Dissonant Ballerinas and Crafty Carrots: A Comparative Multi-modal Analysis of Italian Brain Rot

Anca Dinu[1,2,*], Andra-Maria Florescu[1,3], Marius Micluța-Câmpeanu[1,3,4], Stefana-Arina Tăbușcă[1,3], Claudiu Creangă[1,3,4] and Andreiana Mihail[1,4]

[1]*University of Bucharest, 90 Panduri Road, Bucharest, 050107, Romania*
[2]*Faculty of Foreign Languages and Literatures, 5-7 Edgar Quinet St, Bucharest, 010017, Romania*
[3]*Interdisciplinary School of Doctoral Studies, 36-46 Mihail Kogălniceanu, Bucharest, 050107, Romania*
[4]*Faculty of Mathematics and Computer Science, 14 Academiei St, Bucharest, 010014, Romania*

Abstract

This paper presents a comparative multi-modal analysis of Italian and Romanian brain rot memes, investigating the factors that contribute to its appeal and the linguistic and cultural distinctions between the two versions. To conduct this analysis, we introduce a multi-modal brain rot dataset named CRIB (**C**ollection of **R**omanian and **I**talian **B**rain rot), a manually curated collection of 240 TikTok videos stratified by language (Italian, Romanian) and popularity, on which we examine textual, acoustic, and visual features. Our findings indicate that popularity is not significantly correlated with textual elements like sentiment, absurdity, or rhyme, or acoustic elements such as vocal features or sentiment of the sound. Instead, in Romanian language, video-level dynamics, specifically faster cutting speeds and a more rapid overall pace, are strong predictors of a video's success. The cross-linguistic analysis reveals significant differences. Italian brain rot is textually more negative, exhibits higher perplexity, and uses more rhyme, while its sound is characterized by higher melodic range and loudness. Romanian audio is spectrally brighter with more erratic pitch variations.

Keywords

data set, brain rot, multi-modal, Italian, Romanian

1. Introduction

The first recorded use of 'brain rot' is in Henry David Thoreau's book *Walden*, published in in 1854, which criticizes society's tendency to devalue complex ideas in favor of simple ones, indicating a general decline in mental and intellectual effort: "While England endeavors to cure the potato rot, will not any endeavor to cure the brain-rot – which prevails so much more widely and fatally?" [1].

Brain rot was the Oxford word of the year 2024.[1] Its primary sense is the supposed deterioration of a person's mental or intellectual state as the result of overconsumption of low-quality, trivial, or non-challenging online content. Its secondary sense, acquired over the last year, is the online content itself likely to lead to such deterioration. In particular, the term came to be used in the last months to refer to a certain type of multi-modal short content, intentionally created to be absurd, nonsensical, dissonant and funny, by content developers using generative AI. One of the earliest examples is *Nothing, Forever* in December 2022[2] (shortly after the launch of ChatGPT), while Italian brain rot is a more recent trend that gained popularity in early 2025. As a side note, short animations like *Skibidi toilet* series or *only in Ohio* memes series are not usually created using generative AI, though they are also considered a form of brain rot.

The creator of some of the first Italian brain rots, like Ballerina Cappuccina, supposedly a Romanian,[3,4] describes them as a satiric artistic experiment that both mocks and celebrates pop culture and kitsch. The characters in these creations are childlike, weird, and often grotesque blends of humans, animals, plants and various objects, named with Italian-sounding names, like *Ballerina Cappuccina* in figure 1. The Italian brain rot phenomenon gained traction especially among Gen Z and Gen Alpha communities by means of social media platforms like TikTok and Instagram and has quickly spread to other languages, including Romanian, with characters such as *Morcoveață* in figure 2, a human-like carrot character adapted from Romanian nursery rhymes.

A common trait of brain rot is the uncertain or lack of authorship. As Roland Barthes argued in *The Death of the*

CLiC-it 2025: Eleventh Italian Conference on Computational Linguistics, September 24 — 26, 2025, Cagliari, Italy
*Corresponding author.
anca.dinu@lls.unibuc.ro (A. Dinu); andra-maria.florescu@s.unibuc.ro (A. Florescu); marius.micluta-campeanu@s.unibuc.ro (M. Micluța-Câmpeanu); stefana.tabusca@s.unibuc.ro (S. Tăbușcă); claudiu.creanga@s.unibuc.ro (C. Creangă); andreiana.mihail@s.unibuc.ro (A. Mihail)
0000-0002-4611-3516 (A. Dinu); 0009-0007-1949-9867 (A. Florescu); 0009-0002-3197-170X (M. Micluța-Câmpeanu); 0009-0003-3376-6416 (S. Tăbușcă); 0009-0005-5443-5506 (C. Creangă); 0009-0003-0640-6319 (A. Mihail)

[1] Oxford University Press Word of the Year 2024
[2] Hacker News post, 2nd December 2022
[3] New York Times article, 30th April 2025
[4] Interview in Romanian, Mindcraft Stories, 27th May 2025

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073
published 2025-11-30

Author [2], removing the author takes away a traditional source of authority and shifts the focus to the reader and their interpretation. In a similar way, brain rot content is mostly anonymous or pseudonymous, remixed and layered, reflecting an anarchic and collective form of creation.

Moreover, brain rot has many common traits with the Dadaism movement, which was on the same path of anti-art and anti-meaning, it mocked the art rules and traditions, and it embraced nonsense and absurdity. It celebrated chaos and irrationality, at the same time being psychedelic in the sense of artistic production from the 70s and 80s when using vivid colors, fractals, surreal images, and neon gradients. Both forms of art were born in an age of post truth and information overload.

Finally, the brain rots' original purpose seems to be primarily amusement, but some of them were also used for manipulation, commercial purposes or even political propaganda.

This study aims to explore whether brain rot manifests differently across cultures and languages. We chose to begin with Italian brain rot, as it represents one of the most visible and influential starting points for this phenomenon. The decision to compare it with Romanian brain rot is based on both similar and contrasting historical, linguistic, and cultural characteristics. The main common grounds between the two languages and cultures are that they both belong to the Romance language family, and that they have experienced authoritarian regimes in the 20th century that shaped their collective imagination, creative approaches and imagery. The differences between the two lie in their geopolitical contexts, religious traditions (Catholic and Greek Orthodox), and features of cultural production, which influence the tone, style, and content of their digital aesthetics. At the same time, both Italian and Romanian online cultures remain underrepresented in cultural and computational studies, which represents an opportunity to examine how the upper similarities and differences are reflected in brain rot popular manifestation.

The main research questions are: (1) What makes a brain rot go viral (besides algorithm recommendation/-manipulation or pure chance)? and (2) Are there any cultural or language differences between the form or content of brain rots in these two languages?

2. Related Work

Numerous studies are currently interested in understanding what are the effects of digital content overconsumption. Most of them focus on psychological, neuro-biological, or meta-analytical perspectives [3, 4].

This type of digital content certainly breaks conventional norms of art, narrative, and symbolism, aligning

Figure 1: Example of Italian brain rot character *Ballerina Cappuccina*

Figure 2: Example of Romanian brain rot character *Morcoveață*

well with both surrealism and absurdism, while also embracing the chaos and disjuncture of postmodernity. The chimera-like mixing of humans, animals, plants, and objects is not a new phenomenon, since anthropomorphism can be traced back from Egyptian, Greek, and Roman antiquity, going through Medieval art, all the way to the surrealist movement at the beginning of the 20th century (Max Ernst, Salvador Dali, or René Magritte).

To the best of our knowledge, no computational research has been conducted specifically on Italian brain rot. However, internet culture, including memes and viral content received a significant amount of attention. The detection of meme toxicity was investigated in [5]. Multimodal sentiment analysis was conducted by integrating text, image, and audio for improving sentiment detection from vlogs, spoken reviews, and human-machine interactions in [6]. A comprehensive survey which categorizes advances in multimodal sentiment analysis can be found in [7]. [8] introduced a new benchmark for detecting hate speech from multimodal memes. [9] combined LLM-generated debates and fine-tuned judge models to detect harmful memes with improved interpretability and performance. [10] proposed a template-based approach for meme clustering by employing multi-dimensional similarity features.

3. Data

The data set was constructed manually from TikTok videos by searching for candidate examples through various methods: direct search queries, tags, the *discover* feature, trending pages, compilation and analysis of video clips, related videos (*You may like*), and the *For You* recommendations. This process cannot be reliably automated due to misleading tags being frequently used to influence the recommendation algorithm.

The extracted samples are stratified across two dimensions: language (Italian and Romanian) and popularity (popular and unpopular). The dataset is well balanced: 120 brain rots per language, with 60 viral examples and 60 less viewed examples, the typical threshold between them being 100k views or at least 10k likes.

Given that we are interested in all aspects of communication, especially creative language use, we filtered out posts with extremely low lexical diversity, as well as re-uploads, translations, and repetitive songs.

Moreover, some of the tools used to generate these videos can be traced by watermarks, although we note that some users specifically crop the content or blur such indicators. In alphabetical order, clips have been created or adjusted with: CapCut, ChatGPT, Hailuo AI, Kling AI (version 1.6), PixVerse.ai, Runway, VEED. TikTok also offers its own tools for video creation.

We collected four subcategories of brain rots for each language. The general category in each language includes the notorious Italian brain rot characters with local adaptations (with 60 examples per language). For Italian, the *skeleton* category consists of 20 videos with a poignant tone. The *Matteo* category comprises 20 memes that exhibit more positive attitudes, and the *Politicians and Celebrities* class consists of 20 political satire videos. For Romanian, we followed the same count and structure. We have selected *schelet*, the corresponding category of Italian *schletro*, *Regina brain rot* (brain rot queen), featuring longer stories, and conversely *Morcoveață*, consisting of very short clips. Throughout this process, we have extracted a total of 240 videos with subtitles and metadata.[5] We name the dataset CRIB (**C**ollection of **R**omanian and **I**talian **B**rain rot).

4. Text

4.1. Sentiment Analysis of Text

To obtain sentiment analysis scores for the textual transcripts of the brain rots, we employed cardiffnlp/twitter-xlm-roberta-base-sentiment pre-trained model [11], which returns a percentage of negative, neutral, and positive sentiments associated with each text in both languages, for optimal comparison. We used ChatGPT4o for coding assistance and Python in Google Colab to obtain the graphical illustrations and to perform statistical tests.

Overall, for all brain rots, the negative sentiment was predominant. There were minimal variations in sentiments in popular versus unpopular brain rots in both languages, as we can see in figures 3a and 3b for Romanian and in figures 3c and 3d for Italian. However, the Italian ones contained more negative sentiments than their Romanian counterpart.

We tested for statistical significance of these findings and the results are listed in table 1 from the Appendix.

For Romanian differences of positive, neutral, and negative sentiments between popular and unpopular texts, we performed a multivariate analysis of variance (MANOVA). The results indicate no statistically significant multivariate effect. To further investigate potential differences at the level of individual sentiments, given potential concerns about normality and homogeneity of variance, Mann-Whitney U tests were performed for each emotional score. The results confirmed that there are no statistically significant differences.

The same tests on the Italian brain rots obtained the same results: no statistical differences between the sentiments of popular and unpopular texts.

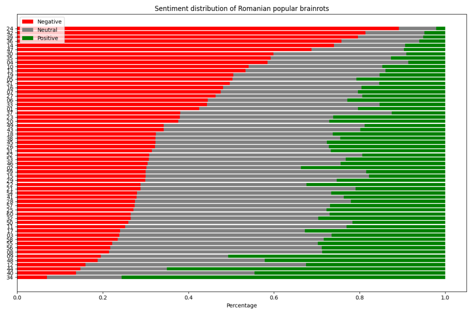
We further performed, for each language, the same statistical tests only on the *general* category of brain rots, which varies less than the combination of all four brain rot categories (*general*, *regina brain rot*, *morcoveață*, and *schelet* for Romanian, and *general*, *matteo*, *politicians and celebrities*, *scheletro* for Italian) to test the differences between popular and unpopular categories. The results showed again no significant differences.

Finally, a multivariate analysis of variance (MANOVA) was conducted to examine whether the distribution of emotional sentiment scores (positive, neutral, and negative) differed between Romanian and Italian brain rots. This time, results were statistically significant, indicating a robust multivariate effect of language on sentiment composition. To further explore individual sentiment differences between languages, Mann-Whitney U tests were applied separately for each sentiment category. Results revealed that positive sentiment was significantly higher in Romanian brain rots ($p < 0.0001$; mean: 0.377 vs. 0.159), that neutral sentiment was also significantly higher in Romanian memes ($p < 0.0001$; mean: 0.385 vs. 0.171), and that negative sentiment was significantly higher in Italian ones ($p < 0.0001$; mean: 0.670 vs. 0.238).

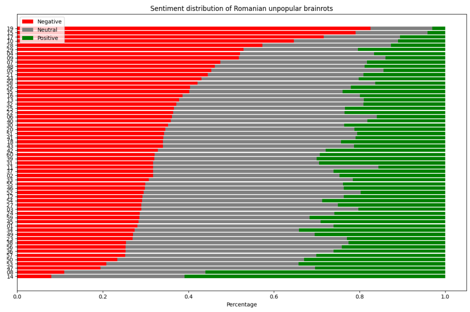
4.2. Semantic Similarity and Perplexity

To measure the degree of “absurdity” or “unpredictability” of the brain rots, we employed two complementary measures: semantic similarity and perplexity. Seman-

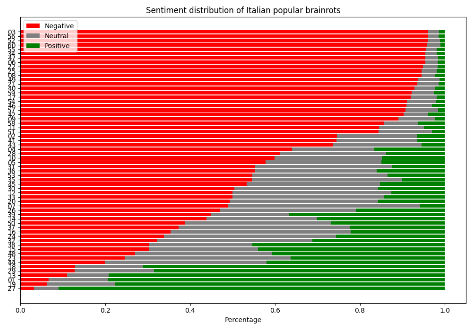
[5]Using yt-dlp version 2025.04.30.



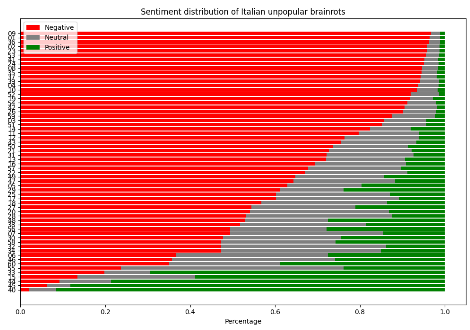
(a) Sentiment scores for Romanian popular brain rots



(b) Sentiment scores for Romanian unpopular brain rots



(c) Sentiment scores for Italian popular brain rots



(d) Sentiment scores for Italian unpopular brain rots

Figure 3: Sentiment scores (Negative, Neutral, Positive) for each language and popularity group of the brain rot texts

tic similarity was computed for a given text of a brain rot as the average pairwise cosine similarity between all word embeddings obtained with the sentence transformer *paraphrase-multilingual-MiniLM-L12-v2* [12]. This reflects how semantically cohesive the vocabulary is — higher values indicate that the words tend to belong to similar semantic fields or contexts, suggesting internal consistency or coherence, while lower values indicate some incoherence or inconsistencies. Perplexity quantifies how unpredictable a text is from the perspective of a pre-trained language model (GPT-2 [13]). Lower perplexity values indicate that the model finds the sequence more predictable and fluent, while higher values suggest syntactic or lexical irregularities, or content that deviates from typical language patterns.

For both semantic similarity and perplexity scores, we tested the differences in means between popular and unpopular brain rots per language and we also tested the difference in mean between the two languages. We used one-way ANOVA and we also conducted non-parametric Mann-Whitney U tests. The results of the statistical tests are summarized in table 2 from the Appendix.

For Romanian brain rots, we evaluated whether the two metrics, semantic similarity and perplexity, differ significantly between brain rot texts categorized as popular versus unpopular. The results of the ANOVA test revealed no statistically significant differences.

The Mann-Whitney tests results also show that there is no significant difference between popular and unpopular brain rot texts for neither semantic similarity or perplexity scores. However, the p-value for semantic similarity was very close to the significance threshold (0.068), indicating a slight preference for more irregular/nonstandard language in popular brain rot texts (M = 168.59 for popular, M = 159.22 for unpopular).

We also performed statistical tests on the Romanian *general* subcategory of brain rots that varies less than the whole Romanian dataset. The ANOVA results confirmed the ones obtained with the the same statistical tests on all the data, with similar p-values that did not surpass the 0.05 threshold. This time, the semantic similarity p-value of the Mann-Whitney test revealed a statistically significant difference in semantic similarity ($p < 0.05$), suggesting that popular brain rots are slightly more semantically coherent (with a difference in mean of 0.002) than unpopular brain rots.

For Italian, the statistical test results suggest that neither semantic similarity, nor perplexity differs significantly across the popular and unpopular brain rot texts. As in the case of Romanian, we also tested the statistical significance of the differences between popular and unpopular texts w.r.t. semantic similarity and perplexity scores, for Italian *general* category of brain rots, but the results were not statistically significant.

Cross-linguistically, we tested whether semantic similarity and perplexity differ significantly between Romanian and Italian brain rot texts. The difference in mean semantic similarity score between Italian (M = 0.59) and Romanian (M = 0.58) was very small. The perplexity scores that reflect language uncertainties or language inconsistencies were substantially higher for Italian brain rot texts (M = 218.74) than for Romanian ones (M = 163.91). While one-way ANOVA test revealed no significant difference in semantic similarity between Italian and Romanian, for perplexity it yielded a statistically significant effect of the language category with a p-value < 0.001, suggesting that the Italian brain rot texts are indeed more language-unpredictable than their Romanian counterparts. The Mann-Whitney tests confirmed the ANOVA results showing no significance for semantic similarity ($p = 0.06$), and a highly significant difference between the two groups for perplexity ($p < 0.0001$).

4.3. Rhyme

We estimated the rhyme density by the following methodology. We employed a computational method based on sub-string similarity at word endings in two and three letters. Since the speech to text automatic transcription did not identify correctly the verses with end-line, we checked for all rhyme pairs locally, within a distance of two verses. The rhyme coefficient was calculated as the ratio between the total number of distinct, non-adjacent word pairs that share the same suffix of 2 or 3 letters and the total number of words in the text.

We tested the differences in rhyme scores of popular versus unpopular brain rot texts for both languages, and also the differences in rhyme scores between the two languages with Mann-Whitney tests. The results are shown in table 3 from the Appendix.

The difference of the means between Romanian popular (M = 0.19) and unpopular (M = 0.16) brain rots suggests a slight preference towards more rhymed ones in the popular group. However, for our sample of 60 (30 popular and 30 unpopular brain rots), this difference is not statistically significant ($p = 0.135 \geq 0.05$).

For Italian, there is also no statistical difference between the rhyme coefficient of the popular and unpopular groups, since their means are very close (0.215 vs. 0.228) and the p-value is 0.6880 (>0.05).

We also compared Romanian and Italian rhyme coefficients (120 each). The mean for Italian rhyme coefficient is 0.221, higher than the mean for Romanian which is 0.177. This time, the Mann-Whitney test returned the p-value of 0.0001 (<0.05), which means that Italian brain rots do rhyme more than their Romanian counterparts.

5. Sound

5.1. Audio Processing and Separation

We developed a multi-stage audio pipeline to isolate and characterize the vocal component of Italian and Romanian brain rot, with the goal of quantifying attributes that may influence their popularity, as well as perform comparisons between the two languages. First, each video file was converted to a high-resolution WAV format and processed with a state-of-the-art source-separation model (Demucs) [14] to obtain separate vocal and music stems. Vocal tracks were then normalized for consistent loudness, ensuring that subsequent analyses would not be affected by variations in recording level.

5.2. Speech Features

After isolating the vocal stems, we computed a comprehensive set of acoustic descriptors. Pitch contours were extracted via *librosa*'s [15] `pyin` algorithm, yielding mean F0, F0 variance, 95th–5th percentile range, and slope-entropy to quantify melodic movement. On 25 ms/10 ms-hop frames, we computed Mel-frequency cepstral coefficients (MFCCs) 1–13 (means and variances), spectral centroid, bandwidth, roll-off, zero-crossing rate, and spectral-flux (means and variances) to capture brightness, noisiness, and timbral dynamics. Rhythmic patterns were quantified by onset detection—calculating syllable-rate and pause-duration statistics—and by RMS (root mean square) energy envelopes (temporal centroid, mean, and variance). Each clip's features formed one row in a master data matrix.

We then compared these matrices in two ways: within each language (popular vs. unpopular memes) and across languages (Italian vs. Romanian). All continuous features were tested with Mann–Whitney U and corrected for multiple comparisons using the Benjamini–Hochberg procedure at $q < 0.05$ [16].

After correction, for the analyses of the popular vs. unpopular groups within each language, no features remained significant, likely due to modest sample sizes and the number of comparisons. Reporting raw $p < 0.05$ findings (visible in Table 5 from the Appendix) highlights, though, that in the Italian dataset, popular brain rots exhibited significantly greater fundamental-frequency variance (var_f0) and a wider pitch range (range_f0) than their unpopular counterparts. These results suggest that a more expressive, melodic content, characterized by larger and more varied pitch intervals, is associated with higher popularity in Italian clips. In the Romanian corpus, popular audios differed from unpopular ones in three respects: they featured a marginally faster onset rate (syllable_rate), a more stable mid-high spectral texture (mfcc_9_var), and a darker high-frequency timbre (mfcc_11_mean). These patterns imply that a slightly quicker rhythm, smoother structure in the mid-high band, and less pronounced very high frequencies tend to cooccur with popularity. None of these comparisons survived false discovery rate (FDR) correction at q < 0.05, reflecting exploratory sample size considerations.

In addition to our broad comparisons, we also examined popular versus unpopular clips within each thematic category—namely *general*, *scheletro*, *matteo*, and *politicians and celebrities* in Italian and *general*, *schelet*, *morcoveață*, and *regina brain rot* in Romanian. After applying Benjamini–Hochberg correction, only the Romanian *general* subset yielded features that remained significant. In those 60 clips, one clear acoustic hallmark of popularity is a smoother timbre in the mid-to-high frequencies. Specifically, popular memes have a noticeably lower mfcc_9_var (181 vs. 232 in unpopular clips), meaning the shape of the spectrum around 3–4 kHz stays more consistent rather than flitting up and down. This steadiness makes the audio feel more even and less "choppy". There is also a trend toward lower mfcc_11_var (170 vs. 193), which means fewer abrupt jumps in the very high frequencies (around 5–6 kHz), so sibilance and hiss are kept under tighter control. These findings suggest that Romanian *general* brain rots become more engaging when their soundtrack maintains a steady, coherent tone.

When we examine the full set of 120 Italian and 120 Romanian brain rots side by side (as can be seen in Table 4 from the Appendix), a clear pattern of prosodic contrast emerges. Italian vocals move through a broader and more dynamic pitch range, peaking and dipping over a span that is markedly larger than in the Romanian samples, which makes them feel more melodically engaging. Romanian clips, on the other hand, swing their pitch contours in a less predictable fashion, with higher entropy of slope suggesting sudden twists in intonation that can come across as more spontaneous or volatile. In the spectral domain, Romanian brain rot tracks push energy into the upper frequencies: their average spectral centroids, bandwidths, and roll-off points all sit higher than in Italian. Yet these high-frequency bands in Romanian audio are less stable over time, fluctuating more from moment to moment and lending a grainier, more restless timbral texture. Italian clips display a darker, more subdued high-end, but they deliver their muted tones with greater consistency and, at the same time, add sharp bursts of change, especially in those same upper bands, so that the audio doesn't feel monotone. They also sound uniformly louder, amplifying their broad, dramatic pitch gestures and deep spectral shifts, whereas Romanian tracks embrace a leaner, quieter tone. These findings point to two distinct audio "signatures" that likely reflect both the underlying voice synthesis models and the cultural styles of meme production in each language.

5.3. Sentiment Analysis of Sound

Full-audio (voice + music) sentiment was assessed using the pretrained speech-emotion classifier superb/wav2vec2-base-superb-er [17]. It categorizes short audio clips into four emotion labels (*angry*, *happy*, *neutral*, *sad*). Each brain rot's waveform was resampled to 16 kHz and fed into the model, after which we captured the full probability distribution over three coarse categories, Negative (-1), Neutral (0), and Positive (+1). These labels were mapped from the original model outputs as such: Negative (-1) for any emotion other than *happy* or *neutral* (i.e. *angry* and *sad*), Neutral (0) for labels predicted as *neutral*, and Positive (+1) for labels predicted as *happy*. This mapping yields a better comparison point for our study, in alignment with the sentiment analysis performed for the brain rots' texts. Separate sentiment files were generated for four groups: Italian-Popular, Italian-Unpopular, Romanian-Popular, and Romanian-Unpopular. The distributions for each audio in the corresponding groups is shown in Figure 4.

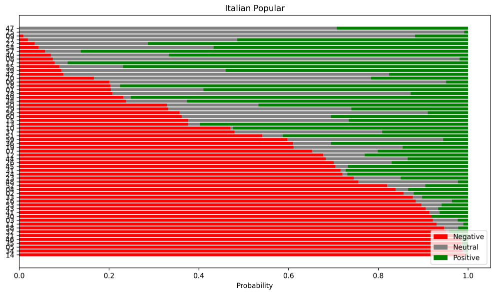
When we treat the three sentiment probabilities (Negative, Neutral, and Positive) as a multivariate outcome, neither Italian nor Romanian shows a significant difference between popular and unpopular brain rot audios. A MANOVA on the Italian data yields Wilks' $\lambda = 0.992$ ($F(3,116) = 0.30$, $p = 0.826$), and the same test on Romanian returns Wilks' $\lambda = 0.966$ ($F(3,116) = 1.37$, $p = 0.257$).

However, comparing Italian versus Romanian clips reveals a modest but statistically significant language effect on the combined sentiment vector (Wilks' $\lambda = 0.965$, $F(3,236) = 2.87$, $p = 0.037$). In other words, the full-audio sentiment differs more by language than by popularity within a language.

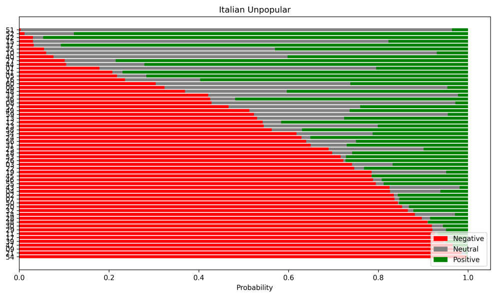
We also used Mann–Whitney tests to look at each sentiment dimension in isolation and to account for potential concerns about normality and homogeneity of variance. Comparing Italian against Romanian across all 240 clips, the sentiment score distributions show however no robust separation: Negative (U=6,536; p=0.217), Neutral (U=8,088; p=0.099), and Positive (U=6,878; p=0.550) all lie above the conventional 0.05 threshold. The Neutral score comes closest ($p \approx 0.10$), hinting at a possible tendency for Italian memes to register slightly higher neutral-vibe probabilities than Romanian ones, but this trend remains statistically inconclusive.

6. Video

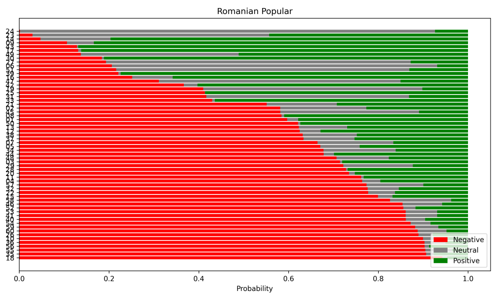
To analyze the visual content of the 240 videos, we employed the Gemini Flash 2.5 multimodal model. Each video was individually processed by the model with a prompt instructing it to perform a visual-only analysis and extract a series of predefined attributes. The model returned its analysis for each video as a structured JSON



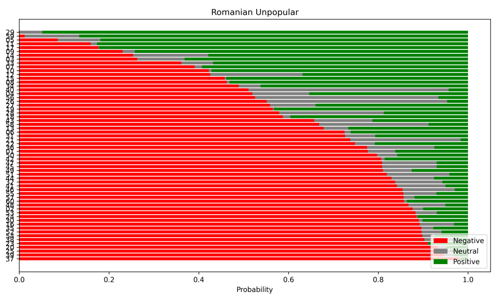
(a) Italian Popular Brain Rot Audios



(b) Italian Unpopular Brain Rot Audios



(c) Romanian Popular Brain Rot Audios



(d) Romanian Unpopular Brain Rot Audios

Figure 4: Continuous-sentiment distributions (Negative, Neutral, Positive) for each language and popularity group of the brain rot audios

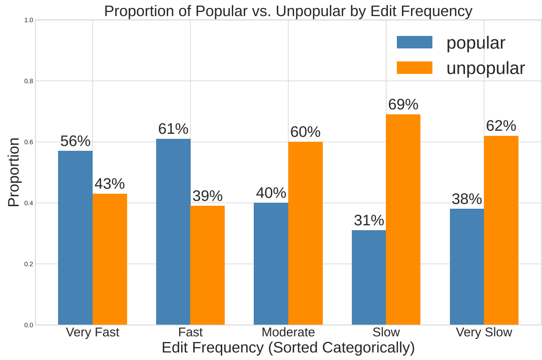


Figure 5: Faster video editing speeds show a stronger correlation with popularity than slower editing speeds

object, which allowed for the systematic collection of data for our study. In an initial test, the model demonstrated a notable capability in discerning the language of origin from visual cues alone, achieving an accuracy of 71.67% (Romanian vs. Italian) by identifying culturally specific items (e.g., the Carpathian Mountains). However, it struggled significantly to predict a video's popularity, achieving an accuracy of only 47.92% (random chance). This initial finding suggests that a video's success is not determined by straightforward, immediately classifiable visual markers of appeal. A deeper statistical analysis was therefore performed to uncover more subtle visual attributes that may correlate with popularity (Table 6 in the Appendix).

A key finding emerged from the analysis of the videos' dynamic properties, specifically the rate of the shot transitions. The data indicates a clear and statistically significant tendency for popular videos to feature a much faster cutting speed. Videos categorized with *Very Fast* or *Fast* transitions were substantially more prevalent among popular content, as it can be seen in figure 5. This observation is supported by a positive Pearson correlation of 0.1712 between cutting speed and popularity, which was found to be statistically significant ($p=0.0231$). This suggests that as the frequency of cuts increases, so does the likelihood of a video being popular, pointing to a dynamic, high-energy visual style as a key component of audience engagement.

Further reinforcing the importance of dynamism, the overall pacing of the videos also proved to be a significant differentiator. A statistically significant difference was found in the distribution of pacing levels between popular and unpopular videos, as confirmed by a Mann-Whitney U Test ($p=0.0492$). Popular videos were most frequently described as having a *Fast* overall pace. Moreover, when the correlation between pacing and popularity was analyzed by language, a significant difference was observed.

For the Romanian videos, we found a positive Pearson correlation of 0.2791, indicating that as overall pacing increases, videos in Romanian tend to be more popular. In stark contrast, the Italian videos showed a negligible correlation of 0.0047, revealing virtually no linear relationship between pacing and popularity. This language-specific finding suggests that the overall positive trend observed in the combined dataset is almost exclusively driven by the Romanian-language content. While the perceived tempo is an important factor, its influence on popularity appears to be culturally moderated.

In contrast to the clear influence of dynamism, thematic elements such as absurdity and narrative structure showed no significant correlation with video popularity. Although it was hypothesized that surreal, chaotic, or illogical content might be a good indicator of popularity, the analysis did not bear this out. Perceived absurdity levels yielded very low Pearson correlation coefficients and statistically insignificant p-values in the Mann-Whitney tests. For instance, the overall absurdity level, despite being *High* in the majority of videos, had no discernible statistical relationship with popularity (p=0.2937).

In conclusion, the image-level analysis indicates that while the kinetic and rhythmic aspects of a video's construction are influential, their effect on popularity is not universal. The visual dynamism, characterized by rapid cutting and a fast overall tempo, appears to be a powerful element in capturing and retaining audience attention, but this trend is highly dependent on the cultural context of the video. Our data shows this relationship is strong for Romanian-language content but non-existent for Italian-language content, suggesting that the preference for such a style is culturally specific rather than a general driver of engagement.

7. Manual Analysis

One of the most striking brain rot characteristic revealed by the manual scrutiny of all the brain rots represents their core element: the dissonance between the tone, the music and the content. The tone is neutral - an AI generated voice- but the text is often deviant, triggering some extreme emotions. There is no harmony or coherence whatsoever between text, speech, music and image. Another notable trait is the unsettling blend of baby talk, nursery rhymes, and children inappropriate content (topics) and language (slang, pejorative jargon, NSFW words).

Overall, the brain rots seem a good example of E(xtended)-creativity, rather than of F(ixed)-creativity, since they tend to break the rules and not only to create new content using existing ones [18].

The manual analysis of the textual content revealed the prevalent topics and characters used in both languages. Some Romanian brain rot content includes political propaganda related to the presidential elections, often carrying extremist nationalist undertones. This suggests that such content goes beyond seemingly harmless absurdist humor and may serve as manipulative material.

The Romanian characters are inspired from the Romanian folklore, such as the Balaur (a dragon-like creature), Morcoveață (a carrot shaped boy inspired by Jules Renard's Poil de Carotte), from Romanian historical figures such as rulers or poets (Mihai Viteazul, Ion Creangă, Mihai Eminescu), or from global pop culture such as Hatsune Miku, Sonic the Hedgehog, or Disney characters, portrayed in explicit real-life situations like relationship with siblings or modern dating.

The topics in Italian brain rot also revolve around daily routine and politics, with celebrity characters such as Volodimir Zelensky or Emmanuel Macron, portrayed in funny and ironical ways, frequently with misspelled names in order to undermine their authority and to deglamorize them. There are also very popular characters like Ballerina Cappuccina, a tragic but very graceful figure featuring sometimes grotesque ballet poses, the ironical and dreamy Skeletro, with emo fragility and self deprecating depression and out dated romanticism.

The music is mostly depressive both in popular and unpopular Italian or Romanian brain rot, written in minor tonality which gives to the brain rots a serious, sad and dark sounding. We identified with the Shazam tool a variety of musical pieces, from classical music by Frederic Chopin or Max Richter, through very popular tunes by Ennio Morricone or Bobby McFerrin all the way to some trap and rap pieces. Most of the titles of these pieces include the words *spooky*, *scary*, *depressive*. We noticed that for the popular category the music is slightly less dark than for the unpopular. Italian popular brain rots showcase a wide array of soundtracks from various genres, while unpopular ones typically feature the same three most used tracks used by viral brain rots. At the same time, the Romanian unpopular samples contain diverse soundtracks from obscure sources and underground artists. Conversely, a single track is present in more than half of all Romanian popular brain rots. This suggests that more effort was required at the beginning of this trend in Italy, but after gaining notoriety, the key to viral clips was to exploit the same soundtracks as established by the initial wave, making them more recognizable.

The visuals are overloaded, kitschy, anti-narrative, absurd, over-sized, ironic, cynical, and, in a way, self-destructive. The popular ones are more animated, present more complex and colorful imagery, the cutting is obviously more dynamic, full of narrative.

8. Conclusion

This study conducted a multi-modal analysis of Italian and Romanian brain rot, seeking to identify the factors driving popularity and to map the cultural and linguistic differences between the two languages. Our findings show that the popularity of these memes is not primarily determined by their textual content. Features like sentiment, narrative absurdity, or rhyme density showed no significant link to a video's success. The same can be said about standard audio features and speech sentiment. Instead, the analysis revealed that popularity is strongly correlated with visual dynamics, specifically a faster cutting speed and overall pace (for Romanian language).

The research also uncovered clear distinctions between the Italian and Romanian versions of brain rot. Italian brain rots were textually more negative, more unpredictable, and used rhyme more frequently. Acoustically, their vocals were characterized by greater melodic range and loudness. Conversely, Romanian content was more neutral, while its vocals were spectrally brighter and showed more erratic pitch changes. These differences extend to thematic content, with each language favoring culturally specific characters and references.

As this is a global phenomenon, we plan to extend this study to other Romance languages in future work. Regarding the small sample size, we intend to include samples that would not fit in either popular or unpopular categories, thus rephrasing the popularity aspect as a continuous problem rather than as binary classification.

In essence, the success of brain rot appears to depend on a combination of universal and culturally-specific elements. While fast-paced, dynamic visuals serve as a universal driver for engagement, the content itself is distinctly shaped by the linguistic, acoustic, and thematic norms of its target culture. What makes the genre unique is the strange mix of message, image and sound.

9. Limitations

One limitation of this study is that the dataset, while carefully curated, is relatively small and manually collected, focusing only on Italian and Romanian content. This may limit the generalization possibilities of our findings to brain rot in other languages or on a larger scale. Secondly, our analysis identifies strong correlations, such as the link between cutting speed and popularity, but does not establish causation. Other confounding factors not examined here may be at play. Finally, brain rot is a rapidly evolving digital phenomenon, and the features that define it today may change over time, potentially dating our specific observations.

Acknowledgments

This research is supported by the project "Romanian Hub for Artificial Intelligence - HRIA", Smart Growth, Digitization and Financial Instruments Program, 2021-2027, MySMIS no. 334906 and by CNCS/CCCDI UEFISCDI, SiRoLa project, PN-IV-P1-PCE-2023-1701, within PNCDI IV, Romania.

References

[1] H. Thoreau, Walden, Mercer University Press, 2011.

[2] R. Barthes, S. Heath, M. Dove, The Death of the Author, 1977.

[3] A. M. F. Yousef, A. Alshamy, A. Tlili, A. H. S. Metwally, Demystifying the new dilemma of brain rot in the digital era: A review, Brain Sciences 15 (2025). URL: https://www.mdpi.com/2076-3425/15/3/283. doi:10.3390/brainsci15030283.

[4] S. A. Satici, E. G. Tekin, M. E. Deniz, B. Satici, Doomscrolling scale: its association with personality traits, psychological distress, social media use, and wellbeing, Applied Research in Quality of Life 18 (2023) 833–847. URL: https://doi.org/10.1007/s11482-022-10110-7. doi:10.1007/s11482-022-10110-7.

[5] D. S. M. Pandiani, E. T. K. Sang, D. Ceolin, Toxic memes: A survey of computational perspectives on the detection and explanation of meme toxicities, 2024. URL: https://arxiv.org/abs/2406.07353. arXiv:2406.07353.

[6] M. Soleymani, D. Garcia, B. Jou, B. Schuller, S.-F. Chang, M. Pantic, A survey of multimodal sentiment analysis, Image and Vision Computing 65 (2017) 3–14. URL: https://www.sciencedirect.com/science/article/pii/S0262885617301191. doi:https://doi.org/10.1016/j.imavis.2017.08.003, multimodal Sentiment Analysis and Mining in the Wild Image and Vision Computing.

[7] A. Gandhi, K. Adhvaryu, S. Poria, E. Cambria, A. Hussain, Multimodal sentiment analysis: A systematic review of history, datasets, multimodal fusion methods, applications, challenges and future directions, Information Fusion 91 (2023) 424–444. URL: https://www.sciencedirect.com/science/article/pii/S1566253522001634. doi:https://doi.org/10.1016/j.inffus.2022.09.025.

[8] D. Kiela, H. Firooz, A. Mohan, V. Goswami, A. Singh, P. Ringshia, D. Testuggine, The hateful memes challenge: Detecting hate speech in multimodal memes, 2021. URL: https://arxiv.org/abs/2005.04790. arXiv:2005.04790.

[9] H. Lin, Z. Luo, W. Gao, J. Ma, B. Wang, R. Yang, To-

wards explainable harmful meme detection through multimodal debate between large language models, 2024. URL: https://arxiv.org/abs/2401.13298. arXiv:2401.13298.

[10] T. Bloem, F. Ilievski, Clustering internet memes through template matching and multi-dimensional similarity, 2025. URL: https://arxiv.org/abs/2505.00056. arXiv:2505.00056.

[11] F. Barbieri, L. Espinosa Anke, J. Camacho-Collados, XLM-T: Multilingual language models in Twitter for sentiment analysis and beyond, in: Proceedings of the Thirteenth Language Resources and Evaluation Conference, European Language Resources Association, Marseille, France, 2022, pp. 258–266. URL: https://aclanthology.org/2022.lrec-1.27.

[12] N. Reimers, I. Gurevych, Sentence-BERT: Sentence embeddings using Siamese BERT-networks, in: K. Inui, J. Jiang, V. Ng, X. Wan (Eds.), Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Association for Computational Linguistics, Hong Kong, China, 2019, pp. 3982–3992. URL: https://aclanthology.org/D19-1410/. doi:10.18653/v1/D19-1410.

[13] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever, et al., Language models are unsupervised multitask learners, OpenAI blog (2019). URL: https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf.

[14] S. Rouard, F. Massa, A. Défossez, Hybrid transformers for music source separation, in: ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2023, pp. 1–5. doi:10.1109/ICASSP49357.2023.10096956.

[15] B. McFee, M. McVicar, D. Faronbi, I. Roman, M. Gover, et al., Librosa, 2025. URL: https://doi.org/10.5281/zenodo.15006942. doi:10.5281/zenodo.15006942.

[16] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing, Journal of the Royal Statistical Society: Series B (Methodological) 57 (2018) 289–300. URL: https://doi.org/10.1111/j.2517-6161.1995.tb02031.x. doi:10.1111/j.2517-6161.1995.tb02031.x.

[17] S.-w. Yang, P.-H. Chi, Y.-S. Chuang, C.-I. J. Lai, K. Lakhotia, Y. Y. Lin, A. T. Liu, J. Shi, X. Chang, G.-T. Lin, et al., SUPERB: Speech processing Universal PERformance Benchmark, 2021. arXiv:2105.01051.

[18] A. Dinu, A.-M. Florescu, Testing language creativity of large language models and humans, in: M. Hämäläinen, E. Öhman, Y. Bizzoni, S. Miyagawa, K. Alnajjar (Eds.), Proceedings of the 5th International Conference on Natural Language Processing for Digital Humanities, Association for Computational Linguistics, Albuquerque, USA, 2025, pp. 426–436. URL: https://aclanthology.org/2025.nlp4dh-1.37/. doi:10.18653/v1/2025.nlp4dh-1.37.

Appendix - Supplementary Statistical Tests Tables

Table 1
Statistical tests for Sentiment Analysis of the texts

pair	MANOVA	Mann-Whitney for neg.	Mann-Whitney for neu.	Mann-Whitney for pos.
Ro pop. - Ro unpop.	λ = 0.99, F(3, 116) = 0.55, p = 0.65	p = 0.79 (M: 0.38 vs. 0.36)	p = 0.57 (M: 0.39 vs. 0.40)	p = 0.990 (M: 0.24 vs. 0.23)
It pop. - It unpop.	λ = 0.97, F(3, 116) = 1.15, p = 0.33	p = 0.36 (M: 0.62 vs. 0.67)	p = 0.54 (M: 0.19 vs. 0.17)	p = 0.39 (M: 0.19 vs. 0.16)
It - Ro	λ = 0.40, F(3, 116) = 56.74, p < 0.001	p < 0.0001 (M: 0.38 vs. 0.16)	p < 0.0001 (M: 0.39 vs. 0.17)	p < 0.0001 (M: 0.67 vs. 0.24)

Table 2
Statistical tests for Semantic Similarity and Perplexity of the texts

Semantic Similarity			Perplexity		
Pair	**ANOVA**	**Mann-Whitney U Test**	**Pair**	**ANOVA**	**Mann-Whitney U Test**
Ro pop. - Ro unpop.	F(1, 118) = 2.35, p = 0.128	p = 0.068 (M = 0.58 vs. 0.57)	Ro pop. - Ro unpop.	F(1, 118) = 0.17, p = 0.684	p = 0.927 (M = 168.59 vs. 159.22)
It pop. - It unpop.	F(1, 118) = 2.38, p = 0.126	p = 0.182 (M = 0.58 vs. 0.59)	It pop. - It unpop.	F(1, 118) = 0.78, p = 0.380	p = 0.821 (M = 228.92 vs. 208.20)
It - Ro	F(1, 238) = 1.88, p = 0.172	p = 0.0614 (M = 0.59 vs. 0.58)	It - Ro	F(1, 238) = 11.19, p < 0.001	p < .0001 (M = 218.74 vs. 163.91)

Table 3
Statistical Tests for Rhyme Coefficient

Pair	Descriptive Statistics	Mann-Whitney
Ro pop. - Ro unpop.	M = 0.195 vs. 0.161	p = 0.135
It pop. - It unpop.	M = 0.215 vs. 0.228	p = 0.688
It - Ro	M = 0.222 vs. 0.178	p = 0.0001

Table 4
Comparison of key audio features between Italian and Romanian brain rot, after FDR correction that differ at q<0.05

Feature	Italian	Romanian	q-value
var_f_0 (Hz^2)	4570.8	2234.0	9.90×10^{-6}
range_f_0 (Hz)	162.0	119.1	3.50×10^{-6}
entropy_f_0_slope (bits)	1.79	2.05	1.96×10^{-4}
rms_mean	0.141	0.123	1.39×10^{-2}
spec_centroid_mean (Hz)	1961	2161	1.39×10^{-8}
spec_centroid_var (Hz^2)	1.24e6	1.67e6	5.99×10^{-18}
spec_bandwidth_mean (Hz)	1936	2059	3.05×10^{-8}
spec_bandwidth_var (Hz^2)	2.45e5	2.84e5	4.71×10^{-6}
spec_rolloff_mean (Hz)	3625	4000	1.78×10^{-7}
spec_rolloff_var (Hz^2)	3.50e6	4.45e6	4.42×10^{-15}
zcr_mean	0.0975	0.1098	6.58×10^{-7}
zcr_var	0.00868	0.01162	2.46×10^{-12}
flux_var	8.84	9.73	6.64×10^{-4}
mfcc_1_mean	-197.4	-220.7	4.75×10^{-3}
mfcc_1_var	1.79e4	2.30e4	7.29×10^{-10}
...	...	...	...

Feature	Italian	Romanian	q-value
mfcc_2_mean	93.7	86.4	1.60×10^{-4}
mfcc_2_var	3.73e3	4.24e3	1.23×10^{-4}
mfcc_4_mean	25.3	20.7	4.42×10^{-5}
mfcc_4_var	892	790	3.12×10^{-5}
mfcc_5_var	627	499	5.61×10^{-11}
mfcc_6_mean	-4.48	-2.52	1.17×10^{-2}
mfcc_7_mean	-11.42	-8.57	4.75×10^{-6}
mfcc_7_var	318	271	1.38×10^{-10}
mfcc_9_mean	-13.44	-11.34	6.97×10^{-7}
mfcc_9_var	218	210	2.70×10^{-2}
mfcc_10_mean	-9.10	-7.59	5.97×10^{-3}
mfcc_11_mean	-5.82	-4.92	8.36×10^{-4}
mfcc_11_var	203	167	4.71×10^{-10}
mfcc_12_mean	-3.49	-2.40	8.03×10^{-3}
mfcc_12_var	112	98	3.91×10^{-6}
mfcc_13_var	105	101	2.24×10^{-3}

Table 5
Raw p < 0.05 comparisons of audio features between all popular and unpopular brain rots within Italian and Romanian corpora

Language	Feature	Popular	Unpopular	p-value
Italian	var_f0 (Hz^2)	6657	2484	0.0160
	range_f0 (Hz)	189	135	0.0028
Romanian	syllable_rate (onsets/s)	4.29	4.09	0.0260
	mfcc_9_var	202	217	0.0062
	mfcc_11_mean	−5.65	−4.18	0.0300

Table 6
Image attributes that might have been a good indicator of video popularity

Attribute	Pearson Correlation	Mann-Whitney P-value	Significant (<0.05)
Guessed Popularity (Visual)	0.0138	0.6001	No
Cutting Speed	0.1712	0.0162	**Yes**
Overall Pacing	0.1226	0.0492	**Yes**
Subject Movement	0.1053	0.0858	No
Narrative Logic	0.0711	0.2485	No
Non-Sequitur Visuals	-0.0703	0.2785	No
Overall Absurdity Level	0.0584	0.2937	No

Declaration on Generative AI

During the preparation of this work, the author(s) did not use any generative AI tools or services.