RELEVANCE-THEORETIC MODEL FOR THE ANALYSIS OF MULTIMODAL COMMUNICATION

UDC 81'33:316.776

Nina Ž. Manojlović

Faculty of Philology and Arts, University of Kragujevac, Serbia

ORCID iD: Nina Ž. Manojlović   https://orcid.org/0000-0003-0238-6413

Abstract. *The paper examines the interpretation procedure of multimodal messages from a relevance-theoretic perspective. The aim is to test the potential of the RT apparatus to various types of multimodal messages – advertisements, political posters and memes. The chosen modes of communication have the objective of providing genre diversity for the sample analysis conducted in this paper. Our analysis has the following specific goals: 1) analyzing the explicit content of the message (visual and textual, when applicable), 2) analyzing the implicit content of the message (including possible weak implicatures), 3) potential differences in interpretation that are related to cognitive environment in which they are processed and 4) how the informative and communicative intention can be achieved. The results confirm that RT is a well-grounded framework for future research regarding multimodal communication.*

Key words: *multimodal communication, relevance theory, genre, explicature, implicature, context*

1. INTRODUCTION

The main focus of this paper is to examine the possibilities of application of Relevance Theory (henceforth RT) to the analysis of multimodal communication. Since RT aims at explaining all aspects of human communication (both within verbal and non-verbal communication) it is to be expected that multimodal communication be explicable in terms of RT principles. All the different approaches to multimodal communication share some of the features. One of those features is that "[t]hey draw on disciplines that originally focused on language in use, that is disciplines concerned with what people do with language in their everyday life, notably how they construe the social world through language" (Jewitt, Bezemer, and O'Halloran 2016, 7). Pragmatics is precisely focused on language in use and RT aims at explaining communication in all its aspects. Therefore, it is not surprising that multimodal communication has become a research subject of RT scholars.

When doing research in the field of multimodality it is important to state the theoretical and methodological stance, since the field is extremely varied and diverse (Jewitt, Bezemer, and O'Halloran 2016, 1). Namely, the term *multimodality* has been mainly used to denote that people use different modes or meaning making means in order to communicate. However, the idea that people communicate in various ways is by no means new (Jewitt, Bezemer, and O'Halloran 2016, 2). What really marked a turning point within the field is the introduction of informational technologies which made the process of dissemination easier (Jewitt, Bezemer, and O'Halloran 2016, 3). The theoretical framework of the research conducted in this paper is relevance-theoretic, which is a cognitive pragmatic approach to communication, so the term used here will be *multimodal communication* (synonymous with *multimodal discourse* or *multimodal interaction*). Even when the subject of a linguistic study is a text that appears to be dominated by writing or speech alone one cannot ignore other modes that contribute significantly to the overall meaning. As Jewitt, Bezemer, and O'Halloran (2016, 29) point out: "[w]hether you study a political 'speech' or a 'written' post on a blog, the other parts of the whole of which language is a part are worthy of as detailed an analysis as language itself."

On the other hand, there is no definite and unified approach to analyzing multimodal documents in an empirically and scientifically adequate way. Bateman (2008, 13) emphasizes the necessity for a framework that is empirically sound and able to separate right interpretations from wrong ones. Kress and van Leeuwen (2001, 121), however, state that the point of their analysis is not whether the explanation is right or not, but that they have used principles of provenance to create an interpretation. They further explain that "[i]n the absence of a 'grammar' to help us along, we have used, to put it another way, modes of interpretation oriented towards 'cultural studies' and 'phenomenology'." (Kress and van Leeuwen 2001, 121). Here, Bateman (2008, 13−14) identifies two major problems:

> First, there is the problem that relatively superficial interpretations become obscured behind apparently technical discourse, whereas in fact the analyses may not have gone beyond what is readily available to inspection. Second, there is the further problem that applying rich interpretative schemes from other areas of study − including multimodal linguistics − may all too easily swamp the rather weak signal that we are currently capable of receiving from the multimodal artefacts themselves. […] Boosting the strength of this signal using linguistic methods will involve uncovering, systematically and empirically, precisely to what extent multimodal documents construct meanings analogous to, and different from, those now established for language.

Multimodal documents have the advantage of combining various semiotic resources which do not merely contribute to meaning by adding more meanings of the same kind but creating meaning by bringing together different semiotic artefacts (Bateman 2008, 54). This, we believe, while adding an additional layer of meaning, further complicates the analysis of such documents. Understanding the need for a more precise and theoretically sound approach to multimodal documents, the aim of this paper will be to present and apply RT to different genres of multimodal documents, examining how different aspects of meaning are constructed and interpreted within multimodal communication.

Having all this in mind, the paper is organized as follows: in the next section we shall present the theoretical framework. This section will introduce the key notions related to multimodal research as well as basic ideas of RT. Section 3. will focus on application of RT to multimodal communication. Finally, section 4. will present a sample analysis of different multimodal documents – advertisements, political posters and memes.

2. Theoretical Framework

Since multimodality is a very diverse field it is important to state basic notions important for the research conducted in this paper. Furthermore, as this paper presents an attempt to apply RT to different genres of multimodal documents, it is significant to relate the notion of genre to multimodal research, as well as present the tenets of the cognitive-inferential approach to communication – RT.

2.1. Multimodality

The key premises of multimodality can be summarized as follows (taken from Jewitt, Bezemer, and O'Halloran 2016, 3): 1) meaning is made with different semiotic resources, each offering distinct potentialities and limitations, 2) meaning making involves the production of multimodal wholes and 3) if we want to study meaning, we need to attend to all semiotic resources being used to make a complete whole. Therefore, any theory of language needs to address all the modes of communication, including RT, which will be discussed further in the next section of the paper. As Bateman (2008, 38) notes, it is completely natural for linguistic theories investigating function rather than form in language to broaden their attention to units other than verbal texts in the narrower sense. What, in fact, makes a study multimodal is the approach and the methodological framework that accounts for all (or at least some) modes or semiotic resources at work (D'Angelo and Marino 2004, 178). Having in mind that RT aims at explaining different modes of communication, i.e. every ostensive act of communication – it is expected that this framework will be a valuable and insightful approach to multimodal communication.

Since the corpus of this research comprises multimodal documents that include images, special attention has to be paid to the role and the *meaning* images carry. For instance, in their study of image, Kress and van Leeuwen (2006) showed that image not only has the equivalent of what linguists call lexis, it also has a 'grammar'. Visual *grammar* describes the way in which the elements of an image (people, places and things) combine to form more or less complex visual *statements*, in the same way that language grammar describes how words combine in clauses, sentences and texts (Kress and van Leeuwen 2006, 1). Furthermore, images are more often than not paired with other modes of communication, mainly with text. The relationship between these different modes can vary. Speaking about exploring the *grammars* of different modes within the area of multimodal discourse analysis, van Leeuwen (2005, 73) notes:

> Comparisons between the various 'grammars' are especially needed. We need to know more both about the extent to which different modes can do the same kind of communicative work and about the extent to which they differ in their semiotic potential, and we need to know more about the difference between doing the same kind of communicative work through one mode or through another. But what is also needed is research on the way different modes are integrated in multimodal texts.

Martinec and Salway (2005) present examples of similar relations between image and writing in multimodal texts. This relation can be that of equality, i.e. when image and writing complement each other (e.g. diagrams and written text in a science textbook). On the other hand, this relationship can be quite unequal (e.g. photographs of artworks and their captions in an art catalogue). Moreover, van Leeuwen 2005 (2005, 74) points out that texts themselves can "(1) foreground visuality, not just by using images, but also by using

composition (layout), typography and colour; and (2) are deliberately designed to allow multiple ways of reading".

On the other hand, certain parts of pictures in multimodal documents will be recognized only if they are picked up cross-modally – e.g. mentioned or explained by the accompanying text or explicit labels (Bateman 2008, 41−42). This idea is close to the notion of various degrees of explicitness of a picture that will be discussed within the RT framework.

The question of genre is crucial to multimodal analysis. Bateman (2008, 10) states that genre is increasingly talked about and studied within the field of multimodal research – genre of websites, newspapers, homepages, etc. However, genre is considered not to be a loose collection of text types, but rather a region in an entire space of genre possibilities (Bateman 2008, 10). Miller (1984, 155) notes that classification of genres should be based around pragmatic, rather than semantic or syntactic properties. She further explains that this classification should be organized around situated actions and remain open to change. Bateman (2008, 189) agrees with this notion of determining genre based on the social action that is being achieved as opposed to form of discourse. He further explains that "[r]ecurring social situations give rise to recurring problems that can be approached through recurrent communicative solutions. These recurrent communicative solutions can then themselves be recognised as such by virtue of their recurrent form − but it is the social action that drives the account." Hence, a rigid taxonomy of genres is not possible, since they evolve, change and disappear (Miller 1984, 163). Genre would be better defined as being nearer or farther from another type of genre along various dimensions of comparison (Bateman 2008, 223). Still, the issue of genre remains important since it has the potential to function predictively – knowing the genre that a particular multimodal document belongs to helps make predictions concerning its form, function and meaning (Bateman 2008: 196).

Attempting to apply systemic-functional linguistic approach to genre (that puts the emphasis on stages) to multimodal communication van Leeuwen (2005) concluded that this approach must be modified, especially since reading path becomes less relevant. Likewise Hiippala (2014, 112) notes that the notions of linearity and stages traditionally tied to genre analysis become problematic when extended from a purely linguistic environment to a multimodal one (linearity might still be important for certain types of multimodal artefacts, such as serial graphics).

Since multimodal documents comprise various different segments, these can be thought of as mini-genres – units that capture re-usable fragments for fulfilling particular discourse purposes (Baldry and Thibault 2006, 42). These mini-genres are similar to the notion of primary genres – prefabricated text-making resources, while secondary genres (e.g. novels, advertisements, scientific articles, etc.) use and combine these resources for their own purposes (Baldry and Thibault 2006, 42). Furthermore, with the advancement of technology numerous new multimodal genres have emerged, calling for new analytical frameworks and marking a multimodal turn in genre analysis (D'Angelo and Marino 2024).

2.1. Relevance theory

RT has a slightly different perspective but the same goal when it comes to the analysis of different modes of communication. We will first present the basic notions of the RT account and then how it can be modified for multimodal analysis.

Relevance theory (Sperber and Wilson 1986/95, 1987, Carston 2002) is a cognitive-inferential approach to communication that proposes the existence of two main principles:

cognitive and communicative principle of relevance. According to the cognitive principle of relevance human cognition is geared towards the maximization of relevance (Carston 2002, 45). Relevance, in turn, cannot be defined in quantitative terms. It represents the relation between positive cognitive effects and the cognitive effort needed for the interpretation – the greater the positive cognitive effects (or the lesser the cognitive effort) the greater the relevance (Sperber and Wilson 1987, 732). On the other hand, the communicative principle of relevance states that each ostensive stimulus carries the assumption of its own optimal relevance, i.e. that it is relevant enough to be worth the processing effort needed for its interpretation and that it is the most relevant stimulus according to speaker's abilities and preferences (Carston 2002, 45). In other words, each time an utterance is communicated the speaker expects that the interpretation of the said utterance will be worth the cognitive effort needed, and that the utterance will be optimally relevant (even though this is not always the case, the expectation of relevance is obligatory).

Communication is normally ostensive-inferential (Sperber and Wilson 1986/95, 50−54), i.e. it is clear to both the sender and addressee that the sender wants the addressee to be aware that she directs a message to him and that he is to infer relevant information from this message. This intention is two-layered – we can speak of informative and communicative intention[1]. Informative intention is the intention to make manifest or more manifest to the audience a set of assumptions (Sperber and Wilson 1986/95, 58) while the communicative intention is the intention to make it mutually manifest to audience and communicator that the communicator has this informative intention (Sperber and Wilson 1986/95, 61). The communicative intention is, in fact, the speaker's intention to get one or more addressees to accept certain assumption(s) as (probably) correct.

Another important distinction is that between explicit and implicit layer of communication. Explicature is a proposition communicated by an utterance iff it represents pragmatically enriched logical form (or one of the logical forms) of the utterance (Mišković-Luković 2018, 131). Implicature is a proposition communicated by an utterance that is not explicature, i.e. a proposition generated solely by means of pragmatic inference (Mišković-Luković 2018, 133). Implicatures can be implicated premises and implicated conclusions, and they can be weak and strong. Strong implicature is a premise or conclusion that the addressee must arrive at in order to understand the speaker's meaning (Mišković-Luković 2018, 136). Weak implicature, on the other hand, is an implicature that is not essential for precise interpretation of the speaker's utterance and the hearer's responsibility for the truth of weak implicatures is greater (Mišković-Luković 2018, 152). Explicatures can also be more or less strongly communicated. The closer the explicature is to the encoded content of the utterance the stronger it is.

When it comes to application of RT to multimodal communication Forceville (2014, 2020) has done the bulk of labor. To the best of our knowledge apart from his work not many studies have been conducted that further test the application of RT to multimodal research (some examples include Hu 2022, Ifantidou and Tzanne 2006, Jagoe and Smith 2016). However, Forcewille (2014, 59) states:

> My central point is that RT can well model visual communication – whether or not accompanied by language or other modes – but that this requires taking into account the affordances and constraints of the visual mode as well as the fact that most visual and multimodal communication is of the mass-medial kind.

---

[1] The latest revision of the notions *informative* and *communicative intention* involves distinguishing two forms of ostension: *basic ostension* and *mentalistic communication* (see Sperber and Wilson 2025).

The adaptation of RT proposed by Forceville (2014, 2020) can be summarized as several main points. Firstly, it is important to note that visuals and multimodal texts can be ostensive stimuli that trigger inferential processes much like verbal communication. Each mode (e.g. visuals, layout, typography, etc.) contributes to meaning and can convey explicatures and implicatures. Since visual communication lacks the explicitness of verbal communication and often lacks conventional codes which then bring strongly into focus the RT's emphasis on inferencing over decoding. The presumption of relevance guides the interpretation process even in the cases of complex multimodal artefacts where the cognitive effort is greater. All of the processes are in line with the application of RT to 'regular' verbal communication. There is one major difference – the importance of relevance to an individual becomes more central. Namely, multimodal message (e.g. presidential speech or TV advertisement) in mass communication can be interpreted differently by the recipients depending on their cognitive environment (background knowledge, experience, beliefs, attitudes, emotions, goals). However, this this shift "does not affect the central RT tenet that relevance is always relevance to an individual" (Forceville 2020, 111). These individuals are grouped in aggregates of individuals who share certain domains of knowledge, interests, attitudes, and/or beliefs so we can speak of range of dominant, negotiated and oppositional interpretations. Finally, it is important to address the issue of communicative failures and misinterpretations. Namely, due to unfamiliar codes, ambiguous cues or divergent cognitive environments, multimodal messages may fail if the intended explicatures or implicatures are not derived as intended.

However, the notion of genre can be extremely helpful in guiding the inferential process (see Forceville 2020, 118−132). Genre guides the expectations and helps limit the range of relevant (and possible) inferences recipients make. Knowing a visual belongs to the genre of an advertisement, political cartoon, or comic strip directly influences how viewers derive explicatures and implicatures.

Acknowledging certain challenges RT faces when applied to multimodal communication, Forceville (2020) proposes integrating RT with disciplinary insights from other fields such as semiotics, iconography, metaphor theory, and narratology. These will be addressed in more detail within the analysis of sample multimodal artifacts in section 4 of this paper.

To conclude this section, as reading paths and stages (see van Leeuwen 2005) become less relevant in multimodal communication and since the goal is to achieve optimal relevance, all of the aspects will be processed – albeit some require additional cognitive effort (perchance depending on the 'grammar' of the mode itself or on the context needed for the interpretation). They need not be biased toward linearity. This does not mean that they don't pose any constraints upon the recipients (something can be central, bold, bigger, etc.). If certain elements are more prominent and available as *direct perceptions* (cf. Bateman 2008 on *pre-attentive perception*) they can be said to be highlighted or emphasized. In relevance-theoretic terms this could affect the derivation of certain implicatures but does not impact the explicit content of the message.

3. METHODOLOGY

The present paper does not aim at conducting a systematic corpus analysis. The aim is rather to examine the possibilities of applying RT to multimodal communication and possibly provide guidelines for future research. Hence the corpus selection implied the random sample method, where the main concern was that the selected documents belong to different genres.

The reason for this is that it allows for testing the applicability of the selected framework to different types of documents. Furthermore, as D'Angelo and Marino (2024, 181) note even small-scale multimodal corpus analysis (collecting five to ten documents) allows researchers to conduct a fine-grained analysis and pose numerous research questions. All the documents analyzed in this paper contain both image and text and belong to the genres of advertisements, political posters and memes. First two genres have persuasive purpose, albeit with a different aim, while memes have a basic purpose of creating a humorous or entertaining effect. Memes can in fact be considered a genre in itself which is in line with Bateman's GeM Model according to which the genre base is „a representation of the grouping of elements from other layers into generically recognisable configurations distinctive for particular genres or document types" (Bateman 2008, 108). It is expected that the interpretation mechanisms will be the same across these different genres, i.e. that no special cognitive principles or pragmatic processes will be required hence highlighting the applicability of RT in future linguistic research of multimodal documents.

Furthermore, the selected genres differ in another important aspect – advertisements and political posters are usually created by professionals while memes represent content that is created (often using templates or image macros) by anyone who wishes to do so. This is important because non-professional design is extremely valuable since it represents examples closest to natural or spontaneous verbal graphic language production (Bateman 2008, 29). However, even for trained professionals, designing documents presents a difficult task and it may happen that professionally produced documents fail to adequately show their consumers how they are to be segmented into meaningful elements (Bateman 2008, 37−38). Within RT this will be analyzed in terms of different available contexts, manifestness of the assumption to an individual (strongly and weakly communicated propositions) as well as informative and communicative intention.

In line with the semiotic systems typology proposed by Matthiessen (2007, 24−25) the cognitive-inferential analysis in this paper will be applied to the following layers: written language, visual paralanguage (font type, color, size, layout, etc.) and the visual (pictorial) system (drawings, paintings, photographs, maps, graphs, charts, etc.). Hence, the research questions can be summarized as follows: What is the explicature of the image or text? How do they affect each other? In what ways do different types of context (immediate context, linguistic context, situational context, background knowledge – communal common ground / personal common ground) affect the interpretation? Are the communicative and informative intentions recognized and fulfilled?

4. SAMPLE ANALYSIS

In this section we will apply RT methodology to selected multimodal artifacts from three different genres – advertisement, political poster and meme.

When it comes to advertisements, Forceville (2020, 150) points out that the most important feature of this genre is the intent to present the product or service in a positive light. The interpretation of such multimodal documents is guided and constrained by information in their cognitive environment pertaining to different aspects of the said genre. Therefore, if the audience notices an advertisement as such and pays attention to it the communicative intention is recognized and fulfilled. When it comes to informative intention the situation is slightly more complex – it is recognized if the addressee understands the claim made about the product and

fulfilled only if the addressee accepts the claim made about the product, which is beyond the reach of linguistic research. In Figure 1, for example, fulfilling the informative intention would entail the recipient accepting the communicated assumption and integrating it into his/her system of beliefs – that the advertised company makes its product solely from tomatoes (possibly grown by them) further attesting the quality of the advertised product.

Fig. 1 Heinz commercial[2]

Fig. 2 Colgate commercial[3]

Let us start from the explicit content of the chosen advertisements. In Figure 1 we have a central image of the product and visual code for ketchup bottle. Another visual code is stacked slices of tomato. This visual metaphor is then recognized as an ostensive stimulus linking the advertised product with fresh tomatoes, giving rise to implicatures about the quality of the product. The accompanying text further reinforces the explicit and implicit message of the visual, containing the same metaphor in itself (*No one grows Ketchup like Heinz*). Hence, the image is central with the text merely elaborating (in line with Halliday's (1994) concepts of *elaboration* and *extension*). We can see that RT can lean on and use not only insight from visual grammar (pertaining to layout, perspective and color) but visual metaphor as well. To achieve relevance this visual must in some way encourage the derivation of a positive claim about the advertised product. Since the explicature of the image contains two encoded concepts – KETCHUP and SLICED TOMATO – we believe that the use of visual metaphor is redundant. The inference here follows the same path of least effort[4] as in the case of verbal communication, creating an *ad hoc* concept (see Carston 2002b, 2010). This entails that the encoded concepts would be loosened and strengthened in the process of mutual adjustment so that the constituent of the explicature of the image is an *ad hoc* concept (e.g. TOMATO** that is broader than the lexical concept TOMATO in that it encompasses other substances such as ketchup, and possibly

---

[2] Source: https://www.marq.com/wp-content/uploads/2023/10/image11-1.webp
[3] Source: https://www.marq.com/wp-content/uploads/2023/10/image12.webp
[4] The least-effort strategy assumes that: a) (following the path of least effort) the addressee forms the interpretative hypotheses according to the order of their accessibility, b) the process ends when the addressee satisfies his expectation of relevance or when such satisfaction is not attainable (Mišković-Luković 2018: 144).

narrower than the lexical concept TOMATO so as to exclude all tomatoes that are not high quality, ripened or tasty). We believe that there is no need for using cognitive linguistic terminology of cross-domain mapping, since RT apparatus is able to explain the interpretation process with the use of the pragmatic processes involved in verbal communication. Implicatures would go along the line of persuasion: If one buys/uses this product he/she is guaranteed quality/healthy product etc.

In Figure 2 the image of a woman is central, visual codes include centrality of white teeth biting on a cloth. The accompanying text puts emphasis on two concepts – SMILE and STRENGTH. Visual and textual explicatures complement each other and give rise to implicature that using the advertised product will lead to the user having beautiful smile and strong teeth. If the recipient arrives at these assumptions the communicative intention is recognized and fulfilled and if the recipient believes these assumptions, i.e. adopts them – the informative intention is fulfilled. Furthermore, depending on the cognitive environment of the recipient, they might derive additional weak implicatures pertaining to strength and sport activities, vitality and so on (since the image encodes also a woman doing sport-like activity).

Political posters share certain traits with printed advertisements – they are usually stand-alone images that are almost always accompanied by text (Forceville 2020, 167). Again, as was the case with advertisements, both the decoding and inferring process are guided by the genre itself, since political posters make persuasive claims about the real world or certain state of affairs. Another characteristic of political posters is that the cognitive effort needed for their processing needs to be minimized since the recipients do not tend to dwell long on their interpretation (Forceville 2020, 169). However, the same can be said for advertisements as well – they also aim at communicating clear message without overwhelming the recipient.

Fig. 3 Obama "HOPE" poster[5]   **Fig. 4** Anti-fascist poster[6]

[5] By Shepard Fairey, 2008.
[6] Source: https://d1hhug17qm51in.cloudfront.net/www-media/2021/01/19103623/PGAD17.033_01_H02-Large-JPEG_5x7-554x1024.jpg

In processing the poster in Figure 3 there is low processing effort since the primary colors were used, the face is recognizable, and the composition is simple. The poster explicitly communicates the word "HOPE" and presents a visually idealized version of Barack Obama looking upward and outward which is suggestive of vision, future, and leadership. This interpretation relies on the recipient's cognitive environment, especially background knowledge and knowledge about the genre of political posters. The explicit content of the picture is here reinforced by the use of the encoded concept HOPE. This prompts the visual connotations of vision, future and leadership. Strong implicature is that Obama brings or symbolizes hope. Through the inferential process other weaker implicatures can be arrived at, e.g. that the country can look forward to a better future or that Obama is a moral, forward-looking leader. The poster's visual economy (few words, intense gaze) relies on the audience to fill in meaning, making the message highly inferential, yet powerful. The color scheme encodes the concept of patriotism, and the style of the portrait turns a political figure into a cultural symbol. The encoded concept of (UPWARD) GAZE (depending on the recipient's abilities and preferences) gives rise to the communicated concepts VISION and ASPIRATION – what Forceville (2020) would call a visual metaphor LOOKING UP is LOOKING FORWARD, TOWARD PROGRESS. However, as previously stated, we believe that conceptual metaphors in RT are redundant and that the meaning can be explained by the notions *encoded* and *communicated* (*ad hoc*) *concept*. Typography also encodes meaning – all capital letters *HOPE* encodes CERTAINTY and GRAVITY, but all additional meaning will depend on the recipient's abilities and preferences, emphasizing once more the relevance of message to an individual.

In Figure 4 three main sections can be identified immediately for processing: the text *Are we next*, the image of a swastika with the pattern of the American flag and the warning below *Be aware*. The poster's stark, provocative visual (American flag transforming into a Nazi symbol) and bold textual imperatives clearly indicate an urgent communicative intention. Again, we have low processing effort (simple graphic design and highly familiar symbols – swastika and US flag) and high cognitive effects (profound political warning, historical allusion, moral appeal). The explicatures of the textual segments cannot be generated without the context provided by the image. Namely, for the first segments at least two pragmatic processes are needed for the logical form to become fully propositional – reference assignment (*we*) and saturation (*next [for what]*). The same stands for the bottom text – who is to be aware and of what. The image, on the other hand, has a strongly communicated explicature from the encoded concepts of SWASTIKA and AMERICAN FLAG. The strong implicature is that US might be heading towards fascism, and weaker implicatures can include the following: American identity is being co-opted or distorted, Citizens must stay alert to political shifts resembling those of Nazi Germany and so on. Since the ironic juxtaposition of American stars and stripes with the Nazi swastika generates high cognitive effects and emotional salience (shock, fear, moral urgency) we can also conclude that the relevance of the poster in Figure 4 lies also in the non-propositional effects tied to emotions. Non-propositional effects are important in verbal communication as well. In certain cases, where less specific meaning is being communicated by the utterance/expression relevance lies in non-propositional effects. Unlike the speaker's meaning, non-propositional effects are expected to be differently paraphrased by different hearers. No final paraphrase encompasses all the nuances of these effects, so they are considered to be open-ended effects (Wilson and Carston 2019, 32). These will be discussed further within the analysis of memes as multimodal documents.

Turning to the meanings that will probably not be recognized and inferred by all recipients of the analyzed multimodal document we can note that the typography used reflects 1970s protest, the vertical organization from the question at the top of the page to the command/warning at the bottom mimics the logic of a deductive argument in visual form and so on. In sum, this is a highly inferential multimodal artefact. Its effectiveness stems from combining minimal text with visually dense symbols, prompting viewers to supply contextual assumptions about fascism, history, and national identity.

Finally, we will briefly analyze examples from the genre of memes. The idea of memes is tied to the notion of *cultural unit* and is formed from a word with Greek origin - *Mimeme* (Dawkins 1976, 192). These units have the characteristic to *infect* the human mind and are hence replicated quickly by imitation (Blackmore 1999, 2001). The goal here is different than it was with the advertisements and political posters, since the primary aim is the humorous effect. When it comes to humor the speaker/creator has a great level of control over the manner in which the inference takes place – which contextual information should be accessed, which implications should be arrived at etc. (Yus 2017). Even though the relevance of jokes does not lie in new information, the cognitive effect expected – that of comic relief and entertainment – is still worth the processing effort (Yus 2017, 189). In exploring humor within internet communication Yus (2018) put an emphasis on the social and emotional dimensions of the use of humor in internet-mediated communication. Humor online generates the already mentioned positive non-propositional effects, such as community bonding, identity shaping, and self-esteem boosts. Yus (2018) takes precisely memes as an example since they often rely on shared cultural knowledge, fostering a sense of belonging among those users who "get" the joke.

Fig. 5 Are you going to sleep[7]

Fig. 6 Distracted boyfriend meme[8]

Memes make use of the image macro which is, in certain groups, a highly recognizable picture that already has explicit meaning gained through its repeated use. In Figure 5 we have *Are you going to sleep* macro which is an exploitable webcomic where a brain asks a sleeping person *Are you going to sleep?* to which they reply *Yes, I am. Now Shut Up*. Finally the brain

then says something that causes the person to stay awake. This macro creates the expected interpretation before the textual signals are even processed. This four-panel comic invites the narrative reading interpreting textual segments in a correct order. Explicature is easily recoverable, but the relevance here lies in the implicatures and the positive humorous effect. The implicatures may encompass the following: Everyday anxiety emerges right at the moment of rest, disrupting peace of mind; the brain is a malicious/chaotic internal monologue, undermining well-being; this represents the shared human experience of nighttime overthinking or regret, etc. Memes usually require minimal cognitive effort due to the use of image macros drawing on familiar context and they provide cognitive effects of relatable insight, emotional resonance, and humor.

Figure 6 is an example of *Distracted boyfriend* meme which is an object labeling stock photo series in which a man looks at the backside of a woman walking by while another woman, presumably his romantic partner, looks on disapprovingly. Similar to the previous analysis, the expectation of inference is pre-existing. The recipient is expecting that the text labels the first woman as something the man (again needs to be labeled) wants/ is interested in and a second woman as something he has or has to do. Again, the explicature is easily recoverable – that the speaker is looking at memes when he/she should be working. The relevance is in the implied and relatable assumptions related to procrastinating at work. In fact, these image macros can be considered as mini-genres (see Baldry and Thibault 2006) since they represent units often intended to capture re-usable fragments for fulfilling particular discourse purposes. These image macros are usually straightforward and easily decoded even if the recipient is not familiar with the mini-genre (image macro) in question.

Finally, it is worth noting that, unlike advertisements and political posters whose aim is persuasive, informative intention is often fulfilled when it comes to interpretation of memes. The reason for this is that the target audience are individuals and groups that share common interests and find the content relatable. This is in line with the conclusions made by Yus (2018) – for social and emotional dimensions, such as community bonding, to be fulfilled the recipient needs not only to recognize the informative intention but to accept the communicated assumption(s).

5. CONCLUDING REMARKS

In this paper we have applied RT framework on several selected multimodal artefacts. In order to better test the approach, we have selected multimodal documents belonging to different genres. This was important for two main reasons. Firstly, the analysis incorporates documents with persuasive (advertisements and political posters) and humorous intention (memes) which can affect the interpretation process when it comes to the notions of informative and communicative intention. Secondly, the analysis conducted in this paper comprises artefacts created both by professional content creators and amateurs (memes). This distinction is especially important when it comes to multimodal communication, since the use of different modes (in)adequately makes the interpretation process more complex and arriving at the intended interpretation becomes less warranted than is the case with verbal communication.

The analysis shows that RT approach is adequate for explaining the process of interpretation of these artefacts, both the explicit and implicit content of text and image(s). Furthermore, it is adequate for accounting for how different types of context (immediate

context, linguistic context, situational context, background knowledge – communal common ground / personal common ground) affect the interpretation process. This further reinforces the relevance of input to an individual −interpretation arrived at will depend heavily on the addressee's abilities and preferences.

RT has been previously adapted for multimodal research and this analysis confirms its adequacy. Still, Forceville's model (2020) was not adopted completely in our research. Namely, the notion of different contexts as well as recipients' abilities and preferences can account for numerous aspects of other fields integrated by Forceville into RT (e.g. semiotics). However, the greatest emphasis in our paper was on the redundancy of conceptual metaphor in a cognitive-inferential pragmatic approach. Since RT has provided valid treatment of the loose use of language (including metaphor) our goal was to test whether these mechanisms work when it comes to multimodal communication. The sample analysis in this paper demonstrates that cross-domain mappings are not needed in explaining the interpretation of multimodal artefacts from a relevance-theoretic perspective. This does not mean that conceptual metaphors are not a useful tool for other approaches, but if cognitive-inferential approach is to be fully applied, the mechanisms this approach already offers should not be abandoned, especially if they are applicable. These conclusions should be tested further on a larger and more diverse body of multimodal artefacts.

References

Baldry, Anthony, and Paul J. Thibault. 2006. *Multimodal Transcription and Text Analysis*. Textbooks and Surveys in Linguistics. London and New York: Equinox.

Bateman, John. 2008. *Multimodality and Genre: A Foundation for the Systematic Analysis of Multimodal Documents*. London: Palgrave Macmillan.

Blackmore, Susan. 1999. *The Meme Machine*. Oxford: Oxford University Press.

Blackmore, Susan. 2001. "Evolution and Memes: The Human Brain as a Selective Imitation Device." *Cybernetics & Systems* 32 (1–2): 225–255.

Carston, Robyn. 2002a. *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Oxford: Blackwell Publishing.

Carston, Robyn. 2002b. "Metaphor, Ad Hoc Concepts and Word Meaning—More Questions than Answers." *UCL Working Papers in Linguistics* 14: 83–105.

Carston, Robyn. 2010. "Metaphor: Ad Hoc Concepts, Literal Meaning and Mental Images." *Proceedings of the Aristotelian Society* 110 (3): 295–321.

D'Angelo, Lorenzo, and Francesco Marino. 2024. "Multimodal Analysis." In *Conducting Genre-Based Research in Applied Linguistics*, 172–196. London: Routledge.

Dawkins, Richard. 1976. *The Selfish Gene*. Oxford: Oxford University Press.

Forceville, Charles. 2014. "Relevance Theory as Model for Analyzing Visual and Multimodal Communication." In *Visual Communication*, edited by David Machin, 51–70. Berlin and Boston: De Gruyter Mouton.

Forceville, Charles. 2020. *Visual and Multimodal Communication: Applying the Relevance Principle*. Oxford: Oxford University Press.

Halliday, M.A.K. 1994. *An Introduction to Functional Grammar*. London: Edward Arnold.

Hiippala, Tuomo. 2014. "Multimodal Genre Analysis." In *Interactions, Images and Texts: A Reader in Multimodality*, 11:111–124.

Hu, Qinghua. 2022. "Meaning Construction of Multimodal Metaphor from Relevance Theory." *Academic Journal of Humanities & Social Sciences* 5 (1): 105–109.

Ifantidou, Elly, and Antonis Tzanne. 2006. "Multimodality and Relevance in the Athens 2004 Olympic Games Televised Promotion." *Revista Alicantina de Estudios Ingleses* 19: 191–210.

Jagoe, Catherine, and Mary Smith. 2016. "Relevance in the Context of Multimodality and Aided Communication." In *The Silent Partner: Language Learning and Language Use in Aided Communication*, edited by Mary Smith, 229–246. London: J&R Press.

Jewitt, Carey, Jeff Bezemer, and Kay O'Halloran. 2016. *Introducing Multimodality*. London: Routledge.

Kress, Gunther, and Theo van Leeuwen. 2001. *Multimodal Discourse: The Modes and Media of Contemporary Communication*. London: Arnold.

Kress, Gunther, and Theo van Leeuwen. 2006. Reading Images: The Grammar of Visual Design. 2nd ed. London: Routledge.

Martinec, Radan, and Andrew Salway. 2005. "A System for Image–Text Relations in New (and Old) Media." *Visual Communication* 4: 337–371.

Matthiessen, Christian M. I. M. 2007. "The Multimodal Page: A Systemic Functional Exploration." In *New Directions in the Analysis of Multimodal Discourse*, edited by Terry D. Royce and Wendy L. Bowcher, 1–62. Mahwah, NJ: Lawrence Erlbaum Associates.

Miller, Carolyn R. 1984. "Genre as Social Action." *Quarterly Journal of Speech* 70: 151–167.

Mišković-Luković, Mirjana. 2018. *Pragmatika*. Kragujevac: Filološko-umetnički fakultet.

Sperber, Dan, and Deirdre Wilson. 1986. *Relevance: Communication and Cognition*. Oxford: Blackwell Publishing.

Sperber, Dan, and Deirdre Wilson. 1987. "Précis of Relevance: Communication and Cognition." *Behavioral and Brain Sciences* 10 (4): 697–710.

Sperber, Dan, and Deirdre Wilson. 2015. "Beyond Speaker's Meaning." *Croatian Journal of Philosophy* 15 (2 (44)): 117–149.

Sperber, Dan, and Deirdre Wilson. 2025. "Rethinking Ostensive Communication in an Evolutionary, Comparative, and Developmental Perspective." *Psychological Review*. Advance online publication. https://doi.org/10.1037/rev0000565. Also available at ResearchGate: https://www.researchgate.net/publication/392330396_Rethinking_Ostensive_Communication_in_an_Evolutionary_Comparative_and_Developmental_Perspective (accessed July 15, 2025).

Van Leeuwen, Theo. 2005. "Multimodality, Genre and Design." In *Discourse in Action*, 73–93. London: Routledge.

Wilson, Deirdre. 2018. "Relevance Theory and Literary Interpretation." In *Reading Beyond the Code: Literature and Relevance Theory*, 185–204.

Wilson, Deirdre, and Robyn Carston. 2019. "Pragmatics and the Challenge of 'Non-Propositional' Effects." *Journal of Pragmatics* 145: 31–38.

Wilson, Deirdre, and Dan Sperber. 2006. "Relevance Theory." In *The Handbook of Pragmatics*, 606–632.

Yus, Francisco. 2017. "Relevance-Theoretic Treatments of Humor." In *The Routledge Handbook of Language and Humor*, 189–203.

Yus, Francisco. 2018. "Positive Non-Humorous Effects of Humor on the Internet." In *The Dynamics of Interactional Humor*, 283–303. Amsterdam: John Benjamins.

PRIMENA TEORIJE RELEVANCIJE U ANALIZI MULTIMODALNE KOMUNIKACIJE

U radu se ispituje interpretacija multimodalnih poruka iz pozicije teorije relevancije. Cilj je testirati mogućnosti kognitivno-inferencijalnog aparata kada su u pitanju različiti tipovi multimodalnih dokumenata – reklamni sadržaj, politički posteri i mimovi. Odabrani modaliteti komunikacije imaju za cilj da obezbede žanrovsku raznovrsnost u okviru uzorka analiziranog u ovom radu. Analiza ima sledeće posebne ciljeve: 1) analiza eksplicitnog sadržaja poruke (vizuelnog i tekstualnog), 2) analiza implicitnog sadržaja poruke (uključujući i potencijalne slabe implikature), 3) moguće razlike u interpretaciji koje se odnose na kognitivno okruženje u kojem se poruke interpretiraju i 4) način na koji se mogu ostvariti informativna i komunikativna namera. Rezultati potvrđuju da je teorija relevancije dobro utemeljen okvir za buduća istraživanja multimodalne komunikacije.

Ključne reči: *multimodalna komunikacija, teorija relevancije, žanr, eksplikatura, implikatura, kontekst*