



Analysis of Ambiguity in Humor: A Pragmatic Analysis of Wordplay in YouTube Channel "We Got The Chocolates" (*Analisis Ambiguitas dalam Humor: Analisis Pragmatik Permainan Kata di Saluran YouTube "We Got The Chocolates"*)

Moh. Kifly S. Rasyid¹, Hanisah Hanafi², Fahria Malabar³

^{1,2,3}English Language Education Department, Faculty of Letters and Culture, Universitas Negeri Gorontalo
kiflisrasyid01@gmail.com¹, hanisah.hanafi@ung.ac.id², fahria@ung.ac.id³

Article Info	Abstract
Article history: Received: 27 Juli 2026 Revised: 11 Agustus 2026 Accepted: 13 Agustus 2026 Keywords: Phonetic Ambiguity Wordplay Pragmatics Maxim Flouting Co-text Kata Kunci: Ambiguitas Fonetis Permainan Kata Pragmatik Pelanggaran Maksim Ko-teks	This study is grounded in the linguistic phenomenon of phonetic ambiguity as a strategy to create verbal humor in digital spoken discourse. The objective is to analyze how segmental sound structures are realised in phonetic wordplay and elucidate how these structures trigger dual meaning interpretations through a pragmatic framework. A descriptive qualitative research design was employed, using spoken dialogues from the "We Got The Chocolates" YouTube channel as the primary data source. Data were collected through systematic observation and transcription of five purposively selected videos, then analysed using a co-textual pragmatic approach in conjunction with International Phonetic Alphabet (IPA) transcriptions and Grice's Cooperative Principle. The findings reveal two primary operational patterns: Perfect Puns (13 instances) relying on absolute homophony, and Imperfect Puns (17 instances) relying on phonetic similitude. These phonetic manipulations consistently trigger humor through a calculated strategy of multiple maxim flouting, where the flouting of the Maxim of Manner via acoustic camouflage is systematically paired with the flouting of the Maxim of Quality or the Maxim of Relation. The study concludes that spoken humor operates as a cognitive resolution process where structural acoustic illusion and narrative incongruity mutually reinforce each other. However, limitations include the restriction to segmental sound manipulation (excluding suprasegmental features) and the reliance on interpretative qualitative pragmatic methodology. Abstrak Penelitian ini dilandasi oleh fenomena linguistik ambiguitas fonetis sebagai strategi untuk menciptakan humor verbal dalam wacana lisan digital. Tujuannya adalah untuk menganalisis bagaimana struktur bunyi segmental direalisasikan dalam permainan kata fonetis (phonetic wordplay) dan menjelaskan bagaimana struktur-struktur ini memicu interpretasi makna ganda melalui kerangka kerja pragmatik. Desain penelitian kualitatif deskriptif digunakan dalam studi ini, dengan memanfaatkan dialog lisan dari kanal YouTube "We Got The Chocolates" sebagai sumber data utama. Data dikumpulkan melalui observasi sistematis dan transkripsi dari lima video yang dipilih secara purposif, kemudian dianalisis menggunakan pendekatan pragmatik ko-tekstual bersama dengan transkripsi Abjad Fonetik Internasional (IPA) dan Prinsip Kerja Sama Grice (Grice's Cooperative Principle). Temuan penelitian mengungkapkan dua pola operasional utama: Plesetan Sempurna / Perfect Puns (13 data) yang bergantung pada homofoni mutlak, dan Plesetan Tidak Sempurna / Imperfect Puns (17 data) yang bergantung pada kemiripan fonetis. Manipulasi fonetis ini secara konsisten memicu humor melalui strategi terencana berupa pelanggaran maksim ganda (multiple maxim flouting), di mana pelanggaran Maksim Cara

(Maxim of Manner) melalui kamufase akustik dipasangkan secara sistematis dengan pelanggaran Maksim Kualitas (Maxim of Quality) atau Maksim Hubungan/Relevansi (Maxim of Relation). Penelitian ini menyimpulkan bahwa humor lisan beroperasi sebagai proses resolusi kognitif, di mana ilusi akustik struktural dan inkongruensi naratif saling memperkuat satu sama lain. Namun, keterbatasan dari penelitian ini meliputi pembatasan pada manipulasi bunyi segmental saja (tidak termasuk fitur suprasegmental) dan ketergantungan pada metodologi pragmatik kualitatif yang bersifat interpretatif.

Corresponding Author:

Moh. Kifly S. Rasyid
Faculty of Letters and Culture
Universitas Negeri Gorontalo
kiflisrasyid01@gmail.com

1. INTRODUCTION

Language possesses a unique flexibility that allows it to be shaped into wordplay as a manifestation of creative verbal communication. In the digital era, Generation Z massively exploits this creativity on social media to build distinctive digital interactions (Putri & Sabrina, 2024). Virtual communication dynamics frequently utilize slang and unfamiliar idioms that potentially trigger ambiguous interpretations (Nisa & Rizal, 2025). Among these phenomena, one of the most complex strategies in creating spoken discourse humor is phonetic ambiguity. Its urgency lies in its strict reliance on the physical realization of speech sounds. Unlike written lexical or grammatical ambiguity, phonetic ambiguity manipulates the segmental sound structures of homophonous words to trigger dual interpretations. Understanding how vowel and consonant segments interact with context is crucial for uncovering the mechanics of audio-based humor in modern digital communication.

Previous linguistic humor research has primarily focused on lexical-syntactic ambiguity and maxim flouting in textual or visual discourse. Prior studies consistently analyzed lexical and syntactic ambiguity as the main incongruity sources generating humor in conventional texts, riddles, and internet memes (Charina, 2017; Huluqiah et al., 2025). Dual interpretations from grammatical engineering are also common in commercial slogans and social media comedy videos (Candra et al., 2024; Faridah et al., 2022). Furthermore, various pragmatic studies have proven that flouting Grice's cooperative principles effectively triggers humorous effects (Cahyaningtias & Kurniawan, 2026). Despite this broad spectrum, a distinct research gap remains. The specific role of phonetic ambiguity particularly how spoken text demands dictate micro-level manipulations of vowels, consonants, and word boundaries (junctures) to execute pragmatic floutings remains overlooked. Exploring this anatomy of speech sound manipulation constitutes this study's primary novelty.

Based on this gap, this research focuses on the mechanics of sound-based wordplay by elucidating how textual demands and segmental sound structure manipulation interact to execute pragmatic floutings. It examines how sentence context dictates specific script demands, which are then phonetically manipulated to flout conversational maxims and trigger dual meanings. Elucidating this cooperative principle is crucial for testing the effectiveness of Grice's maxims in everyday conversation (Hossain, 2021). This is particularly relevant regarding the flouting of the Maxim of Relation in spoken communication. The phenomenon of sudden topic reversal by speakers is proven to be a main catalyst for performative humor that massively entertains audiences (Firda et al., 2021). To capture these dynamics realistically, this study examines digital spoken discourse from casual chat videos on the YouTube channel "We Got The Chocolates".

The unit of analysis is limited to spoken utterances containing phonetic ambiguity, classified into perfect puns (absolute homophony) and imperfect puns (near-homophony). This approach aims to explain spoken discourse through its physical form and communicative effects. Systematically, this article outlines the research methods, analyzes the phonological manipulation mechanisms and their pragmatic execution, and concludes with suggestions for future research.

2. RESEARCH METHODS

The research design used in this study is a descriptive qualitative approach applying a pragmatic framework. The pragmatic application in this research is strictly limited to the realm of internal linguistic context or co-text, which is examining the logical and syntactic relationships between elements within the

utterance itself. This approach tracks how the initial sentence (setup) builds a specific script demand and how the phonetic keyword (punchline) manipulates sound to execute pragmatic floutings and resolve the clash, thus explicitly excluding external sociocultural analysis, social relationships between speakers, and cognitive reactions from the audience. As a descriptive qualitative research, the presence of the researcher acts as the main instrument who independently designs, collects, classifies, and analyzes all discourse data. The location of this research was conducted virtually on the digital media platform YouTube, with the duration of document data collection limited to the content release period up to the thesis writing year in 2026.

The research subject in this study is a popular YouTube channel named "We Got The Chocolates," while the research object analyzed in-depth is the embodiment or manifestation of the phonetic ambiguity phenomenon in the spoken utterances produced by the speakers. The research target or data population encompasses all forms of spontaneous spoken interaction containing sound-based wordplay found on the channel. The research target sampling technique utilized the purposive sampling method, where the selection of source materials was based on the high intensity and density of jokes based on segmental sound similarity. Through this technique, a research sample of 5 casual chat videos was determined, with an accumulated viewing duration of approximately 57 minutes. The selection of the "We Got The Chocolates" channel as the research subject was based on the unique characteristics of the casual interactions among its speakers, who have a high intensity in producing spontaneous speech jokes, particularly in the format of dad jokes or try not to laugh challenges. In these naturally occurring interactions, the sound-based wordplay is not prepared through a written script but emerges instantaneously, thus providing a data corpus rich in variations of vowel, consonant, and word boundary articulation manipulation.

On the other hand, the audio-to-text transcription process in this study faced a significantly high level of micro-linguistic complexity and challenges. This is because the natural speech rate of native speakers frequently triggers assimilation, elision, or segmental sound decay at the end of words such as the weakening of the rhotic /r/ sound which inherently blurs phonemic boundaries. Therefore, the researcher conducted repeated and in-depth acoustic observations to capture these temporary spoken sound waves, and then precisely converted them into International Phonetic Alphabet (IPA) symbols, relying on the Cambridge English Dictionary standard to ensure data validity and reliability.

Regarding the research instruments, besides utilizing text documentation observation sheets, the development of analytical instruments and the assurance of spoken data accuracy standards strictly refer to the Cambridge English Dictionary as an auxiliary tool for sound transcription conversion. The data collection procedure in this study was systematically carried out through qualitative observation methods and documentation techniques. The initial step began with comprehensively observing the casual video materials selected as samples. The data collection technique proceeded with conducting an audio-to-text transcription process of all dialogues within the videos to convert temporary spoken acoustic wave inputs into a permanent written text format. Once the written transcripts were available, an identification and cataloging procedure was conducted to filter and isolate 30 spoken utterances that purely contained phonetic ambiguity characteristics. From this total population of findings, advanced data reduction techniques were applied to select 10 representative utterance data, which were evenly divided into 5 perfect puns and 5 imperfect puns. The selection of these final 10 data prioritized the richness of contextual aspects and the most complex variations of sound manipulation for in-depth analysis.

The data analysis technique applied in this study follows the interactive analysis model developed by Miles and Huberman (1994). The analysis process operates cyclically through three main workflows, namely data reduction, data display, and drawing conclusions or verification. During the data reduction stage, lexical expressions were converted into International Phonetic Alphabet (IPA) symbols, then classified using the sound similarity level parameters from (Hempelmann, 2003) to determine the absolute homophony group (perfect puns / homophonic puns) or sound proximity (imperfect puns / heterophonic puns). These phonological parameters were maintained because the concept of phonological distance is proven to remain the primary taxonomic foundation in measuring the humorous degree of spoken ambiguity in contemporary linguistic studies (Palmann & Miller, 2025). The data display stage was carried out structurally using an instrument in the form of an analytical matrix table juxtaposing the components of context script demands, vowel-consonant sound structure transcription (IPA), and the types of conversational maxims flouted. The conclusion drawing stage was conducted by constructing an integrated causal narrative explaining how textual structure triggers a clash of meanings resolved by the speaker's acoustic manipulation. To ensure the validity of the research results and the reliability of the findings, a data validity checking technique was performed by restricting the analysis to objective physical evidence in the form of IPA phonetic symbol transcriptions, anchoring meaning interpretation to internal word relationships (co-text), and rigorously testing logical consistency using Grice's Cooperative Principle theoretical framework, whose application is proven consistent with current trends in comedy discourse and digital media ambiguity analysis (Candra et al., 2024; Putri & Sabrina, 2024).

3. RESULTS AND DISCUSSION

3.1 Phonetic Ambiguity and Maxim Flouting in Humor

This study identifies 30 spoken utterances that specifically exploit phonetic ambiguity as a means of generating humor within the "We Got The Chocolates" YouTube channel. Out of these total findings, the sound manipulation mechanisms are classified into two primary operational categories: 13 utterances categorized as perfect puns (absolute homophony) and 17 utterances categorized as imperfect puns (near-homophony). From a pragmatic standpoint, all data (100%) exhibit a flouting of the Maxim of Manner, which then operates dually (multiple flouting) alongside the flouting of the Maxim of Quality in 20 utterances and the Maxim of Relation in 10 utterances. To dissect how these findings were obtained and interpreted, Table 1 below presents 10 representative samples that illustrate the varying complexities of sound manipulation within the data corpus.

Table 1. Representative Classification of Phonetic Ambiguity and Maxim Flouting

Code	Sentence (Setup & Punchline)	Word Pair (IPA Transcription)	Category	Maxim Flouting
D-01	"Did you hear Donald Trump is banning shredded cheese? he Wants To make America grate again."	Grate /gret/ VS Great /gret/	Perfect Pun	Manner & Quality
D-02	"Did you hear about the new music group Cellophane? They mostly wrap."	Wrap /ræp/ VS Rap /ræp/	Perfect Pun	Manner & Quality
D-03	"What happens when a frog's car breaks down? It has to be towed away."	Towed /toud/ VS Toad /toud/	Perfect Pun	Manner & Quality
D-04	"Knock knock. Who's there? Cargo. Cargo who? Car go beep beep."	Cargo /'kɑ:rgoʊ/ VS Car Go /kɑ:r goʊ/	Perfect Pun	Manner & Relation
D-05	"What do lawyers wear to work? Lawsuits."	Lawsuits /'lɔ:su:ts/ VS Law Suits /lɔ: su:ts/	Perfect Pun	Manner & Quality
D-06	"How many tickles does it take to make an octopus laugh? tentacles."	Tentacles /'tɛntəkəlz/ VS Ten tickles /tɛn 'tɪkəlz/	Imperfect pun	Manner & Quality
D-07	"Dad what is competitive horse riding?... Is that equestrian?"	Equestrian /ɪ'kwestriən/ VS A question /ə 'kwɛstʃən/	Imperfect pun	Manner & Relation
D-08	"How do cows count? With a cow-culator"	Cow-culator /kəʊkjələtər/ VS Calculator /kælkjələtər/	Imperfect pun	Manner & Quality
D-09	"Why do cows have hooves and not feet? Because they lactose"	Lactose /'læk.tous/ VS Lack Toes /læk touz/	Imperfect pun	Manner & Quality

D-10	"What do you call a fake noodle? An <i>impasta</i> ."	Impasta /ɪmˈpɑːstə/ VS Impostor /ɪmˈpɑːstər	Imperfect pun	Manner & Quality
------	---	---	---------------	------------------

To empirically demonstrate how these mechanisms are captured through International Phonetic Alphabet (IPA) transcriptions, an in-depth dissection of two representative data samples is presented below:

3.1.1 Representation of *Perfect Puns* (Data: *Grate* vs. *Great*)

- Utterance: "Did you hear Donald Trump is banning shredded cheese? He wants to make America grate again."
- Analysis: Textually, the lexicon "shredded cheese" necessitates the presence of the culinary verb *grate*. However, the syntactic structure borrows a political slogan that requires the adjective *great*. The speaker exploits absolute homophony, wherein both words are realized with a 100% identical acoustic shell: /greɪt/. Pragmatically, this referential ambiguity flouts the Maxim of Manner. The humor is successfully executed because the speaker deliberately flouts the Maxim of Quality by fabricating a factual falsehood (Donald Trump banning shredded cheese) purely as a narrative vehicle to produce the sound /greɪt/.

3.1.2 Representation of *Imperfect Puns* (Data: *Cow-culator* vs. *Calculator*)

- Utterance: "How do cows count? With a cow-culator."
- Analysis: This sentence collides the livestock script ("cows") with the mathematical script ("count"), which inherently demands the word *calculator*. The speaker performs lexical engineering (a portmanteau) by substituting the initial syllable /kæl/ with the phonetic form of a cow, /kaʊ/, resulting in the manipulated form /kaʊkjələtər/. Despite the segmental distortion, the phonetic integrity of the final three syllables /kjələtər/ is maintained, allowing its auditory resemblance to deceive the listener's cognition. The absurdity of the narrative (a cow that can count) compels the listener to tolerate the phonetic flaw for the sake of comedic resolution.

3.2 Phonological Tolerance in Spoken Discourse

The empirical distribution of the research data demonstrates a significant structural asymmetry, wherein the *Imperfect Puns* category appears more dominantly (17 data items) compared to *Perfect Puns* (which only amount to 13 data items). This statistical reality must not be viewed merely as a temporal distributional coincidence; rather, it should be interpreted as a reflection of the complex human cognitive-auditory mechanisms involved in processing spoken discourse. The dominance of *Imperfect Puns* provides robust empirical evidence for the existence of a highly elevated phenomenon of "phonological tolerance" within human auditory perception when confronted with casual and rapid spontaneous conversational streams (*connected speech*).

Theoretically, human speech perception does not operate like a rigid linear scanning machine that demands one hundred percent acoustic feature alignment to identify a lexicon. Instead, human cognition operates using an *auditory prototype matching* model based on feature goodness-of-fit. In dynamic digital spoken discourse, such as the casual interactions on the "We Got The Chocolates" YouTube channel, natural speech production undergoes various phenomena including co-articulation, *vowel reduction*, and sound assimilation due to the speaker's articulatory speed. Under these rapid suprasegmental stream conditions, the *acoustic distance* between phonemes frequently blurs. It is this very phenomenon that speakers consciously exploit to construct comedic traps.

The listener's high phonological tolerance can be microlinguistically dissected through the success of paronymy (*near-homophone*) data in deceiving cognition. For instance, in the data manipulation of *tentacles* (/ˈtɛntəklz/) into *ten tickles* (/tɛn ˈtɪklz/), there is a physically evident flaw or vocalic discrepancy between the mid-central neutral vowel (*schwa*) /ə/ and the short-front vowel /ɪ/. Similarly, in the transcription data where *equestrian* (/ɪˈkwɛstriən/) is distorted into the phrase *a question* (/ə ˈkwɛstʃən/), it involves a radical substitution of the plosive-liquid cluster /tr/ with the affricate /tʃ/. Acoustically, both word pairs possess substantial articulatory flaws. However, why are listeners still able to decode the hidden meanings behind them and grasp the humor? The answer lies in the integrity of the consonantal framework and the retention of syllable rhythm maintained by the speaker. The human brain possesses a cognitive tendency to prioritize the consonant sequence (*consonant skeleton*) and the syllable stress pattern as primary perceptual anchors in word recognition. As long as the consonantal constants such as the /t-n-t-k-l-z/ sequence in the word *tentacles* and the stress rhythm on the primary syllables (/ˈtɛn/ and /ˈkwɛs/) remain within a parallel corridor, any variations or distortions in the vocalic segments will be automatically tolerated and independently reconstructed by the listener's auditory system.

Furthermore, this empirical interpretation asserts that phonological tolerance does not operate autonomously within a vacuum; rather, it relies absolutely on the coordination of the surrounding textual

architecture (*co-text*) which triggers semantic priming. The *setup* within the comedic structure acts as a cognitive stimulant that aggressively projects specific semantic domains into the listener's working memory. When the speaker delivers priming words such as "octopus" or "cow," the domains of marine biology and zoology are instantaneously and fully activated within the listener's brain. The activation of these domains radically lowers the cognitive threshold required for the brain to retrieve the target words (*tentacles* or *calculator*).

Consequently, when the speaker delivers a flawed or manipulated acoustic stimulus at the *punchline*, the semantic proximity established by the *co-text* acts as an explanatory bridge. The listener's brain compensates for the phonetic distortion received by the auditory senses by forcing the flawed acoustic input into the semantic slot previously prepared by the *setup*. It is at this precise point that incongruity-resolution occurs the humor culminates exactly when the listener realizes a dual compliance, wherein a single, phonetically distorted acoustic shell is contextually forced to lock and reconcile two discourse scripts that initially clashed in extreme opposition. This empirical conclusion affirms that the flexibility of digital spoken discourse provides a substantially broader linguistic playground for humor creation, as the threshold of human auditory tolerance proves to be highly elastic when supported by robust contextual scaffolding.

3.3 Comparison and Modification of Media Ambiguity Theory

The analytical juxtaposition between the findings of this study and previous literature initiates a fundamental theoretical modification to the locus of media ambiguity studies. Several comprehensive previous studies, such as those conducted by Charina (2017), Faridah et al. (2022), and Huluqiah et al. (2025), have consistently posited that the humor and appeal of media discourse are predominantly driven by lexical and syntactic ambiguities. These conclusions are deeply bound to written communication media or static imagery, such as newspaper headlines, television advertising slogans, and internet memes. This study modifies this perspective by providing empirical evidence that when the medium shifts to the realm of digital audio-visual spoken discourse (YouTube), the epicenter of meaning production no longer relies on dictionary-based semantics. Instead, the primary driving engine of comedy is engineered purely through micro-linguistic manipulations at the segmental phonetic level. This modification affirms that the analysis of spoken discourse demands a phonological approach, rather than a merely lexical one, in dissecting the anatomy of humor.

This finding simultaneously presents a theoretical novelty that sharply addresses a cross-linguistic research gap, particularly when compared to the findings of Candra et al. (2024). In their analysis of Indonesian-language TikTok comedy discourse, Candra et al. explicitly confirmed the absence of phonetic ambiguity cases, thus restricting their findings purely to grammatical and lexical deviations. The fact that phonetic traps thrive so abundantly within English spoken content yields a new theoretical formulation regarding the impact of Orthographic Depth on humor creation. English possesses a deep orthography, characterized by an extreme discrepancy between its written spelling and its actual phonetic realization. Consequently, the English auditory stream provides an exceptionally rich terrain for engineering acoustic camouflage a structural phenomenon that naturally cannot manifest easily in languages with a shallow orthography and consistent spelling, such as Indonesian.

Furthermore, this study reconstructs the sociolinguistic perspective regarding the function of ambiguity in digital social interaction. Previous studies focusing on Generation Z social media, such as the research conducted by Putri & Sabrina (2024), posit ambiguity as a catalyst for communicative dysfunction, obscurity, and misinterpretation in everyday discourse. In the realm of digital comics, Nashshar and Mulyono (2021) similarly observed that ambiguity (predominantly of the lexical type) frequently results in unforeseen plot shifts for the readers. Contrary to the paradigm that views verbal ambiguity as a barrier to literacy, this study demonstrates that within the podcast comedy ecosystem (YouTube), semantic incongruity and the obscuration of phonetic references do not constitute a communicative failure. On the contrary, phonetic ambiguity is exploited deliberately, cooperatively, and transparently as an instrument of rhetorical aesthetics. In this context, the "semantic disorientation" experienced by the listener is a conscious, mutually agreed-upon design aimed at achieving a comedic shock effect.

As a culmination of this comparative analysis, the modification of media ambiguity theory asserts that a shift in the communication medium inherently necessitates a corresponding shift in linguistic instruments. Previous theories relying heavily on polysemy and homonymy have proven highly effective for analyzing static texts; however, they become less relevant when confronted with the flexibility of rapid spoken conversational streams (*connected speech*). This study establishes a new postulate: within digital spoken media, word boundaries (*juncture*) and articulatory features (vowels/consonants) act as highly elastic variables that can be freely manipulated by the speaker. Therefore, the integration of phonetic analysis into media discourse studies is no longer a mere supplementary approach, but an absolute necessity to uncover how acoustic illusions can engineer reality and instantly manipulate the listener's perception.

3.4 Synergy of Maxim Flouting

The pragmatic exploration in this study fundamentally departs from and simultaneously expands upon the architecture of Grice's Cooperative Principle, the validity of which has been affirmed by Hossain (2021). Within the constellation of contemporary pragmatic research, maxim flouting is generally positioned as a singular instrument driven by specific functional motives. For instance, the study by Fitriani and Sudarwati (2025) found that maxim manipulation on social media operates specifically as a tool for psychological manipulation (*breadcrumbing*), while the Maxim of Relation is dominantly and exclusively flouted for the purposes of topic avoidance or the negotiation of cultural politeness (Firda et al., 2021; Hasanah, 2021). This body of previous literature has cultivated a theoretical tendency to view maxim flouting as a linear anomaly that operates autonomously and in isolation within a single communicative framework.

The present study dismantles this assumption of autonomy by providing empirical findings that demonstrate a structural anomaly within phonetic spoken comedy discourse. The research data (100%) indicates the occurrence of obscured sound articulation, which flouts the Maxim of Manner. However, the most crucial theoretical finding is this: the flouting of the Maxim of Manner never possesses comedic explosive power if left to operate independently. In contrast to talk show discourse, which is dominated by the isolated flouting of the Maxim of Quantity (Al-Mazari & Rababah, 2024), or medical podcast communication, which relies heavily on flouting the Maxim of Quality, acoustic illusions in humorous discourse strictly demand the presence of an accompanying maxim flouting. Acoustic obscuration without the support of a manipulated context would merely result in genuine linguistic confusion (*miscommunication*), rather than a cathartic effect or laughter.

Therefore, this study formulates the theoretical *novelty* that phonetic-based spoken comedy is driven by the mechanism of *Multiple Flouting Synergy*. Within the anatomy of this synergy, the speaker must first engineer narrative tension by fabricating absurd factual falsehoods (flouting the Maxim of Quality, 20 data) or by forcefully deviating from the conversational context (flouting the Maxim of Relation, 10 data). In alignment with the expanded incongruity humor framework proposed by Cahyaningtias and Kurniawan (2026) and Saefudin et al. (2023), this incongruity is designed as a *setup* to instill false cognitive expectations within the listener. The subsequent flouting of the Maxim of Manner at the end of the utterance (*punchline*) then serves as a surprising resolution, wherein the phonetic flaw is forced to reconcile two reality scripts that initially collided.

As the culmination of this theoretical modification, the concept of Multiple Flouting Synergy challenges and deconstructs traditional pragmatic views, which frequently perceive multiple maxim non-observance as communicative residue or interactional failure. Within the landscape of digital audio-visual comedy, this cross-maxim collaboration proves that flouting is not merely a grammatical accident, but rather a rhetorical architecture constructed with profound precision and calculation. This synergy establishes a new postulate: textual narrative tension (flouting the Maxim of Quality/Relation) and acoustic illusion (flouting the Maxim of Manner) operate interdependently like two sides of the same coin. The degree of humor in a phonetic spoken discourse is proven to be directly proportional to how sophisticatedly and cohesively a speaker can stack (*layering*) various maxim flouts simultaneously within a single conversational breath.

4. CONCLUSION AND SUGGESTIONS/RECOMMENDATIONS

4.1 Conclusion

Based on the research results, the creation of sound-based wordplay (phonetic wordplay) in digital spoken discourse is proven to operate as an integrated mechanism between textual demands and phonetic manipulation. Phonologically, humor is constructed through the precise manipulation of vowel phonemes, consonants, and word boundaries (junctures) which produces absolute homophony (perfect puns) or near-homophony (imperfect puns). Pragmatically, the engineering of this sound anatomy is consistently executed as a strategy to flout the Maxim of Manner, which operates multiply alongside the initial sentence (setup) containing factual lies (flouting of the Maxim of Quality) or topic leaps (flouting of the Maxim of Relation). The resolution of the humor ultimately depends heavily on the listener's phonological tolerance and cognitive reinterpretation in unraveling the acoustic illusion.

4.2 Suggestions/Recommendations

Future studies are suggested to expand the scope of analysis to suprasegmental aspects such as intonation, word stress, and sound rhythm. Furthermore, exploration utilizing a multimodal approach that combines visual comedy elements and sociocultural dynamics is highly recommended to dissect the mechanisms of spoken humor discourse more comprehensively in the future.

REFERENCES

Al-Mazari, W. Y., & Rababah, L. M. (2024). Flouting Grice's Maxims in a Jordanian TV talk show. *Jordan Journal of Modern Languages & Literatures*, 16(2), 443-460.

- Cahyaningtias, A., & Kurniawan, M. A. (2026). Humor, language, and visual incongruity: A discourse analysis of the Sengklek Family comic by Julian Syahputra. *Journal of Pragmatics and Discourse Research*, 6(1), 115-130.
- Candra, M., Poernomo, K., Maharani, F., & Marwan, I. (2024). Ambiguitas Makna Tren Tiktok Drama Resep Makanan Untuk Menciptakan Humor. *LINGUISTIK: Jurnal Bahasa dan Sastra*, 9(3), 577-587.
- Charina, I. N. (2017). Lexical and syntactic ambiguity in humor. *International Journal of Humanity Studies (IJHS)*, 1(1), 120-131.
- Faridah, L. A., Akhir, M., & Muliana, H. (2022). Ambiguitas Makna dalam Slogan Iklan Makanan dan Minuman di Televisi: Suatu Tinjauan Semantik. *DEIKTIS: Jurnal Pendidikan Bahasa dan Sastra*, 2(2), 193–200. <https://doi.org/10.53769/deiktis.v2i2.269>
- Firda, I. N., Hidayat, D. N., Alek, A., & Nurhalimah, N. (2021). An analysis of flouting maxim in a talk show program in Indonesia. *EDUVELOP: Journal of English Education and Development*, 4(2), 107-118.
- Fitriani, R. N., & Sudarwati, E. (2026). Pragmatics Of Breadcrumbing: Implicature And Maxim Violations In Manipulative Communication. *Language Literacy: Journal of Linguistics, Literature, and Language Teaching*, 1(1), 439-446.
- Hasanah, H. (2021). Pelanggaran Maksim dalam Menciptakan Humor: Studi Perbandingan Antara Humor Madura dan Humor Jawa (Violation Of Maxims In Creating Humour: A Comparative Study Between Madurese And Javanese Humour). *SAWERIGADING*, 27(2), 205-218.
- Hempelmann, C. F. (2003). *Paronomasia and orthographic puns: The linguistic and cognitive bases of sound-based imperfect puns* [Doctoral dissertation]. Purdue University.
- Hossain, M. M. (2021). The application of Grice maxims in conversation: A pragmatic study. *Journal of English Language Teaching and Applied Linguistics*, 3(10), 32-40.
- Huluqiah, D. A. U., & Nurhapitudin, I. (2025). Lexical Ambiguity in English Puns on Memes Posted by X Account@ Punhubonline: English. *Journal of English Development*, 5(1), 73-86.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Sage.
- Nashshar, M. N., & Mulyono, M. (2021). Ambiguitas dalam Komik Strip Pepekomi: Kajian Semantik. *Bapala*, 8(03).
- Nisa, N. C., & Rizal, G. A. (2025). Analysis of Ambiguity and Multiple Interpretations in Language on Social Media: A Semantic Study. *DIALEKTIKA: Jurnal Pendidikan Bahasa Indonesia*, 9(1).
- Palmann, A., & Miller, T. (2025). What's in a pun? Assessing the relationship between phonological distance and perceived funniness of punning jokes. *HUMOR*, 38(4), 643–658. <https://doi.org/10.1515/humor-2024-0060>
- Putri, D. A., & Sabrina, A. H. N. (2024). Ambiguitas Berbahasa pada Sosial Media dan Percakapan Sehari-Hari Generasi Z. *Enggang: Jurnal Pendidikan, Bahasa, Sastra, Seni, dan Budaya*.
- Saefudin, D. P., Mulyadi, & Santosa, P. P. P. (2023). The Analysis of Flouting Maxim in the @Pepekomi comic strip. *SALEE: Study of Applied Linguistics and English Education*, 4(2), 367–379. <https://doi.org/10.35961/salee.v4i2.764>