X"). Due to its risk, this strategy was used sparingly and only at the final stages of the game, where the number of players was reduced. In real-world scenarios, gaslighting can be a highly effective tool for manipulation, as it exploits trust and causes inner doubt. However, in the context of the Mafia game, where every player is actively searching for lies and inconsistencies, executing this strategy is difficult. The intense scrutiny within the game makes it much harder to convince others of false claims, and any misstep is likely to result in immediate elimination. Therefore, mafia players avoid this strategy as much as possible and only use it as a desperate measure.

While gaslighting can occasionally turn the tide in favor of the mafia, its high risk and low success rate in the observed games make it one of the least preferred strategies among mafia players. Instead, players appeared to rely more on subtler methods that required less confrontation and carried a lower risk of exposure.

7. Conclusion

The findings of this study highlight the complex nature of verbal communication in deception. Rather than adhering to fixed linguistic markers, deception is shaped by context, social roles, and individual adaptability. Humans are remarkable adaptors, capable of seamlessly fitting into roles expected of them, making the task of understanding or predicting their behavior very difficult.

Despite the complexity of language and deception, some linguistic patterns were observed in this study as indicators of deception. The results suggest that deceptive players prefer to use fewer words and sentences to limit the flow of information, use less subjunctive forms to appear more reliable and influence others' beliefs, and finally use negative commands to warn and create a sense of danger to shape others' perception. On the other hand, strategies such as avoiding leadership roles, shifting the blame, the Serial Position Effect, and gaslighting were observed as common themes of deception. While these findings align with Interpersonal Deception Theory (IDT), certain expected leakage cues, such as the use of abnormal emotions or stuttering, were not significant. This could be due to players practicing deception as a skill, which raises an important question: Can deception be honed to the point where it becomes indistinguishable from truthful speech?

There is evidence of the effect of lying on memory. Experiments show that lying about an experience can distort the memory in a way that makes it difficult for people to recall the full truth (Battista et al., 2020). This effect was observed in the games, where players, preoccupied with sustaining their deception, inadvertently forgot critical facts, including their own lies, suggesting that such lapses may be more common than players later acknowledged.

There are studies of how belief in receiving pain relief medicine can manifest as a genuine reduction in pain, with corresponding changes in brain activity (Amanzio et al., 2013). Considering that many previously identified linguistic markers of deception, such as pauses and explainers, did not appear in this study, it may be speculated that with sufficient practice or psychological conditioning, individuals can minimize or even eliminate unconscious leakage. Trained liars, such as actors or con artists, may internalize their roles so thoroughly that their fabricated narratives appear entirely authentic, even to themselves.

This study contributes to ongoing discussions on how deception operates in interactive discourse. Rather than viewing deception as a static linguistic phenomenon, these

findings emphasize its context-dependent nature and potential to be learned as a strategic skill. Future research could explore how speakers develop and refine deceptive verbal strategies over repeated interactions and whether these strategies can be systematically categorized across different social deception games.

References

Abouelenien, M., Perez-Rosas, V., Mihalcea, R., & Burzo, M. (2014). Deception detection using a multimodal approach. *Proceedings of the 16th International Conference on Multimodal Interaction* (pp. 58–65). Association for Computing Machinery. https://doi.org/10.1145/2663204.2663229

Amanzio, M., Benedetti, F., Porro, C. A., Palermo, S., & Cauda, F. (2013). Activation likelihood estimation meta-analysis of brain correlates of placebo analgesia in human experimental pain. *Human brain mapping*, 738–752. https://doi.org/10.1002%2Fhbm.21471

Ansari, S., & Gupta, S. (2021). Customer perception of the deceptiveness of online product reviews: A speech act theory perspective. *International Journal of Information Management*, 57, Article 102286. https://doi.org/10.1016/j.ijinfomgt.2020.102286

Bajaj, N., Rajwadi, M., Constance, T. G., Wall, J., Moniri, M., Laird, T., et al. (2023). Deception detection in conversations using the proximity of linguistic markers. *Knowledge-Based Systems*, 267, Article 110422. https://doi.org/10.1016/j.knosys.2023.110422

Battista, F., Mangiulli, I., Curci, A., Riesthuis, P., & Otgaar, H. (2020). Can we believe in our own lies? *The inquisitive Mind*(43). https://www.in-mind.org/article/can-we-believe-in-our-own-lies

Bedwell, J. S., Gallagher, S., Whitten, S. N., & Fiore, S. M. (2011). Linguistic correlates of self in deceptive oral autobiographical narratives. *Consciousness and Cognition*, 547–555. https://doi.org/10.1016/j.concog.2010.10.001

Bond Jr, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review, 10*(3), 214–234. https:// doi.org/10.1207/s15327957pspr1003_2

Buller, D. B., & Burgoon, J. K. (1996). Interpersonal deception theory. *Communication Theory, 6*(3), 203–242. https://doi.org/10.1111/j.1468-2885.1996.tb00127.x

Chittarajan, G., & Hung, H. (2010). Are you a werewolf? detecting deceptie roles and outcomes in a conversational role-playing game. *2010 IEEE International Conference on Acoustics. Speech and Signal Processing* (pp. 5334–5337). IEEE. https://doi.org/10.1109/ICASSP.2010.5494961

Ebbinghaus, H. (1913). *Memory; A contribution to experimental psychology.* Teachers college, Columbia university. https://doi.org/10.1037/10011-000

Fay, M. (2009). Linguistic cues for deception detection in online mafia forums. https://sites.socsci.uci.edu/~lpearl/CoLaLab/papers/Fay2009_DeceptionDetection.pdf

Ibraheem, S., Zhou, G., & DeNero, J. (2022). Putting the con in context: Identifying deceptive actors in the game of mafia. *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 158–168). Berkeley: Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.naacl-main.11

Kheirabadi, M. & Kheirabadi, R. (2024). The linguistic structure of fake news in advertising posts of social media. *Journal of Linguistic Studies: Theory and*

Practice, *3*(1), 85–117 [In Persian] https://doi.org/10.22034/jls.2024.141086.1100

Kim, J. M., Park, K. K.-c., Mariani, M., & Wamba, S. F. (2024). Investigating reviews intentions to post fake vs. authentic reviews based on behavioral linguistic features. *Technological Foecasting & Social Change, 198*, 122971. https://doi.org/10.1016/j.techfore.2023.122971

Malmir, A., Yaghoubi, R., Ameri, H., Dabir-Moghaddam, M. & Aghagolzadeh F. (2023). Nominalization as grammatical metaphor and ideological representation in political discourse of JCPOA. *Journal of Linguistic Studies: Theory and Practice, 1*(2), 21–44 [In Persian] https://doi.org/10.22034/jls.2023.62729

Niculae, V., Kumar, S., Boyd-Graber, J., & Danescu-Niculescu-Mizil, C. (2015). Linguistic harbingers of betrayal: A case study on an online strategy game. *ArXiv*, abs/1506.04744. https://doi.org/10.48550/arXiv.1506.04744

O'Gara, A. (2023). Hoodwinked: Deception and cooperation in a text-based game for language models. *ArXiv*, abs/2308.01404. https://doi.org/10.48550/arXiv.2308.01404

Robertson, M. (2010, February 4). Werewolf: How a parlour game became a tech phenomenon. *Wired UK*(10). https://www.wired.co.uk/article/werewolf

Schumann, J., Zufferey, S., & Oswald, S. (2019). What makes a straw man acceptable? Three experiments assessing linguistic factors. *Journal of Pragmatics 141*, 1–15. https://doi.org/10.1016/j.pragma.2018.12.009

Spinde, T., Rudnitckaia, L., Mitrovic, J., Hamborg, F., Granitzer, M., Gipp, B., et al. (2021). Automated identification of bias inducing words in news articles using linguistic and context-oriented features. *Information Processing and Management, 58*(3), 102505. https://doi.org/10.1016/j.ipm.2021.102505

Zhou, L., & Sung, Y.-w. (2008). Cues to deception in online chinese groups. *Proceedings of the 41st Annual Hawaii International Conference on System Sciences* (pp. 146). HICSS. https://doi.org/10.1109/HICSS.2008.109