

CROSS-MODAL SCRIPT OPPOSITION IN THE CONSTRUCTION OF MULTIMODAL HUMOR IN PAKISTANI EDITORIAL CARTOONS

Nasir Ali
nasir@ntu.edu.pk

Lecturer, National Textile University, Faisalabad, Pakistan / PhD
Scholar at Government College University, Faisalabad, Pakistan.

Muhammad Asim Mahmood
masimmahmood@gcuf.edu.pk

Professor, Government College University, Faisalabad, Pakistan.

Abstract

Script Opposition is the fundamental mechanism of humor; however, its formulation has been studied mainly in verbal texts, and its application to multimodal humor remains underexplored. This study addresses this gap by investigating the realization of Script Opposition in Pakistani editorial cartoons. To examine its multimodal realization, a corpus of 500 editorial cartoons was selected through stratified random sampling from five Pakistani newspapers—Dawn, The Express Tribune, The Nation, Daily Times, and Daily Jang—published over a one-year period from May 2023 to April 2024. The data were first classified using Samson and Huber's (2007) multimodality typology, which was then integrated with Attardo's General Theory of Verbal Humor (GTVH) framework for the analysis of multimodal Script Opposition. The results indicated that 92.2% of the cartoons in the corpus were multimodal, and 89.8% of this exhibited interactive multimodality, requiring both verbal and visual channels to produce humor. Qualitative analysis further illustrated that visual and verbal scripts oppose and interact to generate humorous meaning. This cross-modal organization challenges the largely monomodal assumptions underlying verbal humor theory. The study confirms that cartoons, as a humorous newspaper text structure, are predominantly realized cross-modally through Script Opposition, and extends the Script Opposition Knowledge Resource of the GTVH to incorporate the cross-modal distribution of scripts, thereby advancing the theoretical framework of multimodal satirical discourse.

Keywords: *Editorial Cartoons, Multimodal Humor, Multimodality, Pakistani Print Media, Script Opposition.*

Corresponding Author: Muhammad Asim Mahmood (Professor, Government College University, Faisalabad, Pakistan)
Email: masimmahmood@gcuf.edu.pk

1. Introduction

Editorial cartoons occupy a particular position in humor studies. They are read as jokes are read, texts which give nonliteral, satirical meanings, yet they do not give their humorous meanings through single mode. A single panel cartoon usually consists of a caricatured image, a caption or speech bubble, and often a headline, and is intended to be a satirical point in the blink time of any newspaper page. This is a communicative situation which the main linguistic theory of humor, which was constructed for humor analysis, was not meant to describe. The General Theory of Verbal Humor (GTVH) was developed by Attardo and Raskin (1991) and elaborated by Attardo (1994, 2001, 2020) in the case of verbal jokes, which are texts in which both scripts are recoverable from language. Attardo (2020) has admitted that the framework's applicability is not limited to language or verbal mode alone and has suggested that the terminal Knowledge Resource, Language, could be interpreted as a more comprehensive Semiotic Strategy. However, the key conceptual element of the theory, "Script Opposition," has yet to be rigorously explored for genres where the two opposing scripts are not both linguistically given (where one is verbal, the other is visual).

Tsakona (2009) was the first to explore that Script Opposition can manifest in the verbal and visual components of a cartoon – rather than in each of them alone. Her analysis of Greek newspaper cartoons showed that the text can activate one script whereas the image activates the other one which are opposing and thus produce humorous meanings. It's an observation that has theoretical importance, but it was made in a rather small sample and presented mainly as one configuration of several instead of a phenomenon that needs a separate, corpus-based investigation. The question of whether or not cross-modal Script Opposition activation is on the periphery of Script Opposition activation in the form of editorial cartooning, or its organizing principle is yet to be determined at the scale necessary to answer it.

This paper fills this gap directly. First, the modal composition of the corpus is determined using the Samson and Huber's (2007) typology of editorial cartoons, which classifies cartoons into three types: in a cartoon where the text is the only source of the humor (illustrative); in a cartoon where the image plus the verbal part of the cartoon are equally necessary for the humorous effect (interactive); in a cartoon where the image by itself is sufficient to carry the joke (descriptive). It then uses GTVH's Script Opposition Knowledge Resource to describe the distribution of the two opposing scripts across these modes, focusing especially on the interactive mode that involves cross-modal activation in the structure. The paper advances a focused and testable argument that, in a genre-representative corpus of Pakistani editorial cartoons, cross-modal Script Opposition is the dominant rather than an occasional pattern of humor construction. This finding suggests that the GTVH should formally recognize cross-modal activation as a distinct

configuration of Script Opposition, rather than treating it as a straightforward extension of the single-channel model proposed by Raskin (1985) and Attardo (1994, 2001).

The argument advanced in this paper is deliberately narrow in scope, reflecting a conscious methodological choice. It focuses exclusively on Script Opposition because it is the foundational Knowledge Resource of the General Theory of Verbal Humor (GTVH) and the primary mechanism underlying humor construction. This focused approach enables a detailed examination of how Script Opposition is distributed across verbal and visual modes in Pakistani editorial cartoons and what this reveals about its role in multimodal discourse. Rather than proposing a comprehensive revision of the GTVH, this study advances a specific theoretical claim: when opposing scripts are consistently realized across independent semiotic modes, Script Opposition requires a formal extension to account for multimodal humor. This focused analysis proposes an evidence-based refinement of the Knowledge Resource that preserves the core architecture of the GTVH while accommodating the distinctive communicative organization of multimodal editorial cartoons.

2. Literature Review

2.1. The SSTH, the GTVH, and Script Opposition

Raskin's (1985) Semantic Script Theory of Humor (SSTH) proposes a concise and falsifiable hypothesis: a text is humorous when it is compatible, fully or partially, with two scripts that stand in opposition to one another. A script, in Raskin's (1985, p. 81) formulation, is "a large chunk of semantic information" organized around a word, object, or event—a structure with clear intellectual debts to Bartlett's (1932) schema theory and to Schank and Abelson's (1977/2013) scripts in artificial intelligence, and one that activates not only through explicit lexical mention but through inference, presupposition, and spreading activation (Attardo, 2020). This inferential property of script activation matters considerably for multimodal analysis: a script need not be named to be present, and an image can activate a script that the accompanying caption never states.

For Raskin (1985), the opposing scripts are one of these three categories: actual/non-actual, normal/abnormal and possible/impossible. Attardo's (1997) refinement of local antonymy specifies that opposed scripts need only differ in one structural node or link while the remainder of their content overlaps—a minimal condition that, notably, does not require the opposition to be lexically fixed in advance. This detail opens theoretical space, largely unexploited in the SSTH's original verbal-joke corpus, for an opposition established through non-linguistic means: caricature exaggeration, incongruous spatial juxtaposition, or symbolic iconography that places a figure in a role inconsistent with the reader's expectations.

The General Theory of Verbal Humor is an extension of the SSTH that proposes the use of a framework of six Knowledge Resources (KRs): Script Opposition (SO),

Logical Mechanism (LM), Situation (SI), Target (TA), Narrative Strategy (NS), and Language (LA), hierarchically ordered by their contribution to a text's identity and with Script Opposition as the condition for the presence of humor (Attardo, 1994, 2001, 2020; Attardo & Raskin, 1991). Script Opposition is thus a foundational construct in the GTVH's own architecture which assumes that the two opposed scripts which the text calls up are identified in a Knowledge-Resource-based analysis of a humorous text. This is exactly the construction that is questioned for its fittingness in a multimodal setting as is done in the present paper.

Attardo's (2020) retrospective reassessment of the GTVH's sixth Knowledge Resource is instructive for the argument developed here. He proposes that, as it was originally called, the theory of Language is better known as Semiotic Strategy, given that the theory was first introduced in 1991 as a linguistic theory, and that this has led to a misconception that the GTVH cannot provide analysis of non-linguistically realised humor, something which he repeatedly has attempted to address with cartoons and other multimodal or non-verbal texts. This rethinking allows for the research of visual codification as a valid object for GTVH analyses. However, Attardo's (2020) proposal only tackles the surface-realization Knowledge Resource of the terminal, and leaves the question of foundational construct of Script Opposition (as a structure of the framework) unaddressed when both scripts are not available in the same semiotic channel. It is this question for which this paper gives an answer.

2.2. Multimodal Humor and Image–Text Relations

Multimodal discourse analysis offers a vocabulary that is not based on humor theory to characterize the combination of image and text in the production of meaning that is jointly created. Barthes (1977) points out that anchorage and relay play different roles: anchorage multiplies the range of possible interpretations that might be placed on the image, but relay adds to it a new content of its own that can only be understood in combination with the image. Nikolajeva and Scott (2000) continue this division along a spectrum from symmetrical redundancy, through enhancement and counterpoint, to outright contradiction, where as they state, word and image seem to be opposing each other, and the reader is tasked with making a judgment between the two. The systematic account of multimodal document structure that Bateman (2008) develops in this book is a corpus-based one, and is linguistically grounded, as is any account of how a cartoon's visual and verbal channels are structurally related, but it formalizes this work more generally than Bateman's account of humor does.

Forceville's (1996, 2005, 2008, 2016) program on pictorial and multimodal metaphor is the most extensively developed account of how visual and combined visual–verbal resources construct non-literal meaning outside language, and Bounegru and Forceville's (2011) analysis of editorial cartoons on the global financial crisis demonstrates

directly that this apparatus is productive for the political cartoon genre specifically. Though multimodal metaphor and multimodal Script Opposition are related and both theoretical objects, the latter is the foundational construct of the GTVH, while the former explains how understanding of a target domain is structured across modes by means of a source domain. A cartoon can take advantage of both—the visual metaphor can be a Script Opposition itself, while the mapping between the two literatures has not, to date, been made explicit in the case of editorial cartoons, as Section 5 reiterates below in the interpretation of the present findings.

The social-semiotic theory of visual grammar by Kress and van Leeuwen (1996, 2006, 2020) has been the most widely used framework for the description of how images make sense, which they distinguish as representational, interactive and compositional metafunctions based on the systemic functional linguistics of Halliday (1978, 1985). It is important, however, to state clearly what this framework is not meant to account for if it is applied to a political cartoon: how the image is composed, what it denotes, and how the viewer is oriented in relation to it do not explain what makes the cartoon funny, how two conflicting scripts can be co-activated, or how the incongruity is resolved as humor and not just as confusion. Visual grammar and Script Opposition analysis are complementary approaches, the former deconstructing the semiotic structure and the latter cognitive mechanism. Considering the visual channel as an independent, structurally organized meaning-making channel that can stand alone, as a second analytical apparatus, rather than as a second analytical frame parallel to the GTVH, Kress and van Leeuwen's (1996, 2006, 2020) apparatus is therefore invoked here.

2.3. Tsakona's Multimodal Application of the GTVH

The paper's most relevant precedent is that of Tsakona (2009) who studied 565 Greek newspaper cartoons. She categorises cartoons according to the predominant mode of humor production (dominantly verbal, interactive, and dominantly visual) and demonstrates the analyticity of GTVH in all three modes. Three of her findings are pertinent to the argument presented here. First, she demonstrates that the Language Knowledge Resource needs to be expanded to account for the visual mechanisms that bring about the humor in a cartoon, which the Semiotic Strategy later rephrased would label Attardo (2020). Second, she observes that hyperdetermination (Attardo, 2001), in which two or more sources of humor are activated at once, is particularly prevalent in cartoons in which the visual metaphor is the organizing principle. Third, and most importantly for purposes of this paper, she shows in a close examination that in truly interactive cartoons, Script Opposition is often realised cross-modally—that is, the denial made by the journalist, "no disaster" is directly contradicted by the image, "disaster is occurring," and the activation of the former script by the caption is realised as Juxtaposition as Logical Mechanism.

Theoretically, the demonstration made by Tsakona (2009) is important because it points to the existence of cross-modal Script Opposition as a possibility within the GTVH's current framework, but without making any formal changes to the categories. What her study cannot and by design could not determine is the relative frequency of this configuration to other equally common configurations (script opposition, verbal or visual) in a representative corpus of cartoons, or whether this configuration is sufficiently common in the genre to be classed as the modal default for the genre rather than among the other, equally common, configurations. In the analysis of newspaper comic strips by Fallianda et al. (2018), the theoretical possibility is confirmed again in the sense that Script Opposition in cartoon humor can be found in text alone, between text and image or in image alone, though the frequency of these three locations is not quantified with respect to a systematically sampled corpus. The present study adds to this ongoing research in several ways: it applies the full architecture of the GTVH to a corpus of nearly as large as that used by Tsakona (2009), but from a non-European, linguistically and politically different context; it also, importantly for the present argument, provides the vast distributional evidence (500 cartoons, five newspapers, twelve months) that neither Tsakona (2009) nor Fallianda et al. (2018) offer, evidence that can help determine the peripheral or dominant nature of cross-modal Script Opposition activation within the genre.

2.4. Samson and Huber's (2007) Multimodal Typology

Samson and Huber (2007) identify three classification types for the combination of verbal and visual resources in cartoon humor: In the first type, the image itself does not contribute to the humor, but illustrates a joke that is completely contained in the verbal language; in the second type, the image provides information that is not expressed in the text so that both the verbal and the visual language contribute to the humorous effect; and in the third type, the image is the source of the joke, and the verbal language serves only as a description of the image. As in the multimodal humor literature, the three categories are called here illustrative multimodality, interactive multimodality and descriptive multimodality, respectively.

This typology takes one very well-defined and restricted task, to classify the modal composition. It's neither a theory of humor, nor does it claim to be a theory of how incongruity is built and resolved cognitively. It is not competitive but complementary, sequential in relation to the GTVH: Samson and Huber's (2007) classification is the pre-analytical layer that determines which cartoons are structural candidates for cross-modal Script Opposition, the GTVH is the analytical layer that determines whether and how the opposition is realized. Script Opposition activation is most plausibly realized in the verbal channel alone (in illustrative cartoons), in the visual channel alone (in descriptive cartoons), and in the cross-modal channel (as a cross-over of the two channels, which is

structurally expected, though not logically required, in interactive cartoons). This is a structural assumption that is tested by the present paper's quantitative and qualitative analysis.

2.5. The Research Gap

Three literatures converge on a single unresolved question. The GTVH developed by Attardo and Raskin (1991) and subsequently refined by Attardo (1994, 2001, 2020), theorizes Script Opposition as a construct realized within a text, without formal provision for scripts distributed across independent perceptual channels. Tsakona (2009) demonstrates, through close qualitative analysis of a modest sample, that cross-modal realization is a genuine possibility within this architecture. Samson and Huber's (2007) typology, adopted widely in subsequent multimodal humor scholarship including Fallianda et al. (2018), provides the descriptive apparatus for identifying which cartoons are candidates for this configuration, but has not itself been paired with a systematic, corpus-scale test of how often the candidacy is realized. To date, no study has integrated these three streams of research to determine if cross-modal Script Opposition activation occurs in multimodal editorial cartooning more often than it does less often, or if it is the rule or the exception; and, in turn, if the existing single-channel account of the GTVH's foundational Knowledge Resource is formally extendable or is theoretically adequate as formulated. This paper provides that proof.

3. Methodology

3.1. Corpus and its construction

The corpus comprises 500 editorial cartoons drawn from five Pakistani newspapers: Dawn, The Express Tribune, The Nation, and Daily Times, all published in English, and Daily Jang, published in Urdu. The newspapers were selected purposively based on three criteria: had a national circulation (in at least one of the countries studied), published the editorial cartoons during the study period, and had access to the editorial cartoons in the official digital archive. Both English and Urdu newspapers are included to make sure that the analysis is not limited to only one print tradition but has a wider linguistic scope.

The selected time frame for this sampling of cartoons was from May 2023 to April 2024, which covers the period of ongoing political turmoil in Pakistan, including the riots that followed the arrest of the former Prime Minister Imran Khan in May 2023, the prolonged caretaker government, the severe economic crisis due to inflation and IMF-backed economic reform, the general election of 8 February 2024, and the establishment of a new federal government. The adoption of a continuous full-year window was done to ensure that this timeframe captures a full cycle of editorial cartoon activity, from pre-election, through election and post election.

Systematic search from the five newspapers' official digital archives identified a sampling frame of 2,056 cartoons (Dawn: 357; The Express Tribune: 349; The Nation: 355; Daily Times: 333; Daily Jang: 662). A stratified random sample of 500 cartoons was drawn from this population, with each newspaper receiving a random sample of 100 cartoons, independent of each newspapers' proportions of the sampling frame. To be honest, equal distribution was chosen as it is a key goal of the larger research program that this paper represents, rather than proportional population estimation. Simple random sampling was used to select cartoons in each newspaper stratum, using randomly generated numbers, and the first 100 cartoons by the random order selected for each stratum were kept for analysis. This procedure minimizes selection bias that could otherwise arise from selecting cartoons on the basis of their perceived humorousness or political salience.

3.2. Coding Procedures

Two coding layers are relevant to the present paper. The first is Samson and Huber's (2007) three-category typology, applied to every multimodal cartoon in the corpus to classify it as illustrative, interactive, or descriptive. Cartoons constructed exclusively through a single semiotic mode—overwhelmingly visual in this corpus—fall outside this typology by definition and were coded separately as unimodal; they remain, however, part of the corpus analyzed for Script Opposition, since script activation is not, in principle, dependent on the number of modes through which a text is realized.

The second coding layer is the GTVH's Script Opposition Knowledge Resource, operationalized following Raskin (1985) and Attardo (2020) through a two-level framework. At the abstract level, each cartoon's Script Opposition was classified into one of Raskin's (1985) three categories—actual/non-actual, normal/abnormal, or possible/impossible—using coding criteria fixed in advance of full-corpus analysis: a normal/abnormal opposition was assigned when humor depended on the violation of an expected social, political, or institutional norm; an actual/non-actual opposition, when humor depended on a contrast between an existing state of affairs and an imagined or counterfactual alternative; and a possible/impossible opposition, when humor depended on a logically or physically impossible scenario. At the concrete level, coders recorded the specific scripts activated in each cartoon using descriptive labels reflecting its particular political or social content, following the same two-level procedure Tsakona (2009) adopts for her Greek cartoon corpus.

Critically for the present paper, this operationalization was devised from the beginning to process the two opposing scripts verbally, visually, and/or combined verbally-visually, as the overwhelming multimodality of the corpus meant that criteria based on a single modality were not sufficient for the material. In addition to abstract and concrete Script Opposition, coders identified the channel (verbal, visual, or both) along which each one of the two scripts was anchored for each cartoon. The cross-modal analysis

of this channel level identification in conjunction with the classification of each cartoon by Samson and Huber (2007) is the basis for the qualitative analysis.

3.3. Inter-Coder Reliability

Because coding both the Multimodal Category and Script Opposition requires interpretive judgment, a formal inter-coder reliability procedure was conducted prior to full-corpus analysis. A stratified random sample of 50 cartoons (10% of the corpus), with ten cartoons drawn from each newspaper, was independently coded by a second coder—a linguistics professor with postdoctoral humor-studies training, prior collaboration with Salvatore Attardo, and native Urdu proficiency alongside advanced professional English—working from the same coding manual and without access to the primary coder's decisions. Agreement was evaluated using Cohen's Kappa (κ), selected because, unlike raw percentage agreement, it adjusts for agreement expected by chance (Neuendorf, 2017).

Agreement for the two categories central to this paper was almost perfect by the Landis and Koch (1977) benchmark: $\kappa = 1.00$ for the Multimodal Category and $\kappa = .96$ for Script Opposition. These coefficients indicate that the coding criteria for both modal classification and Script Opposition identification could be applied with a very high degree of consistency by an independent coder working from the same manual, supporting the reliability of the quantitative results reported below.

3.4. Analytic Procedure

The analysis proceeds in two stages, consistent with the study's concurrent mixed-methods design. The quantitative stage reports descriptive frequencies and percentages for modal composition (unimodal versus multimodal; and, within multimodal cartoons, interactive, illustrative, and descriptive), together with chi-square tests of independence assessing whether the distribution of these categories varies significantly by newspaper. The qualitative stage undertakes close multimodal discourse analysis of a purposively selected set of interactive cartoons, drawn from among the corpus's existing coded examples, to demonstrate how Script 1 and Script 2 are respectively anchored across the verbal and visual channels, how the two scripts are processed as a simultaneous rather than sequential conflict, and how this pattern differs from single-channel Script Opposition in the illustrative and unimodal cartoons analyzed as a contrast set.

Selection of the qualitative examples followed a purposive rather than random logic, consistent with standard practice in qualitative multimodal discourse analysis (Creswell & Clark, 2017): cartoons were chosen because their coded Script Opposition channel-attribution made them clear and representative instances of one of the three configurations identified in Section 4.2—mixed, textual, or visual modality—rather than because they were ambiguous or contested cases. This selection criterion means the examples analyzed in Section 4.3 are illustrative of the corpus's dominant pattern rather than of its interpretive edge cases; edge cases in which channel-attribution required

extended coder deliberation are not the focus of the present paper, though their existence is acknowledged as a source of the residual interpretive variability inherent to multimodal discourse coding generally.

Consistent with the paper's scope, the analysis does not extend to Logical Mechanism, Situation, Target, Narrative Strategy, or Language except where necessary to illustrate a specific cross-modal example, nor does it report the cross-newspaper comparative statistics for Knowledge Resources unrelated to modality.

4. Results

4.1. Modal Composition of the Corpus

Table 1 reports the distribution of unimodal and multimodal cartoons across the five newspapers.

Table 1. *Distribution of Modality Across Pakistani Newspaper Cartoons*

Newspaper	Unimodal	Multimodal	Total
Dawn	4	96	100
The Express Tribune	11	89	100
The Nation	18	82	100
Daily Times	5	95	100
Daily Jang	1	99	100
Total	39	461	500

Note. $\chi^2(4) = 25.42$, $p < .001$.

Of the 500 cartoons in the corpus, 461 (92.2%) are multimodal, establishing multimodality as the corpus's defining modal characteristic rather than a stylistic option exercised by some cartoonists and not others. The proportion of multimodal cartoons is high across all five newspapers but not uniform: Daily Jang records the highest proportion (99/100), followed by Dawn (96/100) and Daily Times (95/100), while The Nation records the corpus's highest unimodal proportion (18/100), distinguishing it from the other four newspapers. A chi-square test of independence confirms that this cross-newspaper variation is statistically significant, $\chi^2(4) = 25.42$, $p < .001$, indicating that the distribution of unimodal and multimodal cartoons is not independent of newspaper affiliation.

Table 2 presents the distribution of multimodal cartoons across the five newspapers according to Samson and Huber's (2007) typology of illustrative, interactive, and descriptive multimodality.

Table 2. *Distribution of Types of Multimodality in Multimodal Cartoons*

Newspaper	Interactive	Illustrative	Descriptive	Total
Dawn	96	0	0	96
The Express Tribune	86	3	0	89
The Nation	46	34	2	82
Daily Times	95	0	0	95
Daily Jang	91	7	1	99

Total	414	44	3	461
--------------	-----	----	---	-----

Note. $\chi^2(8) = 129.09$, $p < .001$.

Because Samson and Huber's (2007) typology applies only to cartoons in which verbal and visual modes co-occur, Table 2 restricts the classification to the 461 multimodal cartoons. Interactive multimodality—the configuration in which both channels are jointly necessary to the humorous effect—is overwhelmingly dominant, accounting for 414 of the 461 multimodal cartoons. This pattern is categorical in Dawn and Daily Times, where every multimodal cartoon is interactive, with no illustrative or descriptive instances at all, and is nearly so in The Express Tribune and Daily Jang. The Nation is the clear exception: interactive multimodality remains its most frequent type (46 of 82), but illustrative multimodality accounts for a substantial 34 cartoons, and The Nation contributes the corpus's only two descriptive-multimodal cartoons besides one in Daily Jang. A chi-square test confirms that this cross-newspaper variation in multimodality type is statistically significant, $\chi^2(8) = 129.09$, $p < .001$.

Table 3 summarizes the overall distribution of multimodality types across the corpus.

Table 3. *Overall Percentage Distribution of Multimodality Types (N = 461 Multimodal Cartoons)*

Multimodality Type	Frequency	Percentage
Interactive	414	89.80%
Illustrative	44	9.54%
Descriptive	3	0.65%
Total	461	100%

Table 3 collapses these newspaper-level distributions into a corpus-wide summary. Interactive multimodality accounts for 89.8% of the 461 multimodal cartoons (414 instances); illustrative multimodality, 9.54% (44 instances); and descriptive multimodality, 0.65% (3 instances). This 89.8% figure is the empirical foundation for the analysis that follows: it establishes that in nearly nine of every ten multimodal cartoons in this corpus, humor construction structurally requires both the verbal and the visual channel, since removing either mode—by Samson and Huber's (2007) own definitional criterion for the interactive category—disrupts the cartoon's communicative and humorous effect.

4.2. Script Opposition Realization Across Modal Configurations

Because interactive multimodality is defined by the joint necessity of both channels, it creates the structural precondition for cross-modal Script Opposition: if neither mode alone produces the humorous effect, and if Script Opposition is the mechanism responsible for that effect, then the two opposing scripts are, in the great majority of these cartoons, distributed across the two channels rather than concentrated within either one.

This is not a logical necessity—an interactive cartoon could, in principle, realize both scripts verbally while relying on the image only for a Logical Mechanism such as Exaggeration—but the coding of Script Opposition channel-attribution across the corpus (Section 3.2) confirms that, in the dominant pattern among interactive cartoons, one script is anchored in the caption, headline, or speech bubble and the other is anchored in the image, corresponding directly to the 89.8% interactive-multimodality figure reported above. Three configurations recur across the corpus.

In mixed-modality Script Opposition, one script is carried by the verbal channel and the other by the visual channel, and the viewer perceives the resulting contradiction as an immediate cross-channel collision rather than as a tension internal to either mode taken alone. In textual-modality Script Opposition, both scripts are generated entirely within the verbal channel—through lexical ambiguity, contradictory statements, or a reversal of expectation stated in language—with the image, where present, functioning as contextual illustration that does not itself contribute to the incongruity. In visual-modality Script Opposition, both scripts are encoded through the image alone, without verbal anchoring, drawing on caricature, exaggerated scale, symbolic iconography, or incongruous spatial juxtaposition; such cases occur both among the corpus's 39 unimodal cartoons and among the small number of descriptive multimodal cartoons in which text, if present, is purely labeling.

Because no single frequency table in the source coding scheme separately cross-tabulates Script Opposition channel-attribution against newspaper as an independent variable, the analysis below does not report a supplementary chi-square robustness check disaggregating this specific cross-tabulation by newspaper; the corpus-level correspondence between the 89.8% interactive-multimodality figure and the dominant mixed-modality Script Opposition pattern is instead established through the qualitative channel-coding described in Section 3.2 and demonstrated through the close analysis that follows.

4.3. Qualitative Analysis: Multimodal Configuration of Script Oppositions

This section presents close multimodal discourse analysis of seven purposively selected cartoons: three illustrating mixed-modality Script Opposition, drawn from the corpus's interactive cartoons; two illustrating textual-modality Script Opposition, drawn from illustrative and unimodal cartoons functioning as a contrast set; and two illustrating visual-modality Script Opposition, drawn from unimodal cartoons completing that contrast set. Each analysis identifies the abstract and concrete Script Opposition, the channel anchoring each script, and, where relevant, the resulting difference in how the incongruity is cognitively processed.

4.3.1. Mixed-modality Script Opposition



Figure 1. *Cartoon from Dawn depicting Finance Minister with begging bowl claiming 'no prospect of default' (26 May 2023)*

A cartoon published in Dawn on 26 May 2023 depicts then-Finance Minister Ishaq Dar in formal dress, holding a begging bowl, above a speech bubble reading "no prospect of default." Script 1 is activated verbally: the minister's statement invokes the governmental frame of official economic confidence. Script 2 is activated visually: the begging bowl, a conventional symbol of financial desperation, directly contradicts that claim. The abstract Script Opposition is Normal/Abnormal; the concrete opposition is Truth/Deception. Crucially, neither channel alone produces the cartoon's satirical effect: the caption alone would read as an unremarkable official assurance, and the image alone would depict an ambiguous, uncontextualized figure. The incongruity is available only once caption and image are perceived together, which—because a single-panel cartoon presents both channels simultaneously rather than sequentially—occurs as an immediate cross-channel collision rather than as the setup-then-punchline structure characteristic of the verbal jokes on which the GTVH's original model of Script Opposition was built.



Figure 2 Cartoon from *The Express Tribune* showing three-panel social media advertisement satire (21 November 2023)

A second example, published in *The Express Tribune* on 21 November 2023, presents a three-panel cartoon titled "Social Media Advertisements," satirizing the declining credibility of digital marketing endorsements. Script 1 is activated by the panels' conventional textual framing, which invokes the ordinary expectation that advertisement endorsers embody the benefits their products promise—healthy hair, successful smoking cessation, youthful skin. Script 2 is activated visually: a bald man promotes shampoo, an active smoker advertises smoking cessation, and an elderly woman endorses an anti-aging product, each image directly contradicting the corresponding textual claim. The abstract Script Opposition is Actual/Non-actual, realized concretely as Truth/Deception. As in the first example, the humor depends on the collision of a script established through conventional textual framing with a script established through contradictory visual evidence—neither is redundant, and neither could generate the incongruity unassisted by the other.



Figure 3 Cartoon depicting PTI leader at electoral starting line with legal weight impediment (Dawn, 13 September 2023)

A related pattern is evident in a cartoon published in Dawn on 13 September 2023, in which Imran Khan is depicted as a sprinter at a starting line, his ankle shackled to an iron weight labeled "CASES," beneath the caption "NO BAR ON PTI FROM CONTESTING POLLS, SAYS PM." Script 1—unimpeded electoral participation—is carried by the caption's official claim; Script 2—physical restraint and immobility—is carried entirely by the visual depiction of the shackle and weight. Here again, the caption in isolation would read as an unremarkable political statement, and the image in isolation would depict an athlete facing an unspecified obstacle; the specific incongruity between official discourse and depicted reality is constituted only by their joint, simultaneous presentation.

4.3.2. Textual-modality Script Opposition



Figure 4 Cartoon from *The Nation* depicting job interview rejecting 'workaholic' applicant for government position (6 May 2023)

A cartoon in *The Nation* depicts a job interview in which a self-described "workaholic" candidate is told, "I really don't think you are right for a government job." Script 1—diligence as a desirable professional attribute—is activated by the candidate's self-description; Script 2—diligence as institutionally disqualifying—is activated by the interviewer's reply. The abstract opposition is Normal/Abnormal, concretely Diligence/Indolence, and both scripts are entirely verbal: the accompanying image of a seated interviewer and candidate provides contextual illustration but contributes nothing to the incongruity, which would survive unchanged if the image were removed altogether.



Figure 5 Cartoon from *The Nation* depicting student listing TikTok, Instagram, and Snapchat as summer 'sights' visited (28 August 2023)

A second example from the same newspaper depicts a teacher asking a student what sights they visited over the summer, to which the student replies, "TikTok, Instagram and Snapchat." Script 1—vacation sightseeing as physical travel—is activated by the teacher's question; Script 2—"sights" reinterpreted as digital platforms—is activated by the student's answer, the incongruity turning entirely on the lexical ambiguity of "sights." Both scripts, again, are constructed in language alone. These two cartoons illustrate a categorically different cognitive operation from the mixed-modality examples above: the reader processes a single channel and resolves the incongruity through sequential linguistic inference, in a manner directly comparable to a verbal joke of the kind the GTVH was originally built to describe.

4.3.3. Visual-modality Script Opposition

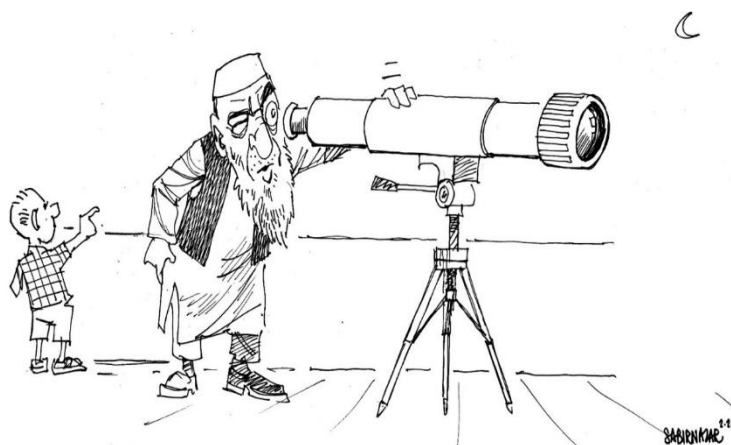


Figure 6 *Cartoon from The Express Tribune depicting religious scholar struggling with telescope while child identifies visible crescent moon (10 April 2024)*

A unimodal cartoon published in The Express Tribune on 10 April 2024 depicts a religious scholar squinting through an oversized telescope beside a child who casually points to a crescent moon plainly visible overhead. Script 1—that trained religious authorities require elaborate instrumentation to determine moon-sighting, referencing Pakistan's official Ruet-e-Hilal Committee procedure—and Script 2—that the moon is, in fact, directly and effortlessly visible—are both activated visually, through the contrast between the scholar's laborious posture and the child's effortless gesture; no verbal element is present.

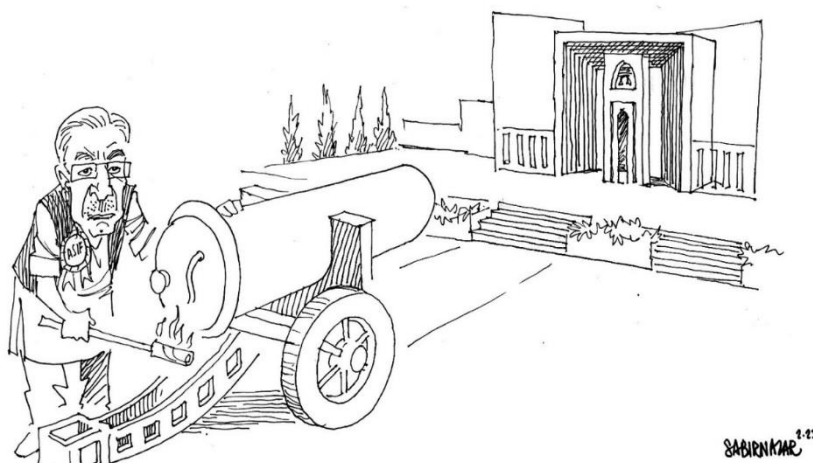


Figure 7 *Cartoon from The Express Tribune depicting PML-N politician firing cannon at Supreme Court (25 June 2023)*

A second unimodal cartoon from the same newspaper and published on 25 June 2023 depicts a senior political figure lighting the fuse of an antiquated cannon aimed at

the Supreme Court building. Script 1—the normative expectation that political actors respect judicial independence—is inferred entirely from the reader's background political knowledge, without any textual cue; Script 2—the depiction of literal armed aggression against a judicial institution—is constructed entirely through the image. The abstract opposition is Normal/Abnormal, concretely Lawful Adherence/Defiant Protest. These two cartoons confirm that Script Opposition, as a cognitive mechanism, does not require linguistic realization at all: both scripts, in each case, are constructed and resolved within a single visual field, with the reader's background socio-political knowledge substituting for an explicit verbal statement of Script 1.

These cartoons make a structural point that quantitative distribution alone cannot: the three Script Opposition configurations identified in Section 4.2 are not stylistic variants of a single underlying process but distinct cognitive operations. In textual-modality Script Opposition, the reader processes one channel and resolves a sequential linguistic incongruity, exactly as the GTVH's founding verbal-joke model describes. In visual-modality Script Opposition, the reader processes a single visual field in which both scripts are simultaneously present, one of them frequently supplied by inference rather than explicit representation. In mixed-modality Script Opposition—the configuration realized in 89.8% of this corpus's multimodal cartoons—the reader processes two distinct perceptual channels at once, each anchoring one of the two opposing scripts, such that the incongruity registers as an immediate cross-channel conflict rather than as the outcome of processing within either channel alone. It is this third configuration, structurally dominant across the corpus, that the GTVH's existing formulation of Script Opposition does not explicitly theorize.

5. Discussion

5.1. Cross-Modal Script Opposition as the Genre's Dominant Mechanism

Given the fact that 89.8% of the multimodal cartoons in this corpus are interactive, and that across the multimodal configuration Script Opposition is activated at the level of the verbal or pictorial channel, rather than the other, then the following theoretical claim can be made: Cross-modal Script Opposition activation is the default cognitive process by which editorial cartoonists create humor in the genre, not just a peripheral possibility. This claim is not quantitative but qualitative, and is of a different nature than the claim that Tsakona (2009) was able to prove. The cross-modal Script Opposition configuration was shown to be theoretically feasible in the GTVH architecture and was found to be empirically present in individual instances of cartoons in Tsakona's (2009) analysis, but was not, nor could it be, systematically compared with other single-channel configurations in a systematically sampled set of cartoons. The importance of the present study, the point at which five hundred cartoons, five newspapers, twelve months, with a near-perfect inter-coder reliability for both the Multimodal Category ($\kappa = 1.00$) and the Script Opposition (κ

= .96), converge is precisely this: that, within this genre, cross-modal Script Opposition is not the exception that needs special explanation but the rule against which single-channel Script Opposition is the marked, comparatively infrequent case.

This distributional dominance also impacts on the Script Opposition—multimodal metaphor relationship seen in Section 2.2. The program on pictorial and multimodal metaphor developed by Forceville (1996, 2005, 2008, 2016) and its use for the specific purpose of creating political commentary by Bounegru and Forceville (2011) show that the production of political commentary is typically achieved through the use of cross-modal metaphorical mappings. The examples in Section 4.3, however—denial of default with begging bowl or unimpeded electoral participation with shackled sprinter—are not just multimodal in the sense of the examples above, but also instances of Script Opposition in the specific sense outlined above: two mutually incompatible interpretive frameworks are co-activated by a single text. The two literatures refer to similar, but non-identical phenomena, and the present results indicate that a full account of editorial cartoon humor must make use of both: multimodal metaphor, to account for the way the source domain of the visual channel facilitates the reader's understanding of the figure domain; and cross-modal Script Opposition, to account for the way the resulting juxtaposition is experienced as a mismatch that must be resolved through humorous rather than figurative means.

5.2. Why Cross-Modal Activation Is a Qualitatively Different Cognitive Operation

The GTVH's original account of Script Opposition was derived from verbal joke data, which assumes sequential processing, in which a setup provides one script and a punchline restates the information in a way that suggests it must be reinterpreted within the context of the opposing script, the classic incongruity-resolution structure formalized in the SSTH. As the examples of textual-modalities in Section 4.3 illustrate, this sequential structure of text is implemented in the workaholic job interview and the TikTok “sights,” though they are nested in a cartoon and not in the form of a standalone text.

Unlike Mixed Modality Script Opposition, it does not have this sequential structure. A single-panel cartoon presents both the verbal and visual elements in one bounded visual field, so that both are available, in a strict perceptual sense, simultaneously and not sequentially, as in the construction of a verbal joke, which has a “setup” that must precede a “punchline. In these instances, the incongruity is sensed as a sort of cross-channel conflict; the reader doesn't read the caption, make an interpretation and then alter it once the image arrives, it is called as one cognitive event. This fits into and extends Tsakona (2009) who concludes that mechanisms suitable for perceptual processing are particularly appropriate to the simultaneous visual field presented in the cartoon, mechanisms such as Exaggeration and Juxtaposition fall in this category. The present analysis suggests that the same simultaneity that favors the use of these particular Logical

Mechanisms is also what makes Cross-modal Script Opposition the unique cognitive operation it is, instead of a multimodal restatement of the sequential, within-text mode.

5.3. A Proposed Formal Extension of the GTVH

The above evidence indicates a specific, viable modification of the GTVH's foundational Knowledge Resource. As for the multimodal nature of Script Opposition realization, a single-channel model, such as that of Raskin (1985) and Attardo (1994, 2001) developed for verbal jokes, should be formally distinguished in the three cases of Script Opposition realization within the same semiotic channel (the textual case in Section 4.3 and the visual case in Section 4.4) and across different channels (cross-channel Script Opposition, in which the two opposing scripts are realized in structurally different channels of the semiotic system, and are processed as simultaneous, non-sequential conflict).

This is not just a descriptive difference but rather has a testable implication for the coding and reporting of the Script Opposition Knowledge Resource in future multimodal analysis based on GTVH. In cases where Script Opposition is achieved cross-channel, an adequate analysis should also include the type of channel that anchors each of the two scripts as identified by Raskin (1985): an analysis of the opposition cross-channel would, in this instance, depend on the type of channel that anchors each of the two scripts, a decision which was necessary in order to make this present quantitative claim possible at all. If no such channel-attribution step is taken, a corpus-level analysis of Script Opposition in a multimodal genre is unable to distinguish a multimodal genre on which cross-channel activation is dominant (as in the present corpus, Pakistani editorial cartooning) from a multimodal genre where it is marginal (as in, say, the production of a news broadcast), and the resulting genre characterization would therefore be correspondingly underspecified. The extension outlined here is intentionally limited: it investigates how the interpretation of Script Opposition is realised in multimodal texts, without affecting the other five Knowledge Resources of the GTVH, the relationship of whose interpretation with the modal interpretation of multimodal texts lies outside the scope of the present paper and is treated in the broader research program from which this corpus is drawn: the interpretation of Language as Semiotic Strategy, Narration Strategy categorisation, Target identification, Situation identification and Logical Mechanism interpretation.

The proposed extension additionally specifies the theoretical contribution of the typology of Samson and Huber (2007) for the analysis, which is based on a GTVH. That typology, in its own terms, characterizes the typology of composition without claiming any particular type of humor mechanism. The present findings confirm that, in the present corpus, the interactive category is a good albeit not infallible structural predictor of the activation of cross-channel Script Opposition: the interactive category is the structural category that corresponds to the necessity of both modes to the humorous effect, and the

mechanism of cross-channel Script Opposition is what is responsible for that effect, as it appears in the overwhelming majority of cases found in the present study. This convergence provides just the operational link between descriptive modal classification and cognitive-semantic humor analysis that the two approaches, working separately, cannot provide alone: descriptive modal composition specifies what configuration of Knowledge Resource one should expect in a GTVH analysis and the analysis of the GTVH determines whether this expectation is met in a given text.

5.4. Situating the Finding Against Existing Multimodal Humor Scholarship

The 92.2% multimodality figure and the 89.8% interactive-multimodality figure reported in Section 4.1 are also directly relevant to a broader question in the multimodal humor literature: whether editorial cartooning should be understood as a genre in which visual and verbal resources are alternative, substitutable means of constructing humor, or as a genre structurally organized around their joint necessity. The corpus-wide dominance of interactive multimodality, uniform in Dawn and Daily Times and strong in The Express Tribune and Daily Jang, supports the latter characterization for four of the five newspapers examined here. The Nation is the clear exception, with a substantially larger illustrative-multimodal share (34 of 82 multimodal cartoons) and the corpus's largest unimodal share (18 of 39 unimodal cartoons corpus-wide)—a genuinely distinct editorial approach to modal composition that a companion analysis, examining cross-newspaper variation across the GTVH's full Knowledge-Resource architecture, treats in detail. For the argument developed in this paper, The Nation's departure is theoretically informative in a specific way: it demonstrates that interactive, cross-channel Script Opposition is not a structural inevitability of the editorial cartoon genre as such, but a modal choice that most, though not all, Pakistani newspapers in this sample exercise consistently—reinforcing, rather than undermining, the claim that where interactive multimodality is chosen, cross-channel Script Opposition follows as its structurally expected cognitive correlate.

5.5. Implications for Coding Practice in Multimodal GTVH Research

A further implication of these findings concerns coding practice rather than theory as such. Script Opposition is also qualitatively identified in much previous work on GTVH (such as Tsakona's 2009 study), which presents a typology of individual cases studied closely and individually rather than follows a standardized procedure to record the relationship between a semiotic channel and a script. Although it allows for methodologically proving that cross-modal activation is possible, which was Tsakona's contribution, it is not enough to prove the distributional claim of this paper, which involves coding the channel-attribution of every cartoon in a representative corpus in a systematic and reliable manner. In this study, the agreement of independent coders on the Script Opposition measure was near perfect ($\kappa = .96$), despite the interpretive judgment needed to apply the measure, the result suggests that channel-attribution can be sufficiently

explicit to be applied consistently by independent coders. It has a direct methodological implication for future multimodal GTVH research: if there is no step of channel attribution that is comparable with the one used here, subsequent research would be unable to differentiate a genre where cross-modal Script Opposition was structurally predominant from another in which it was occasional, and any claims about characteristic humor mechanism for the genre would be based on impression rather than on corpus level evidence.

5.6. Boundary Conditions Reserved for Related Work

Two additional patterns identified in the wider research program from which this corpus is drawn bear on the interpretation of the present findings without being part of this paper's own argument, and are noted here as explicit boundary conditions. The first is that the GTVH's formal vocabulary lacks a systematic taxonomy for the description of the visual realization of Script Opposition and its accompanying Logical Mechanisms, which are often carried out in terms of a variety of compositional features, as demonstrated by the visual-modality examples in Section 4.3, e.g. relative scale, spatial arrangement, caricature proportion, figure-ground organization of pictures. The extension proposed in section 5.3, which looks at the distribution of scripts across channels, does not solve this gap; closing it would involve making the schema of visual grammar (e.g. that developed by Kress and van Leeuwen, 2020) explicitly systematic with the linguistic taxonomy of the GTVH, which is already in place, but is beyond the scope of the present paper. This second question is about the reception of humor: the GTVH is a production-oriented framework, which does not attempt to explain why some structurally similar Script Opposition configurations may be received with vastly different affective responses depending on the Targeted and the power relations that hold between them. This issue is not dealt with by the present corpus-composition analysis, nor can it be resolved by additional coding of Script Opposition alone. Here, for the sake of over-doubling the theoretical effect, we leave unmentioned the distribution and cognitive structure of Script Opposition as such, but not the wider array of respects in which the GTVH's multimodal application calls for additional theoretical development.

6. Conclusion

This paper aimed to examine, for the first time in such a large scale, whether the activation of Script Opposition is a peripheral option or the primary construction mode of humor in the genre of a text image combination. The analysis, which was performed on a corpus of 500 editorial cartoons written and produced in the Pakistani context, revealed that 92.2% of the cartoons in the corpus were multimodal, 89.8% of multimodal cartoons were interactive (both modes needed), and, within this dominant mode, the two opposing scripts the GTVH requires are characteristically anchored in different modes, with one in the verbal mode and the other in the visual mode, processed by the reader as simultaneous

rather than sequential opposition. This pattern occurs in the context of a contrast set of examples of mixed-, textual-, and visual-modality jokes, but is not a variant realization of the single-channel Script Opposition model developed from verbal-joke data by Raskin (1985) and Attardo (1994, 2001), but a structurally different cognitive operation that the GTVH's existing formulation does not explicitly theorize.

On this evidential basis, the paper suggests a formal distinction between cross-channel and within-channel Script Opposition, to be coded as channel-attribution in any context where the GTVH is used for a multimodal text, and a formal elaboration of the concept of cross-channel activation as a configuration that deserves to be theorized separately. This proposal is intentionally limited to the Script Opposition Knowledge Resource and the others five Knowledge Resources, and how they relate to modal composition, are left for further research.

The results presented here are subject to certain limitations. This corpus is genre- and language-specific, because it contains editorial cartoons from English and Urdu newspaper newspapers during one politically eventful year (of 12 months) and does not represent editorial cartooning as a universal field. Of course, similar cross-modal rates of Script Opposition activation are not necessarily found in other national and linguistic traditions, other historical periods, or other multimodal humor genres (such as internet memes, comic strips, multi-panel graphic satirical cartoons), a point that can only be answered by a comparable corpus-scale investigation using the same channel-attribution methodology that is proposed here. The most immediate and fruitful extension of this line of inquiry will be to test the present claim against the genres that lie adjacent to it, where the formal relationship between image and text is different from the single-panel editorial cartoon.

References

- Attardo, S. (1994). *Linguistic theories of humor*. Mouton de Gruyter.
- Attardo, S. (1997). The semantic foundations of cognitive theories of humor. *Humor: International Journal of Humor Research*, 10(4), 395–420.
- Attardo, S. (2001). *Humorous texts: A semantic and pragmatic analysis* (Vol. 6). Walter de Gruyter.
- Attardo, S. (2020). *The linguistics of humor: An introduction*. Oxford University Press.
- Attardo, S., & Raskin, V. (1991). Script theory revis(it)ed: Joke similarity and joke representation model. *Humor: International Journal of Humor Research*, 4(3–4), 293–348.
- Barthes, R. (1977). *Image–music–text* (S. Heath, Trans.). Hill & Wang.
- Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge University Press.

- Bateman, J. A. (2008). *Multimodality and genre: A foundation for the systematic analysis of multimodal documents*. Palgrave Macmillan.
- Bounegru, L., & Forceville, C. (2011). Metaphors in editorial cartoons representing the global financial crisis. *Visual Communication*, 10(2), 209–229.
- Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. Sage Publications.
- Fallianda, F., Astiti, R. Y., & Hanim, Z. A. (2018). Analyzing humor in newspaper comic strips using verbal-visual analysis. *Lingua Cultura*, 12(4), 267–271.
- Forceville, C. (1996). *Pictorial metaphor in advertising*. Routledge.
- Forceville, C. (2005). Addressing an audience: Time, place, and genre in Peter Van Straaten's calendar cartoons. *Humor: International Journal of Humor Research*, 18(3), 247–278.
- Forceville, C. (2008). Metaphor in pictures and multimodal representations. In R. W. Gibbs Jr. (Ed.), *The Cambridge handbook of metaphor and thought* (pp. 462–482). Cambridge University Press.
- Forceville, C. (2016). Pictorial and multimodal metaphor. In N.-M. Klug & H. Stöckl (Eds.), *Handbuch Sprache im multimodalen Kontext* (pp. 241–260). De Gruyter.
- Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning*. Edward Arnold.
- Halliday, M. A. K. (1985). *An introduction to functional grammar*. Edward Arnold.
- Kress, G., & van Leeuwen, T. (2020). *Reading images: The grammar of visual design* (3rd ed.). Routledge.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174.
- Neuendorf, K. A. (2017). *The content analysis guidebook* (2nd ed.). Sage Publications.
- Nikolajeva, M., & Scott, C. (2000). The dynamics of picturebook communication. *Children's Literature in Education*, 31(4), 225–239.
- Raskin, V. (1985). *Semantic mechanisms of humor*. D. Reidel Publishing Company.
- Samson, A. C., & Huber, O. (2007). The interaction of cartoonist's gender and formal features of cartoons. *Humor: International Journal of Humor Research*, 20(1), 1–25.
- Schank, R. C., & Abelson, R. P. (2013). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Psychology Press. (Original work published 1977)
- Tsakona, V. (2009). Language and image interaction in cartoons: Towards a multimodal theory of humor. *Journal of Pragmatics*, 41(6), 1171–1188.