

MARTIN OJA

Semiotics of Multimodal Conflict:
Modes, Modalities and Audio-Visual
Incongruence in Feature Film



MARTIN OJA

Semiotics of Multimodal Conflict:
Modes, Modalities and Audio-Visual
Incongruence in Feature Film



UNIVERSITY OF TARTU

Press

Department of Semiotics, Institute of Philosophy and Semiotics, University of Tartu, Estonia

The council of the Institute of Philosophy and Semiotics of the University of Tartu has, on November 4, 2024, accepted this dissertation for the defense for the Degree of Doctor of Philosophy (in Semiotics and Culture Studies).

Supervisor: Professor Peeter Torop, PhD, University of Tartu

Opponent: Professor Kay Lynette O'Halloran, PhD,
University of Liverpool, UK
Research Professor Pia Tikka, PhD, Tallinn University

The thesis will be defended at the University of Tartu, Estonia, on December 16, 2024, at 14.15 University Senate Hall and online.

This research was supported by the University of Tartu ASTRA Project PER ASPERA (2014-2020.4.01.16-0027), which is financed by the (European Union) European Regional Development Fund; Estonian Science Foundation research grant RANS Professorship of Semiotics of Culture (Department of Semiotics, University of Tartu).



European Union
European Regional
Development Fund



Investing
in your future

ISSN 1406-6033 (print)
ISBN 978-9916-27-737-9 (print)
ISSN 2806-2582 (pdf)
ISBN 978-9916-27-738-6 (pdf)

Copyright: Martin Oja, 2024

University of Tartu Press
www.tyk.ee

ACKNOWLEDGMENTS

My very special gratitude goes out to the people who are not very often remembered on such occasions – all anonymous reviewers of journal articles whose efforts bring them no visible glory but are absolutely essential in developing the minds of young researchers.

I am utterly thankful to my supervisor Peeter Torop who had every reason to be worried about my crazy ventures into farm keeping and neurosciences, but despite of that kept his unrelenting support that has always reflected politeness, honesty and the spirit of bygone ages – the chronotope of Juri Lotman.

The journey I undertook could not have been the same without the invisible presence of Kalevi Kull whose living example have taught me lessons not only about being a person of academy, but a Person in general.

I am grateful to Ulvi Urm for her kindness, Ene-Reet Soovik for ekphrasis and to Timo Maran for satyric mimicry, and my “roommates” Silvi Salupere and Elin Sütiste for their concern. I want to thank Kaie Koppel and Andreas Ventsel, who have helpfully reached out their hands during the expeditions through the jungle of Applied Semiotics, and my colleagues Katre, Ott, Eleni, Lona, Silver and Tiit for their little but crucial pieces of wisdom that have helped to understand things better.

Partly, I want to dedicate this work to my dear Father whose loving and unrelenting support is always combined with brilliant expertise in hard sciences. At any moment, he happily wants to lend his keen analytical mind and his trickster pranks.

On this journey, a lot of my thoughts were dedicated to the memory of my Mother who was always ready to give everything to her own PhD students. I remember her saying many times the same that Juri Lotman has said: “In the University – the most important persons are the students.” She never lost belief in me, keeping her dream of me “graduating” and becoming one of the torchbearers to the last moments of her life.

Last but not least, I want thank my friends for the inspiration, and my adversaries for the challenges. Sometimes it will take years to realize how many reasons you have to be proud of. Perhaps the greatest of those is the opportunity of spending time in the middle of people and traditions of Tartu.

CONTENTS

LIST OF ORIGINAL PUBLICATIONS	8
0. INTRODUCTION.....	9
0.1 Multimodal objects	9
0.2 The research questions and the structure of the dissertation.....	15
0.2.1 The aims and topics of the research articles	15
0.2.2 The structure and the focus of the cover chapter	21
1. MULTIMODALITY, MULTIMEDIA AND INTERSEMIOTICS IN THE CONTEXT OF TRANSLATION: IS MULTIMODAL MEANING- MAKING MEDIATED BY AMODAL INVARIANTS?	23
1.1 Multimodality and intersemiotic translation.....	23
1.2 Multimodality and multimedia studies	29
1.3 Crossmodal integration and amodal invariants.....	32
1.3.1 Paradigm change in neurosciences	32
1.3.2 Multimodal / crossmodal integration	35
1.3.3 Amodal invariants in the context of integrated experiences and transmodal modeling systems	39
2. SEMIOSIS AS INCOMPATIBILITY: THE FUNCTION OF CONFLICTS AND THEIR SPATIAL MODELING.....	44
2.1 Habits and deautomatization. From established codes to multiple- choice systems	44
2.2 Audiovisual conflicts: practices and perspectives	50
2.2.1 One time and two spaces: an artistic model for multimodal conflict	50
2.2.2 Inducing humor by multimodal conflicts.....	52
2.2.3 Multimodal efficiency: avoiding blunders and annoyances.....	54
2.3 Space as modelling system for evaluating multimodal conflicts.....	62
3. CONTRAPUNTAL SOUND AND INTERNAL DIEGETIC SOUND (IDS): VARIOUS USES IN CINEMA	69
3.1 Buñuel's "L'Age D'Or": From contrapuntal to internal diegetic sound.....	69
3.2 Olivier's "Hamlet": IDS marking narrative's turning point	71
3.3 Truffaut and Godard: "Shoot the Piano Player" and "Bande à part"..	72
3.3.1 "Piano Player": a sign consisting of signs	73
3.3.2 "Bande à part": a minute of silence	76
3.4 A Clockwork Orange.....	78
3.4.1 A multimodal translation	78
3.4.2 Singing in the Rain.....	79
3.4.3 Beethoven's 9th	80

3.5 Kieślowski and Scorsese.....	81
3.5.1 To whom the notes belong to?	82
3.5.2 Bringing out the borderline.....	83
3.6 IDS and contrapuntal music in the context of the two-layered framework.....	84
CONCLUSION AND FUTURE DIRECTIONS	86
REFERENCES.....	91
SUMMARY IN ESTONIAN.....	102
PUBLICATIONS	109
CURRICULUM VITAE	211
ELULOOKIRJELDUS.....	214

LIST OF ORIGINAL PUBLICATIONS

- I Oja, Martin 2023. Semiotic mode and sensory modality in multimodal semiotics: Recognizing difference and building complementarity between the terms. *Sign Systems Studies* 51 (3/4), 604–637.
<https://doi.org/10.12697/SSS.2023.51.3-4.05>
- II Oja, Martin 2024. Conflicting modalities in feature film: From contrapuntal editing to internal diegetic sound. *Semiotica* 259, 95–126.
<https://doi.org/10.1515/sem-2023-0068>
- III Oja, Martin 2019. On the concept of the deceptive trailer: Trailer as paratext and multimodal model of film. *Sign Systems Studies* 47 (1/2), 177–204. <https://doi.org/10.12697/SSS.2019.47.1-2.07>

0. INTRODUCTION

0.1 Multimodal objects

“Could you give me an example of an accident that happened when different sensory modalities provided conflicting data?” I asked the ChatGPT. The machine took no time to answer. “Yes, accidents can occur when different sensory modalities provide conflicting data, leading to misinterpretations and incorrect responses. On June 1, 2009, Air France Flight 447 was flying from Rio de Janeiro to Paris. The Airbus A330-203 aircraft crashed into the Atlantic Ocean, resulting in the deaths of all 228 passengers and crew on board.”

Recalling an article I had read about the accident, I searched for additional sources, wishing to get the relevant details right. The cause of the crash remained mystery until the flight recorders or “black boxes” were recovered almost two years later. The troubles for the Airbus started when entering a thundercloud, ice crystals clogged plane’s Pitot tubes, the devices that measure air flow and indicate the speed of the aircraft. First, when the reading of the airspeed suddenly fell from 275 to 60 knots, the autopilot disconnected and shut down most of the built-in protections. Due to the false reading of airspeed, altimeter corrections were automatically but incorrectly calculated (see e.g. Palmer 2013). It was 2.10 a.m. and the total darkness prevented the pilots from visually confirming the distance from the ocean. The pilots mistakenly believed that the plane was nearing the surface and forcefully “pulled up”.

Due to the sharp angle of climb, the plane rapidly gained altitude but lost speed at the same time, eventually stalling and starting to fall back. While losing altitude, pilots still applied “nose up” inputs. When they eventually realized the need for assuming “nose down” position for restoring steerability, it was too late and the plane crashed into the ocean (Salmon et al. 2015: 6). During the ascend, the stall warning system had been repeatedly activated, sounding an audible alarm. Still, the physical sensation of the aircraft’s nose-up attitude (tactile and visual feedback) did not align with the expected behavior or the instruments’ stall warnings (auditory feedback). The pilots received conflicting auditory signals (stall warnings) and unreliable visual data (inconsistent airspeed) while dealing with a loss of external visual references (night flight over the ocean).

As Salmon and his colleagues suggest, the error should not be attributed to the pilots alone, but to the whole sociotechnical system where situational awareness can be seen as an emergent property (Salmon et al. 2015: 3). Situational awareness, in terms of semiotics, can be seen as a type of meaning that relies upon different factors and agencies, as well as contextual influences. In the dramatic semiosis of Flight 447, the interpretation of various cues provided through multiple channels and sensory modalities, was an act of meaning-making evidently more complex than decoding a message in a single language. Semiotic analysis for complex situations or systems (see Engelbrecht 2021, Rickberg 2023)

calls for a multimodal approach, focusing on the synthetic meaning-making where different sign systems or languages combine and mutually influence the interpretation.

Building theoretical frameworks for analysing complex or so-called poly-semiotic activities (see e.g. Louhema et al. 2019) benefits from simplified models where variables or parameters appear somewhat reduced and stable. In comparison to real-life events, such simplified models are provided by artistic texts, first of all because of their relative delimitation: the borders of those texts (compared to socio- or biotexts) are ostensibly demarcated already before the approach of a researcher. Furthermore, artistic texts are more distinctly rooted in a certain medium or multiple media, allowing us to better distinguish emerging codes from the established ones. All this stabilizes the much-vaunted “experimental conditions”, although the context of humanities advises us to take the phrase at least partly metaphorically. To sum it up, it should be noticed that the problems discussed in this dissertation, in my view, do not belong specifically to the fields of cultural-, social- or biosemiotics, but ask for an integrated understanding of all those subdisciplines of semiotics. Such a requirement is shaped by the characteristics of the research object: the interpretation of multimodal conflict is, on the one hand, influenced by the socially shared practices and on the other, by psycho-biological procedures of perception and cognitive processing.

While I will soon return to the development of semiotics from monomodal towards multimodal objects, the notion of “multimodal” calls for a further clarification by the means of theoretical background and some simple examples. Linguist Ellen Fricke has considered, in my view, one of the most fundamental distinctions in the framework. She identifies two types of multimodalities, broad and narrow one. According to her, the broad multimodality entails multiple codes in a single ‘sense modality’, the narrow multimodality employs more than one ‘sense modalities’ (Fricke 2013: 740)¹. The allocation of the adjectives ‘broad’ and ‘narrow’ may appear counterintuitive at the first glance, because the ‘narrow’ seems to embed the ‘broad’. Nevertheless, taking into account the scope of application of those two multimodalities, there are more semiotic situations where only the rules of broad multimodality apply – i.e. the meaning-making relies upon the combination between elements that are ‘carried’ by a single sense or sensory modality. In other words, ‘narrow’ multimodality events (multiple sensory modalities or channels) usually entail the dynamics of ‘broad’ multimodality (single channel but various sign systems), but not vice versa. Following Fricke’s distinction, this dissertation underlines two fundamental levels of description that are not exclusive but mutually compatible; both are needed for a comprehensive description of multimodal processes.

Most commonplace examples of multimodal cultural texts in the broad sense (employing a single sensory modality but multiple sign systems) are encountered

¹ An illustrative comparison can be drawn from the discussions of multimodal animal communication. Correspondingly, the distinction is made between multi-component single-channel signals and the co-occurrence of multiple channels (Partan 2004: 225).

in the compositions that rely on the relationship of words and pictures. While images are better for conveying spatial arrangements, (verbal) language is used for other things, for instance telling stories that have a reach in temporal dimension. Images and words as two semiotic modes “are given complementary tasks, just like the photographer and the writer” (Kress, Van Leeuwen 2001: 18). A prominent contemporary approach addresses such relations with the help of a document- and layout-centered framework (see Bateman et al. 2017, Wildfeuer 2014) that applies not only to static semiotic objects as printed pages featuring graphs, diagrams and verbal texts, but also to dynamic texts as social interactions and audiovisual media.

Compared to the theoretical advances, multimodal cultural texts are substantially older. Singing and storytelling, perhaps the oldest forms of expression, are evidently multimodal, involving various verbal and bodily components that constitute whole meanings. As for recorded texts, the papyrus of scribe Ani from Egypt, 1300 BC, depicting the god Anubis weighing a dead man’s heart, exemplifies the images with a written prayer: “My heart of my ever-changing form /.../ Do not turn against me before the master of scales” (Hagen, Hagen 2005: 11). The Bayeux Tapestry, half a meter tall and 70 meters long medieval comic strip, tells the story of William the Conqueror’s victory at Hastings in 1066 (Hagen, Hagen 2005: 20, 24). Intermittently above the images, capital letters specify what event is represented in particular places. “hAROL D REX INTERFEC TUS EST” over a wounded knight on a horseback informs the viewer about the poor fate of English King Harold.

In contrast, modern comic books tend to convey the identity of characters by visual details that might have been somewhat impossible in the medium of woven carpet. Now, verbality, first of all, strives to represent the dialogue. Still, an additional function or purpose emerges: when looking at Batman punching his opponent “POW”, “SOCK”, “BONK” or “KRAKK”, the power of onomatopoeia makes us suspect a mild case of transmodality: one sensory modality may begin activating another. In Batman’s punching, visual modality (that carries or entails the mode of verbal text) has potential to activate auditory modality to some extent (if we consider inner speech and activated brain areas that process auditory information as sufficient criteria).

Potentially, any multimodal component may rely upon multiple codes, or as Umberto Eco has said, “the reader of the text knows that every sentence and every trope is ‘open’ to a multiplicity of meanings which he must hunt for and find” (Eco 2007[1981]: 51). One code may occur harder to decipher than another. Acronyms that work as labels for some visual representation offer interesting examples. Perhaps most of the people who have ever entered a building of Roman Catholic or Protestant Church are familiar with the letters INRI above the head of crucifixes. While its dominant meaning is “Iesus Nazarenus, Rex Iudaeorum”, it also has some alternative, resistant readings, for

instance “Igne Natura Renovatur Integra”². Here we see how a shift in one component brings along the semiotic transformation in other, transferring the figure of Jesus from one discourse into another, even more symbolic one.

Concerning various types of texts, the question of dominance is preeminent. All the previous examples addressed the compositions where image or depiction was somewhat dominant and verblivity had a specifying role. In illustrated fiction books, the relationship of dominance is usually reversed: images are subordinated to narrative. There, *images* modify the result of *imagination* that has been built by the means of words. The development of the story organizes the placement of illustrations that aim to provide answers to such questions as “what did captain Nemo’s submarine *Nautilus* look like?”, or, “what kind of clothes were worn by the three musketeers?”, if we recall, for instance, the novels of Jules Verne and Alexandre Dumas.

In graphic design, the relationship of dominance often gets negotiated, especially when different components or modes work together to provide emergent meanings. In 1977, Milton Glaser designed the famous “I Love New York” logo that reportedly earns the state \$30 million each year (Müller, Wiedemann 2018: 103, 184)³. The red pictogram of a heart has the same height and similar rounded shapes as the three black letters, communicating the equality and cooperation between sign systems.

Likewise, most album covers attempt to associate the informational content with emotional connotations. The names of artist or band allow the swift identification, resembling brand names on consumer goods packages. Despite this, some of the most outstanding album covers have averted this principle, leaving out verbal modes and keeping only images. Peter Saville, designing the cover for Joy Division’s “Unknown pleasures” (1979), used the visual trace of the first recorded signal of radio pulsar – white waveforms on a black backdrop. He considers the pattern both technical and sensual; simultaneously tight and fluid, resembling a heartbeat (see Grundy 2011). Here, the absence becomes significant and casts its influence (see also Deacon 2013); in order to enter the semiosis, an absent mode must have a *position* for it in the whole.

With pop songs we do not yet enter the realm of Fricke’s narrow multimodality. While all the previous examples used various modes in visual modality, i.e. the meaning-making relying upon visual channel and corresponding processing patterns, the semiotic modes of musical text (see about the notion in Tagg 2013: 115) belong to auditory modality. A multimodal analysis of a song may distinguish as its semiotic modes the melody and the rhythm, the voice characteristics as timbre, and also verbal components. In order for a pop song to become multimodal in the narrow sense (that is, combining different sensory

² It was a Rociucean interpretation that associated INRI with the initials of an Hermetic (that is, alchemical) principle “By fire, nature renews itself” or, “By fire, nature is perfectly renewed” (Laos 2023: 338).

³ How typography is employed in multimodal meaning-making is discussed e.g. in Van Leeuwen 2006 and Norgaard 2009.

modalities), a listener has to find the lyrics of the song and start to read them while listening, or just watch the music video.

Illustrating the conflict between semiotic modes I discussed René Magritte's renown "The Treachery of Images" (1929) in two of the articles of this dissertation. Here I would like to shortly refer to Magritte's somewhat similar work with the title "The key to dreams" (1930). In the six-segment panel the Belgian artist depicts six objects with undertitles that do not correspond to the images, e.g. the image of a men's hat is paired with the undertitle *la Neige* (the Snow). As semiotician Louis Hébert, analysing the same piece, seeks the iterations of motifs or repeated meanings (calling them *semic molecules*), he differentiates the possibility of repetitions on the intrasemiotic (inside a single mode) and intersemiotic level (Hébert 2019: 160, 162). Still, focusing on possible connections between the elements is just one feasible approach. Some authors have emphasized Magritte's wish to engage in *metapainting*, that is, enjoying a semiotic game where signification is subverted and the relationship between word, image and object is experimented with (see Foucault 1973, Lipinski 2019)⁴.

According to Lisa Lipinski, between 1927 and 1931 Magritte painted altogether forty-two paintings where words appear. The period is characterized by Magritte's relationship with various writers, engaging in discussions over language problems, and artistic experiments in the medium of collage. The period is further exemplified by Magritte's essay "Les mots et les images" (The Words and images, 1929), where the painter acknowledges a flexible relation between the object and its name (Lipinski 2019: 37). While Lipinski interprets Magritte's experiments as playful and humorous, Marcus Silverman looks upon his multimodal disruptions in a darker context. Silverman believes that for Magritte, the reality was inherently cracked and meaningless because of his childhood trauma. Supposedly, Magritte took his mother's suicide by drowning as a total abandonment and therefore, his paintings "became the exercises of denying reality" (Silverman 2012: 79). It must also be noticed that Magritte "outspokenly condemned Freud and all psychoanalytic thought" (Silverman 2012: 79), so his disrupting "interpretation of dreams" can be understood as an irony.

As we later see when discussing Stanley Kubrick's "A Clockwork Orange" among other films in the third subchapter of this cover article, cueing ironical interpretations is often a goal of multimodal conflict, considering both semiotic modes and sensory modalities. With turning to film, we enter the sphere of Fricke's narrow multimodality. Whereas the majority of semiotic works on multimodality proceed from Michael Halliday's sociolinguistics and its development in Gunther Kress's sociosemiotics, considering *semiotic mode* as their pivotal concept, this dissertation will focus more on the "narrow" multimodality. First, it means that the objects of the discussion belong to media that have multiple

⁴ In the words of Lipinski, "although Magritte had not read Saussure's *Cours de linguistique générale* (Course in General Linguistics), published in 1916, he illustrated a central tenet of structuralism, namely the arbitrary nature of the sign and the artifice of the link between form (signifier) and content (signified)" (Lipinski 2019: 53).

channels or apply to multiple senses, and film is, in that sense, an exemplary medium where the relationship between visual and auditory modality can be modeled in a relatively straightforward manner. Second, this means that from the viewpoint of this dissertation, the organization principles of multimodal texts are firstly analysed regarding sensory modalities, and semiotic modes are taken into account in second order.

In filmic expression, there are various organized forms of conflicts between image and sound. The most eminent theoretical background for such practices can be found in Sergei Eisenstein's writings about contrapuntal editing, articulating the mechanisms for the synthesis in complex, multi-linguaged sign systems. A general typology would make a distinction between conflicts that concern dialogue or spoken word, and film music. An established notion in film theory is *internal diegetic sound* (IDS): the character appears "visually silent" but talks in voiceover, as though the sound would be coming "from inside the mind of the character" (Bordwell et al. 2019: 291). Usually, it expresses character's thoughts as they happen in the time and place of the diegesis (Horton 2017: 195). In the article "Conflicting modalities in feature film", I discussed a scene from Laurence Olivier's "Hamlet" (1948), where IDS is employed as a narrational device, marking an important turning point in hero's progress. There, the disruption between auditory and visual modalities correlates with Hamlet's indecision, conveying a liminal, transformative phase of the journey.

Incongruent film music offers an example of multimodal conflict where visual component is mismatching with some part of film score. As David Ireland observes, "the lack of shared audiovisual properties in an incongruent pairing can challenge the automatic perceptual encoding that more congruent constructions tend to invite" (Ireland 2018: 13). Thereupon, multimodal conflicts can also be seen in the light of Russian formalists's idea of defamiliarization or *ostranenie*. The device, as was formulated by Viktor Shklovsky, was fulfilling one of the most important functions of art, heightening perception and breaking automated responses in viewer or reader (see e.g. Stam et al. 1992: 10).

Film critic Laurent Jullier has contemplated that contemporary film theory is shaped by two contrasting principles: ecological and constructivist one. Defamiliarization, in his opinion, appears as an important locus for negotiating between those principles (Jullier 2010: 139). Here, environmental psychology has also its implications on the film viewing. While sensing the surroundings, the inputs from different modalities are integrated into consistent perceptual experiences (Pennartz 2009: 720). Therefore the mutually supporting type of modality relation is prevalent for evolutionary reasons and the use of conflicting modalities relation has its own special purposes as, for instance, the forementioned defamiliarization. As I have indicated in one of my earlier articles (Oja 2014) that discussed art film, the devices that break narrative flow or embodied simulation, have a potential to induce a subjective, more active reception mode in viewers.

Besides irony, narrative punctuation and poetic estrangement, multimodal conflicts can draw attention to the "mechanics" or the construction principles of the text. Although Roman Jakobson's language functions were applied in the

realm of verbal messages, with some reservations they can be called upon in discussing audiovisual media. Thus, multimodal disruptions can fulfill Jakobson's poetic and metalingual functions that are oriented towards the message and the code⁵ (see Jakobson 1985[1956]: 115). In this regard, Jean-Luc Godard's experiments in the 1960ies were exemplary. In "Bande à part" (1964), a character proposes: "If there is nothing to say, lets have a minute of silence". Subsequently, the whole soundtrack switches off. Not just the characters keep silent, but the ambient sounds in the environment are muted: the auditory modality is totally excluded. As a result, a quite different reception regime is emerging. It doesn't only call attention to visual details (that are highlighted by the absence of audio) but to the textual whole: the integrity of the storyworld is punctured and the audience is invited to reflect on the new narrative rules that have just been manifested.

The development of semiotics in the 20th century can be broadly seen as a dynamics between theories of signs and theories of sign systems. In my view, one of the most central problems of the 21st century, for semiotics and for the whole range of sociocultural disciplines, is the problem of synthesis between multiple sign systems. In actual semiosis, the presence of Umwelt and Innenwelt already suggest some default multimodality: more often than not, the meaning is negotiated, and dependent on various sources. On the one hand, a further integration between neurosciences and humanities will certainly shed more light on the exceptionally interesting question of amodal invariants – a common ground for the synthesis. On the level of the brain, it is supposed that modality-specific input can be "translated" by the means of "multimodal" neurons (see Stein, Meredith 1993). On the other hand, i.e. what has to do with cultural texts, I propose that the multimodal processes, especially modality conflicts can be approached in the light of Juri Lotman's idea of two incompatible, but still complementary languages that comprise a mechanism for generating novel information.

0.2 The research questions and the structure of the dissertation

0.2.1 The aims and topics of the research articles

This dissertation aspires to offer its contribution to the development of the semiotic theory of multimodality. Primarily, it will address the central concepts of the theory – *mode* and *modality*. By various authors, these terms are used interchangeably and without sufficient specification of their meanings. The

⁵ The term "message" has less use in contemporary semiotics; it originates from the structuralist, cybernetics-influenced paradigm where the problems of coding and decoding had a central position. While the "message" tends to connote a certain "correct" result, the "text" presupposes more freedom on the side of interpretation.

situation that is portrayed as “messy character of the field” (Stöckl 2014: 275) at least partly stems from the intermingling of two essential perspectives. The question is, what should be considered the elements of a multimodal whole or composition, and consequently, how should multimodal theory segment its object? As multimedia scholar Lars Elleström puts it, “in [...] media studies and linguistics, ‘multimodality’ sometimes refers to the combination of text, image and sound, and sometimes to the combination of sense faculties (the auditory, the visual, the tactile and so forth)” (Elleström 2020: 41). The attempts of coordinating between those perspectives have often led to various hybrid categories that contain elements of both types, e.g. treating verbal and visual both as modes (e.g. Forceville 2020), but in my view, such attempts result in classification fallacies. Therefore, following Elleström’s delineation, I propose that those two systematizations should be knowingly distinguished. As a result, the position of the basic terms will instantly clarify. *Mode* will apply to the first combination and designate *semiotic mode*. *Modality*, in turn, will refer to *sensory modality*.

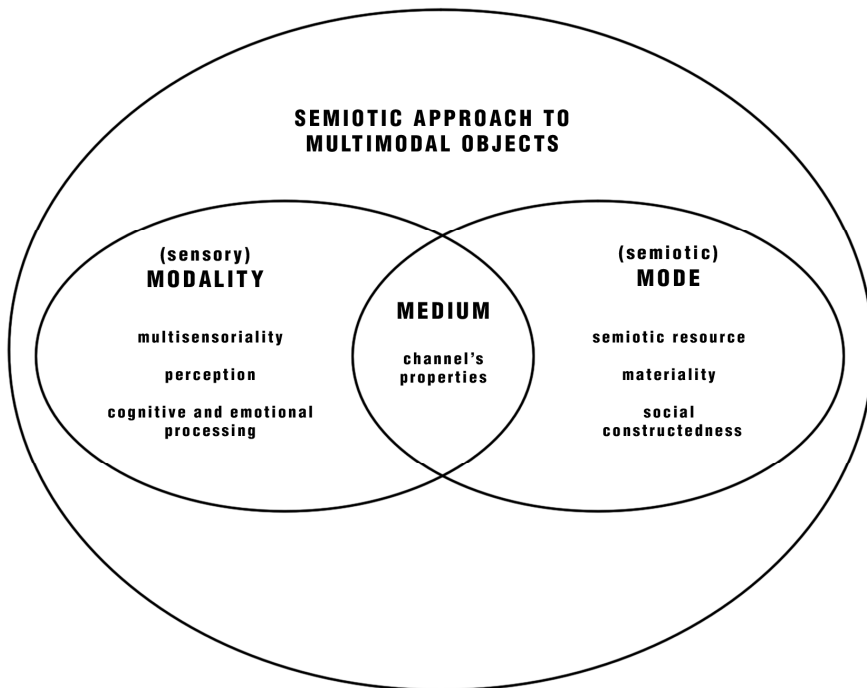


Figure 1. Modality-oriented approach is based on multisensoriality and corresponds to Fricke’s *narrow* multimodality. Mode-oriented approach is concerned with semiotic resources and corresponds to Fricke’s *broad* multimodality. This dissertation proposes an integrated framework where these two approaches are complementary.

Although the *mode*-approach and the *modality*-approach have encountered above-mentioned issues of compatibility, this dissertation forwards the standpoint that they are not mutually exclusive. Moreover, the research of multimodal objects will benefit if the arrangement of modes and the arrangement of modalities are joined in a single framework as complementary systems. Such an integration would presuppose a task allocation: the system of modes can be seen as concerned with textual combinations of signs and the semiotic potential of the materiality (see Bateman 2016: 46) that allows various objects function as affordances (see Gibson 2015[1979]: 119, Kress 2010: 157); the system of modalities addresses the side of meaning-making that is dependent on reception, cognitive-emotional processing and subjectivity⁶.

Although most of the time it is not advisable to proceed from such oppositions as “natural vs culturally constructed”, there is an intrinsic difference between modes and modalities. Sensory modalities are strictly limited and subordinated to biological organization, being “classically distinguished based on the type of physical stimulation that they are most sensitive to: light for vision, sound for hearing, skin pressure for touch, molecules in air for smell, etc” (Bertelson, De Gelder 2004: 141). Modes, by contrast, are influenced by their historical and socio-cultural background; therefore they vary across cultures (Kress 2010: 130). Moreover, what is a mode will significantly depend on the subject’s “microculture”, i.e. her personal background, education and competence with various media. The problem of a subjective segmentation and the multiplicity of possible submodes is eloquently articulated by Carey Jewitt and her colleagues:

Gesture and gaze, image and writing seem plausible candidates, but what about color or layout? And is photography a separate mode? What about facial expression and body posture? Are action and movement modes? You will find different answers to these questions not only between different research publications but also within. To avoid potential confusion, it is important to make a deliberate decision on what categories and terms to use when engaging with multimodal research. (Jewitt et al. 2016: 12)

To overcome such problems in the research of multimodality, this dissertation proposes a two-level framework that combines the constancy of modalities and the flexibility of modes. The organization of sensory modalities can be seen as a foundational layer under the system of modes; any mode is always carried by one, or possibly even by multiple sensory modalities. For instance, the verbal or

⁶ Let us notice, as a parallel, that Jordan Zlatev, in the context of cognitive semiotics, speaks about ‘monosemiotic’ vs ‘polysemiotic’ communication (employing single and multiple sign systems). Zlatev’s categorization is somewhat an alternative to the mode- or modality centered views. His basic ‘unit’ is ‘sign system’ in a rather general sense. Zlatev classifies sensory modalities (calling these ‘perception modalities’ as properties of sign systems. For instance, Zlatev indicates that actual human communication, especially in face-to-face contexts, is as a rule polysemiotic (Zlatev, Möttönen 2022: 285).

textual mode in film can be simultaneously advanced as sound in auditory modality, and in the form of subtitles that employ visual modality. The reverse dynamics is even more frequent: usually several modes work inside the boundaries of one modality. Again, in film medium, the auditory modality may carry the “core” modes of dialogue, music and sound effects (with submodes as foley art, digital sound effects etc); the visual modality has even much broader range of possible modal components.

The second objective of the dissertation is to construct an explorative typology of modality relations that also takes account of modes. Whilst the typologies of modes (see e.g. Stöckl 2004) fulfill a prominent task to specify what modes there generally are in a specific sign system, and how can the intermodal borders be drawn (i.e. where one mode ends and the other begins, or what mode is more fundamental and what, in turn, can be called a submode), the division of sensory modalities is inflexibly organized and does not have to address those problems. Accordingly, this categorization is not the typology of modalities but modality relations. As such, it attempts to envision a preliminary “grammar” of synthesis or combinatory rules of sensory modalities. In the typology, I have distinguished between *supporting*, *modifying*, *conflicting*, *substituting* and *cross-activating* relations. Under each category, I aimed to provide a short explanation about the dynamics of the relation, and discuss some corresponding examples.

The overall orientation of this dissertation follows the progress from general to particular. Moving from the typology of modality relations to a more detailed level, I will focus on a single specific relation type, namely *conflicting modalities*. As is my subjective evaluation, this appears to be an extremely interesting and revealing category. Conflicting modalities correspond to several semiotic strategies and purposes, e.g. the artistic devices that defamiliarize or self-reflexively indicate the constructedness of the text. Also, conflicting modalities are occasionally employed in deceptive meaning-making both in human social interactions (see e.g. Docan-Morgan 2019) and animal communication (see Howse, Allen 1994; Partan 2004), where contradictory signals form the basis of the strategy called satyric mimicry (Howse, Allen 1994: 111). This dissertation mainly concentrates on the multimodal conflict in audiovisual media, especially in film.

With the purpose of building a more comprehensive treatment, in the article of “Conflicting modalities in feature film”, I examined the idea of synthetic meaning-making in the intersection of semiotics and film theory, starting from Sergei Eisenstein’s concepts of *contrapuntal editing* and the *montage of attractions*, followed by Roland Barthes’s concept of the *third meaning* and Roman Jakobson’s *syncretic messages*. Concerning Jakobson, the idea of dominant is estimated as crucial when discussing multi-component or multi-language sign systems. The hypothesis on the domination of visual modality over auditory modality in audiovisual media was observed in the context of cognitive neuroscience that has demonstrated that the attention in different modalities is not mutually independent (Shams, Kim 2010: 272) and the reception of multimodal texts is more of a holistic process, where perception, cognition and bodily simu-

lation of action are closely integrated (see e.g. Gallese, Guerra 2015: 151). Last but not least, this section of the dissertation proposed that synthetic meaning-making and particularly multimodal conflicts can be examined in the light of Juri Lotman's idea of two partly incompatible languages that comprise a mechanism for creating new information. When Lotman denotes that "juxtaposed units that are incompatible in one system force the reader to construct an additional structure in which the incompatibility is eliminated" (Lotman 1977a: 283), he also touches upon the artistic mechanism that empowers the interpreter, inviting her towards the active participation in meaning-making.

Discussing multimodal conflicts, one cannot avoid an evident question that could even turn out *too* evident – it can hide itself in plain sight. What is actually a multimodal conflict? In an attempt to provide an answer, in the same article, I turned to the theoretical problem of detecting the presence of the conflict. Concerning visual-auditory relationship, I initially evaluated *redundancy* as a possible scale parameter (hypothesis: the conflict appears when redundancy is broken). However, it became clear that the level of redundancy is not a sufficient measure for the scope of multimodal conflict, because, in short, auditory and visual modality are not mutually redundant. The combination of image and sound always creates somewhat an emergent meaning (See Altman 1980; Chion 2016: 161). Next, *synchrony* was evaluated (hypothesis: the conflict appears when synchrony is broken) but also dismissed. Paradoxically, synchrony turns out to be a necessary premise for multimodal conflicts, bringing the components together and enabling the interpreter to take notice of their connection. Thirdly, I turned to the idea of *space* as a modeling system, that in Juri Lotman's semiotics, besides verbal language, is the fundamental device for artistic modeling (Lotman 1990a: 239). Space (see also Randviir 2010) functions as the precondition for any semiotic activity (Lotman 1990b: 123). As such, spatial modeling allows us to evaluate a great range of dynamic phenomena on the scale of proximity. It is viable to set the parameter of *distance* in spatial models into correspondence with semantic distance in multimodal conflicts.

The third and the most object-centered article of this dissertation ("On the concept of a deceptive trailer") focused on a single example of multimodal conflict, namely the deceptive film trailer. Film genre can be seen as a dominant type of information communicated by trailer (see Kernan 2004, Johnston 2009). So-called deceptive trailers tend to present misleading cues about genres; one mode or modality may refer to one genre, the other to another. The aim of such practice is to increase the amount of possible viewers, but the result can bring disappointment in viewers. The discussion of film trailers as paratexts (see Genette 1997, Dusi 2015b) indicates towards an important sphere of problems that concern the relation of meaning and textual borders. Here the notion of *transtextual multimodality*⁷ would be suitable to illustrate the situation where

⁷ There can be found no usage of the term in the literature yet, but despite that, implementing the notion of *transtextual multimodality* would accurately characterise the type of semiosis described above, as such contributing to transmedia studies.

the meaning of a text is influenced from its immediate surroundings; it turns out that some multimodal components should not necessarily belong into textual borders, but can have their impact on the level of the discourse.

To sum it up, the core of this dissertation incorporates three research articles: **“Semiotic mode and sensory modality in multimodal semiotics: making difference and building complementarity between the terms”**, **“Conflicting modalities in feature film: from contrapuntal editing to internal diegetic sound”** and **“On the concept of the deceptive trailer: trailer as paratext and multimodal model of film”**. The first article strived to fulfill the first and the second research task, aligning the terms *mode* and *modality*, and constructing the typology of modality relations. Behind the paper is the intention to build a proper theoretical foundation for addressing the main topic of the thesis.

The second article was written last of all. As such, it has the central position in the dissertation. Firstly, here are mapped some of the essential ideas that characterize the shift of semiotics’ focus towards the so-called complex or multimodal systems where the meaning emerges at the intersection of different types or semiotic resources or sign systems. In addition, the viewpoint of cognitive film theory is engaged, setting the multimodal meaning-making in the light of embodied simulation and the evolutionary perspective, illustrating why modality conflicts are rather exceptions than parts of “default” semiosis. Secondly, the prospect of evaluating a multimodal conflict is discussed, and as the topic is complicated one, only some provisional conclusions are outlined. Above all, the modality conflict achieves its artistic or rhetoric power only when it is a semantic conflict. In order to model semantic divergence between different modalities, the space as a universal modeling system is proposed. Finally, the functions of audiovisual conflict in feature films are discussed, mainly focusing on the incongruent film music and internal diegetic sound as devices in filmic expression.

The third article addressing film trailers as models of film was written the earliest. Following, once again, Jakobson’s concept of dominant, the article examines how film’s genre is communicated by film’s trailer. Comparing the dominants in the genre-rhetoric of film and its trailer, it is possible to conduct a closer analysis of misleading model-making and to predict possible disappointment in viewers. The multimodal construction of the trailers of Alejandro G. Iñárritu’s *Birdman* and Nicolas Winding Refn’s *Only God Forgives* is analysed. It has to be noticed that because of author’s youthful inexperience, the article bears the traces of exactly similar terminological inconsistencies that he criticized later (that is, in the two articles that follow “Deceptive trailer” in time but precede it in the organization of this dissertation). Although occasionally lacking sufficient clarity in the categorization of modes and modalities, considering the requirements of the dissertation, the article is not rewritten but presented in its unchanged form.

0.2.2 The structure and the focus of the cover chapter

In addition to the present introduction, the framing article consists of three chapters. Their goal is not to duplicate the content of the research articles, but offer a more detailed background to the topics discussed in there. Doing that, some specific issues related to conflicting modalities are examined. Those problems have been in author's mind while writing the research articles, but the limitations of the publication process and the concern for the works' structural clarity encouraged to postpone observing them.

The first chapter will complement the discussion about the synthetic meaning-making, or the semiotics' focus shift towards complex sign processes. The problem of demarcating the research units is highlighted in the context of inter-semiotic translation. The main question that is discussed is the possibility of translation between sensory modalities as different sign systems. In turn, this brings along the question about the common basis that can make this type of translation possible. I will propose that amodal invariants that are mediated by the pre-linguistic modelling system of neuronal connections can be seen as the basis for such translation. Also, the first chapter will touch upon the relation between multimodality and multimediality, clarifying the point of view that issues of media can largely be addressed in the frameworks of semiotic modes and sensory modalities, but not vice versa. Lastly, the first chapter explores amodal invariants in the context of crossmodal integration. The paradigm change in neurosciences emphasizes a more integrated network where different brain areas are in mutual "dialogue"; this has important implications on the theories of meaning-making. Consequently, semiotic theories of multimodality should also be considered on the background of the idea that cognitive processes are already inherently multimodal.

The second chapter has two main goals: firstly it addresses systems' incompatibility as the basis for semiosis, secondly, it weights the possibility of spatial modeling of multimodal conflicts. The dynamics of multimodal conflicts are also observed in the light of Viktor Shklovsky's concept of *ostranenie* and Juri Lotman's idea of explosive processes. The idea that semiosis happens in the force field of two opposite dynamics – habituation and deautomatization – has a central position. While one of the basic functions of multimodal conflict is to activate the interpreter, renewing the practices of perception and creating new information, eventually habituating forces deter the explosive potential of the conflict and the novel meanings are established in codes. A good example of such a process is internal diegetic sound (IDS) that presents modality conflict only technically, otherwise it has become a conventional device of filmic expression. As IDS and contrapuntal editing are substantially observed in the article of "Conflicting modalities in feature film" and also in the third chapter of the cover article, I will, in this place, briefly turn to some alternative examples of the use of multimodal conflict in audiovisual media. Some practices and

perspectives concerning subtitles and the retrospective scoring of silent films are addressed, as well as the mechanism behind creating humor with the help of multimodal conflict.

Discussing spatial modeling of conflicts is supported by the findings of neurosciences – while processing various types of information in the brain, overcoming neuronal distances can be seen in correlation with the levels of energy expenditure; thus, a part of spatial modeling is already done on the level of brain architecture. Here, the ‘small-world’ type networks are crucial: at least metaphorically, the two connection types (abundant, short-distance and cost-effective *vs* infrequent, long-distance and costly) reflect two complementary semiotic strategies: automatizing for energy saving and inducing conflicts for creating novel information and refashioning systems themselves.

The third chapter functions more or less as an appendix, presenting several analyses of multimodal conflicts in feature films, organized by the chronological principle. For the start, the chapter looks upon Luis Buñuel’s scandalous “L’Age D’Or” (1930) as an early example of sound film (here the new medium enabled one of the first experiments with audiovisual conflict); secondly, the use of internal diegetic sound in Laurence Olivier’s “Hamlet” (1948) is discussed more comprehensively than in the article “Conflicting modalities in feature film”. An experimental use of modality conflicts is observed in the films of French New Wave authors, namely François Truffaut’s “Shoot the Piano Player” (1960) and Jean-Luc Godard’s “Bande à part” (1964). The next stop will be at Stanley Kubrick’s “A Clockwork Orange” (1971) where audiovisual conflict creates ironical comments. Lastly, Krzysztof Kiesłowski’s “Three Colours: Blue” (1993) and Martin Scorsese’s “Bringing Out the Dead” (1999) are visited in order to illustrate some delicate narrative functions that are achieved by the use of multimodal conflicts.

1. MULTIMODALITY, MULTIMEDIA AND INTERSEMIOTICS IN THE CONTEXT OF TRANSLATION: IS MULTIMODAL MEANING-MAKING MEDIATED BY AMODAL INVARIANTS?

1.1 Multimodality and intersemiotic translation

In 1906, Charles Peirce wrote a passage that anticipated Google Translate and other machine translation services, those that possibly operate with signs⁸ but without any need to engage imagination or interpretants.

What does it mean to speak of the “interpretation” of a sign? Interpretation is merely another word for translation; and if we had the necessary machinery to do it, which we perhaps never shall have, but which is quite conceivable, an English book might be translated into French or German without the interposition of a translation into the imaginary signs of human thought. (Peirce 1998 [1906]: 388)

However, from the viewpoint of this dissertation, more interesting than this farseeing prediction is Peirce’s likening of interpretation to translation. Dinda Gorfée explains that

Peirce characteristically used “translation” and related terms in a very broad sense. For him, translation is the same as sign interpretation. /.../ In the infinite process which Peirce commonly called semiosis, signs are being translated, or interpreted. The truth of reality which is the ultimate goal of semiosis, is approached through what is an infinite series of translations, or interpretant-signs, one following the other and elaborating on it. (Gorfée 1994: 153)

In transmission of Peirce’s ideas towards contemporary semiotics no one’s work has been as essential as Roman Jakobson’s. Nevertheless, his reading of Peirce is influenced by his structuralist and linguistic background; according to Gorfée “Jakobson used Peirce’s general theory of signs in special reference to one subdiscipline of linguistics, namely verbal arts” (Gorfée 1994: 147). Jakobson’s essay “On Linguistic Aspects of Translation” (1959), a pivotal writing for semiotic translation studies, introduces the typology of *intralingual*, *interlingual* and *intersemiotic* translations. Doing that, it widens the perspective on the concept of translation (see Sütiste, Torop 2007), incorporating non-verbal signs as the objects of translation studies. Still, it has to be noted, Jakobson’s position remains rooted in verbality and the intersemiotic translation or *transmutation* (that concerns multimodal processes the most) is envisioned as unidirectional.

⁸ Is the machine semiosis possible without human agents, is the question that is not addressed here. When Jesper Hoffmeyer asserted that life “is based entirely on semiosis, on sign operations” (Hoffmeyer 1993: 24) and “...fundamentally grounded in semiotic processes” (Hoffmeyer 2008: 3), he saw semiosis as the precondition for life, but not necessarily vice versa.

Jakobson's claim that "the meaning of any linguistic sign is its translation into some further, alternative sign, especially a sign "in which it is more fully developed", as Peirce, the deepest inquirer into the essence of signs, insistently stated," (Jakobson 1971[1959]: 261) is critically commented by Umberto Eco. Eco affirms that it would be tempting to identify the totality of semiosis with a continuous process of translation (Eco 2001: 68), but "the universe of interpretations is vaster than that of translation proper" (Eco 2001: 73). Jakobson, in this question, follows Peirce's vocabulary that is "notoriously protean and often impressionistic" (Eco 2001: 69) and as Eco suggests, leaves the question of equality between translation and interpretation somewhat unspecified (Eco 2001: 71). Discussing various post-Jakobsonian approaches, Eco concludes that compared to interpretation, there are limits to translation and they appear with "diversity in the purport of expression" (Eco 2001: 73) – that means, when a would-be "translator" encounters parallel meanings or "open-ended semantics" in the source text.

To sum it up, Jakobson's stance remains somewhat controversial: on the one hand, any interpretation can be counted as translation on the basis of Peirce's infinite semiosis; on the other, the category of intersemiotic translations⁹ still remains a one-way avenue and contains only non-written adaptations of literary works. Thus, such examples as film remake or the "translation" of classic music piece into a more popular form do not count here. Moreover, the claim that "any human communication of non-verbal messages presupposes a circuit of verbal messages, without a reverse implication" (Jakobson 1970: 425) can be considered obsolete or even evidently misleading in the context of "Sebeokian turn" that convincingly established the evolutionary primacy of non-linguistic communication.

An obvious shortcut to conclude the discussion is to apply to the level of generality of the notion of *translation*, maintaining that translation should not necessarily entail linguistic component. As Susan Petrilli sees it,

in the light of today's semiotic or biosemiotic or, better, global semiotic perspective (Sebeok 2001), it is now obvious that translation does not only concern the human world, anthroposemiosis, but emerges far more extensively as a constitutive modality of semiosis or, more exactly, biosemiosis in general. Translational processes pervade the living world, the great biosphere in its entirety. (Petrilli 2014: 255)

Apart from the biosphere, *culture text* as a general notion that does not necessarily include linguistic foundation, has been increasingly associated with translative

⁹ It is important to notice that Jakobson uses the notions of *transformation*, *transmutation* and *transposition* in parallel with *translation*. As Elin Sütiste explains, sometimes *transposition* appears synonymously with translation in a narrower sense, but sometimes Jakobson implies with transposition that "the change of medium in the process of transformation brings about a different set of rules that govern the setup and functioning of the target text" (Sütiste 2021: 176).

processes. While the studies that address the translatability of nonverbal sign systems with other nonverbal systems are still not frequent (Kourdis 2015: 317), it can be concluded that in the last decades, Jakobson's original concept of intersemiotic translation has been expanded and converted into a two-way model, including transmutations e.g. between cinema and theatre or painting and cinema (see Dusi 2015a for a concise overview of terminology). Such bidirectionality also means that *ekphrasis*, that was originally understood as "verbal representation of visual representation" (Mitchell 1995: 152) but as a notion, later expanded to designate verbalizations of texts in any non-verbal sign systems (see Sager Eidt 2008: 18, 19), can be approached as intersemiotic translation (see also Aktulum 2017).

As Kay O'Halloran and her colleagues indicate, intersemiotic translation forms the basis of cultural communication, where the notion of grammar, at least metaphorically, applies to the organization of semiotic resources as the units of multimodality (O'Halloran et al. 2016). This comprehensive view is commensurate with the paradigm that sees the whole culture as *total translation*, according to the concept that was introduced by Peeter Torop in 1995. A central idea to total translation is that textual units of different levels and sizes can be seen as mutually translatable – i.e. conceiving a title for a story or writing an abstract for an article can be seen as an act of translation. Total translation also enables us drawing parallels between Peirce's and Lotman's totalizing concepts of sign processes (Torop 2003: 272). Thus, "intersemiotic translation is a complex "form of action," not a simple transcodification but a transcultural, dynamic and functional event" (Dusi 2015a: 183).

From the viewpoint of this dissertation, the primary question is how the paradigm of intersemiotic translation helps to understand multimodal processes, more exactly how the translative mechanisms facilitate the synthesis between multimodal components. Before discussing the synthesis in a more detailed way, the scope and characteristics of the units should be addressed. It was previously presumed that sensory modalities and semiotic modes (as resources) can be seen as the primary elements in multimodal meaning-making. On the one hand, multimodal communication can be conceptualized by the premise that modes and modalities are the main "things" that are sent, received and interpreted; on the other, the notion of text must be kept in a complementary position altogether with modes and modalities.

As Helen Purchase conceives,

...when considering texts that use more than one modality, it becomes possible (and indeed, sometimes essential), for more than one message to be transmitted at once, as the receiver can now receive messages through each of the differing modalities used. The messages in the different modalities may be considered independent of each other from a perceptual point of view, although the receiver will make a cognitive link between them (e.g., in watching a video with a soundtrack) /.../ A *composite text* is a text which comprises two or more *minitexts*, each of which is a text in its own right. (Purchase 1999: 253)

Here, the textual segmentation into messages and lower-level texts (minitexts) is maintained; the modalities (i.e. sensory modalities that are wide but concrete, biologically anchored, and therefore strictly limited) are conceived as something that can be used for the transmission of messages. The position of this dissertation is close to Purchase's, conceptualizing modalities as carriers of semiotic modes. It is also interesting to notice that according to Purchase, a text seems to maintain its characteristic properties both when decomposed to a certain level (minitexts) and when combined into a bigger whole (composite texts).

For a more comprehensive understanding of the problem of segmentation and units, I would suggest a brief excursion to the concept of *intersemiosis* that underlies intersemiotic translation but perhaps a little paradoxically has been constructed afterwards. As Torop draws attention to, the notion of *intersemiotics* (as the study of intersemiosis) relies upon several concepts in the history of semiotic thinking, namely Julia Kristeva's Bakhtin-inspired concept of *intertextuality* and its reader-centered interpretation by Roland Barthes, Gerard Genette's development of the ideas on various types of *transtextuality*, *metatextuality* and the previously mentioned *total translation* that allows the integration between textual and metatextual interpretation (Torop 2023). In a nutshell, intersemiotics can be defined as the "circulation of meaning between different sign systems" (Gignoux 2005: 98), or asserted that "intersemiotics /.../ deals with two or more completely different codes, e.g., linguistic one vs. music and/or dancing, and/or image ones. Thus, when Tchaikovsky composed "Romeo and Juliet", /.../ he "translated" Shakespeare's play from the linguistic code into a musical one" (Aktulum 2017: 34).

A useful differentiation that allows to be more specific is sometimes made between intersemiotics and polysemiotics (though the notion is infrequently used). As explained by Torop, "if polysemiotics analyses coexistence, for example visual and auditory coexistence in an audiovisual text, intersemiotics creates analysability in cases where the visual is transposed into auditory or the auditory into visual" (Torop 2023: 38). Here, a strong parallel can be drawn: *polysemiotics* relates to *intersemiotics* as *multimodality* relates to the *cross-modality*, i.e. impact of one sensory modality on the perception of another (see Spence et al. 2009: 108)¹⁰.

While one should be cautious when taking those distinctions as absolute, it is possible to clarify the relation between multimodality and intersemiotics on the basis of those relations. In this situation, the key is to see the functions characterized by multimodality and intersemiotics in a spatiotemporal framework. As O'Halloran and Lim Fei note when discussing systemic functional multimodal discourse analysis (SF-MDA), "SF-MDA takes into account the semiotic resources, metafunctions, systems (at different levels), system choices and the inter-semiotic relations which unfold across space and/or time in multimodal

¹⁰ An example of a strong cross-modal relation is *haptic visuality*: here visuality functions like the sense of touch (Marks 2000: 22). When someone sees another person touch something, the mirror neuron system activates the processing circuit of touch in spectator's body.

texts” (O’Halloran, Lim Fei 2014: 148). According to a tentative delineation of mine, a multimodal process presents the interpreter modes and modalities in a parallel fashion (regardless of the fact that the interpretation of a text is not instant but often requires allocation of interpreter’s attention to various modes and even modalities in turns), letting possible transformations and crossmodal dynamics happen only when the emotional and cognitive processing have started. Intersemiotic mechanisms, on the other hand, presume that the text, before finding a contact with an interpreter, is already cross-modally pre-processed by the means of transferring modalities into other modalities or modes into another modes. Taking all this into account, could we still apply the concept of *translation* to multimodal processes? I suggest that the positive answer, at least as a hypothesis, is worthful to consider.

For that consideration, I would suggest to pause for a moment at the concept of intrapersonal communication. In short, it is a complex phenomenon, covering the notions of inner and private speech, self-talk, inner dialogue and also imaginary companions (Brinthaupt et al. 2020: 3). Roman Jakobson’s famous observation that *interpersonal communication bridges space and intrapersonal communication bridges time* (Jakobson 1985 [1974]: 98) encourages us also to look towards Lev Vygotsky, whose concept of inner speech has been widely influential¹¹. In the words of Lauri Linask, when Jakobson draws the distinction between interpersonal and intrapersonal communication, he substantially relies on Vygotsky’s analyses of egocentric and inner speech (Linask 2024: 31). Linask also proposes that Juri Lotman’s discussion of autocommunication (where culture as a self-reflective unit has replaced individual person) draws heavily on Vygotsky’s psychology (Linask 2024: 33).

According to Vygotsky,

Inner speech is to a large extent thinking in pure meanings. It is a dynamic, shifting, unstable thing, fluttering between word and thought, the two more or less stable, more or less firmly delineated components of verbal thought. Its true nature and place can be understood only after examining the next plane of verbal thought, the one still more inward than inner speech. That plane is thought itself. (Vygotsky 1934: 264)

For the research of multimodality, the pivotal questions are, “how multimodal is inner speech?” and, “is inner speech qualitatively more multimodal than regular speech?”. Contemporary neurocognitive research¹² relates inner speech predominantly to verbality, introducing it as verbal thinking, covert self-talk,

¹¹ For an overview of contemporary neurocognitive research of inner speech see Grandchamp et al. 2019.

¹² According to Alderson-Day and Fernyhough, “despite its apparent importance for human cognition, inner speech has received relatively little attention from psychologists and cognitive neuroscientists, partly due to methodological problems involved in its study. Nevertheless, a large body of empirical work has arisen relating to inner speech /.../ in rather disparate research areas” (Alderson-Day, Fernyhough 2015: 931).

internal monologue or internal dialogue (Alderson-Day, Fernyhough 2015: 932). Despite that, referring back to Vygotsky's dynamic vision of the entity between thought and word, a hypothesis of a more integrated semiosis can be conceived. An inner-speech word can automatically trigger non-verbal images, memories or even sounds, and vice versa. In this dynamics of meaning-making, the brain areas that process language are in a constant and mutual interaction with areas that process visual, auditory, and other modalities. If this hypothesis is correct, it is possible to return to the concept of translation and ask in turn, how translation works between different brain areas and between different modalities? Does it rely upon a certain common code or an internal *lingua franca* – for instance, amodal invariants?

An interesting key for understanding amodal invariants can be pointed out in the context of zoosemiotics. In his discussion of mimicry, Timo Maran refers to the possibility of interspecial protocodes. These are transindividual communication systems that are based on the premise that various living beings have similar preferences that organize their perceptions. This protocode would comprise, for example, symmetry, contrast, movement etc (Maran 2008: 22). In human cognition, some intriguing experiments also highlight the possibility of meaning transfer across modalities. As related by Ramachandran and Hubbard¹³, a classical pre-World War II experiment consisted of...

two drawings, originally designed by psychologist Wolfgang Köhler. One looks like an inkblot and the other, a jagged piece of shattered glass. When we ask, "Which of these is a 'bouba,' and which is a 'kiki'?" 98 percent of people pick the inkblot as a bouba and the other one as a kiki. Perhaps that is because the gentle curves of the amoebalike figure metaphorically mimic the gentle undulations of the sound "bouba". (Ramachandran and Hubbard 2003: 58)

The results of experiments that pair meaningless words and register the semantics they evoke in relation to oppositions as *small vs big* or *angular vs rounded*, in many recent studies are referred to as the 'Bouba/Kiki effect' (see Spence et al. 2016: 134). It is highly probable that such effects are based on the common protocodes: contrast, for example, characterizes both phonetic compositions and visually experienced figures. The abstract property of jaggedness is probably coded in a higher convergence zone, a region close to the area called TPO (that stands for the junction of temporal, parietal and occipital lobes) (Ramachandran, Hubbard 2003: 54). In a similar fashion, the change of pitch of sound would correspond to the movement of a body (e.g. a popular canine cartoon character who falls off a cliff); dropping in space feels natural if paired with a sound of falling pitch (see McCormick et al. 2018 about crossmodal correspondence and the relation between auditory pitch and visuospatial elevation).

¹³ Ramachandran and Hubbard address Kiki and Bouba effect in the context of synaesthesia, "– from the Greek roots *syn*, meaning "together," and *aisthesis*, or "perception" – is a condition in which otherwise normal people experience the blending of two or more senses" (Ramachandran and Hubbard 2003: 54).

It also seems natural that large objects emit lower sounds and smaller objects emit higher sounds. The subfield of neuropsychology takes as one of its research objects the *sonic seasoning*¹⁴, i.e. how music corresponds to the perception of taste. As Bruno Mesz and his colleagues report, taste words elicit

very reliable and consistent musical patterns: ‘bitter’ improvisations are low-pitched and legato (without interruption between notes), ‘salty’ improvisations are staccato (notes sharply detached from each other), ‘sour’ improvisations are high-pitched and dissonant, and ‘sweet’ improvisations are consonant, slow, and soft. (Mesz et al. 2011: 209)

It can be suggested that behind common image schematics that allows to connect various sensory information is the biosemiotic significance of transmodal properties as *contrast*, *symmetry* etc, developed early in the evolution. In addition, the sense-making in the brain’s higher convergence areas enables to overrule the individual signification of the medium, or in this context, the specifics of sensory modalities. Approaching semiotically and applying the previously discussed translation model to crossmodal processes, we can articulate a *cross-modal signification principle*: information in modality ‘a’ signifies information in modality ‘b’. Then the main task for the researchers of that process is to ask about the consistency of that information: is it the same or has it been transformed in some way?

1.2 Multimodality and multimedia studies

Before the return to amodal invariants and a discussion of crossmodal modulation, it would be proper to briefly address the concept of media, which is inevitably contiguous to the notions of mode and modality. Up to this point, this dissertation has deliberately evaded the topic of multimodality. As I uphold in the article of “Conflicting modalities in feature film”, there can be envisioned a framework that is detached from issues of media and focuses only on modes and modalities. While the notion of medium can be seen somewhat closer to semiotic mode and more distinct from sensory modality, the position of media (as already its name suggests, and the pun is not intended here) can be conceived between modes and modalities. For this reason, theories of modes and theories of modalities can almost fully divide up the topic of media, both discussing their “sides”: semiotic modes rely heavily on a certain medium’s materiality and the social organization of it; modality-centered approach would firsthand deal with the interpreter, including her neurocognitive processes that underlie the meaning-making.

Similar idea is expressed by Schneiderman in the context of user interface design: “Multimedia seems to be defined by the hardware required /.../ rather

¹⁴ About cross-modality in sonic seasoning, see also Reinoso-Carvalho et al. 2020.

than by the user's experience" (Shneiderman 1992: 418). Building on Schneiderman and taking a semiotic angle, Helen Purchase demonstrates a way to build a classification of media upon the classification of semiotic representational systems and communication devices (Purchase 1999: 247). It is also significant that Purchase's taxonomy of representational systems and the strategies of syntactic arrangement of signs takes place in the general context of sensory modalities, divided, in this case, between visual and aural.

While it is easier to distinguish media from sensory modalities, media – the “means of distribution of messages” (Bezemer, Kress 2008: 169) – and modes share significant common characteristics, e.g. both have material and social aspects¹⁵, and also affordances (ibid.). For instance, while Bezemer and Kress consider image, writing, layout, speech and moving image modes, they state that “medium is the substance in and through which meaning is /.../ realized and through which meaning becomes available to others (cf. “oil on canvas”). From that perspective, print (as paper-and-print) is medium; by extension, the book is medium /.../ the screen another” (Bezemer, Kress 2008: 171, 172). Accordingly, it can be claimed the socio-cultural forces that shape a book as a medium must do so by acting on the practices that concern writing and layout as modes.

The focus on the materiality, or, in other words, “the stuff in which communication is inscribed” (Bateman et al. 2017: 103) in communication and media studies of the last decades, have been called *material turn* (Bateman 2021: 35, 36). Building on Michael Halliday's theory of language functions, Kress and van Leeuwen expanded the idea of language as a system of functional resources of meaning-making to non-linguistic media, e.g. gestures and audiovisual entities. The concept of semiotic mode that emerged in this dynamics reflected the use of materiality for the achievement of these metafunctions (Bateman et al. 2017: 49). Following Marshall McLuhan's media analysis and Bruno Latour's Actor-Network Theory, Verena Fuchsberger and her colleagues also draw attention to the fact that “considering materials as actors might help us to describe and understand their active role in the interplay with the user, as well as with the designer” (Fuchsberger et al. 2013: 2853). Conceptualizing material aspects as non-human actors leads to envisioning a hybrid network where human and non-human actors have mutual influence; “a behavior can be imposed onto humans by non-human delegates, e.g. technology” (Fuchsberger et al. 2013: 2857). Regarding the material turn, there can be spotted an approximate parallel with the research of intersemiotic translation, where Jakobson's linguistic-oriented concept was expanded to non-verbal sign systems and refashioned as multidirectional, network-like dynamics.

When switching the attention from the level of object to the level of study, it partly becomes the matter of viewpoint how to align the discourses that

¹⁵ As Gunther Kress has put it, “every community has a range of resources for making meanings evident: speech, gesture, gaze, writing, and others; that is, the modes of social semiotic multimodal theory” (Kress 2020: 28) or as Paul Thibault claims, “material and semiotic processes are always crosscoupled in meaning-making activity” (Thibault 2004: 193).

converge around *multimediality* and *multimodality*. For illustrating this tendency, we can bring up two somewhat alternative approaches. Multimedia scholar Lars Elleström tends to integrate the discussion of both semiotic modes and sensory modalities into his framework of media studies, consequently also bringing along a shift in terminology, coining a notion of *media modality*. According to Elleström, the four media modalities are *material*, *sensorial*, *spatiotemporal* and *semiotic* (Elleström 2020: 20). Elleström interprets senses basically as channels – “media products must reach the mind through at least one sense” (ibid.). Regarding the semiotic mode, Elleström does not identify it similarly to Kress, but instead refers to Peircean sign types:

A certain media product must be realised through at least one material mode (as, say, a solid or non-solid object), at least one spatiotemporal mode (as three-dimensionally spatial and/or temporal), at least one sensorial mode (as visual, auditory or audiovisual) and at least one semiotic mode (as mainly iconic, indexical or symbolic). (Elleström 2020: 46)

Hence we see that Elleström has reserved the notion of *modality* to signify a wider category of *media modality* and speaking of sensorial aspects via *modes*. At the same time, it remains inexplicable why he initially addresses the category of media modalities but then, when turning to some of these in detail, suddenly switches to modes, e.g. as *material mode* (see above – 2020: 20 vs 2020: 46). Perhaps this mix-up is accidental, but as I have indicated in the introduction, this problem is widespread in the study of multimodality. Driven by the problem, my article “Semiotic mode and sensory modality in multimodal semiotics” aims to bring some clarity to such terminological entanglements.

A somewhat reversed perspective to Elleström’s is offered by John Bateman, Janina Wildfeuer and Tuomo Hiippala, who assert that media must be seen as particular bundles of semiotic modes (Bateman et al. 2017: 71). In their proceeding, a salient note of criticism towards the majority of media-centered approaches is communicated. Bateman and his colleagues note that although analysing media platforms gives knowledge about the participants of the communicative situation, it does not provide information about communicative strategies or why the situations are occurring at all (Bateman et al. 2017: 69). Therefore, they maintain that the notion of the medium can be reconstructed as generalised materialities of modes or canvases (Bateman et al. 2017: 103). Media, as they see it, is socially and historically situated practice that works as an “incubator” for new mode combinations. According to them, no meaning can be realised in a medium, “meaning can only be realized in a semiotic mode participating in that medium” (Bateman et al. 2017: 124).

Accordingly, it has become evident that the notion of media is “too wide” for being sufficiently instrumental in theoretical frameworks. In post-McLuhanean media studies where the field is already charted – at least in broad terms – and filled with theories that address medium’s specific properties, discussing “just media” has become unfruitful. Media as a notion has fallen back in order to function mostly as an umbrella term. To sum up the relation of multi-

modality to multimedia, we can see multimodality as the fundamental working principle or property of multimedia. It is possible to extrapolate that modes can be treated both as properties of media and the means of researching media. Here, a certain hierarchy of primacy emerges: we cannot seriously research the medium of books without addressing the mode of writing, but we can address writing without necessarily studying the medium of books.

Regarding researcher's viewpoint, there is freedom to decide what to consider as constant and what to consider as variable. When leaving modes and modalities for a moment and returning to the notion of *text*, we can construct models where we fixate on a certain medium and observe how it influences a variety of different texts. Alternatively, we can build models where a single text is focused on, tracing its actualization across different media. It seems that such models are more successful when the observed media are relatively similar. Examining how a certain novel is transmuted (or, intermedially translated) into a film or a comic book would prevent the researcher to delve into certain details; moreover, it continuously poses the question about the identity of the text: is it still the same text in a substantially different medium?¹⁶ On the other hand, investigating "micro-transmutations" allows to postulate more specific research questions. For example, comparing the reception of a certain film in cinema *versus* on smartphone successfully retains the text as the constant and thus permits a better knowledge of variability.

1.3 Crossmodal integration and amodal invariants

1.3.1 Paradigm change in neurosciences

In his recent book "The Entangled Brain", neuroscientist Luiz Pessoa discusses comprehensively the current shift of the paradigm in neurosciences. Generally, until now, a fairly modular view has prevailed; it has emphasized the importance of area-specific processing (attempting to allocate specific functions to specific areas) and hierarchical models of the brain (differentiating between lower-level areas, e.g. where sensorial data is gathered, and high-level areas where a more integrational processing is done, including multimodal integration). The recent tendency is towards a more holistic, network-based view. According to that perspective, a function (or meaning, as semioticians could paraphrase) arises in a collective dynamics of networks' elements¹⁷. A function

¹⁶ Answering that question is possibly facilitated by the distinction between the notions of *intersemiotic translation* and *transposition*. The first notion would imply the unchanged identity of the text, transposition would suggest that an independent and novel text is generated in another medium.

¹⁷ For instance, in prey animals, "the decision to take flight is not just triggered by threat detection and involves computations that rely on multiple external and internal variables. Together, escape behaviors are far from simple stimulus-driven, stereotypical reactions" (Pessoa 2022: 212).

or a meaning can be seen as emergent, i.e. having properties that are absent at the level of the individual elements (areas) but can be observed in the system as a whole.

The entanglement or integratedness of the brain refers to the idea that brain parts dynamically assemble into coalitions that support complex behaviors (Pessoa 2022: 14). Despite that much of the communication in the brain occurs locally (Pessoa 2022: 27), according to the “small-world” networks model by Duncan Watts and Steven Strogatz (Watts, Strogatz 1998; see also Barabasi 2002 for a thorough exploration), in a system that has local clusters of multiple nodes *and* a few of those nodes linking externally, pretty much any node (or neuron, in the case of the brain) is capable to achieve connections to remote parts of the network, with just a small number of intermediary steps. It is curious, as Pessoa notices, that the brain is even more interconnected than would be necessary for a small-world system; there are sufficiently medium and long-range connections for information to reach around well (Pessoa 2022: 171).

As Pessoa also remarks, “nodes that work as connector hubs are distinctly interesting because they have the potential to integrate diverse types of signals (if they receive connections from disparate sources) or to distribute signals widely (if they project to disparate targets)” (Pessoa 2022: 184). Such connector hubs may offer an intriguing implication for the semiotic viewpoint, concerning the most abstract level of translation models. Different languages (understood generally as organized sign systems) that are mutually translatable, belong to the same network; they are fundamentally differentiated by the means of space (and therefore translation is needed). The premise of any translational activity is connectedness, access to different parts of the network. Thus, two languages, or in our case, sensory modalities, must not necessarily be conceptualized as partly overlapping, but employing the same base code (that enables them to be on the same network). In this model, translating, firsthand, means connecting and overcoming the distance¹⁸.

In the semiotic discussion of multimodal processes we should also consider another paradigm change that foregrounds the role of predictive coding. In the words of philosopher Andy Clark, “brains build human experience only by combining their own predictions with sensory evidence” (Clark 2023: 34, see his whole book for the discussion of the brain as a “prediction machine”). According to Pessoa, in the predictive framework traditional logic is flipped: perception is directed by action. For instance, speaking of vision, traditional view considered it rather passive, similar to a camera that registers “the world”. In the predictive framework, “vision is active and guided by endogenous computations that try to

¹⁸ Applying the small-world model to translation proper, e.g. interlingual translation, we can conceptualize a language corpus in source language as a local community of nodes, the corpus of all possible words in the target language as other, distant community of nodes. Necessary connections via the hubs are based on already known correspondences; unknown and uncertain correspondences are built by the means of connecting first to the distant cluster and then searching for a best matching correspondence moving from node to node.

anticipate the most valuable future information for the animal” (Pessoa 2022: 125). As Laurence Aitchinson and Máté Lengyel specify:

Activity in the visual cortex is also strongly modulated by the spatial and temporal context of stimuli — to the extent that /.../ neurons in primary visual cortex (V1) even respond to illusory contours, stimulus features that are not physically present in the input but must be inferred from context. Overall, there is much evidence that perception and, correspondingly, neural responses in sensory cortical areas are as influenced by predictions and expectations about stimuli as by the actual stimuli themselves. Indeed, while ascending feed-forward connections convey stimulus-related information, long-range horizontal and feed-back connections within and between different cortical areas provide a natural anatomical substrate for conveying such ‘contextual’ effects. (Aitchinson, Lengyel 2017: 219)

The framework of predictive coding is based on the idea that for the brain, representing sensorial direct input is too costly; much more energy-efficient¹⁹ is to operate with prediction errors, i.e. differences between sensory input and prediction (ibid.). For the same reason, neurotransmitter dopamine that has been somewhat inaccurately titled as the “reward molecule”, does not function as the sign of reward itself. Instead, it represents reward prediction error. Most dopamine neurons in the midbrain of humans, monkeys and rodents are activated when the reward is bigger than predicted, remain at baseline activity for accurately predicted rewards, and show depressed activity when reward is lesser than predicted (Schultz 2016: 23).

It is justified to suppose that predictions assume certain world-models, which are not based on the activity of a single brain area but rather on networks that connect multiple locations. Therefore, those models can be inherently multimodal, resembling to “maps” and “images” – units what Antonio Damasio reckons as the main content of our minds (Damasio 2012) and which I will briefly discuss below. I propose that such integrated models and predictive processing indirectly support the view that multimodal, composite meanings resemble dynamic networks that partly reach out of the boundaries of interpreter’s mind, relying, on the one hand, upon the network-like brain architecture, and on the other, upon the network-like multiplicity of external semiotic resources, both loosely scattered in the environment and strictly organized into the borders of culture texts.

When Paul Thibault (2004: 192), drawing on Gregory Bateson and Jay Lemke (see Bateson 2000 [1972] and Lemke 1998), underlines that in multimodal processes, “different semiotic modalities co-contextualize each other” and the character of this meaning-making is rather multiplicative than additive, the observation of co-contextualization is well-put. Nevertheless, a degree of caution should be maintained with the “multiplication”. Applying strict mathe-

¹⁹ In addition to energy efficiency, evolutionary initiative seems to favor systems with better predicting abilities.

matical “metaphors” to the meaning-making in network-like systems can turn out reductionist already for the simple reason that “neurons influence each other not only in an excitatory fashion but also through inhibition. In the latter case, when a neuron fires, it makes the neurons connected to it less likely to generate an action potential” (Pessoa 2022: 27). While I don’t deny the possibility to attribute the characteristic of computability to multimodal processes, I suggest, firsthand drawing on Pessoa, that the components of a multimodal whole (i.e. the clusters of information from different sensory modalities) rather than simply add up or multiply, have various degrees of modulating influences on each other. In the words of Lisa Feldman Barrett:

...when you kiss someone, you are enveloped in a unified experience that combines the sight of a face, the sound of breathing, the feel, taste, and scent of luscious lips, and the racing of your heart. Your brain assembles these sensations into a cohesive whole. Scientists call this process sensory integration. (Feldman Barrett 2020: 56)

1.3.2 Multimodal / crossmodal integration

Although the concepts of multimodality and crossmodality are sometimes used interchangeably, it would be reasonable to follow the practice that bounds multimodality with a more general perspective. According to this, multimodality refers to the situations where two or more semiotic modes or two or more sensory modalities (predominantly in the literature of psychology and neuroscience) are incorporated in a single system. Crossmodality, in turn, refers more specifically to mutual interaction of sensory modalities, e.g. their possible integration in multimodal “wholes”. Therefore, in a sense, the mechanics or the grammar of crossmodality is, on a very abstract level, a basis for the mechanics of semiotic synthesis in complex situations or culture texts that engage two or more different sign systems. Hence, multimodality as a wider term embeds crossmodality as a more specific notion. It is also relevant to make a distinction between the processes that take place in the environment (that includes culture texts) and the processes inside an interpreter – crossmodality is associated with “internal” processes, i.e. reception and interpretation.

In their concise overview of the research on multisensory perception and crossmodal integration, Jon Driver and Charles Spence strongly reflect the tendencies discussed above. According to them, from the 1960s to the 1980s, the main goal of the study was identifying separate modules in the mind/brain; 1990s led to an increasing realization that the interplay between components of extended networks were equally crucial (Driver, Spence 2000: 731). The concept of feedback started to play more important role. The old models relied upon the idea that information builds up from sensory receptors and is integrated in “higher” convergence zones, e.g. in superior colliculus (see especially Stein and Meredith 1993). Now it was found that the areas of multimodal convergence can send feedback to the “lower”, unimodal zones (Driver, Spence

2000: 731). This brings along some interesting implications. The primary or so-called unimodal areas can be influenced by another sensory modality (ibid); that means in some cases, information in one sensory modality can pre-influence the interpretation from another modality in various ways. In other words, subjective experience within one modality can be affected by stimulation within another; e.g. changing the color of food may alter the perception of taste.

As even more intriguing example, functional magnetic resonance imaging (fMRI) studies have demonstrated that watching speaker's lips during a face-to-face conversation, linguistic visual cues can activate auditory cortex in normal hearing individuals even when they are purposefully deprived of auditory speech sounds in the experiment. According to Gemma A. Calvert and her colleagues, psychophysical evidence suggests that audiovisual integration of linguistic signals occurs before the stage of word identification; possibly it happens at the prelexical level, more specifically at the stage of phonetic categorization (Calvert et al. 1997: 593). Similar substitution can happen when on a video, musical instruments are being played, and despite the missing audio track, an experienced viewer is able to "hear" the melody.

Marcela Perrone-Bertolotti and her colleagues point out that silent reading often activates auditory cortex and involves an imaginary speech component. It has also been suggested that while children learn to read loudly, adults continue to pronounce written text privately (Perrone-Bertolotti et al. 2012). Research also suggests that the brain uses top-down connections to stimulate sensory regions and neuronal populations. In this manner, silent reading creates an imagined experience where auditory modality is simulated, involving emotions. Direct speech in a literary text can activate auditory areas that respond to voices; because the people in literature are imagined, it seems that a more general 'inner' voice than the voice of a specific individual is simulated (Petkov, Belin 2013).

It is important to note that there are strong parallels between modality relations in brain-level processing and in culture texts. In the typology of modality relations that is featured in the article "Semiotic mode and sensory modality in multimodal semiotics", I referred to the "substituting" relation and brought as an example the artistic device of *ekphrasis*, i.e. verbal representation of visual representation (see e.g. Mitchell 1995). For example, in several films of Ingmar Bergman, a character verbally describes a picture that is not shown. This semiotic game is especially refined as the medium of film would originally favor iconic representations. As Maaret Koskinen has pointed out, there are lot of ekphrastic descriptions of photographs and paintings found in Bergman's notebooks, drafts, and manuscripts (Koskinen 2021: 62).

While one sensory modality can "substitute" for another, a type of synthesis has also been observed in speech perception where visual and auditive information contradict each other and create a third, fused meaning. In the much-quoted experiment that gave the name to McGurk effect, a film was shown of "a young woman's talking head", in which repeated utterances of the syllable [ba] (auditory modality) was dubbed to lip movements for [ga] (visual modality).

Instead of selecting a variant, normal test subjects reported hearing [da] (McGurk, MacDonald 1976: 746). Presumably, this primary model of fusion would work to illustrate the mechanics of meaning-making in more complicated situations and also culture texts. While addressing the fusion-type meanings in audiovisual art is possibly an intriguing object for multimedia- and neurosemiotics, it is beyond the limits of the current dissertation.

However, still another crossmodal relation ought to be highlighted. In an experiment conducted by Jean Vroomen and Beatrice de Gelder, the effect of auditory modality on the perception of visual modality was examined. In the words of Vroomen and de Gelder, “In the present study, we /.../ investigated an illusion that occurs when an abrupt sound is presented during a rapidly changing visual display. Phenomenally, it looks as if the sound is pinning the visual stimulus for a short moment so that the visual display freezes. In the present study, we explored this freezing phenomenon” (Vroomen, de Gelder 2000: 1584). The task to detect a visual target was performed better when the target matched with a unique sound (high tone in contrast to the default low tone) in auditory modality (ibid). As Driver and Spence interpreted this, “Popout of the unique sound among its auditory stream could then lead to accompanying popout of the concurrent visual display from its visual stream” (Driver, Spence 2000: 733).

There can be proposed some intriguing semiotic interpretations for the modalities punctuating and underscoring each other. On the first hand, it is pertinent to recall Paul Thibault’s observation that multimodal components can offer a context to each other (Thibault 2004: 192). On the other hand (and this interpretation is not contrasting, but rather close to Thibault’s), Lotmanian tradition allows to model a situation where one modality begins to work as other’s metatext, accentuating or bringing forth specific elements. With allocating one modality the role of text and the other the role of metatext, this mechanism enables a multimodal whole to become self-referential and engage in the activity of self-description²⁰. In other words, the experiment of Vroomen and De Gelder demonstrated how an element of auditory modality rendered an element of visual modality *marked*²¹ (comp. marked and unmarked elements of film language in Lotman 2004: 53).

Nevertheless, while the substituting and modulating crossmodal relations allow discussing intriguing artistic and cultural mechanisms, most of the time, modalities still provide information about the same property of the external world. As Driver and Spence acknowledge, crossmodal integration has to be the

²⁰ In “Culture and Explosion”, Juri Lotman discusses a culture model where various dynamic processes develop in a parallel manner, but have different velocities. In such a case, slower processes can provide a description of the more rapid processes. This constitutes one possible model for culture’s self-description (Lotman 2009 [1992]: 12).

²¹ However, it would be correct to note that Lotman’s *marked–unmarked* opposition was at least partly inspired by Russian formalist concept of artistic function and *ostranenie* as estrangement; thus *marked–unmarked* predominantly correlated to the oppositions of *natural – artistic* and *regular – exceptional*.

rule rather than the exception in real-world perception: “It seems entirely adaptive that the multiple sources of information /.../ should be combined to yield the best estimate of the external property in such cases” (Driver, Spence 2000: 731). In the article “Semiotic mode and sensory modality in multimodal semiotics” this recognition is extended to culture texts with an understanding that the evolutionary initiative is reflected in the composition of multimodal culture texts where the mutually supporting relation type is predominant.

To sum it up about the crossmodal integration and the possibilities to model it mathematically, Dominic Massaro’s view has been considered as one of the most promising. Building upon the discussions of McGurk effect, Massaro maintains that pattern recognition is central to cognition; the perception of speech by eye and ear can be seen as a prototypical case of pattern recognition. In the framework of his Fuzzy Logical Model of Perception, Massaro states that perceivers form perceptual judgments “by evaluating and integrating multiple ambiguous sources of information, in an optimal manner based on relative goodness of match” (Massaro 1999: 310). As Driver and Spence interpret this, one modality can be weighted more heavily than another in a multimodal combination, especially if it produces a greater reduction of uncertainty (for example when visual localisation is easier than auditory localisation (Driver, Spence 2000: 732).

It is also important to notice that vision does not always dominate other modalities. In accordance with the paradigmatic turn towards a more holistic and network-based view, Ladan Shams and Robyn Kim point out that attention in different modalities is not independent (Shams, Kim 2010: 272); the weighting of visual cues is affected by how consistent is the visual cue with the signals from other modalities (Shams, Kim 2010: 277). According to Bertelson and De Gelder we cannot state that any modality is dominant *per se*. Instead, we have to consider the characteristics of the specific act of meaning-making. For example, experiments have indicated that for detecting the temporal resolution of events, audition is superior over vision (Bertelson, De Gelder 2004: 149).

Concludingly, from the point of view of semiotics’ intersection with cognitive neurosciences, it is highly significant to recall that dominance of different modalities depending on the specific semiotic situation was already foreseen by Roman Jakobson. When Jakobson admitted the “prevalence of icons among purely spatial, visual signs and the predominance of symbols among purely temporal, auditory signs” (Jakobson 1971b[1967]: 701), he also affirmed that

all five external senses carry semiotic functions in human society. /.../ Within the systems of auditory signs never space but only time acts as a structural factor, namely, time in its two axes, sequence and simultaneity; the structuration of visual signantia necessarily involves space and can be either abstracted from time. (ibid)

Taking account of the non-modular approach to the brain architecture and the central unit of network instead of an area, overlapping of networks and the dynamic nature of the signaling processes where the signification/meaning is an

emergent property, let us return to the concept of translation, at least metaphorically, and consider amodal invariants as possible units in a primary modeling system that underlies beneath both primary modeling systems in Lotman's and Sebeok's theories.

1.3.3 Amodal invariants in the context of integrated experiences and transmodal modeling systems

In the semiotic theory of Tartu-Moscow school (TMS), the conceptualization of modeling systems as belonging to a hierarchy on the basis of their characteristics, has an important place. For TMS, due to its background of structuralism, natural language has the primary place in this hierarchy. Secondary or supralinguistic systems (e.g. artistic "languages" as literature) were conceptualized as a strata above natural language and concurrently relying upon it (see Lotman et al. 2013[1973]). Thomas Sebeok, in the context of biosemiotics, notably challenged the TMS's position about the primacy of natural language. Taking into account animal communication, he indicated that zoosemiotic "layer" was present before the language was acquired by humans or their evolutionary ancestors. Moreover, in contrast to what TMS claimed, language is established upon the pre-linguistic systems of cognition and communication; this pre-linguistic and zoosemiotic component is present and highly influential also in contemporary humans (see Sebeok 1991: 334, 335). Therefore, a reposition in the forementioned typology was called upon. While in many situations "thinking without language" is primary, the natural language comes only as a secondary modeling system, and supralinguistic systems have tertiary position.

Approaching the hierarchy of modeling systems in the light of the findings in neuroscience's paradigm change suggests, first of all, to be cautious with strict categorizations and hierarchies. Meanings are, more often than not, network-based and multi-component at least in the sense of brain areas. Some of those components can be language-related, for instance expressed by activation of neurons in Broca's or Wernicke's areas (that enable speech production and speech comprehension, respectively), but apparently those language-related components are also integrated with visual or auditory components, altogether constituting a certain mental image that Peirce referred to as interpretant.

Hence, an alternative way to position the problem is, can we assume a fundamental modeling system that underlies all these not-so-hierarchical but network-like models of internal and external reality, being not a primary, but instead the *zero level* system? Presumably, such a system would position towards the other, higher-level systems (if we still assume their presence in our description despite the above-mentioned holistic disposition) as the lower structural level to the higher structural level in the theories of double articulation of natural language. In double articulation, the base level or, according to André Martinet, the second level (sic! – but if we assume that Martinet followed not the synthetic but the analytic logic, i.e. deconstruction of language into components, this

direction is well explained) is independent of meanings – it is language’s articulation into distinctive units or phonemes. The first level of articulation for Martinet is into larger elements that already have meanings and are subject to syntactic organization (see Martinet 1984: 31, 37)²². In a similar manner, our zero-level internal system can be seen essentially as a Saussurean system of differences (relying upon the binary choice between a neuron generating action potential or not; this means, transmitting a signal via electrochemical mechanism or staying passive).

Still, the binary pattern of neuronal states is just a conceptual part of that system that is, in essence, already complicated by the fact that neurons influence each other not only in an excitatory fashion but also through inhibition. Moreover, neurons are not exactly standardized semiconductor type “switches” that populate the brain; among estimated 100 billion neurons in a human’s nervous system there is a variety of distinct cell types categorized by morphology, molecular identity, and physiological activity (Purves et al. 2010: 10). Furthermore, an extended version of that abstract language model could probably include various types of synapses, endocrine system and neurotransmitters as chemical sign vehicles. As goes for a signal of a single neuron, neither a specific type of neuron nor neurotransmitter does not have a fixed meaning. The message is neither in the electric impulse nor in a molecule: these elements could only be the components of a signalling network where meaning – the level of semiosis that builds upon signalling – appears as emergent. However, the confines of this dissertation would prevent any attempt to build another “grand theory” and therefore, we can only discuss such base-level system in a knowingly hypothetical manner²³.

For the viewpoint of this dissertation, the crucial idea is that this zero-level modeling system can be envisioned as cross- or transmodal; i.e. its elements universally employed by all sensory modalities. As Barry Stein and Alex Meredith point out in their classic treatise on multisensory convergence, in some ways the sensory information processed by individual modalities is “amodal” or interchangeable across modalities. It can be assumed that such amodal features of sensory information are *intensity, form, number and duration* – because of their amodal characteristics they are efficiently translatable across modalities. “The constancies among amodal stimulus characteristics also provides basis for

²² See also Chandler 2017: 181 for a very compact overview of the idea of double articulation in semiotics.

²³ Despite of the tentative nature of this model, it affords at least one clear answer, and namely to “a key semiotic debate used to be whether or not semiotic systems such as photography, film, or painting have double articulation” (see Chandler 2017: 181 once more). Assuming such base level modeling system allows us to claim that double articulation characterizes *any semiotic activity*, including iconic processes. Once again, a tentative designation of the levels could be “*signalling or pre-semiotic articulation*” and “*semiotic articulation*”.

perceptual cohesiveness when cues from multiple sensory modalities are present at the same time” (Stein, Meredith 1993: x).

Stein and Meredith suggest that mechanisms for multisensory integration in higher organisms were laid down quite early in the evolution. For example unicellular animals acted out a primitive supramodal system by the very fact that they were one-celled beings responding to more than one sensory stimulus. So the paramecium’s membrane is a multisensory integrator of the most elemental form. A significant step forward was the development of structures where a bimodal neuron type emerged that could receive input from different sensory areas, converging the impulses. For instance, coupling tactile and chemical stimuli facilitates feeding in marine invertebrate *Aplysia* (Stein, Meredith 1993: 24, 28, 31). It can be also suggested that channeling of multiple sensory cues into the same neurons provides a mechanism for substituting one cue for another (Stein, Meredith 1993: 32) – this contributes to the explanation of crossmodal processes.

Here, once again, naturally comes up the question how this primeval language of *relations, gradients, differences* and *contrasts* with its proto-meanings of *intensity, form, number, duration* and also *rhythm* – the parameters that strangely resemble the components of music – should be discussed in the framework of semiotics?²⁴ I suggest that we can briefly return to Roman Jakobson and his concept of invariance. Paraphrasing Elin Sütiste and Peeter Torop who maintain that *invariance* is one of the fundamental concepts of semiotics that enable us to draw parallels between different areas of study, for example narrative and translation (Torop, Sütiste 2007: 189), it can be claimed that the dynamics between variability and invariability allows also to grasp the homology between sensory modalities, and their shared contents²⁵.

Jakobson referred to scholastic terms *signum, signans* and *signatum*, treating them in Saussurean paradigm where they refer, respectively, to bilateral linguistic sign, signifier and signified. Jakobson asserts that the basic difference between signans and signatum is that *signans must necessarily be perceptible* and *the signatum must be translatable*. Combining Saussurean duality and Peirce’s idea that “the basic property of any verbal sign lies in its capability of being translated into another verbal sign, either a more developed, explicit sign, or, on the contrary, a more elliptical sign, of the same language system or of a different one”, Jakobson appears to suggest that variance correlates with the

²⁴ In the context of film theory, it is also pertinent to recall that Sergei Eisenstein highlighted *movement* as the basic element of art, and as Greg Smith points out, “similarities in the movement of pictures and music provide the foundation to create correspondences between the two elements” (Smith 2004: 307).

²⁵ A vivid example how some properties can be transmodal not only in the sense of sensory modalities but semiotic modes, is provided by Van Leeuwen. Discussing typesets and scripture, he points out that “‘curvature’ and ‘regularity’ are not restricted to typography – They can also be realized in other semiotic modes – though of course by means of material signifiers that are specific to those modes” (Van Leeuwen 2006: 152).

level of signifiers and the invariance can be detected on the level of signifieds (Jakobson 1985c: 30). At this place it is also relevant to recall Umberto Eco's assessment that Jakobson "could not focus on the laws of language without considering the whole of their behavioral background /.../ in Jakobson's work every discussion on verbal language is always connected with other communicative phenomena" (Eco 1987: 113, 114).

The direction I would like to point is also well characterized by Guagnano and Mininni, who explore the relation between Peirce and translation studies. They suggest the 'psychosemiotic perspective', claiming that "translation is the psychosemiotic operation par excellence, because translating is thinking in two languages"²⁶. Psycho-semiotic perspective can also be seen as "a complex mode of knowing capable of combining the logic of the interpretant with the dialogic of the interpreter" (Guagnano, Mininni 2018: 136, 137). Extrapolating from the discussions above, a hypothetical sign model can be envisioned where the signans or sign vehicle is unimodal and the interpreter is always multi-, trans-, or supramodal as a whole mental image.

An interesting parallel can be drawn from the work of Antonio Damasio, a Portuguese neuroscientist who most notably has explored the relation between the brain, emotions and consciousness. I propose that Damasio's thinking is remarkably semiotic, regardless of the fact that he almost never uses the 's-word'. In Damasio's theory of signification, also two spheres or levels are set into mutual correspondence. Here, the *sign vehicle* is omitted; Damasio's *reality* and its *image* would roughly correspond to *object* and *interpretant* of Peircian tradition. The object can be easily overshadowed by the interpretant: "We often fail to notice this simple reality because the mental images of the objects and events that surround us, along with the images of the words and sentences that describe them, use up so much of our overburdened attention" (Damasio 2002: 3). The *image* of Damasio denotes the dynamic whole that emerges in the multimodal intersection of cognitive and emotional processes²⁷:

²⁶ Such "thinking in two languages" can also be interpreted as thinking in metalanguage.

²⁷ It is important to notice that besides the holistic 'image', Damasio also speaks about 'maps' as units of the mind. Maps and images correspond to somewhat separate brain spaces. Maps have risen earlier in the evolution and in a sense, they are constituents of images. Thus, Damasio posits two somewhat separate "brain spaces." As he claims, /.../ "one space constructs explicit maps of objects and events during perception and reconstructs them during recall. In both percept and recall, there is a manifest correspondence between the properties of the object and the map. The other space holds dispositions rather than maps, that is, implicit formulas for how to reconstruct maps in the image space. The explicit image space is constituted by the aggregate of early sensorimotor cortices. /.../ The two spaces point to different ages in brain evolution, one in which dispositions sufficed to guide adequate behavior and another in which maps gave rise to images and to an upgrade of the quality of behavior. Today they are seamlessly integrated" (Damasio 2012: 153).

I regard the problem of consciousness as a combination of two intimately related problems. The first is the problem of understanding how the brain inside the human organism engenders the mental patterns we call, for lack of a better term, the images of an object. By object I mean entities as diverse as a person, a place, a melody, a toothache, a state of bliss; by image I mean a mental pattern in any of the sensory modalities, e.g., a sound image, a tactile image, the image of a state of wellbeing. Such images convey aspects of the physical characteristics of the object and they may also convey the reaction of like or dislike one may have for an object, the plans one may formulate for it, or the web of relationships of that object among other objects. Quite candidly, this first problem of consciousness is the problem of how we get a “movie-in-the-brain,” provided we realize that in this rough metaphor the movie has as many sensory tracks as our nervous system has sensory portals—sight, sound, taste, and olfaction, touch, inner senses, and so on²⁸. (Damasio 1999: 9)

Applying Damasio’s understanding of “movies-in-the-brain” to real movies, we can see that the first are naturally the interpretants of the second: movies, as any objects in the environment, arouse these metaphorical “movies” in the brain that contain feelings. These images can also arise from memory, represent and facilitate various actions, helping to plan, model, or simulate the happenings of our future selves. For Damasio, meaning-making is image-making, and the holistic unit that emerges as the result of this process is essentially multimodal.

If our aim is to treat the system of multimodal meanings as language with its specific grammar, and address crossmodal processes as translations from one modality to another, the first task would be to choose the reference system, i.e. to agree what we consider as invariants, because only on the background of *invariance* we can examine *variance*. For example, in cognitive film research, such reference system could be remarkably general, as general as the narrative itself. The scholars of neurocinematics, Pia Tikka and Mauri Kaipainen suggest a remarkable integratedness that would empirically support Damasio’s concept of image: “Several fMRI studies have revealed that narrative sense-making overrides the effect of the storytelling medium /.../ the broad socio-emotional context of narratives engages “higher-level” brain areas in a holistic manner, i.e., beyond the “lower-level” sensory modalities that activate based on media-specific visual or auditory qualities of the stimulus” (Tikka, Kaipainen 2023: 163).

To sum it all up, the discussion of amodal invariants indicates that the synthetic meaning or the result of the combination of different sensory modalities does not arise so much from the intrinsic difference of language elements but from their position, or more specifically, the connection pattern. This pattern somewhat resembles a Saussurean system of differences, where the location of an element and its relations towards other elements determine its meaning and contribute to the meaning of textual whole.

²⁸ For all who started wondering about the second problem of consciousness, here it is: “This is the problem of how, in parallel with engendering mental patterns for an object, the brain also engenders a sense of self in the act of knowing” (Damasio 1999: 9).

2. SEMIOSIS AS INCOMPATIBILITY: THE FUNCTION OF CONFLICTS AND THEIR SPATIAL MODELING

2.1 Habits and deautomatization. From established codes to multiple-choice systems

In 1935, American psychologist John Ridley Stroop published an article that became one of the most cited works in the history of experimental psychology. Stroop's aim was to look deeper into the opposing tendencies of association and interference. Despite that Stroop did not directly talk about 'signs', I give myself permission to interpret his position semiotically. Therefore, it can be claimed that on the basis of his forerunners' works, for Stroop it was clear that the stronger associative bonds that hold a sign together, the easier was recalling the absent part of the sign on the basis of the present part. In the similar manner, when a sign component was already associated with another specific component, it required more effort to create a new 'bond' or association. To further the discussion, Stroop turned towards the factors that could weaken this association. He looked at situations where two stimuli were presented simultaneously, not in mutually supporting relation but interfering with another (see Stroop 1935), and devised an experiment that has been called 'The Stroop task' since then. As eloquently related by Luiz Pessoa,

the Stroop task /.../ is attractive given its combination of simplicity and effectiveness. Imagine a page with words used for colors, such as "red," "green," and so on, written in black ink. Now imagine writing the words with pens of different colors, including one that matches the word and one that doesn't—for example, the word "red" written in red ink (called congruent) or the word "red" written in green ink (called incongruent). Your task is to tell the color of the ink used to write the word. /.../ Every stimulus contains two properties (word meaning and color), and the participant must attend selectively to the appropriate attribute to perform adequately. The key comparison is to contrast a person's reaction time to answer incongruent and congruent trials—one is a little slower in the former. (Pessoa 2022: 121)

Such tasks are often called 'conflict tasks' because the focusing on a particular stimulus is complicated by a competing stimulus. As Pessoa indicates, here cognitive control comes into play, allowing to select task-relevant response in the face of competition from a stronger but task-irrelevant one. Researchers believe that especially prefrontal cortex contributes to this endeavor (Pessoa 2022: 122). While habit formation makes certain actions efficient and automatic patterns of conduct are harder to break, conflict tasks enable to model situations where the detachment between "sign components" is required. In turn, this demands an increased activity or even a heightened agency of an interpreter.

From a general viewpoint, sign processes always depend on the tension between the 'thing' that is present and 'thing' that is absent. Any reference or

representation involves a separation, a gap that rules out the absolute sameness. It can be claimed that the same understanding was expressed by Peirce's idea of endless semiosis, inspiring Jakobson's concept of intersemiotic translation and Eco's consequent criticism, in turn. The idea of conflict as the basis for semiosis is outstandingly summed up by Anti Randviir. He indicates, on the basis of Juri Lotman's structural theory of text, that creation of any text (and, I suggest it would apply to the creation of any *sign*, as well²⁹) involves drawing a borderline between a certain unit and everything that surrounds it; this automatically creates at least two entities with different structures or types of internal coherence. Consequently, any text and any meaning can emerge only in the field of tension between systems with different structures (Randviir 2010: 68).

Boldly paraphrased, semiosis happens in the tension field between habits and habit breaking, or, between established associations and the new, prospective ones that have not yet been fixed into a recognizable relation. Speaking of habits, it is now proper to recall that both for Peirce and contemporary semiotics, 'habit' is a fundamental concept (see Kull 2016: 624). According to Peirce, symbols incorporate habits, and habits, in turn, associate with knowledge (CP 4.531). While indexes are vehicles that enable bringing in novel information (as the footprint that Robinson Crusoe found on the sand), for Peirce the symbols "rest exclusively on habits already definitely formed but not furnishing any observation even of themselves, and since knowledge is habit, they do not enable us to add to our knowledge" (ibid.). It is evident that Peirce interprets both habits and symbols as something *fixed*, non-dynamic as the results of some previous dynamic processes and simultaneously as the preconditions to some upcoming dynamic processes – notably similar to what has been later seen as *code* in mathematical theory of communication³⁰.

While Peirce associated symbols with habits and knowledge, modern, biosemiotic and neuroscientific views allow to address the relation of both makings – *habit-making* and *meaning-making* independently of Peirce's sign typologies. Habit-making, primarily, happens on the level of synapses or neuronal connections, evolutionally preceding any symbolic- or language-mediated semiosis. In such manner, it would firsthand underpin the primeval *lingua franca* of amodal invariants. In 1949, Canadian psychologist Donald Hebb famously stated that neurons that fire together, wire together³¹ and also somewhat later – not very surprisingly but yet significantly for the theories of meaning-making –

²⁹ In the "Theses on the semiotic study of cultures", Lotman and his colleagues argue that the text can be seen both as an integral sign and as a sequence of signs (2013[1973]: 58).

³⁰ Such fluctuation between static and dynamic phases, and especially development and utilization of fixed structures as codes are also characterized by the concept of *semiotic scaffolding* in biosemiotics (see Hoffmeyer 2007, Kull 2015b).

³¹ In detail, the statement that Hebb himself titled 'neurophysiological postulate', is this: "When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased" (Hebb 2002[1949]: 62).

it was postulated that neurons which wire together, fire longer (see Mittal et al. 2018). Apparently, the binary opposition of *old vs new* is culturally so significant because it reflects the effort and energy expenditure that is required to re-route already established networks. Both natural environment and culture have mechanisms for encouraging such types of actions; a remarkable mechanism is facilitated by the perception of *choices*.

In the context of biosemiotics, Kalevi Kull discusses choice as a crucial condition for semiosis: “Only in case if there is a certain problem to solve, if there is more than one option (which requires choice and decision-making), the process related to it is in semiotics commonly called interpretation, or semiosis” (Kull 2015b: 225). Kull associates problems with confusion, i.e. the situations where a living system has several options to choose and these options are mutually incompatible, in others words, contain logical inconsistency. Kull highlights multimodal conflicts among other examples of such inconsistency: “For instance, perceptions from two sense organs order the opposite actions of the same effector” (ibid.). Therefore, semiosis happens on the condition of code plurality and consequent incompatibility. In semiosis, there is “not determined what happens next /.../ the parts of the model do not fit each other” (Kull 2015b: 227). We can conclude with the observation of Tyler Bennett, who unites biosemiotic principles with discussing Roman Jakobson’s poetic function, that “contradiction gives rise to new hypothesis generation” (Bennett 2021: 159).

While relying upon Kull’s understanding of incompatibility as the basis for semiosis, I would remain cautious about the total equalization of semiosis and conflict. It would be conceivable to see semiosis as a dynamic fluctuation between stable and conflictual states. It can also be realized that this fluctuation is similar to Juri Lotman’s model of cultural dynamics that entails the alternation of gradual and explosive processes (see e.g. Lotman 2009 [1992]: 12). Both maintaining and changing codes have their functions that emerge in the interaction of interpreters and their environments. Returning to the pivotal question of this dissertation, we are asking about the functions of explosive processes; first of all in general, as we have done already, and then specifically about the functions of multimodal conflicts.

The standpoint of this dissertation suggests that there is no necessity to see intrinsic differences between sign processes in nature and culture: these can be recognized as having mutual modeling capability. We would probably be accustomed with the view that cultural languages model biological ‘languages’, bringing into mind all the descriptions and metatexts that have been written about nature (including mathematics, a fully abstracted description of nature). However, vivid examples about the reverse dynamics are provided by various bodily turns in culture studies, starting from Merleau-Ponty’s phenomenology and its application upon visuality by Vivian Sobchack (see Merleau-Ponty 1962, Sobchack 1992), ending with branches of film studies or literary theories that highlight body’s role as the predominant interpreter.

For instance, it has been demonstrated that while reading a verbal text, interpreters use motor simulations to clarify meaning: comprehending a word

that indicates some action our motor system represents its meaning (see Barsalou 2009). As such, reading can be considered as an inherently multisensorial activity (Hillesund et al. 2022). As I have also indicated in the typology of modality relations in the article of “Semiotic mode and sensory modality in multimodal semiotics”, reading can be an example of cross-activating type. In the context of cognitive film studies, the role of emotions in meaning-making has been discussed with connection to ecological approach (see e.g. Anderson 1996); the brain/body system has central position in the interpretation theories that rely upon *embodied simulation* or mental simulation (first of all, see Grodal 2009). As neuroscientists Vittorio Gallese and Michele Guerra put it, “with the support of contemporary cognitive neuroscience, it is possible to formulate a new perceptual model in which action, perception, and cognition are closely integrated” (Gallese, Guerra 2015: 151). Such examples are sufficient to illustrate how ‘nature’ can interpret cultural texts, as well³².

Another important part that is played by cognitive / body-oriented culture theories is that they set a necessary background for a contemporary revision of the notions that were used when addressing the dynamics between automatization and deautomatization³³. In the discussions of culture and artistic expression in 20th century that focussed on the overcoming of automatism, a cluster of interconnected notions formed a network that was strongly oriented by Viktor Shklovsky’s concept of *ostran(n)enie*. Several terms were closely affiliated to it, not only in a following, but also preceding manner. As elderly Shklovsky notes in “Tales about Prose” (1966/1983), his *ostranenie* was not original; for instance, Romanticist poet Novalis celebrated art that “makes things pleasantly strange” (Shklovsky 2016b: 272). Among the terms that are partly inspired by *ostranenie*, Brecht’s *Verfremdung* is often highlighted, but also noted that *Verfremdung* was not just a formal device but also a didactic technique for promoting audience’s better grasp of reality and fight for political goal (Tihonov 2005: 688). It has also pointed out that Derrida’s concepts of *deconstruction* and *différance* (as “productive movement of differences”) have close affinity to *ostranenie* as *defamiliarization* (Bogdanov 2005: 56).

Both *defamiliarization* and *estrangement*, as well as *enstrangement* have been proposed as translation equivalents for *ostranenie* (see a comprehensive discussion about the translation of the term in Berlina 2016), although Alexandra Berlina herself leaves it untranslated and uses the original ‘ostranenie’ in her Shklovsky’s translations. According to Shklovsky, in his “Art as a Device”:

³² ...and of course, here we see already how next-level cultural texts model the process of nature modeling previous-level cultural texts. Such a mutual dynamic forms a full circle, and repeats itself.

³³ I have briefly addressed the topic beforehand in an article that observes a small selection of Estonian arthouse films, focussing on the device that creates subjective regime of reception by discontinuing a default regime of bodily character simulation (Oja 2014). Here, it is also relevant to refer to David Bordwell’s discussion of ‘retardation’ of narration (see Bordwell 1985: 38).

...the goal of art is to create the sensation of seeing, and not merely recognizing, things; the device of art is the *ostranenie* of things and the complication of the form, which increases the duration and complexity of perception, as the process of perception is its own end in art and must be prolonged. (Shklovsky 2016[1917]: 80)

For instance, discussing *ostranenie* in Leo Tolstoy's prose, the notion refers "not to the calling a thing or event by its name but describing it as if seen for the first time" (Shklovsky 2016[1917]: 81). In essence, it means "showing an object outside the usual patterns, describing a phenomenon with new words, taken from a different field of relations" (Shklovsky 2016[1917]: 252).

A new view, a novel context entails an alteration of the viewpoint; sometimes it is achieved by moving away from the object. Along these lines, *ostranenie* has also been interpreted just as *distance*, similar to Bakhtin's 'outsideness' (*vnenakhodimost*) (Emerson 2005: 637). A highly captivating viewpoint is also offered by Douglas Robinson in the context of somatics of literature. It presumes that the engagement with art is largely influenced by an evolutionary response called *somatic transfer* or somatic mimeticism (that is, in turn, based upon the activity of mirror neurons and the simulation of the bodily feelings of a character, e.g. the pain, in reader's body) (Robinson 2008 xiii, xiv, 23). Literary and artistic devices supposedly negotiate with such automatic regimes and incite "switches" or scheme changes that interrupt the default, energy-efficient processing method and incite alternative mechanisms. Those mechanisms bring along heightened attention and apparently activate other, more "analytic" loops or networks in the brain, e.g. more specifically in prefrontal cortex.

While Shklovsky was actively engaged in film-making, being a screenwriter for several films (see Baker 2010), his approach to deautomatization should not be taken as limited with literature, but in a dual manner, as Berlina suggests: *ostranenie* can be both "extraliterary, applying to the world, and also intraliterary, applying to "poetic" language, genres, and devices" (Berlina 2016: 24). According to Annie van den Oever, Shklovsky's statements on *ostranenie* were "first and foremost an urgently required and utterly relevant theoretical answer to the /.../ impact early cinema had on the early avant-garde movements" (Van den Oever 2010: 11). Moreover, when neoformalist film theorists as David Bordwell and Kristin Thompson³⁴ in the end of 1970's and the beginning of 1980's looked back to Russian Formalists for inspiration, the nascent cognitive film theory that resulted from such efforts held as its essence the idea that experiencing film presupposes the same mechanisms that are used to perceive life outside of an artwork, or the environment. We can presume that any text

³⁴ For instance, Thompson in her "Eisenstein's "Ivan the Terrible": A Neoformalist Analysis" (1981).

or artwork is, for a human as an animal, first of all a part of an environment³⁵. The reception of a text cannot employ some other, spectacularly different ‘hardware’ systems than our good old brain with its architecture relatively unchanging long before the emergence of, for example, written communication³⁶.

Still, when filmmakers employ defamiliarizing elements, neither recognition nor interpretation of the meaning of those elements is straightforward. According to Laurent Jullier, to detect defamiliarization, one first must have knowledge about what is familiar, and “within the familiar, what is familiar for oneself only, or for a whole body of persons, or for the whole of humanity” (Jullier 2010: 139). Defamiliarization is feasible on almost any level of film – this complexity is first of all addressed by the two-way logic that simultaneously applies to film and other arts. “Cognitive distinction of “perceiving” as a bottom-up and “knowing” as a top-down logic” (Kiss 2010: 165) refers to, for example, that *ostranenie* can happen both on the depiction of the world itself and also on the level of *expectations* towards that world. For instance, the spatial distribution of objects and their properties has been challenged since the time of Georges Méliès by magic-like tricks (this follows the bottom-up logic). On the other hand, an example of top-down logic would include the issues of narrative. When chronological storytelling principles were challenged by Quentin Tarantino in his “Pulp Fiction”, it did not confront basic level models of the “fabric” of the world but the high-level models concerning the interpretation strategies. However, these few examples hardly illustrate the range of possibilities for filmic *ostranenie*³⁷.

Whether *ostranenie* leads to wonder or discovery, provokes tension, norm deviation or choices between variants, triggers the feelings of disturbance or uncanny, works against established codes, arouses positive connotations as Novalis’s romantic sublimity or negative ones as Freud’s *Unheimliche*, charac-

³⁵ It is done despite Tom Gunning’s viewpoint that “art-making (involves) a divergence from everyday cognitive processes” (see the discussion in Van den Oever, Gunning 2020: 17). To keep it simple, a solution is to make a clear difference between *everyday perception* and *environmental perception*. The last category is larger and contains both, *everyday* perception and de-automatized processes.

³⁶ According to Laurent Jullier, “the concept of defamiliarization lies at the heart of a quarrel between/.../ the culturalists-constructivists (who think our perceptive habits themselves are structured by our social habitus, or even our language /.../ and, on the other hand, the universalists-ecologists (who think our perceptive habits are the product of an evolution in which the modern era, even when the appearance of verbal language is included, is quantitatively of little importance” (Jullier 2010: 139). This dissertation aligns itself somewhere in the middle of these extremes, with an inclination towards the second, ecologist position, which in semiotics corresponds to Sebeokian, or biosemiotic approach.

³⁷ Furthermore, as indicated by Tom Gunning, “ostranenie can be applied not only to art-making (technique) but to reception (and therefore criticism as well)” (Van den Oever, Gunning 2020: 20).

terizes a stage in person's psychological development as Piaget's *decentering*³⁸, or draws attention to the working mechanisms of the text³⁹, we can propose that multimodal conflicts in films are generally related to devices with a common denominator of *de-automatization*. As meanings emerge in the relation between world and perception, also deautomatization – as a certain type of meaning on the one hand and a property that can be attached to other meanings on the other – always emerges between reality and the perception of it.

To sum it up, the discussion of Shklovsky's *ostranenie* and related notions offers one possible framework for a further approach to multimodal conflicts. At the same time it does not allow to cover the whole spectrum of issues. The specific viewpoint that is first of all stemming from the notion of multimodal or audiovisual conflict, includes a range of other aspects besides de-automatization as the dominant one. It should be kept in mind that both defamiliarization and multimodal conflict are wide notions; their mutual positioning is not suggesting that one category is necessarily bigger and encompasses other; rather they are partly overlapping. There is a whole range of defamiliarizing events or devices in film that are not, by any means, connected to audiovisual conflicts. On the other hand, many audiovisual conflicts that inherently have defamiliarizing potential do not actually defamiliarize. As I have indicated in the article of "Conflicting modalities in feature film", internal diegetic sound works as a vivid example of *formal* audiovisual conflict, but its interpretation by viewers can rely upon an established convention that it conveys characters' thoughts or inner speech; as such it does not defamiliarize. We will leave the topic now in order to return to it at the end of this chapter, focusing on some specific functions that are fulfilled by audiovisual conflicts in film.

2.2 Audiovisual conflicts: practices and perspectives

2.2.1 One time and two spaces: an artistic model for multimodal conflict

The undeniable dominant of Jonathan Glazer's film "The Zone of Interest" (2023) is its sound design. But not in an isolated manner; it works in the ensemble with other devices, design of the *mise-en-scène*, cinematography and editing. Depicting the everyday life of Auschwitz commandant Rudolf Höss and his family, Glazer did not want to show the atrocities of the camp. Instead, he allocated the

³⁸ In the words of Jean Piaget, a "general decentralisation /.../ means a dissolution or melting down of static perceptual forms in favour of operational mobility, and consequently there is the possibility of constructing an infinite number of new structures which may be perceptible or may exceed the limits of all true perception" (Piaget 2001[1947]: 73).

³⁹ As David Ireland observes, "incongruent film music can often draw attention to elements of the cinematic construction, and thus notions of production and agency" (Ireland 2018: 19).

role of the signifier of the tragedy to the auditory modality. Although the soundtrack works inside the synthetic space that entails visuality, it works in an eerily contrasting manner to those somewhat dull and cosy family scenes. Consequently, there unfolds “another film” on the soundtrack.

In order to achieve the impression, sound designer Johnnie Burn first assembled a 600-page document that related the events at Auschwitz and contained witness testimonies. He created a large map of the camp in order to properly determine the distances and echoes of the sounds. Learning what type of weapons were used in Auschwitz, it was important for him to accurately represent how the shots sounded from a specific distance and angle. In addition, Burn recorded a vast sound library with painful human voices, sounds of crematoriums, industrial furnaces, and the manufacturing machinery that the prisoners had to operate (see Kring-Schreifels 2023, Struss 2023).

In “The Zone of Interest”, the auditory modality achieves the status of a metatext, becoming a separate, organic world, as well as a commenter towards the events and characters conveyed in visual modality. The viewer is partly defamiliarized, partly activated by presenting her a choice between multiple reception strategies. At first, she may try to keep the modalities somewhat apart, focusing on the one or the other, letting the emotional outcome of the narrative be influenced by this choice. Another strategy is to synthesize modalities together, but this may easily provoke the third option, that is, a constant fluctuation between visual and auditory spaces. Remaining in such mental limbo, the viewer could experience the simulation of living simultaneously in two contradictory realities. While the film generates dark dramatic irony towards Höss’s family members, who half-willingly, half-naively deny the auditory meanings, it also addresses the viewers, reminding them of the vast masses of people who, in their multimodal reality, choose to live in the “mode” of comfort and ignore the “mode” of nearby atrocity.

The example of “The Zone of Interest” illustrates a general model for audiovisual conflicts: in the confines of a synthetic chronotope, two contrasting spaces are created – the visual and auditory space. Both are set to comply with the same timescale; the synchrony enables the juxtaposition of various properties and dynamics: the spaces can partly overlap; they can contain similar and even common elements such as the human bones that Höss finds by a creek where he relaxes with his children; the bones have trespassed the border from the world of killing. The distance between spaces can be in a constant change; it can both diminish and expand during the unfolding of a narrative. Lastly, on the level of the synthetic whole of the text, the contrasting spaces are united in a juxtaposing manner. Höss’s character who regularly crosses the borderline between the spaces, guides viewers in the process of this disturbing synthesis.

In the third chapter of this dissertation I will explore various uses of multimodal conflict in a brief historical overview that passes several waypoints in the development of 20th century cinematic expression, but this overview will be limited by the notions of *contrapuntal editing* and *internal diegetic sound* that are also discussed in the article “Conflicting modalities in feature film”. With an

intention to expand the focus, here, in this subchapter, I aim to address some additional functions of the incongruence between sound and image, both intentional and unintentional. While the example of “The Zone of Interest” illustrated a general model of juxtaposition that was powerfully employed for directing viewers’ response, in the following I intend to highlight the comic use of modality conflict, very shortly touch upon the role of incongruent subtitles in a multimodal whole, discuss some less-than-perfect multimodal composition strategies that can obstruct the comprehension of texts, and lastly, address some ideas that can possibly enhance multimodal communication.

2.2.2 Inducing humor by multimodal conflicts

The cybernetics conferences organized by Josiah Macy Jr. Foundation in New York from 1941 until 1960 were remarkable interdisciplinary events. Macy conferences, as they were summarily called, addressed a range of problems where systems theory overlapped with semiotic ideas⁴⁰. In the conference of 1952, Gregory Bateson presented his thoughts on humor and said:

The hypothesis that I am presenting is that the paradoxes are the prototypic paradigm for humor, and that laughter occurs at the moment when a circuit of that kind is completed. This hypothesis could be followed up with an analysis of jokes, but rather than do that, I should like to present to you the notion that these paradoxes are the stuff of human communication. (Bateson 1952: 3)

In his classic treatment of humor, three decades later, Victor Raskin builds a semantic theory that is established on the concept of *script*, “a large chunk of semantic information surrounding the word or evoked by it /.../ – a cognitive schema internalized by the native speaker (Raskin 1985: 81). In general, according to Raskin, the combinatorial specificity of scripts determine the effects that texts elicit upon the readers or hearers. Humor develops when “two overlapping scripts are perceived as opposite in a certain sense, and it is this oppositeness which creates the joke” (Raskin 1985: 100). While some sort of incongruity is humor’s essential condition, Daniel Perlmutter notices that there are different forms of incongruities and the “realization of humor is ultimately in the mind of the hearer” (Perlmutter 2002: 156). More often than not, to detect humor the listener has to make a critical evaluation based on the information that is external to joke (Perlmutter 2002: 161).

Despite that the classic theory of humor is built around verbal semiotic mode, the idea of scripts as cognitive schemas and carriers of information would also apply to non-verbal modes. Therefore, paradoxes and conflicts that breed those paradoxes can also be the “stuff” of multimodal communication, where perceiver’s presumptions are negotiated and something unexpected and

⁴⁰ It would be intriguing to look for the traces of those ideas in the works of Tartu-Moscow school that took cybernetics as one of its influences.

defamiliarizing emerges out of the juxtaposition. It would be functional to make a basic typological distinction here. Hence, multimodal script conflicts can generally belong to three categories:

- 1) two scripts are carried by the same sensory modality but different semiotic modes, e.g. we access image and accompanying text through visual modality;
- 2) two scripts belong to the same semiotic mode (e.g. verbal text) but are carried by different sensory modalities (e.g. in film, where subtitles are presented in visual and dialogue in auditory modality);
- 3) two scripts belong to different semiotic modes that are carried by different sensory modalities (e.g. the mode of actor's gesture that we apprehend visually and the mode of actor's verbal speech that we apprehend via auditory modality).

Here, for a moment, let us recall the undeniable hero of this dissertation, René Magritte. His experiments with dissociating texts and images can be aligned with the first category, regardless of the amount of humor they elicited. More importantly, Magritte's influence on modern advertising design has been recently acknowledged (see Othman 2021). Huda Othman recognizes Magrittean surrealist techniques of "surprise, unexpected juxtapositions, and non sequitur" in the landmark VW Beetle print advertising (1960, DDB advertising agency) that features the image of the car and the title *Lemon*:

The 'Lemon' advertising, and the other adverts in the Volkswagen ads series, used an honest and charming humor that would correspond to the honest and charming car. Though the word 'lemon' (signifier) may seem to have no link to a picture of a Beetle (signified), it was in fact was a reference to the cars' appearance. (Othman 2021: 118)

To be precise, while such combinations of incongruent image and text induce multimodal conflicts, they are not *modality* conflicts. Nevertheless those "multiple modes / single modality" examples offer close analogies to audiovisual texts. One of such is highly popular "Hitler's rant video meme" that is not precisely a single text but a paradigm or a text family; each variant uses the same source material in auditory modality (Hitler's speech) but generates distinctive element inside visual modality (while most of the visual information remains unchanged, the mode of subtitles is altered by meme-makers). The parodies of the bunker scene in Oliver Hirschbiegel's film, "Downfall" (2004) started to emerge a couple of years after film's premiere. The purposefully "wrong" subtitles that were added tended to refer to trivial issues of users' daily lives – most of the time to something that is semantically as distant as possible from the original content and therefore has a highest potential for humor.

Christopher Gilbert discusses the meme in the context of participatory culture that blurs the lines between production and consumption; he notices that playfulness and user-generated content is common to new media environments

(Gilbert 2013: 408). As such, the meme is partly influenced by remix and *re-cut* aesthetics; more specifically being affiliated to trailer remixes or fake trailers. In this fan-generated genre new meanings are produced that add up to the original, but simultaneously the issue of recognizability (of the original) has the central position (see Korsgaard 2020, Deaville 2016, and Dusi 2015b). Still, it is important to notice that unlike film trailers, video memes are not exactly paratexts (see Genette 1997); their status is more independent as they do not depend on one specific text (the understanding of Hitler's rant meme does *not* depend on the fact if the audience has seen Hirschbiegel's film, but depends on the contextual knowledge about who Hitler was).

Galia Hirsch, addressing pragmatic mechanisms behind video memes, notes that memes exist "beyond the notions of author and text", belonging to a network or an "instant community" that is created by Youtube's 'related video' function (Hirsch 2019: 26). She also acknowledges that the balance between repetition and difference is the key element in a successful parody (ibid.). To sum it up, a possible model for humor-generation by an audiovisual conflict would entail a similar balance between shared and diverging parts of Raskin's scripts: the semiotic mode that was initially assumed to be coherent across visual and auditory modalities is unexpectedly "split", duplicated and set into contradiction with itself. It must also be noticed that in the context of remix logic, subtitles are exceptionally easy semiotic mode to technically manipulate (compared to voice-over, for example); and this is one reason among several others behind the success of Hitler's rant meme. In a dominantly English-speaking cultural space, films in other languages rely heavily upon the semantic potential of subtitles, not only a specific form of translation, but a possible semiotic mode with its grammar and pragmatics that include a capability of lying or purposeful mistranslation.

2.2.3 Multimodal efficiency: avoiding blunders and annoyances

An extensive discussion of the delicate role of subtitles in audiovisual communication would clearly benefit from the framework of multimodality that first of all, enables a precise distinction of the basic-level textual components (subtitles relate to the semiotic mode of *verbal text* that is conveyed by *visual* sensory modality), and also allows to address subtitles' relations to other textual elements. While the issues of film subtitles have been predominantly addressed in the context of translation studies (see e.g. Guillot 2010; De Linde, Kay 1999), a somewhat formalist perspective would be complementary, for example, in the task of re-defining the textual status of the subtitles that do not exactly belong to borders of original text (at least when imposed by some other agency than filmmaker's themselves), but are also too closely and organically integrated with the textual elements to be classified as paratexts. I believe that the view-

point that emphasizes the dynamic process of meaning-making with the help of subtitles as a semiotic resource would help here.

In this short section I will not touch upon the topic of subtitles any further, but instead aim to indicate some problems and questions around the multimodal composition in a more general sense. By discussing two examples, I hope to highlight the problem of multimodal efficiency or “literacy”, that is, text’s creator’s awareness about the mechanisms of synthesis between the components of different modes and modalities. In detail, such awareness would influence the resulting meaning of the whole (at least at the stage where we do not yet count in the influence of an interpreter). In some compositions, for example, various information-carrying elements would start to compete with each other for the limited resources of the channel (that depends both on the interpreter’s capacities and the specifics of the medium), threatening to damage each others’ semiotic potential. In such a case, creator’s attempts to fulfill her purposes would be diverted.

Scoring⁴¹ silent films is a highly intriguing semiotic challenge, where adding a whole new modality is required. It can be argued that silent films have never been fully silent because of the musical accompaniment that usually featured the presentation of films (see first of all Tieber, Windisch 2014 about historical practices of silent film sound), but according to a technical simplified model silent films are still monomodal⁴² (only visual) if we consider sensory modalities and do not take into account possible crossmodal influences. Consequently, attaching a whole auditory modality with its various modes is characterized by a range of possible choices that are made between two extremities: keeping total coherence with existing visual modality and achieving total incoherence with it. Essentially, a similar dilemma would illustrate various remix-like new media practices that transform or expand original texts and also translation processes (although totally incoherent or dissimilar target text cannot be usually recognized as a *translation* of the source text at all). While the academic discussion of silent film music follows primarily the historic perspective, a theory of scoring silent films retrospectively, from a remarkable distance in time, can be developed in the oversection of remix studies, musicology and especially film music research, and also intersemiotic translation.

When British Film Institute initiated the program “Save the Hitchcock 9” in order to restore Alfred Hitchcock’s nine surviving silent films, it also commissioned new scores for the films, with an idea of them performed live at the UK premiers in 2012. The undertaking coincided with London Summer Olympics, for that reason it was later christened “Hitchcock Olympiad” (Combs 2012). “The Lodger: A Story of the London fog” (1927) is claimed as Hitchcock’s breakthrough film: while being his first commercially successful direction it also influenced the development of the genre of thriller. The job of scoring that

⁴¹ Differently from the *soundtrack*, the *score* is an original piece of music that is composed for a film as a whole.

⁴² At the same time, silent films are multiodal in the sense of modes as semiotic resources.

film was allocated to Nitin Sawhney, a musician and composer whose work entails influences both from electronic and world music genres (especially Asian musical cultures). The score, with the cooperation of London Symphony Orchestra, was performed live on 21 July 2012 at the Barbican Centre and later, it was featured in the release of the film on digital data carriers, in a way starting to compete with other variants of “The Lodger’s” scores⁴³.

Attentively, Richard Combs has pointed out various intertextual elements of the score that are linked to Hitchcock’s other, perhaps more known films, as well as to some more abstract, emotional and historical clusters of meanings:

Nitin Sawhney’s new score goes partly in the same direction (slashing Psycho strings, a touch of North by Northwest in the final chase and near-lynching of the wrong man). But it also moves backwards. In its playfulness, a lurking brassiness in the menace, and even a little ballad for the burgeoning romance, the score has an air of carnival, of the shows, circuses, and musical entertainments of Hitchcock’s youth. (Combs 2012)

In an interview broadcasted by BFI on Youtube, Sawhney recounts that his task begun with finding a musical vocabulary that corresponds to characters and the psychology of the film. He acknowledges that his strangest choice was to put in songs; he doubted if there is a place for them in a Hitchcock films. Finally, the argument behind the inclusion of the songs was, that Sawhney treated the scoring operatically, understanding the songs as arias (Sawhney 2012). Curiously, there are only two comments under the video (as seen on the 8th of September 2024) and they are not favorable. The user @corinnegroundgate3383 says: “Awful, awful soundtrack; it almost makes the film unwatchable. Some of the music is so jarringly out of place that one prefers to watch the film with the audio muted”. @DinoAgent69 seems likewise disappointed: “I have never heard a worse score in my life. Shame on you for (almost) single handedly ruining Hitchcock’s film” (ibid).

A cultural analyst, encountering such cases, has two prominent tasks. First is to remain as neutral as possible towards the research object, and the second is nevertheless attempting to offer an explanation why the text in focus “failed to work” or “succeeded” in upsetting some of the audiences. It is not entirely possible to measure the coherence between the original visual modality and the added auditory one without the slant of subjectivity; moreover it would be an assignment that would require the preparation of a musicologist. Here, we can limit ourselves with Sawhney’s own notification that adding the songs to the score turned out problematic.

⁴³ In my opinion, significantly more coherent with film’s visual modality is the jazz-influenced score by Joby Talbot and Ashley Irwin (1999) and the melancholic-atmospheric score by Graham Reynolds (2020).

The various comments under “Daisy’s Song” that is presented on the YouTube separately,⁴⁴ clarify that the song itself was beautiful, but inside the film it sounded too modern, out of the place and not fitting the movie. The song begins approximately 23 minutes into the film, with an original intertitle “Next Morning” and the initiation of a new sequence where film’s heroine brings a breakfast into the hero’s room. “I don’t know who you are / but something’s changed...” sings a sensitive and sweet-sounding female voice with a slowish, romantically laid-back pop-song style instrumental accompaniment. Here, we still have to keep in mind that in the historic, i.e. original screenings of silent films, singing was not a taboo. As Claus Tieber and Anna Windisch point out about the cinemas of pre-World-War I Vienna,

...another crucial aspect of music films with regard to their exhibition was the use of the human voice. Viennese cinemas featured soloists, quartets and choirs as accompaniment to film screenings, not only for opera- and operetta-related films. A great number of music films were based on songs and arias as their main attraction. Thus, the singing voice was crucial in Viennese cinemas: on- and off-screen, diegetic, as well as an essential part of the musical accompaniment and multimedia program. (Tieber, Windisch 2014: 94)

Presumably, the main problem in the score’s use of songs was not the general fact that the songs were used, but a specific “semiotic work” that the songs, for example “Daisy’s Song”, performed. The composer rather intrusively pushed the *semiotic mode of verbal dialogue* into the *auditory modality*; the words can be attributed to the smiling heroine, bringing out her “inner speech” that was never intended to be represented in such a simplified and straightforward manner by Hitchcock himself. Perhaps such a meaning-making strategy would work in a film of an author of less detailed and elaborated style of film-making; Hitchcock’s developing “handwriting” was already too sophisticated for cheap and blatant solutions⁴⁵. The audience, somewhat familiar with Hitchcock’s style, would thus take the beginning of the romantic song as a surprise and defamiliarization, but this time, in a negative sense.

The second example will pay a visit to docudrama, both a subgenre between fiction and documentary and a method of film-making⁴⁶. While documentary theorist Bill Nichols leaves docudrama purposefully out of his treatment of documentary genre (Nichols 2001: xvi), Steven Lipkin and his colleagues point out that the status of docudrama is under debate because it, by its very nature, “breaks boundaries” (Lipkin et al. 2006: 11). While the definition of docudrama may verge towards either direction – a fiction film that depicts “real” historical

⁴⁴ Available online <https://www.youtube.com/watch?v=8prhoLrP3LQ>; Accessed 8th of September 2024.

⁴⁵ In 1925, young Hitchcock did some filmwork in Germany where he, in the words of Patrick McGilligan, “was swept away by the controlling style and pictorial mood of German expressionism” (McGilligan 2003: 63).

⁴⁶ See Paget 2011 for a comprehensive discussion of the subgenre.

events and towards documentary that includes staged scenes, it can be stated that in the center of docudrama are actors who personify historical individuals in the context of historical, supposedly true events.

Docudrama tends to provoke disputes about the permissibility, usefulness and even danger of mixing the functions of documentary and drama (Lipkin et al. 2006: 15). As Torben Grodal observes, “the primary goal of mainstream fiction is to entertain, and an evaluation of its reality status is often irrelevant or even undesirable. /.../ Documentaries and docudramas, too, need to entertain, but their main purpose is to provide information about the world (Grodal 2018: 79). According to the cognitive perspective, film viewers continuously assess the reality status of the information that is provided by a film, and for this they need to use many different skills, ranging from a generic understanding to specific knowledge. As Grodal emphasizes, by default we tend to believe and the disbelieving is only secondary⁴⁷:

Even when watching a fantasy film, we may be shocked by the sudden approach of a dinosaur because it is perceptually real. As Joseph Anderson (1996) argues, following James Gibson’s (1979) lead, the basic composition of our brain developed long before motion pictures made it possible to simulate reality; thus, we still are prone to accept such media-borne information as true indices of real phenomena. Seeing is believing. Reality evaluation demands an active process of disconfirmation of what is not real /.../ whether a given film is fiction or docudrama, the viewer must constantly employ the necessary cognitive processes to assess its reality status. (Grodal 2018: 81)

It can be hypothesized (the limits of this thesis do not allow to go further) that viewers tend to prefer the situations where they can rather quickly and energy-efficiently decide either way: take the textual world with which they are presently in contact as the fictional or documentary. This decision allows them to know what type of information they have been provided with and accordingly, which internal reception strategies to employ. Supposedly, when cues about the reality status of the text are mixed, it can provoke dissatisfaction with the material.

In the recent Netflix’s docuseries “Hitler and the Nazis: Evil on Trial” (2024), three basic modes of expression are combined: 1) historians speak in an informal lecture or interview format, 2) archive footage of historical events and characters, including Hitler and other Nazi leaders are presented, and thirdly, 3) the mode of docudrama is added where actors play the same characters, including Hitler, Goebbels, Göring and others. As the miniseries is built around the Nuremberg trials, both archive material and staged footage with actors are

⁴⁷ It has to be kept in mind that perceptual realism cannot strictly be linked to a choice of the genre because of multiple reasons, e.g. “in conventional fiction films concrete locations may have documentary qualities that “serve as a backdrop to the drama”, and “a propaganda documentary or docudrama may provide a far more distorted representation of reality than many a fiction film” (Grodal 2018: 76, 78).

extensively presented to convey the courtroom action. It must also be noticed that the staged footage is stylized in post-production with an intention to make it more similar to archive material. Occasionally, an original audio recording of the trial is played in the auditory modality, while in visual modality, simultaneously the actor impersonating the corresponding defendant in the courtroom, as if speaking, is presented.

Our quick and modest method for studying audience reflexivity encourages to look for some user comments again, this time from Internet Movie Database. *Mohsen_Hadid* says (on the 6th June 2024):

Did not need to watch the trial scenes while watching the footage from the original one. And the actors who do not resemble the original characters who appear constantly. This caused some confusion.

kvnngel states that (on the 8th June 2024):

What I will say as far as production value is that they're casting is absolutely atrocious. The guy they chose for Hitler looks more like Himmler or Goebbles, gaunt, thin and tiny... like a 3rd choice for the role to save money...,

and *souplahoopla* (11th June 2024) finds that

the dramatisations are actually awful, and totally un-needed. A lot of the time the dramatisations of the trials are still voiced by the criminal, but slow mo reenactments, even when there are actual videos? It makes no sense. Still pictures would have been better if the reel was damaged.

Here, two basic multimodal conflicts seem to emerge. The first defamiliarization results from the conflict between the semiotic mode of actors and the characters from historical footage; the second involves also sensory modalities when combining archive audio with staged visuals. It can be suggested, as the continuation of the above-presented hypothesis, that those multimodal conflicts disturbed the viewing process by hindering the viewers' decision about which reception strategies to choose. The documentary footage applied for the "real", but the Goebbels-look-alike Hitler started to function as a parody. Considering that according to Lipkin and his colleagues documentaries' "persuasive power" rests predominantly on their proximity to the real (Lipkin et al. 2006: 18), it can be conceived that the clumsy and over-artificial re-construction of the events depleted the series of its persuasive power; transforming the material to something that cannot be processed neither as real nor fictional.

Hopefully, the research on multimodal meaning-making will also help to improve our 'textual environments' or semiospheres, joining the efforts of culture scholars with the struggles of designers, educators or multimodal composers in the most general sense. What is better? – will be a natural question – and I would suggest first of all, that favoring thought-provoking defamiliarizations over the frustrating ones would partly answer it. Undoubtedly, that wish is

subjective and biased, but prioritizing theoretical explanations of the conflict mechanisms over practical results, the subjectivity can be compensated. Consequently, works on multimodal meaning-making can also offer practical insights about multimodal strategies or tactics on several levels, starting from the choices in the confines of individual's personal conduct and ending with public or corporate strategies that benefit from avoiding cross-modal disturbances.

Some simple principles are reflected in the folk psychology: occasionally, for instance, *less is more*. Film director Jim Jarmusch has recounted about his several trips to Japan, where he has bought a lot of videotapes with Japanese films by Ozu, Mizoguchi and others. The films were not subtitled, but Jarmusch asserts that he likes to watch the films without knowing what the characters say. "The language of acting is not primarily a spoken language," thinks Jarmusch (Hertzberg 2001: 125). It can be suggested that Jarmusch's preference benefits from the multimodal strategy of omission, a semiotic subtraction or clearing up the space⁴⁸. Cutting out the mode of verbal semantics or semantic verballity "makes space" for another modes, both in auditory and visual modalities. While attentional focus in humans and other animals shifts continuously for multiple reasons, a mode or modality omission can be developed to a purposeful reception strategy that benefits the whole de-coding side of semiosis, e.g. learning of specific tasks, or focussing on the elements of complex texts that need to be interpreted separately in order to sufficiently understand them⁴⁹. In film viewing, bypassing verbal mode or auditory modality as a whole, helps to concentrate on the meanings that are conveyed in visual modality. In listening songs, on the contrary, a listener can modify her modal focus when *adding* visual modality by reading lyrics simultaneously. In this manner she strengthens the verbal mode, perhaps at the expense of processing pure *musical information*.

Awareness of multimodal processes would improve "media hygiene" both on the level of textual composing and interpreting (the topic of multimodal composition has come up predominantly in the discourse of education in recent years – see e.g. Kitalong, Miner 2017, and Warshcauer 2017). As a negative example, many authors of educational Youtube videos who supposedly want their content to efficiently reach target groups, overload their compositions' auditory modality with musical background which strongly interferes with their talking. Establishing the personal dominant of the message that depends on the situation and purpose would help any 'multimodal composer' to transmit it with

⁴⁸ It has to be noticed that omitting modes or modalities is somewhat related to, but still essentially different strategy from Juri Lotman's 'minus-device' (see Portis-Winner, Winner 1976: 143; Lotman 1977[1971]: 51 or Terrence Deacon's 'absential influence' (Deacon 2013: 14, 44). Although, the relation between the concepts would be an interesting topic for future research.

⁴⁹ As a more complex method, a temporary separation of an element is necessary for connecting it to other possible syntagmatic (commutation tests) or paradigmatic formations, with a purpose of describing relations between system's elements.

minimum hazard of overshadowing by the non-relevant information in other modes and modalities.

The semiotic discussion of multimodal strategies can only remain scientific when it keeps strong interdisciplinary connections; most of all, it means grounding itself in the findings of cognitive psychology and neuroscience. In these domains, one of the most closely related topics to our theme is the problem of multitasking or simultaneously managing information from various sensory modalities. The keyword of *attention* and the understanding of it as a limited resource that has to be allocated between multiple sensory modalities, i.e. ‘inter-modal attention’, is the unifying axis of those discussions (see Wahn, König 2017, Salmela et al. 2014; Talsma et al. 2006). How fast and efficiently a subject can concentrate on a certain mode or modality, depends on a variety of factors. In the last decades, perceptual load theory has been prevalent with an understanding that the efficiency of selective attention is dependent on both perceptual and cognitive load (see Murphy et al. 2016). This is congruent with the view that both bottom-up and top-down processes have significance in meaning-making.

Among some fundamental findings are also the results suggesting that early sensory competition is exclusively modality-specific and does not extend across senses (Porcu et al. 2014: 224); the research of the simultaneous processing of auditory and visual information has revealed that in recognition tasks, pairing pictures and words together slowed down visual responses and sped up auditory ones. This, in turn, suggests an underlying auditory dominance: “auditory stimuli automatically grab attention and attenuate/delay visual processing” (Robinson et al. 2018: 1). Such observation is compatible with the report that finds the degradation of visual performance in virtual reality when auditory effects are present; more specifically, presentation of temporally congruent but spatially incongruent sound, visual performance significantly weakens in the tasks of detection and recognition (Malpica et al. 2020). I suggest that these and other consecutive findings have also far-reaching implications on the semiotic modeling of multimodal processes.

It can also be suggested that in the dialogue between (neuro)semiotics and neurosciences, semiotics operates with more general concepts and notions (in a sense, translating or transposing neuroscience’s specific terms into a more universalized vocabulary). Doing this, it offers a more abstract level of explanation to the scientific paradigm as a whole. I propose that discussing the issues of multimodality, especially spatial models helps to understand where, for example, sensory modalities are too tightly “packed” with modes, or whether semiotic resources are evenly or unevenly distributed. Also, as I argued in the article “Conflicting modalities in feature film”, spatial models can be applied to the conflicts between modalities with the purpose of indicating semantic incongruencies. In such type of modeling, the parameter of spatial distance can be set into correspondence with semantic distance.

2.3 Space as modelling system for evaluating multimodal conflicts

The following short subchapter is conceived firstly as a general comment on the section discussing spatial modeling in the article “Conflicting modalities in feature film”, and secondly, it aims to conclude the topic that overarches the whole second chapter, namely the dynamics between conflicts as defamiliarizing events and the habituation that leads the emerging, novel meanings to be eventually crystallized into a code⁵⁰. The idea that aligns conflicted meanings with semantic distance is evidently a tentative one, and making it sufficiently solid would require more *space* than the confines of this dissertation allow. Hereby I will only propose a hypothesis that regards the mutual associations between spatial distance, energy management concerns of living beings and the dynamics between defamiliarization, habituation and codes.

It can be claimed that any great semiotic model has relied upon space, from Saussurean compound of *signifiant–signifié* as an opposition grounded in an abstraction of smallest binary space – to Jakob von Uexküll’s *Umwelt–Innenwelt* and the division between perceptual and operational world⁵¹; from Greimas’s logical square-based models to Lotman’s semiosphere⁵². It is also highly revealing that Thomas Sebeok and Marcel Danesi discuss the system of Cartesian coordinate geometry as a modeling system that “makes possible to model geometrical concepts in specific ways” (Sebeok, Danesi 2000: 31–33) and linguistic concepts that are combined by image schemas and represented in mind-space (Sebeok, Danesi 2000: 77, 78). It is relevant to reiterate that the mathematical coordinate system can be used to model not only geometrical concepts but also objects in the geographical, i.e. real-world space.

While the article “Conflicting modalities in feature film” touched upon Lotman’s idea that the semiotic space functions as the precondition for any semiotic activity (Lotman 1990b: 123) and the relations between textual elements can be in correlation with relations between components in a spatial model (Lotman 1977[1971]: 217), it is appropriate to recall Anti Randviir’s bold abstraction that Tartu-Moscow school is predominantly the school of semiotics of space (Randviir 2010: 63). Randviir also observes that the concepts of space

⁵⁰ Kalevi Kull draws attention to the relation between the notion of codes and semiotic scaffolding, that entails “the building for the development of fixed schemata or codes. In fact, scaffolding itself looks almost like a code—codes, too, being always the products of semiosis. What distinguishes codes from scaffolding is their functionality. A code can be described just as a correspondence, whereas scaffolding always has a helping-supporting task or function. A habit, as a product of semiosis, is always, to a certain extent, instructional. Semiosis is a learning process that produces scaffolding or habits, that may ultimately become established as codes” (Kull 2015b: 230).

⁵¹ Uexküll 1987: 168.

⁵² It can be noticed that the idea of the semiosphere is closely connected to Lotman’s opinion that natural language and the language of space constitute two fundamental ‘languages of culture’.

and text are in a mutual relationship (Randviir 2010: 66). It is reasonable and even inevitable, because both creating and interpreting of any type of text are inseparably grounded in timespace; we can claim that any text is spatially organized. If we presume that in a minimal model of text, at least two elements have some relation to each other, this relation is spatial. Even when the minimal model of *dynamic* texts assumes that *element 2* replaces *element 1* at some interval exactly at the same position – in this case, the change is sufficient to mark the spatial position as unchanging; when other system components change, remaining unchanging can be interpreted as significant.

So the fundamental question about spatial modeling is not *what* modeling systems are based on space and what are not (they all are) but the question is about the relation between models that have *different internal structure*. This difference calls for the Lotmanian concept of semiosphere (as the total sum of multiple spatial models) and the notion of borderline as a bilingual mechanism of translation between different structures (see Lotman 2005[1984]: 210). Another question would concern the ‘evolutionary stages’ in the development of sign systems. In the “Conflicting modalities” article I shortly discussed the divergence between Tartu-Moscow semiotic school (TMS) and Thomas Sebeok: TMS considered natural language the primary modeling system and literature with other systems of artistic expression as secondary or supralinguistic systems that rely upon language (Lotman et al. 2013[1973]: 72). This idea was contested by Sebeok who pointed out that in human species, language was adapted only at a certain phase of evolution and therefore, zoosemiotic modeling systems were primary, underpinning the usage of language (Sebeok 1991: 334, 335). Perhaps the most curious or even somewhat paradoxical in Lotman’s semiotics is that both linguistic and spatial modeling systems can be reckoned as primary in a parallel manner; and – from yet another viewpoint, taking Lotman’s and Sebeok’s views altogether, both linguistic and zoosemiotic modeling systems rely upon the spatial one with its more abstract and primeval distinctions.

While Randviir notices the similarity between spatial and linguistic sign systems (Randviir 2010: 203), it can be presumed that the sign- or code systems that precede linguistic ones are spatially organized and this organization strongly influences the linguistic conceptualization of the world. For instance, Lakoff’s and Johnson’s theory of metaphor as the predominant cognitive tool that pervades our language and “governs our everyday functioning” (Lakoff, Johnson 1980: 3) is just one possible intermediate link in the chain between zoosemiotics and language proper⁵³. For just an example, according to Lakoff and Johnson, “given that a bounded physical space is a CONTAINER and that our field of vision correlates with that bounded physical space, the metaphorical concept VISUAL FIELDS ARE CONTAINERS emerges naturally” (Lakoff, Johnson 1980: 30).

⁵³ See also another angle on the topic in Terrence Deacon’s “The Symbolic Species” – a discussion that relates human brain evolution to the development of symbol-type, arbitrary signification (Deacon 1997).

It is feasible to agree that such abstract spatial thinking precedes the symbolic signification and verbalization.

An indirect support for this claim would come on the basis of the cross- or transmodal proto-code we discussed above; the *lingua franca* or zero-level modeling system of neuronal connections. Here, the key factor is that this proto-code is also spatially organized. A classic example of the spatial organization of perception is offered by the concept of ‘somatosensory homunculus’ (in Latin, “the little man”): in human primary somatosensory cortex, there is a neuronal “map” containing arrays of cells that have mapping, i.e. spatially corresponding connections to the skin’s mechanoreceptors that detect touch, pressure and vibrations. Although this correspondence is distorted and liable to be reorganized to a certain extent, the distortion can be explained by certain functions (e.g. hands are represented in the somatosensory map bigger than naturally are because they interact with the environment in a more detailed way than feet) (see e.g. Parpia 2011, Kaas 1991).

In a similar manner, as Luiz Pessoa claims, the responses of anterior insula reflect the internal states of the body “in a very real sense generating a map of “feelings from the body”” (Pessoa 2022: 106). Neuroanatomical research has disclosed that visual perception follows a comparable logic of spatial mapping. As Pessoa sums it up,

Retinal projections to the superior colliculus are topographic, meaning that the spatial layout of light hitting the eye (left/right, up/down) and triggering retinal responses is preserved in the colliculus. Cells in the colliculus thus form a map of the external visual space, allowing the colliculus to “know” where objects are in the world. (Pessoa 2022: 34)

An interesting implication for zoosemiotics is also highlighted by Pessoa:

The stimulus’s position in the visual field plays an important role here. In small rodents, unexpected movement overhead (much like that of a predator) more likely triggers flight, whereas movement in the lower field (possibly a prey) more commonly elicits approach. (Pessoa 2022: 36)

It can be suggested that not only perception, but also the organization of memories is partly map-based. Neuroscience has associated the area of hippocampus in the brain both to episodic memory and spatial navigation; e.g. it contains “place cells” that fire at special spatial locations during navigation. The function of hippocampus has been interpreted as a cognitive map that helps to form predictions about space, and consequently takes part in prospective planning (Ekstrom, Ranganath 2017). It is also important to notice that despite the idea that hippocampus generates a memory map where spatial cues and episodic memories are mutually linked, in hippocampus...

the cells do not always respond in a manner consistent with a spatial representation. Thus, place fields are distributed heterogeneously in space, their locations are determined by nongeometric information, the population of active cells can indicate more than one location in space, and finally, hippocampal cells encode stimuli independent of their spatial location. From these data, we argue that to the extent that the hippocampus encodes a map, it is more simply described as a memory map than a spatial map. Rather than computing spatial locations, the space the hippocampus encodes is better described as a life or a problem space that encodes the patterns of experience (Shapiro, Eichenbaum 1999: 365).

In spatial modeling, the question of proximities and distances comes forth as highly relevant. According to Olaf Sporns,

a network architecture that has received an extraordinary amount of attention over the past decade and is particularly relevant for the brain is the so-called “small world” /.../ Small-world networks are found in many social, technological, and biological systems, and their non-random attributes are of central importance in neural systems as well. In the brain, small-world architectures arise because of characteristic non-random features of the connection topology, specifically the existence of modules and hubs. (Sporns 2011: 3)

While this thesis already briefly touched upon small-world networks above (p. 22), it suffices to recall their pivotal characteristic for the context of our discussion. Namely, in such networks we can discern two fundamental connection types: short-distance connections that are in abundance and link to nearby nodes (i.e. neurons), and long-distance connections that are significantly fewer. Due to this distribution, there are many highly interconnected neuronal clusters or “modules”⁵⁴ in the brain, while these clusters are mutually connected by long-distance links. On the basis of recent analyses of structural connectivity, Danielle Bassett and her colleagues suggest that such networks, evolved by natural selection, have taken their form in physical space cost-efficiently (Bassett et al. 2010). First of all, such cost-efficiency arises from the fact that the highly functional global network satisfies with much reduced number of costly long-distance connections. Also, structurally modular networks are characterised by high adaptivity: “the system can evolve or adapt to new information one module at a time, without risking loss of function in modules that are already well-adapted” (Bassett et al. 2010: 7).

Despite that this step would be remarkably bold and tentative, I would draw an analogy between the two “spatial strategies” (building short- *versus* long-distance connections) and the two “semiotic strategies” (conducting most sign

⁵⁴ It is crucial to point out the difference between structural and functional modularities. Here, speaking of brain’s topography, we refer to structural modularity. While we discussed functional modularity of the brain above, concurrently we recognized a shift of paradigm that entails withdrawing from functional modularity in favor of high functional connectivity. In other words, it can be stated that in the brain, structural modularity supports functional non-modularity.

operations in an automatized, code-driven manner *versus* occasionally shifting to the conflict-centered, explosive semiosis that enables renewing the code and generating new information). The functional strategy of making short-distance connections most of the time, i.e. in the time when they satisfy organisms' need and applying for more costly long-distance connections in special situations follows a biologic algorithm that has developed on the background of constant concern for energy management. It can also be hypothesized that the influences that come from the outside of a clustered subsystem play the pivotal role: the communication between the local modules are required because of the specific conditions (or the change of conditions) in the larger environment. The long-distance signalling is substantially associated with choices that have to be made. The principal binary choice pertains to neuron itself: to fire or not to fire. Inside a module or a cluster of neurons, a "consensus"-based choice "has to be made" about initiating the long-distance connection.

In cultural communication, the requirement for choice is also imposed from the outside of the internal textual sphere (whose internality, by default, is defined by isomorphic organization that strives to avoid contradictions). The encounter between an interpreter and a text can be seen as a contact between two textual spheres. A multimodal conflict in a text requests a choice of an interpreter. In a sense, conflict-laden or defamiliarizing texts simulate the dynamic characteristics of an environment that induce the necessity to update one's existing codes and construct new ones. Generally speaking, it corresponds to Lotman's (1977a: 283) characterization of the process that generates novel information.

A range of semiotic problems would also arise regarding the controversiality of text's borders. Concerning the viewpoint of an interpreter, where a text begins and ends is not always clear. The issue is notably addressed by Gerard Genette's notion of paratextuality (among other concepts of inter-, hyper-, and metatextuality). Paratextual elements are liminal, i.e. they belong to text's border areas: e.g. books's cover designs, film posters and trailers are elements that do not exactly belong to the core of the text but influence the interpretation of text's meaning. According to Genette, "A paratextual element /.../ necessarily has a location that can be situated in relation to the location of the text itself: around the text and either within the same volume or at a more respectful (or more prudent) distance" (Genette 1997[1987]: 4). This passage clearly indicates some possibilities of spatial modelling of "exotextual" dynamics, i.e. how on certain conditions, the cues for textual interpretation come from the outside.

The third article of this dissertation, "On the concept of deceptive trailer", discussed how film trailer as a paratext becomes a functional model of film; in most cases as Lisa Kernan and Keith Johnston indicate, the main goal of a trailer is to communicate the information about film's genre to viewers (Kernan 2004, Johnston 2009). A special attention was paid to the trailer type that has a multimodal conflict (that may involve both semiotic modes and sensory modalities) embedded inside. In the typology of modality relations that was proposed in "Semiotic mode and sensory modality in multimodal semiotics", the first article

of this dissertation, it figures as an example of “conflicting” type. In the analyzed trailers, various trailer’s multimodal elements (for example visual and auditory modalities in the trailer of Alejandro G. Inarritu’s “Birdman”) communicated different genres in a confusing manner. In such occasions, a multimodal conflict outside of a text would invite an interpreter to make a choice about the text itself. When characterizing this process spatially, at least three intersecting spaces can be distinguished: the interpreter’s, paratext’s and the text’s own space. An optional characteristic of the spatial model would be the distance between these spaces. Presumably, when the interpreter is more “close” to the text, a distant paratext has less influence on her choices and decisions. When, inversely, interpreter’s and paratext’s spaces closely intertwine, the distant text has less active role in interpreter’s mind and its meaning is apparently more strongly influenced by the paratext.

Concludingly, open-ended situations of conflicts, choices, semiotic explosions or defamiliarizations would be eventually transformed into more linear structures before new bifurcations happen. In the words of Frank Kessler, “the novelty that characterizes the emerging medium, provokes wonder and dazzles those who encounter it for the first time. Once this effect has worn off, the utilitarian aspect of the media technology comes to the fore, and then it is experienced as some kind of a “transparent” carrier for the semiotic material that it mediates” (Kessler 2010: 77). According to Tom Gunning, “a cinematic device – for instance, the point of view shot in early film or the zoom in 1960s films – moves from having an effect of strong defamiliarization and novelty to becoming routine or even cliché. /.../ Cinema in its first decades moved from appearing as a scientific novelty, an attraction, to becoming a medium involved less in displaying itself than a vehicle for narrative or information, a means of communication rather than an attraction in itself” (Van den Oever, Gunning 2020: 17). In a similar manner, devices such as internal diegetic sound (IDS) that were discussed in the article of “Conflicting modalities in feature film” have been moving from the field of novelty to the realm of habituation, their function and meaning becoming fixed in filmic codes. While IDS technically offers a great example of multimodal conflict – we hear a person speaking but see that her mouth is closed – it is also an example how the habituation has depleted its potential as a semantic conflict; anyone who is familiar with the code interprets IDS firstly as an internal speech of a character.

Furthermore, it has to be noticed that besides activating an interpreter, a conflict-laden text can self-reflexively indicate towards its own construction principles. As Tom Gunning elsewhere remarks, “when a tool works, we pay it no attention; it seems to disappear. However, if the tool breaks down, if in some way it doesn’t function, it suddenly becomes conspicuous” (Gunning 2003: 46). On the one hand, self-reflexivity can also be approached by spatial means, seeing it as distancing that enables new viewpoints; on the other hand, it would automatically require some of the heightened activation on the behalf of an interpreter.

To sum it up, **a simple and tentative hypothesis is forwarded that will function as a framework to the problem of conflicting modalities in film and other audiovisual media.** More widely, it would also address multimodal conflicts in natural and social environments. In a sense, the historical relations between Tartu-Moscow school and cybernetics are reflected in proposing of this idea. Despite of its generality, it would allow a closer interconnection of the basic level theories of meaning-making and experimental sciences, also contributing to a closer dialogue between bio-, socio- and cultural semiotics. The hypothesis consists of five stages:

- 1) The concept of space is bound to energy expenditure: overcoming long distances requires energy, maintaining close distances (when interchanging both information and material components) saves it.
- 2) Because of energy concerns, automatism is a default procedure in nature and culture. It is reflected in sign processes that have the fundamental purpose of refraining from energy expenditure. The strive towards automatism is concurrent with habituation and establishing fixed codes.
- 3) Life and culture as dynamic processes have to adjust with contextual / environmental changes: in order to persist, an internal change is necessary. For this reason, change-driving mechanisms are required; these are contrary to habit-inducing mechanisms. Such processes are automatism-breaking, defamiliarizing and conflictual. Automatism-breaking processes are energy expensive; they have special, explosive characteristics and special purposes – first of all, activating interpreters and creating novel information.
- 4) Compared to automatic practices, explosive or conflictual processes do not last long. The trend towards habituation and crystallization of codes works as a suppressive or deterring force on these.
- 5) Semiosis, both in nature and culture, is characterized by an oscillation between two forces: habit-creating and habit-breaking dynamics.

3. CONTRAPUNTAL SOUND AND INTERNAL DIEGETIC SOUND (IDS): VARIOUS USES IN CINEMA

The third and final chapter of the dissertation mainly functions as an appendix; it is shorter and displays less of a theoretical character. Partly, it is a historic overview that addresses the development of the devices of *contrapuntal sound* and *IDS*. It is also an attempt in film analysis with an aim to touch upon five eras or “stylistic clusters” in a chronological sequence. First stop of the journey will draw attention to Luis Buñuel’s early experiment with IDS; the second will discuss the multimodal construction of Hamlet’s monologue in Laurence Olivier’s “Hamlet”; the third visits some bright examples of experimental sound editing by François Truffaut and Jean-Luc Godard in the context of French New Wave Cinema. The fourth section observes an ironic use of music in Stanley Kubrick’s “Clockwork Orange” and the fifth subchapter looks at a more sophisticated use of the modality conflict, namely how it communicates subtle meanings and advances the narrative in Krzysztof Kieślowski’s “Three Colours: Blue” and Martin Scorsese’s “Bringing Out the Dead”.

For a context of this five-stop expedition, a short manifest-disguised-as-an-article by Sergei Eisenstein, Vsevolod Pudovkin and Grigori Alexandrov can be recalled. “A Statement on Sound”, published in 1928, appeared in the middle of theoretical discussions that followed the introduction of sound to the cinema. The Russian filmmakers, touched by the revolutionary ideology and hegelian-marxist dialectics (see more about Eisenstein’s philosophy of synthetic filmmaking in the article of “Conflicting Modalities in Feature Film”) stated theoretical premises for the future evolution of sound film. Doing this, they expressed criticism towards the use of sound in Western cinema, stating that the sound which is coherent with visuals and represents dialogue in a lifelike manner may lessen the perfection of cinema as an art, even destroying the culture of montage. The authors claimed that only *contrapuntal use of sound* will afford new ways for the development of montage, calling up for *a distinct non-synchronization with the visual images* (Eisenstein et al. 1977[1928]: 257, 258).

3.1 Buñuel’s “L’Age D’Or”: From contrapuntal to internal diegetic sound

Despite of the bold invitation to contrapuntal editing, filmmakers’ enthusiasm towards Eisenstein’s call to dissociate sound from image was rather moderate. Still, the new possibilities and dangers of sound were widely discussed among theoreticians. The multiplicity of viewpoints settled down somewhere between two opposites: the Bazinian approach to sound as a natural ingredient in the illusion of reality and the forementioned Eisensteinian stance. In practice, the first avenue was chosen *en masse*: the use of auditory and visual as mutually supporting modalities, or the “logical codification of sound promoted by Hollywood”

(Richards 2008: 31) established itself as an overwhelming standard. On the other hand, a few directors took interest in non-synchronized sound for the sake of developing film techniques or following some broader artistic programs. First, I turn my attention to one of the first examples of experimenting with *contrapuntal sound*.

Luis Buñuel's scandalous "L'Age D'Or" (1930) marks the period of rapid transition from silent film to sound film⁵⁵. Compared to Hollywood, the French film industry was more resistant to convert, having fears of losing silent film aesthetics, and carried by overall aversion towards any standardization as well (Richards, 2008: 24). For example, in "L'Age D'Or", intertitles still play significant role, supporting, although mockingly, a bulk of the narrative. In the context of this thesis, I just want to point out a single episode of the film as an early use of conflicting modalities⁵⁶, where in audio track we hear a character's speech but visually don't see him or her speaking. Even more, this device requires a camera setup and framing that is proper for a dialogue – the face of the alleged speaker is in close-up and clearly focused. The striking difference is a deliberately closed mouth, a sign of non-speak, even a tear in the fabric of reality. As a synthesis of two confronting sign systems, this device attains a strong meaning-making potential. In film theory, such inner monologue has been called *internal diegetic sound* (IDS) (Bordwell et al. 2019: 291). In Buñuel's film, approx. 00.51.00, not only a monologue, but a dialogue between a man and a woman is conveyed in such a manner.

Man: "Are you cold?"

Woman: "No, I was falling."

Both sit in a garden, romantically caressing while the question and answer are delivered. Taken out of the context, this scene could express romantic mood and mental closeness where a higher level of mutual understanding emerges, not unlike telepathy. Nevertheless, as we could expect from the surrealist Buñuel, events take an unexpected turn. "*I've been waiting so long for him,*" the female voice says, face still not speaking but just smiling. "*What joy! What joy in having killed our children.*" The twist serves a mental dish for an avid psychoanalyst, restoring the irony and subversion towards love and family as integral parts of conventional discourse. Thus the comprehensive meaning of the segment depends on the context: surrealists considered themselves revolutionaries against the norms of society and its approved practices.

⁵⁵ Alan Crosland's "Jazz Singer" (1927) is considered the first feature-length sound film or "talkie", although various ventures of synchronizing sound and image has been made earlier. See also Abel, Altman: 2001.

⁵⁶ As Richards points out, Buñuel's experiments with sound were not necessarily contrapuntal. Eisensteinian montage strived towards dialectical constructions in the context of a coherent diegetic world, but the aim of surrealists was a revolt against stable diegesis itself (Richards 2008: 27).

In most cases, IDS represents character's inner speech, being heard only "in the head" of a single character and, of course, the viewer. As such, it conveys character's thoughts as they happen in the time and place of the diegesis (Horton 2017: 194, 195). As a device of meaning-making in film, it helps to convey personally experienced, subjective, Merleau-Pontiesque space (Huvenne 2017: 51). In that sense, Buñuel goes a step beyond, conveying not only internal monologue or thoughts made audible, but a telepathic dialogue and its satirical subversion, simultaneously.

3.2 Olivier's "Hamlet": IDS marking narrative's turning point

Laurence Olivier's "Hamlet" (1948) was the first sound adaptation of the play in English. Olivier directed the film and also starred as Hamlet. The controversial film was highly decorated with awards, e.g. Golden Lion from Venice and Academy Award for the best picture, but also criticized by purists for leaving out almost half of the Shakespeare's original text, and consequently, some plot elements. I will briefly analyse the scene of Hamlet's monologue, which also uses IDS, but interestingly, not throughout the monologue. The transfer between IDS and "normal" speaking, in this case, functions as narrative punctuation, emphasizing a turning point in the story.

While in auditory modality the monologue (starting approx. 01.02.30) is presented constantly, with its rhythm and intonations, in visual modality the "image" swings between Olivier speaking and acting in a close-mouthed mode, which provides a slightly estranging effect. The style of auditory presentation is soft and rather intimate, adopting to the medium of film and opposing to the declamatory style of theatre, or as Patrick Cook observes, it is "voiced with more attention to meaning than musicality" (Cook 2011: 25).

During the first dozen of seconds of the monologue, camera shows roaring sea waves that quickly dissolve to the eyes of Olivier, first out of focus, then sharp. After this, a sudden cut to medium close-up of Olivier / Hamlet sitting on a rock, both seen and heard speaking. Thus, at this moment, both modalities have still worked in synchrony, supporting each other. Additionally, in the background, the distant noise of waves is accompanying the monologue from the beginning to end. With the lines "*...take arms against a sea of troubles... and by opposing, end them*", Olivier slowly but still demonstratively draws a small dagger from his belt, holding it at the height of his chest. Then, quiet and sombre music starts on the soundtrack. In that moment, the most interesting happens. Approx. 01.03.10 the modality disruption occurs. Olivier shuts his eyes, and with his mouth kept closed now, auditory modality continues to carry on the monologue. Consequently, the words "*To die. To sleep no more*" become internal diegetic sound (IDS), although there already is some internality in the situation of a man speaking loudly to himself in a lonely place.

But now, the camera dollies in, framing more closely Olivier's troubled and dreaming face, eyes closed, small sweatdrops covering forehead, revealed as highlights. Reaching again to the similar close-up as before, namely actor's eyes, with the lines "...it is a consummation devoutly to be wished. To die. To sleep. To sleep". Then, at 01.03.35 the slow music suddenly bursts into a violent quaver of string instruments. This can, depending on the viewer, both eerily and comically suggest an alarm clock. After a couple of seconds of intensive music, Hamlet "wakes up", opens his eyes suddenly, stirs his body and resumes to declamate his monologue in both auditory and visual modality. The music quickly subsides and stops, leaving only the murmur of the sea in the background of words. At 01.05.05 as the monologue sequence nears its end, Hamlet drops the dagger down the cliff, not unlike a conductor lowering his baton after directing a music piece. The film is cut to the next shot, deep downward angle towards the waves and dagger seen falling into the abyss. Then, cut back to Hamlet, who delivers a few more lines, at the same time rising up from the rock, looking briefly down the cliff and walking away from the camera, towards the fog and down the invisible stairs. The music silently starts again, image fades to black, concluding the scene.

As one of possible interpretations, I suggest the staging of the monologue can be seen in the Jungian paradigm of hero's journey (supposedly, more or less intentionally taken into account by Olivier and even by Shakespeare⁵⁷), namely reflecting the stages of *the refusal of the call* and the *crossing of the first threshold* (e.g. Campbell 1993: 59, 77). In that case, the disruption of modalities marks a *critical turning point* in the hero's progress: first the connotations of hesitation and instability; then a brief dreamlike episode where one reality (or view upon reality) transforms into the other, more decisive and active one. We should keep in mind that in the next segment, Polonius announces the arrival of the traveling actors, which offers Hamlet more specific ideas and tools for the revenge. During the liminal, transformative stage, the semantic discontinuity between auditory and visual modalities correlate with Hamlet's hesitation, his wavering between life and death, action and resignation. When the distance has been finally overcome, the hero has fresh objectives, and the narrative has moved into the next phase.

3.3 Truffaut and Godard: "Shoot the Piano Player" and "Bande à part"

The French New Wave is characterized by its stylistic inventiveness; most of its filmmakers were also theoreticians whose critical activities were closely bounded to the magazine of *Cahiers du Cinéma* and connected to the popular French

⁵⁷ In that case reflecting more general principles of drama and storytelling which were later interpreted by Edward Tylor, Lord Raglan, Otto Rank, Carl Jung, Joseph Campbell and others.

tradition of film clubs. For the directors of New Wave, experimenting with the medium was not so much for experiment's sake or with avanguardistic intentions but with a purpose to widen the possibilities of filmic expression. This is exemplified by the frenchmen's curiosity towards Hollywood masters, especially Alfred Hitchcock. As Truffaut relates, while in New York in 1962, American journalists kept asking him, *Why do the critics of Cahiers du Cinema take Hitchcock so seriously? You love "Rear Window" because /.../ you know nothing about New York!* Astonished Truffaut replied that "Rear Window" isn't about New York but cinema. It was obvious to him that Hitchcock has given more thought to the potential of his art than any of his colleagues (Truffaut 1985: 11) and Truffaut's interest, first of all, was directed towards the form of Hitchcock's work, not the content.

3.3.1 "Piano Player": a sign consisting of signs

Truffaut's "Shoot the Piano Player" (1960) can be considered with an intent to highlight a couple of simple but important methodological questions pertinent to multimodal analysis, striving for a tentative path towards answering these. Discussing the film, Anne Gillain remarks that "there seems to be a radical rupture, from the very first scene, between the visual track and the soundtrack /.../ The visual track conveys a story about gangsters, the soundtrack a story about love" (Gillain 2013: 53). This remark will lead us to a circle of interesting problems. One of the most eminent is, how and whether such a statement can be verified? And also, when a modality is naturally seen as consisting of signs, could we say that modality as a whole functions as a sign, too?

As Lotman repeatedly emphasized the double potential of text, e.g. a certain homology between an element of an artistic model and the model itself (e.g. Lotman 2011 [1967]: 261) or texts' hierarchical architectonics – "Culture as a whole may be considered as a text. However, it is exceptionally important to emphasize that this is a complex text, which consists of a hierarchy of "texts within the texts" (2009 [1992]: 77) – we can provisionally envision a modality as a certain type of text within film as well. Film as a bimodal whole can be seen as a text of the higher level of hierarchy. On the other hand, we could see modes as lower-level texts inside a modality. In such case, it is possible to accept a situation where a modality functions as a sign, as well.

The next question would be, what makes a modality to represent something? In the current case, what makes a modality to represent a genre? On the one hand, the answer largely relies upon the genre theories, but these are not in the focus of this dissertation⁵⁸. Consequently, as our interest is towards the methodology of multimodal analysis, I propose the most relevant question would be, how does the internal structure of a modality facilitate representing a genre? Or, more specifically, how are modes structured in this modality, and how this

⁵⁸ It suffices to notice there are various complementary ways to define a film's genre, the content or topic of the film is just one of them.

affects meaning-making? If the soundtrack of “Piano Player” is about love, what modes are present and do they all convey the topic of love? It has to be kept in mind that there is a technical possibility of conflicting modes inside a single modality, as well.

For our preliminary observation, inside the auditory modality of “Piano Player”, a division between music and verbal component would suffice⁵⁹. In some films, the mode of sound effects or foley art has a significant influence of the meaning-making, but not here. Here we may ask, first and foremost, whether film music can convey a narrative or signify a specific topic, e.g. criminality or love. Nevertheless, attaching a solidly fixed meaning to a musical piece or motif is problematic. As musicologist Philip Tagg notices, “precision of musical meaning can never be the same as precision of verbal meaning. Music and language are not interchangeable sign systems” (Tagg 2013: 171). Thus it would suffice to acknowledge the capacity of music to *suggest* genres by connotative means. On the other hand, let us hypothesize that the verbal component has a certain edge over the music in conveying a topic or suggesting a genre.

There is a possibility that in verbal mode or submode of dialogue, different characters represent different “topics”, attitudes or emotions, and the verbal mode can therefore be a synthesis of this multiplicity. Chronological dimension could also be relevant: a character’s text can go through thematic changes. As for our attempt to evaluate “Piano Player’s” auditory modality and its capability to carry the topic of romance, let us consider its dialogue’s relation to the suggested themes.

The central character of the film is Charlie Kohler (Charles Aznavour), a musician who has retired from big stages for tragic personal reasons to play piano in a small nightclub. Indeed, as Gillain above referred, the main conflict in the character (and consequently, the main conflict of the story) lies in the contradictory relationship between Charlie’s love interest and his family ties: his brothers are connected to criminal world. Naturally each character’s lines should follow their own topic in some measure, thus it is impossible to generalize the dialogue to the talk of all characters on equal grounds. Thus the focus should be kept on Charlie as the protagonist.

Yet, right in the beginning of film, two strangers, having met at the night-time street, start discussing marriage immediately. Neither of them is Charlie (who is not yet introduced), and their identity is only partially unravelled by their dialogue and by the fact that one was pursued and the second helped him (this is the visual track that conveys the topic of criminality, as Gillain noticed). They are strangers both to each other and to the viewer. After all, the situation feels natural: *“I don’t know who you are. I may never see you again, so I can*

⁵⁹ In a such type of analysis, if there is a need to get into details, a natural question would be about submodes, for example diegetic or extradiegetic music. The verbality or verbal mode in film could be seen consisting at least of dialogue, voiceover and written text, which in turn can be similarly divided as diegetic and extradiegetic written text. The verbal mode could be supported both by auditive and visual modalities, in turn.

speak freely”, says one of the guys to another, who later turns out to be Charlie’s brother. As the beginning of a text is heavily marked (see Lotman 1976: 9), any information placed here has stronger influence on the meaning of the whole. So, the first minutes of the film seem to assure Gillain’s apprehension.

Although romantic topics regularly recur in various conversations, they do not embrace the whole “dialogic space”; only approximately third of it (it is my estimation, not a result of measuring the duration of the lines). Leaving alone other characters, neither Charlie’s lines are exclusively connected to the topic of love. Depicting him, Truffaut frequently uses inner speech expressed by internal diegetic sound (IDS) which was already discussed above⁶⁰. The first episode with Charlie’s IDS commences about 00.12.15, when he has helped his brother to escape from the bar, proceeding his work behind the piano. Aznavour’s voiceover delivers the lines, while the actor is playing the piano in visual modality, his head in close-up, with serious expression, mouth closed and eyes fixed on the room as trying to look into an unknown space. As a professional, he doesn’t need to look at the keys. On the soundtrack, besides his thoughts, also music and ambient noise of the bar is heard. Worrying about what happened to his brother, who was pursued by two suspicious men, Charlie speculates: “*They looked as if they wanted to talk business, not dispatch him.*”

This brings forth the theme of family ties: these aren’t usually a basis for any specific genre in film narratives, but still a crucial component. Even more, a lot of plot intrigues emerge out of family relations; consequently the topic can function as a connective between various genres or other topics (as in this case, the motifs of criminality and love life). Here we should notice how closely connected can be different topics inside a modality.

Very soon (approx. 00.15.05) follows the next IDS episode where Charlie walks with a girl. On the soundtrack, romantic extradiegetic music accompanies them. Both are initially framed in a medium close up, then we see in detail Charlie’s hand which tries to get hold of the girl’s, but without success. His romantic interest and hesitation are clearly explicated in the visual modality, supported by the music. Then, through inner speech, he addresses himself: “*Your silence must seem odd. Say something, anything, otherwise she’ll think you’re scared of her*”. Charlie proceeds to conceive his options until the girl notices that the same suspicious men are following them. Here again, suddenly enters the gangster topic in both modalities equally.

These examples are sufficient to show that in the filmic meaning-making, the combination of modalities and main topics is not done in the one-to-one principle, but is rather an intricate process where a topic can quickly change or dissolve into another topic inside a single modality or mode. Consequently, such claims as Gillain made have mostly the value of generalization. There is also a certain amount of viewer’s subjectivity that has to be taken account of. In the multimodal organization of film, different viewers may have their personal inclinations towards which textual element to consider significant. Considering

⁶⁰ It is also a way to emphasize character’s status as a protagonist.

this, the connection between modalities and topics is partly the subject to statistical evaluation. Nevertheless, in the meaning-making where one modality represents a certain topic and another modality represents another, we can envisage a powerful mechanism not only for activating, confusing or estranging viewers, but creating hybrid genres and complicated, multi-themed narratives which more effectively reflect the complexity of life.

3.3.2 “Bande à part”: a minute of silence

When reflecting upon the filmic expression of French New Wave, the work of Jean-Luc Godard comprises an entire dimension of that movement. In contrast to Truffaut, Godard was a metafictional author who negotiated the frames of his works and deconstructed the borderline between film and viewer. This is perhaps most vividly exemplified by a scene of “Pierrot le Fou” (1965), where Jean-Paul Belmondo’s character suddenly turns towards the camera and addresses the viewer. And not only so: when Anna Karina’s character asks whom he talked to, he answers, *spectateur*.

Discussing Godard and *Pierrot le Fou*, David Bordwell emphasizes the semiotic aspect and relates: “Filmic storytelling is a matter of signs /.../ viewers are sign-readers. /.../ Hollywood has created codes of character behavior, linear ordering, and smooth shot-matching. Godard has arranged his scene in a way that violates the codes – and perhaps creates a new code of his own” (Bordwell 2012). As the filmmakers of the New Wave were keenly interested in certain American directors whom they claimed creative authors with unified personal vision (e.g. Staiger 2003: 34, 36), it prepared the ground for conscious dialogues with other auteurs and their personal “codes”. Approaching Godard’s style, I propose that even more definitive is his dialogue with certain topics and themes (for example gangster motifs of Hollywood), as well as satirical replays of bourgeois clichés, stereotypical characters and situations. Such content is, if we remind Eisenstein’s notion and the central interest of this article, represented in twisted codes of expression.

Due to the limits of this article, I will only briefly look at two scenes from “Bande à part”⁶¹ that was released a year earlier (1964). Anna Karina, who was Godard’s wife in 1961–1965 and the lead actress in several of his films, here plays Odile, a young woman who meets two guys in an English language class. She discloses that there is a big sum of money hidden in a villa she lodges in, and the three start to plan of stealing it. With its love triangle, good guy / bad guy opposition and criminal suspense, the film could be approached as a lighter sort of entertainment. Nevertheless, some multimodal meaning-making techniques complicate and sophisticate the code, urging the viewer towards unexpected regimes.

⁶¹ In North America, it was released under the title “The Band of Outsiders”.

Both episodes approach the sound and image synthesis not with the method of addition but subtraction. The purpose of Godard's experiment is to cut the sound not just from the image but from the *whole*; in order to discover new meanings that start filling that vacuum. Both scenes belong to the same segment⁶²: they are connected to a single location, a bar or *salon* where film's three heroes enter in the daylight. The passing of the time is not marked, and the mood will get livelier, perhaps turning towards the night, because when the characters leave the bar in the end, it is already dark outside. Nevertheless, Arthur (Claude Brasseur), Franz (Sami Frey) and Odile start to feel bored at a small table just smoking and drinking. "*If there is nothing to say, lets have a minute of silence,*" says Franz. "*You can really be dumb sometimes,*" Odile snaps. "*A minute of silence can be a long time,*" Franz answers. "*A real minute of silence takes forever.*" Odile seems to agree and exclaims: "*Okay. One, two, three...*".

As a result, the whole soundtrack shuts off (it happens approx. 00.46.15). Not only the characters keep their mouths closed, but the ambient sounds, clatter and talking in the bar are totally muted. With auditory modality excluded, *another viewing regime / reception practice starts to emerge*. Now, the viewer can allocate more attention to the visual details, minor gestures of actors and nuances of their body language. A lot of things become significant that weren't significant before: how Franz holds a burning cigarette, how he and Odile look at each other. Nevertheless, this experiment doesn't much affect the perception of the narrative of the current film, but relates to the viewing practices in a general sense. Besides the estranging effect which activates subjective meaning-making in viewer, the subtraction of auditory modality diminishes cognitive load and allows to shift the cognitive effort towards the visual processing. Still, that new regime has not enough time to establish itself. Franz, again, interrupts the silence with "*Okey. That's enough. I'll put a record on*", jumps up, takes his hat and leaves the table. Simultaneously, all background voices from the soundtrack return. The "minute of silence" lasted about 35 seconds. The characters, in a sense, are transformed into something else as casual film characters: they become united with director's intentional, metadiegetic action, with an illusionary power to control aspects that are out of reach of film characters.

Quite soon, around 00.48.00, Godard plays another trick. The guys and Odile have stood up and went in the middle of the floor where is an area free of tables and seats. They stand side by side, Odile in the middle. Focused in a medium shot, they commence Madison Line Dance (Granados 2020: 98) to a tune that is presented as diegetic, mixed with the bar noises, hand-clapping and foot-tapping of the dancers. The camera, set on a normal height of standing person's head, slightly moves along dancer's movements as an empathic spectator: the distance between the viewer and action is diminished, the viewer is

⁶² With the notion of "segment" I keep in mind Christian Metz's syntagmatic units, i.e. "autonomous segments" or sequences that may contain several shots but pertain to the unity of time and location (Metz 1991: 123; 1974: 189).

almost invited to dance along. Around 00.48.32, the music is abruptly cut. This, in itself, does not contradict the acknowledged rules of the narrative world as the previous cut of the total auditory modality did. Maybe someone has shut off the record player? Yet, our three protagonists continue dancing as nothing had happened.

Here, a similar subtractive operation is accomplished. While the previous “trick” cut off entire modality, now just one of the modes is eliminated. I propose this works similarly on the viewer, activating cognitive processes and drawing attention to other modes. The noise of the bar, dancers’s claps and footsteps feel amplified. Although not so strongly as before, such device could heighten the attention towards visual modality, if it would last longer. After the music has been off for three seconds, a voice-over begins. A man narrates: “*Now is the time for digression in which to describe our heroes’ feelings.*” Then, music starts again. After 15 seconds of music and dancing, the music is cut again. Dancers continue, and voice-over continues as well: “*Arthur keeps watching his feet, but his mind is on Odile’s mouth and her romantic kisses.*” Then music again. Such a pattern is repeated a couple of times until another characters’ thoughts are introduced.

To such use of multimodality, I would like to point out two interpretations. First, a viewer could accept the code as a the new normative, at least in the context of this film or the discourse of Godard’s filmic expression. In such a case, the device would not draw attention to itself any more; the code becomes transparent and allows the viewer a direct access to narrative, naturalizing events and characters. As an alternative, the multimodal conflict retains its disturbing potential and works as metalanguage, drawing attention to the construction of the text, reminding that the characters are puppet-like entities without their own agenda; they belong to the mechanism of narrative and express the intentions of the storyteller / *auteur*. What is most important, these interpretations do not necessarily contradict each other. The viewer can fluctuate between the both regimes.

3.4 A Clockwork Orange

3.4.1 A multimodal translation

Anthony Burgess, the author of the eponymous novel Stanley Kubrick adapted, was a self-taught musician and composer himself. The music was recurrent theme in his literary work, and “A Clockwork Orange” can be claimed to be the most music-related source Kubrick ever adapted (Gengaro 2013: 104). A modern fable in the form of picaresque black comedy, ironic of human nature and society, the film is also a metalevel commentary on the visual communication and the meaning of music itself.

In 1972, Kubrick told in an interview to the Rolling Stones magazine,

I'd say that my intention with *A Clockwork Orange* was to be faithful to the novel and to /.../ see the violence from Alex's point of view, to show that it was great fun for him /.../ it was necessary to find a way of stylizing the violence, just as Burgess does by his writing style. The ironic counterpoint of the music was certainly one of the ways of achieving this. All the scenes of violence are very different without the music. (Kubrick in Nelson 1982: 134)

Therefore, the choices in the filmmaking, at least partly, were imposed by the need for the translation of Burgess's distinctive style. As Christine Gengaro observes, "in the novel, music serves a largely symbolic function. It is arguable that music, in Kubrick's adaptation, is more important than it was in Burgess's novel, for it is in the film that the music is transformed into a structural element, an aesthetic standard, or an emotional hook. It ceases to be simply a symbol" (Gengaro 2013: 108). Such intersemiotic or intermodal translation – from a single modality to a multimodal structure – reveals not only the possibilities of storytelling by affective means, but the potentialities of intertextual connections where the pieces of music, dependent on the viewer, refer to the specific works of specific composers, or just to a general notion of "classical" music with its connotations. Equally in both cases, the juxtaposition of classical music with visuals of violence feels unusual and incites a certain semantic conflict.

The contrapuntal use of music is applied on two levels: firstly as a characteristic device for the text as whole, and secondly, as a principle of construction for several specific scenes. I suggest, in both cases the semantical distance (if we apply Lotmanian space-oriented model discussed above) diminishes in the course of the film. The connection between violence and the "musical sublime" establishes itself as a certain code of that specific fictional world, first as unexpected, later more familiar. Next, I will briefly analyse the meaning-making in two key scenes, and where necessary, draw connections outside of the scenes, because Kubrick's film is a remarkably integrated text which components have strong relations to the whole.

3.4.2 Singing in the Rain

The first unit in meaning-making I want to discuss can be approached in two ways, first as an episode, second as a musical leitmotif. Approximately 00.11.10, Alex (Malcolm McDowell) and his gangmates have infiltrated a cosy home of an elderly writer and his wife, starting to abuse the couple. The central piece of this scene is about 9-second long static full shot where McDowell mimicks the dance of American actor and choreographer Gene Kelly. While his companion maltreats writer's wife, Alex sings a tune from Kelly's musical; while dancing, in the place of an iconic feet movement, he kicks his prostrate victim. First and foremost, it is an intertextual link to the musical "Singing in the Rain" (1952, dir. by Stanley Donen and Gene Kelly) via both modalities, auditory with the modes of melody and lyrics, and visual modality via the mode of dance or choreography.

The lyrics of “Singing in the rain”, which originally expressed happiness and creative hope, are now semantically reversed. Here, the song represents a joy of destruction. Alex, still singing, smashes writer’s desk, pulls down his bookshelves, and lastly, rapes his wife. Yet, I suggest that such meaning-making can still be seen deeper than a simple correspondence between the bright audio and dark visual. This multimodal construction is clearly made to function as an ironic commentary. I propose that Kubrick, more or less knowingly, gets down to a ladder of meanings which has several steps or signifiers. Thus, the deconstructive commentary⁶³ works towards, from the most specific to the most abstract: 1) the Gene Kelly’s song < 2) to the musical as such < 3) to the part of (bourgeois) society that cherishes musical < 4) to the habit in culture of creating “light” and popular narratives as the mechanism of forgetting, glossing over violence, death, inequality < 5) to the practice of signification as such.

Besides the sound of a specific scene, “Singing in the Rain” appears again in the film, this time as Gene Kelly’s original version. Gene Phillips has drawn attention that the film also ends with the song, and it symbolizes Alex’s return to violence as the happy ending (Phillips 2001: 157). The song as leitmotif clearly facilitates of widening the concept of irony to the film as whole. Even more, Kubrick has put it in the service of narrative progression. Alex, unsuspectingly humming the same tune later, betrays himself to his former victim, a writer he had tortured.

3.4.3 Beethoven’s 9th

The second juxtaposition under discussion has a similar character: first it is an auditory component of the specific scenes, secondly a repeating motif. Approximately 00.19.20, Alex relaxes in his home, listening to Beethoven 9th symphony. While music is playing, the editing assembles a row of symbolic visuals as a live snake, a life-sized graphics of a naked woman hanging on the wall, and, mostly importantly, underneath the picture, four identical statues of Jesus Christ, naked, with crucifixion wounds and thorn crowns, but without crosses. The Jesuses, one hand raised and another set on the shoulder of the adjacent figurine, allude not only to the passion but the collective rebellion, and also express a sort of rocking pose similar to four guys having fun in a concert or celebration.

Then, the editing spurts and, following the symphony’s tempo, in quickly alternating shots reveals Jesus’s wounds in detail. Then cut to Alex in close-up, and his voiceover, “*Oh Bliss. Bliss and Heaven! It was gorgeousness and gorgeousity made flesh.*” Alex’s facial expression and movements slightly hint that he is masturbating. The binarity of *sacred – sexual* is represented here, and the

⁶³ This interpretation is supported by the mode of costumes. Intruders remind of eerie, weird type of clowns with satyric allusions. A clown as a character can be seen as a trickster. In turn, the trickster is, first and foremost, a disruptor of social order, a mediator between opposites, a wild force that brings new things (e.g. new information) to the world through the destruction of old structures.

music becomes the third component, the mediator. Moreover, in a short while, there follows a sequence of shots depicting violence. The images from Alex's flow of consciousness are presented visually: hanging of a woman, explosions, Alex's own face with bloody fangs, heavy boulders crashing onto helpless people, a volcano erupting (the last one as an allusion to ejaculation). The element of destruction, once more, is added to that construction of mindscape.

In another place, approx. 01.15.30, Beethoven 9th reappears. Here, it more distinctly expresses one of the film's main ideas: the struggle between psychological conditioning and free moral choice as the competing principles of this fictional world. Now, Beethoven's music is transposed to synthesizers and the visual modality features a documentary footage of Nazis, parading and waging war. All this is presented to Alex in one of the "therapy" sessions. The source of the soundtrack is first left ambiguous – a viewer can initially hypothesize it comes from Alex's mind. The distortion of the familiar melody widens the semantic gap; the effect of disruption is amplified when the vocal is heard. Lastly, Alex's voiceover reveals that his favorite music was deliberately attached to the unpleasant visuals. *"Then I noticed, in all my pain and sickness... what music it was that, like, cracked and boomed. It was Ludwig van. Ninth Symphony. Fourth movement."*

As a concluding remark, I want to make a brief comment on the voice-over, which is one of the meaning-making devices in Kubrick's film. A voice-over of the character, who is shown, *but not shown speaking*, can also be considered technically an instance of multimodal conflict. But as the voice-over is rather a well-matured convention of filmic expression, there is no conflict in a functional sense. Here, the similar mechanism works as was referred above: when a multimodal conflict is repeated and established as a code, it becomes a conventional part of filmic expression, losing its potentiality as a device of *ostranenie* or the generator of new information.

3.5 Kieślowski and Scorsese

For the conclusion of the section of multimodal analysis, I would briefly stop at a scene from Krzysztof Kieślowski's "Three Colours: Blue" (1993) and a motif from Martin Scorsese's "Bringing out Dead" (1999), in order to highlight some additional nuances of conflicting modalities.

3.5.1 To whom the notes belong to?

Kieślowski's "Blue" demonstrates more subtle ways of storytelling than previously discussed films, a delicate approach to music⁶⁴ and audiovisual metaphors. Multimodal constructions have important part here. As Marek Haltof notices,

sound effects also play a powerful role /.../ the noise of the garbage truck crushing Patrice's musical score (as if crushing the music – the noise overtakes the music), the sound when Julie breaks the window in the hospital. The loud sounds often accompany the flashes of blue and the flashes of Patrice's concerto /.../ These flashes of blue and the bursts /.../ emphasise the importance of memory. (Haltof 2004: 130)

Although the multimodal meaning-making in "Blue" is worth a thorough discussion, the limits of this chapter allows me to turn attention only to a short audiovisual metaphor.

Julie, the film's protagonist (Juliette Binoche), loses her child and husband in a car accident. The late husband, a renown composer, leaves behind several burdens. An unraveling affair with his secret lover intertwines with the mystery behind his unfinished score of the piece that celebrates European unity. There is a suggestion that the main effort of writing that piece came from Julie herself.

The scene of interest begins ca 00.18.30. In an extreme close-up, we see sheet music and hear a piano playing. Only the center, approximately 1/8th of the screen is clearly visible. This is not an ordinary shallow focus, but achieved with the technique of masking the sides of lens. It allows to see one note at a time, and as the camera moves from left to right, the visual and auditory representation of the notes is delivered in synchrony. This lasts for ten seconds, then a cut to Julie in profile and close-up. On the background, we see a grand piano with its lid open. The interesting nuance here is that Julie does not seem to play at the moment, but seems to hold the partiture in hand and look at it. While the camera lingers on Julie for some seconds, we continue to hear piano in auditory modality. Then cut back to the partiture; the focusing and camera movement repeat what we saw before. After some seconds, cut to Julie again, now camera angled upwards from a lower position. She, in frontal view, looks downwards, not towards the camera, but to the notes. The distant sound of an airplane, hardly recognizable, is mixed into the audio track. Once again, a cut back to the partiture, and the similar movement from left to right, sliding over separate notes.

⁶⁴ Kieślowski's close relation to classical music was facilitated by his long-time cooperation with Polish composer Zbigniew Preisner who worked on a lot of scores of his films. Besides music, Preisner created a fictitious character of Van den Budenmayer, an eighteenth-century Dutch composer "whose" music Preisner himself wrote (Haltof 2004: 114) and who is referred in several Kieślowski's films. Preisner's *Song for the Unification of Europe* is attributed to the character of Patrice in Kieślowski's "Blue" and plays an important role in the story.

At this moment, a remarkable modality conflict is created. The notes in visual modality end; the camera follows empty lines now. In auditory modality, piano sounds continue, now without visual coupling. When the next shots return to Julie, it is indeed revealed that her hands do not play the instrument. Hence, the scene opens up Julie's subjectivity; the modality conflict is in the center of a construction that represents character's consciousness. It indicates that Julie knows the piece extraordinarily well and suggests two further conclusions. First, with certainty, she is able to continue her husband's work. Second, it subtly implies that she has actually written his masterpiece herself.

3.5.2 Bringing out the borderline

Finally, I want to return to a use of IDS in Martin Scorsese's "Bringing Out the Dead" (1999). Here it serves as a modest device, although used by an experienced filmmaker whose knowledge of the styles and approaches of filmic expression is encyclopaedic. "Bringing Out the Dead", as Vincent Lobrutto observes, revisits the moods and matters of "Taxi Driver" (1976), carrying the dark atmosphere of New York City that belongs more to the 70's than the 90's (Lobrutto 2008: 365, 366). The protagonist of the film is paramedic Frank (Nicholas Cage), who falters on the verge of burnout. At least partly his spiritual crisis is induced by a memory of a girl whom he did not manage to save.

Scorsese, a devout Catholic as a young man, was troubled by the themes of guilt and redemption, which were later revisited in several of his films, "Bringing Out the Dead" included (Lobrutto 2008: 37). Director's search for harmony is paralleled with his character's existential journey, which, at first glance, rather runs in circles. Frank's crisis is accompanied by the addiction to alcohol and drugs, and haunted by personal visions. Occasionally, out of stress or addiction, he hears some dead or unconscious people talk to him. This is represented by conflicting modalities. Actors playing such people keep their mouths closed and their voice come from the soundtrack.

Approx. 01.21.15, Frank stops by an emergency ward to see an old guy who is the father of the film's female protagonist. As the context implies, the man is in coma. He lies motionless, with his eyes closed. Frank feels his pulse and looks at the heart rate monitor. When the film is cut back to Frank's face in close-up, a man's voice with a slight echo effect says: "Go to the bank, boy!". Here, a fast cut to the patient follows: he is now staring at Frank as trying to say something. The voice continues immediately: "Take out everything you can!". As the patient's face and the verbal line are represented in synchrony, a conflict is facilitated between visual and auditory modalities. When camera returns to Frank's face, we are assured that he, by chance, wasn't the one who said this. Instead, this is a *reaction shot* of a dialogue which conveys Frank's astonishment and terror. "Mr. Burke?" he answers tentatively, his speaking visually confirmed. Then, cut back to the patient, still staring Frank in the eyes. "I am going. I had

enough,” he continues in auditory modality. The heart rate monitor starts to give emergency signal and a nurse rushes in to start resuscitation.

Here, an important nuance of IDS is well illustrated: the viewer has to know that the verbal text comes from a person who visually does not say it. Usually the necessary information to confirm this relation is provided by the close context or more precisely, the syntagmatics of the text. The verbal lines should be, as a rule, semantically relevant to the specific character, not to speak of the character’s presence in visual modality at the time of speaking. In the scene described above, though, the multimodal meaning-making is not based on the strict and full conflict. The modes or submodes in visual modality are also set into contradictory positions. The submode of actor’s gestures suggests consciousness and agency, if taking into account his glance. Although he is not visually speaking, it cannot be claimed as a conflict between (sub)modes *per se*, because *not speaking* and *glancing* are not contradictory in itself. The conflict fully emerges only in audiovisual whole, where soundtrack demands the visual confirmation of speaking.

Besides of the scene described, the figure of expression where dead people speak in auditory modality and are silent in visual, appears elsewhere in the film. More generally, we should also ask about how this metaphor facilitates the network of meanings in the narrative as a whole. As Frank’s visions are strictly personal, such episodes cannot be interpreted as the straight contacts with “spirits” as in horror genre. Usually Scorsese (as several good directors) leaves the genre of his films loosely specified: the umbrella-term *drama* is often Hollywood’s option to negotiate the tradition of genre film. Here, the uncanny and ghostly, expressed by a modality conflict, suggests a liminal reality, a borderline state not in the narrative world but of Frank’s troubled *psyche*.

3.6 IDS and contrapuntal music in the context of the two-layered framework

The birth of the film medium was surrounded by a cultural environment that was inherently multimodal, filled with sounds and interaction. As Mervyn Cooke points out, live music attended films in most places, especially when the touring motion picture attractions were presented in vaudeville theatres and music halls. The earlier theatrical traditions (opera, ballet and melodrama) had already well established the symbiosis between music and drama (see Cooke 2008). The cultural codes that organized such multimodal synthesis were consequently adapted or modified by filmmakers. While in film, there is an indefinite range of possibilities to combine various sounds, dialogues, melodies etc with moving and also still images, the previous examples of audiovisual conflict focussed on sound film era and discussed two central semiotic modes – verbal expression or natural language (IDS) and music (contrapuntal or incongruent film music). Both modes are evidently carried by auditory modality.

Having methodological concerns in mind, it is important to acknowledge that the borderlines between semiotic modes are somewhat disputable because of their social constructedness. An example of this is provided by the use of the notion of ‘contrapuntal’ sound and Michel Chion’s criticism toward its use. Chion believes that film theory uses musical ‘counterpoint’ as a too loose metaphor, divorcing it from its musical meaning. Therefore, according to Chion, multi-modal conflicts in film should be addressed rather as ‘audiovisual dissonances’ and the notion of ‘true’ counterpoint should be reserved for the special cases (Chion 1994, 36, 37) as the following example:

There exist hundreds of possible ways to add sound to any given image. Of this vast array of choices some are wholly conventional. Others, without formally contradicting or “negating” the image, carry the perception of the image to another level. And audiovisual dissonance is merely the inverse of convention, and thus pays homage to it /.../ For an example of true *free counterpoint* consider the amazing resurrection scene in Tarkovsky's film *Solaris*. The hero's former wife, who committed suicide, comes back to him in flesh and blood on a space station, thanks to mysterious forces summoned forth by a brain-planet /.../ Over these images Tarkovsky had the imagination to dub sounds of breaking glass, which yield a phenomenal effect. We do not hear them as “wrong” or inappropriate sounds. Instead, they suggest that she is constituted of shards of ice. (Chion 1994: 39)

Therefore, in the film analyses of this chapter, the notion of *contrapuntal* was used provisionally, acknowledging that it can characterize various semiotic modes, e.g. somewhat loosely refer both to *contrapuntal sound* and *contrapuntal music*. It can be claimed that one factor that incited Chion’s criticism is hardly noticeable but nevertheless impactful: by constructing the metaphor, the property of being counterpointal was taken out of its original context where it works only in the confines of musical mode (an element of music has a counterpointal relation to another element of music) and transferred into the context where it does not only characterize various modes (sound effects and dialogue in addition to music) but also steps into a juxtaposed relation with the modes belonging to a different sensory modality (visual images). This transfer can be seen as the deeper cause for the dispute. Despite of this plurality, the previous film analyses illustrated that the metaphor of counterpoint can be consistently used across different modes when it is anchored in the solid ground of (auditory) *modality*. As such, these examples hopefully work also as evidences of the flexibility of the two-layered framework of semiotic modes and sensory modalities.

CONCLUSION AND FUTURE DIRECTIONS

The research of multimodality is concerned with objects whose meaning is communicated by multiple sign systems simultaneously. Therefore, the semiotic theory of multimodality⁶⁵, with the task of explaining meaning-making by such objects, has to address a situation where various information channels or codes are processed altogether. What are the exact research units of a multimodal theory, would depend on the researcher's focus. In general, two broad approaches can be distinguished. The sociosemiotic school of Kress and his colleagues was foremost inspired by Halliday's linguistics and Gibson's theory of affordances. Its central units are *semiotic modes* – bundles of material properties and social practices that shape the meaning-making capacity of a text or a medium. In Kress's own words, modes are *semiotic resources*. For example, *written text* and *picture* work together as semiotic modes in the medium of book, newspaper or printed advertisement. This approach can be observed in the context of the material turn in the humanities, as, for example, has done Bateman and his colleagues (Bateman, Wildfeuer and Hiippala 2017), who have developed Kress's theory further, complementing the notion of semiotic mode.

Another viewpoint would concentrate on perception and interpretation – the practices of meaning-making that happen at the end of the communication chain. Any act of interpretation is inseparable from the emotional and cognitive processes of the interpreter. For this reason, a complete theory of meaning-making requires an integration with the knowledge that is offered by psychology and cognitive neurosciences. In these paradigms, when perception is discussed, *sensory modality* is one of the pivotal notions. Therefore, multimodality, seen through the lenses of cognitive disciplines, is expressed through the activity of multiple brain circuits that correspond to different sensory channels. As an alternative to the prioritization of materiality, cognitive approaches have influenced humanities through the bodily turn.

It is important to bring out that there is an intrinsic difference between modes and modalities: sensory modalities are strictly limited and subordinated to biological organization; modes are shaped by their historical and socio-cultural background. Therefore, modes vary across cultures, and there is a constant problem of sub-segmentation of modes, e.g. if we consider *image* as a semiotic mode, should we consider *drawing* and *photograph* separate modes, submodes or something else? As Jewitt and her colleagues point out, "to avoid potential confusion, it is important to make a deliberate decision on what categories and terms to use when engaging with multimodal research" (Jewitt et al. 2016: 12).

⁶⁵ A precursory idea for such a theory can be recognized in the Theses of the Tartu-Moscow School, exemplifying that "individual sign systems, though they presuppose immanently organized structures, function only in unity, supported by one another. None of the sign systems possesses a mechanism which would enable it to function in isolation" (Lotman et al. 2013[1973]: 53). Accordingly, the semiotics of culture can be seen as "the study of functional correlation of different sign systems" (ibid.).

Another source of problems that haunt the field of multimodal methodology is the inconsistent and interchanging use of the terms *mode* and *modality* by many authors.

The initial point of departure for this thesis was the author's interest towards the mechanisms of meaning-making in film. The consequent aspiration of seeking out the up-to-date methodology for film semiotics encouraged to look towards cognitive film theory and the related disciplines, e.g. cognitive neuroscience. As explaining meaning-making in film and other audiovisual texts relies strongly upon the understanding of multimodal processes, the first provisional half of the dissertation explored fundamental issues that concern the framework of multimodal research and also general semiotic theory. The dissertation is characterized by the progression from the theory towards the objects, i.e. multimodal conflicts in film that appear, most of the time, as incongruences between the information in visual and auditory modalities.

The most general goal of this dissertation was to strive for an integration between mode and modality based views; the attempt for the synthesis was the main objective of the research article "Semiotic mode and sensory modality in multimodal semiotics: making difference and building complementarity between the terms", which is written as the second in the chronological order, but in the structure of the dissertation is the first because of its theoretical focus. To overcome the problem of the ambiguity of semiotic modes and the inconsistent use of the terms, the author proposed a two-layered framework that combines the constancy of modalities and the flexibility of modes. The organization of sensory modalities can be approached as a foundational layer under the system of modes; modes can be seen as 'carried' by modalities. In order to illustrate the functioning of the framework, a typology of modality relations with corresponding examples was built in the above-mentioned article. The typology included supporting, modifying, conflicting, substituting and cross-activating relation types between sensory modalities.

The second article that was written last of all, "Conflicting modalities in feature film: from contrapuntal editing to internal diegetic sound", discussed one specific relation type in this typology, namely conflicting modalities and their synthesis in the textual whole. Systemic conflicts are intriguing in several ways: they help to reveal the working principles of mechanisms and have the potential to activate participants, changing their *modus operandi*. Semiotic conflicts can also be approached in the context of Russian Formalist (especially Shklovsky's) concept of *ostranenie* (defamiliarization) and Lotman's idea about mutually incongruent sign systems or languages that constitute a mechanism for creating new information. While Eisenstein talked about the *juxtaposition* of image and sound in the context of contrapuntal film editing, the potential for conflict still greatly varies in such synthetical practices. Because of that, in the article, I weighed the possibilities of detecting and measuring a multimodal conflict with the conclusion that neither *redundancy* nor *synchrony* can be considered as a basis for such measuring (with an initial hypothesis that conflict means breaking redundancy or synchrony between modalities). Following

Altman and Chion, it can be stated that auditory and visual modalities are never mutually redundant, and paradoxically, synchrony is necessary for modality conflicts.

Instead, I found that Lotman's concept of *space* as an alternative primary modeling system can potentially help in modelling multimodal conflicts. The idea was observed in a more detailed manner in the cover article's second chapter. Spatial models, in addition to the fundamental conceptualization of semiotically active borders that are central in defining the spaces, usually contain the parameter of *distance*. Distance, in turn, can be correlated with energy concerns of a subject or a system. Therefore, an analogy can be drawn between two "spatial strategies" (building short and long-distance connections) and the two "semiotic strategies" (conducting sign operations in an automatized, code-driven way and engaging in a conflict-based, explosive type of semiosis that enables renewing the code and generating new information). It can be pointed out that the brain's architecture already follows certain principles of spatial mapping, including the organization of neurons in a principle that is similar to *small-world* networks. While such networks consist of substantially more energy-efficient local connections and significantly fewer costly long-time connections, both connection types are mutually complementary and their proportion is evidently functional.

As a hypothesis, I propose that the background of Lotman's idea of spatial modeling allows us to draw an analogy between above-mentioned particularities of brain architecture and the dynamics in semiotic processes that are characterized by an oscillation between two opposing forces. On the one hand, constant habituation leads the meanings (e.g. the connection between a sign vehicle and an interpretant) to be eventually crystallized in a code that works rather automatically – this has happened, for instance, to the device of internal diegetic sound that has depleted its conflictual potential and become standardized in the context of filmic expression. On the other hand, conflicts as explosive processes can break up codes and cause their renewal.

The third article "On the concept of the deceptive trailer: Trailer as paratext and multimodal model of film" was written first, but as it is the most object-centered one, it has been allocated the last position in the structure of the dissertation. The article discussed an instance of conflicting modality relation type, namely film trailer as a paratext (see Genette 1997) and a multimodal model of film that enables modeling of several characteristics of film. Comparing those characteristics, it can be found that often one characteristic among others is emphasized; following Jakobson's notion, it can be called trailer's *dominant*. Comparing films and their trailers, it can be detected if a film and its trailer share the same dominants. In the cases of the contradiction between trailer's and film's dominants, the trailers can be labeled as deceptive. According to Kernan and Johnston, usually the most important information that a trailer communicates is about film's genre. Deceptive trailers are made with an intention to communicate multiple genres or a more popular genre than the film actually belongs to; despite of the hope that it benefits film's box office, it often leads to the disappointment of viewers.

The cover article of the dissertation attempted to avoid the duplication of the research articles; generally it addressed the same topics but took a complementary angle. While the second chapter addressed the spatial modeling of modality conflicts, as pointed out above, the first subchapter approached multimodal conflict on the background of the concept of *intersemiotic translation*; then it discussed the relation between the notions of multimodality and multi-mediality, and thirdly, touched upon the knowledge about crossmodal integration that has been provided by cognitive neurosciences. While one of the central problems of intersemiotic translation concerns the possibility of “translating” between different sensory modalities, the research of the mechanisms behind crossmodal integration offers explanatory potential in addressing this problem. I propose that transmodal- or *amodal invariants* (as highlighted in Stein, Meredith 1993) comprise a pivotal category of elements that can possibly enable such a translation type. Amodal invariants can also be associated with a hypothetical interspecial protocode that has been indicated by Maran. This protocode would convey meanings by creating distinctions in contrast, movement, symmetry and other characteristics that can be equally communicated through multiple sensory modalities.

To sum it up, the research of complex systems, both biological and cultural, implicates the necessity for interdisciplinary approach. This dissertation, to a certain extent, followed this direction, complying with the conviction that film semiotics benefits from the support of cognitive film research or neurocinematics – and vice versa – cognitive and experimental branches benefit from the semiotic frameworks that focus on the aspect of meaning. Concerning the wider context of perception and cognition studies, conflicting modalities should also be approached in the light of recent research that emphasizes the holistic models where the integrated, top-level meanings sometimes “override” the impact of individual sensory modalities and media. A similar complementarity can be strived for between cultural-, socio-, and biosemiotics. In the discussion of conflicting modalities, biosemiotics can provide a fundamental mechanism for the explanation how culture languages work – while habituation creates codes, conflicts provide a mechanism for deautomatization, i.e. renewing these codes or creating new ones.

For the future directions and developments, I would suggest that introducing the notion of *transtextual multimodality* would benefit the discussion of the situations where meaning-making encompasses the co-influence of multiple texts. The problems surrounding paratexts illustrate this well: a paratext can rely upon one sensory modality and the “core” text another; on certain conditions, they comprise a textual whole in the mind of an interpreter and therefore this whole is the result of *transtextual multimodality*.

Potentially, an intriguing discussion would also concern the situations where multimodal wholes start to function self-referentially, engaging in an internal dialogue or self-description. Another question for the future can be posed about the functional relationship between multimodal components and media: if an additional multimodal component (both mode or modality) comes into play,

does the textual structure, in general, gets more rigid or more flexible across different media? What factors can influence the inclination in either direction? And finally, if certain text types tend to attach themselves to certain medium types, have modes or modalities a role to play in this?

Apparently, a variation between code-driven, highly automatized functioning and occasional explosive defamiliarizations characterizes all semiotic processes. After the bifurcations that were discussed in the last essays of Juri Lotman, or the stage where a system is faced with new choices that determine its perspectives, a period of habituation allows the choices to be repeated and crystallized into codes that benefit energy-efficiency. As a contrast, conflicts are costly. Sometimes they come unexpected; sometimes they are induced with specific purposes in mind. The key of managing multimodal conflicts is to recognize and evade the catastrophic ones. While these strip our agency for ever, beneficial conflicts would take that agency for a while but give it back for good. Perhaps the confidence, regained after choice-making, emerges already from the system of the next level.

REFERENCES

- Abel, Richard; Altman, Rick (eds.) 2001. *The Sounds of Early Cinema*. Bloomington and Indianapolis: Indiana University Press.
- Aitchinson, Laurence; Lengyel, Máté 2017. With or without you: predictive coding and Bayesian inference in the brain. *Current Opinion in Neurobiology* 46: 219–227.
- Alderson-Day, Ben; Fernyhough, Charles 2015. Inner speech: development, cognitive functions, phenomenology, and neurobiology. *Psychological Bulletin*, Vol. 141, No. 5, 931–965.
- Aktulum, Kubilay 2017. What is Intersemiotics? A Short Definition and Some Examples. *International Journal of Social Science and Humanity*, Vol. 7, No. 1: 33–36.
- Anderson, Joseph D. 1996. *The Reality of Illusion. An Ecological Approach to Film Theory*. Carbondale and Edwardsville: Southern Illinois University Press.
- Baker, Rosemari, Elizabeth 2010. *Shklovsky in the Cinema, 1926-1932*. Masters thesis, Durham University. Available at Durham E-Theses Online: <http://etheses.dur.ac.uk/378/>.
- Baldry, Anthony; Thibault, Paul 2006. *Multimodal Transcription and Text Analysis*. London and Oakville: Equinox.
- Barabasi, Albert-László 2002. *Linked. The New Science of Networks*. Cambridge: Perseus Publishing.
- Barsalou, Lawrence, W. 2009. Simulation, situated conceptualization, and prediction. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364: 1281–1289.
- Bassett, Danielle; Greenfield, Daniel L.; Meyer-Lindenberg, Andreas; Weinberger, Daniel R.; Moore, Simon W.; Bullmore, Edward T. 2010. Efficient physical embedding of topologically complex information processing networks in brains and computer circuits. *PLoS Computational Biology*. Vol. 6, Issue 4: 1–15.
- Bateman, John; Wildfeuer, Janina; Hiippala, Tuomo 2017. *Multimodality. Foundation, Research and Analysis*. Berlin and Boston, Walter de Gruyter GmbH.
- Bateman, John 2018. Peircean semiotics and multimodality: Towards a new synthesis. *Multimodal Communication*, 7(1): 1–24.
- 2021. Dimensions of materiality. Towards an external language of description for empirical multimodal research. In: Pflaeging, Jana; Wildfeuer, Janina; Bateman, John (eds.) *Empirical Multimodality Research. Methods, Evaluations, Implications*. Berlin and Boston: De Gruyter, 35–64.
- Bateson, Gregory 1952. The Position of Humor in Human Communication. *Macy Conferences 1952*. Available online: <http://pcp.vub.ac.be/books/Bateson-humor.pdf>.
- 2000 [1972]. The role of somatic change in evolution. In: *Steps to an Ecology of Mind*. Chicago and London: The University of Chicago Press, 346–363.
- Bennett, Tyler James 2021. Incompatibility, unlimited semiosis, aesthetic function. In: Sütiste, Elin; Gramigna, Remo; Griffin, Jonathan; Salupere, Silvi (eds.) *(Re)considering Roman Jakobson*. Tartu Semiotics Library 23. Tartu: University of Tartu Press, 149–163.
- Berlina, Alexandra 2016. Translator’s introduction. In: Shklovsky, Viktor. *A Reader*. New York: Bloomsbury Academic, 1–50.
- Bertelson, Paul; De Gelder, Béatrice 2004. The Psychology of multimodal perception. In: Spence, Charles; Driver, Jon (eds.) *Crossmodal Space and Crossmodal Attention*. Oxford: Oxford University Press, 141–178.
- Bezemer, Jeff; Kress, Gunther 2008. Writing in multimodal texts: a social semiotic account of designs for learning. *Written Communication*, Vol. 25, No. 2: 166–195.

- Bogdanov, Alexei 2005. Ostranenie, kenosis, and kialogue: The Metaphysics of formalism according to Shklovsky. *The Slavic and East European Journal*, Vol. 49, No. 1 (Spring, 2005), 48–62.
- Bordwell, David 1985. *Narration in the Fiction Film*. Madison: University of Wisconsin Press.
- 1991. *Making Meaning. Inference and Rhetoric in the Interpretation of the Cinema*. Cambridge and London: Harvard University Press.
- 2012. The Viewers Share: Models of Mind in Explaining Film. On: *David Bordwell's Website on Cinema*. Available online.
<https://www.davidbordwell.net/essays/viewersshare.php>. Accessed: 01.03.2023.
- Bordwell, David; Thompson, Kristin; Smith, Jeff. 2019. *Film Art: An Introduction. 12th Edition*. New York: McGraw-Hill.
- Brinthaupt, Thomas M.; Morin, Alain; Puchalsky-Wasyl, Malgorzata 2020. Editorial: Exploring the Nature, Content, and Frequency of Intrapersonal Communication. *Frontiers in Psychology*, Volume 11, October: 1–3.
- Calvert, Gemma A.; Bullmore, Edward, T.; Brammer, Michael J.; Campbell, Ruth; Williams, Steven C. R.; McGuire, Philip K.; Woodruff, Peter W. R.; Iversen, Susan D.; David, Anthony S. 1997. Activation of auditory cortex during silent lip reading. *Science* Vol. 276 April: 593–596.
- Campbell, Cary; Olteanu, Alin; Kull, Kalevi 2019. Learning and knowing as semiosis: Extending the conceptual apparatus of semiotics. *Sign Systems Studies* 47(3/4): 352–381.
- Campbell, Joseph 1993 [1949]. *The Hero with a Thousand Faces*. London: Fontana Press.
- Chandler, Daniel 2017. *Semiotics: The Basics. Third Edition*. London and New York: Routledge.
- Chion, Michel 1994. *Audio-Vision: Sound on Screen*. New York: Columbia University Press.
- Clark, Andy 2023. *The Experience Machine. How Our Minds Predict and Shape Reality*. New York: Pantheon Books.
- Combs, Richard 2012. Hitchcock Olympiad. *Film Comment*, Vol. 48, Iss. 6, (Nov/Dec). Available online: <https://www.filmcomment.com/article/hitchcock-olympiad>.
- Cook, Patrick J. 2011. *Cinematic Hamlet. The Films of Olivier, Zeffirelli, Branagh, and Almereyda*. Athens: Ohio University Press.
- Cooke, Mervyn 2008. *A History of Film Music*. Cambridge: Cambridge University Press.
- Damasio, Antonio 1999. *The Feeling of What Happens*. San Diego: Harcourt.
- 2003. *Looking for Spinoza. Joy, Sorrow, and the Feeling Brain*. London: William Heinemann.
- 2012. *Self Comes to Mind*. London: Vintage Books.
- Deacon, Terrence 1997. *The Symbolic Species. The Co-Evolution of Language and the Human Brain*. London: Penguin Books.
- 2013. *Incomplete Nature. How Mind Emerged from Matter*. New York: W. W. Norton & Company.
- Deaville, James 2016. Recut and re-tuned: Fan-produced parody trailers. *Journal of Fandom Studies*, Vol. 4 Number 2, 209–223.
- De Linde, Zoé; Kay, Neil 1999. Processing subtitles and film images. Hearing vs deaf viewers. *The Translator*, Vol. 5, Number 1, 45–60.
- Docan-Morgan, Tony (ed.) 2019. *The Palgrave Handbook of Deceptive Communication*. Cham: Palgrave Macmillan.

- Driver, Jon; Spence, Charles 2000. Multisensory perception: Beyond modularity and convergence. *Current Biology* Vol. 10 No. 20, 732–735.
- Dusi, Nicola 2015a. Intersemiotic translation: Theories, problems, analysis. *Semiotica* 206: 181–205.
- 2015b. Remixing movies and trailers before and after the Digital Age. In: Navas, Eduardo; Gallagher, Owen; burrough, xtine (eds.), *The Routledge Companion to Remix Studies*. New York: Routledge, 154–165.
- Eco, Umberto 1987. The influence of Roman Jakobson on the development of semiotics. In: Krampen et al. (eds.) *Classics of Semiotics*. Berlin: Springer, 109–127.
- 2001. *Experiences in Translation*. Toronto: University of Toronto Press.
- 2007[1981]. *The Role of the Reader. Explorations in the Semiotics of Texts*. London: Hutchinson & Co.
- Eisenstein, Sergei; Pudovkin, Vsevolod; Alexandrov, Grigori 1977 [1928]. A Statement on Sound. In: Leyda, Jay (ed.) *Film Form: Essays in Film Theory*. San Diego: Harcourt, 257–258.
- Ekstrom, Arne D.; Ranganath, Charan 2017. Space, time, and episodic memory: The hippocampus is all over the cognitive map. *Hippocampus*: 1–8.
- Elleström, Lars 2020. The Modalities of Media II: An Expanded Model for Understanding Intermedial Relations. In: Elleström, Lars (ed.) *Beyond Media Borders, vol 1. Intermedial Relations among Multimodal Media*. Cham: Palgrave MacMillan.
- Emerson, Caryl 2005. Shklovsky's ostranenie, Bakhtin's vnenakhodimost' (How Distance Serves an Aesthetics of Arousal Differently from an Aesthetics Based on Pain) *Poetics Today*, 26: 4, 665–696.
- Engelbrecht, Jüri 2021. *Complexity in Social Systems and Academies*. Cambridge: Cambridge Scholars Publishing.
- Feldman Barrett, Lisa 2020. *Seven and Half Lessons About the Brain*. Boston and New York: Houghton Mifflin Harcourt.
- Fricke, Ellen 2013. Towards a unified grammar of gesture and speech. A multimodal approach. In: Muller, Cornelia et al. (eds.) *Body - Language - Communication. An International Handbook on Multimodality in Human Interaction*. Berlin and New York: De Gruyter Mouton, 733–754.
- Foucault, Michel 1973. *Cesi n'est pas une pipe*. Paris: Editions Fata morgana.
- Fuchsberger, Verena; Murer, Martin; Tscheligi, Manfred 2013. Materials, materiality, and media. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2853–2862.
- Gallese, Vittorio; Guerra, Michele 2015. *The Empathic Screen. Cinema and Neuroscience*. Oxford: Oxford University Press.
- Genette, Gerard 1997[1987]. *Paratexts: Thresholds of Interpretation*. Cambridge: Cambridge University Press.
- Gengaro, Christine Lee 2013. *Listening to Stanley Kubrick. The Music in His Films*. Lanham: The Scarecrow Press.
- Gibson, James J. 2015 [1986]. *The Ecological Approach to Visual Perception*. New York: Psychology Press.
- Gignoux, Anne-Claire 2005. *Initiation à l'intertextualité*. Paris: Ellipses.
- Gilbert, Christopher J. 2013. Playing with Hitler: downfall and its ludic uptake. *Critical Studies in Media Communication* Vol. 30, No. 5, December: 407–424.
- Gillain, Anne 2013 [1991]. *François Truffaut: The Lost Secret*. Bloomington: Indiana University Press.

- Gorbman, Claudia 1987. *Unheard Melodies. Narrative Film Music*. Bloomington: Indiana University Press.
- Granados, Marlowe 2020. I Turn My Camera On. *The Baffler*, No. 54 (Nov-Dec), 96–103.
- Grandchamp, Romain; Rapin, Lucile; Perrone-Bertolotti, Marcela; Pichat, Cédric; Haldin, Céline; Cousin, Emilie; Lachaux, Jean-Philippe; Dohen, Marion; Perrier, Pascal; Garnier, Maëva; Baciu, Monica; Loevenbruck, Hélène 2019. The ConDialInt Model: Condensation, Dialogality, and Intentionality Dimensions of Inner Speech Within a Hierarchical Predictive Control Framework. *Frontiers in Psychology*, Volume 10, September: 1–30.
- Grodal, Torben 2009. *Embodied visions. Evolution, Emotion, Culture and Film*. Oxford: Oxford University Press.
- 2018. Docudrama and the cognitive evaluation of realism. In: Brylla, Catalin; Kramer, Mette (eds.) *Cognitive Theory and Documentary Film*. Cham: Palgrave Macmillan: 75–91.
- Grundy, Gareth 2011. Peter Saville on his album cover artwork. *The Guardian*, 29. May 2011. <https://www.theguardian.com/music/gallery/2011/may/29/joydivision-neworder>
- Guagnano, Giacinte Davido; Mininni, Giuseppe 2018. Translation as sign exploration: A semiotic approach based on Peirce. *Semiotica* 225: 129–140.
- Guillot, Marie-Noëlle 2010. Film Subtitles from a Crosscultural Pragmatics Perspective. *The Translator*. Vol.16, Number 1, 67–92.
- Haltof, Marek 2004. *The Cinema of Krzysztof Kieslowski*. New York: Columbia University Press.
- Hagen, Rainer; Hagen, Rose-Marie 2005. *What Great Paintings Say. From Bayeux Tapestry to Diego Rivera*. Köln: Taschen.
- Hebb, Donald O. 2002 [1949]. *The Organization of Behavior. A Neuropsychological Theory*. Mahwah and London: Lawrence Erlbaum Associates.
- Hébert, Louis 2019. *An Introduction to Applied Semiotics. Tools for Text and Image Analysis*. London: Routledge.
- Hertzberg, Ludvig (ed.) 2001. *Jim Jarmusch: Interviews*. Jackson: University Press of Mississippi.
- Hillesund, Terje; Schilhab, Teresa; Mangen, Anne 2022. Text materialities, affordances, and the embodied turn in the study of reading. *Frontiers in Psychology*, March 2022, Vol 13: 1–9.
- Hirsch, Galia 2019. Hitler’s out of Dope: A cross-cultural examination of humorous memes. *Journal of Pragmatics* 149: 25–39.
- Hoffmeyer, Jesper 1996. *Signs of Meaning in the Universe*. Bloomington: Indiana University Press.
- 2007. Semiotic scaffolding of living systems. In: Barbieri, Marcello (ed.), *Introduction to Biosemiotics: The New Biological Synthesis*. Dordrecht: Springer: 149–166.
- 2008. *Biosemiotics: An Examination into the Signs of Life and the Life of Signs*. Scranton and London: University of Scranton Press.
- Horton, Justin 2017. Sound, space, and complex narrative cinema. In: Boni, Marta (ed.) *World Building. Transmedia, Fans, Industries*. Amsterdam University Press: 187–203.
- Howse, P.E., Allen, J.A. 1994. Satyric Mimicry: the Evolution of Apparent Imperfection. *Proceedings of the Royal Society, Biological Sciences*, 257: 111–114.
- Huvenne, Martine 2017. Editing space as an audio-visual composition. In: Wildfeuer, Janina; Bateman, John (eds.) *Film Text Analysis. New Perspectives on the Analysis of Filmic Meaning*. London and New York: Routledge, 46–65.

- Iacoboni, Marco. 2009. *Mirroring People*. New York: Picador.
- Ireland, David. 2018. *Identifying and Interpreting Incongruent Film Music*. Cham: Palgrave Macmillan.
- Jakobson, Roman. 1970. Linguistics. In: *Main Trends of Research in the Social and Human Sciences. Part One: The Social Sciences*. The Hague: Mouton / Unesco, 419–466.
- 1971[1959]. On Linguistic Aspects of Translation. In: *Selected Writings. II. Word and Language*. The Hague: Mouton Publishers, 260–266.
- 1971b [1967]. Language in relation to other communication systems. In: *Selected writings II: word and language*, The Hague & Paris: Mouton, 697–708.
- 1971c [1964]. On the relation between visual and auditory signs. In: *Selected writings II: Word and language*, The Hague & Paris: Mouton, 338–344.
- 1985a [1956]. Metalanguage as a Linguistic Problem. In: Rudy, Stephen (ed.) *Selected Writings VII: Contributions to Comparative Mythology. Studies in Linguistics and Philology, 1972–1982*. Berlin: Mouton, 113–121.
- 1985b [1974]. Communication and society. In: Rudy, Stephen (ed.) *Selected Writings VII: Contributions to Comparative Mythology. Studies in Linguistics and Philology, 1972–1982*. Berlin: Mouton, 98–100.
- 1985c. Sign and System of Language. A Reassessment of Saussure's Doctrine. In: Pomorska, Krystyna; Rudy, Stephen (eds.) *Verbal Art, Verbal Sign, Verbal Time*. Minneapolis: University of Minnesota Press.
- Jewitt, Carey; Bezemer, Jeff; O'Halloran, Kay. 2016. *Introducing Multimodality*. London and New York: Routledge.
- Johnston, Keith M. 2009. *Coming Soon: Film Trailers and the Selling of Hollywood Technology*. London: McFarland & Company.
- Jullier, Laurent. 2010. Should I see what I believe? Audiovisual ostranenie and evolutionary-cognitive film theory. In: Annie van den Oever (ed.), *Ostranenie: on "Strangeness" and the Moving Image. The history, Reception, and Relevance of a Concept*, Amsterdam: Amsterdam University Press, 119–140.
- Kaas, Jon H. 1991. Plasticity of sensory and motor maps in adult mammals. *Ann. Rev. Neurosci.* 1: 137–167.
- Kernan, Lisa. 2004. *Coming Attractions: Reading American Movie Trailers*. Austin: University of Texas Press.
- Kessler, Frank K. 2010. Ostranenie, Innovation, and Media History. In: Van den Oever, Annie (ed.) *Ostranenie. On "Strangeness" and the Moving Image. The History, Reception, and Relevance of a Concept*. Amsterdam: Amsterdam University Press, 61–79.
- Kiss, Miklós. 2010. The Perception of reality as deformed realism. In: Van den Oever, Annie (ed.) *Ostranenie. On "Strangeness" and the Moving Image. The History, Reception, and Relevance of a Concept*. Amsterdam: Amsterdam University Press.
- Kitalong, Karla Saari; Miner, Rebecca L. 2017. Multimodal composition pedagogy designed to enhance authors' personal agency: Lessons from non-academic and academic composing environments. *Computers and Composition* 46: 39–55.
- Korsgaard, Mathias Bonde. 2020. Fake trailers as imaginary paratexts. *MedieKultur* 68, 107–125.
- Koskinen, Maaret. 2021. Bergman, writing, and photographs: the auteur as an ekphrastic ghost. In: Hedling, Erik (ed.) *Ingmar Bergman. An Enduring legacy*. Lund and Manchester: Lund University Press, 56–69.

- Kourdis, Evangelos 2015. Semiotics of Translation: An Interdisciplinary Approach to Translation. In: Trifonas (ed.) *International Handbook of Semiotics*. Dordrecht: Springer, 303–320.
- Kress, Gunther; Van Leeuwen, Theo 2001. *Multimodal Discourse. The Modes and Media of Contemporary Communication*. London: Arnold Publishers.
- Kress, Gunther 2010. *Multimodality. A Social Semiotic Approach to Contemporary Communication*. London and New York: Routledge.
- 2014. What is mode? In Carey Jewitt (ed.) *The Routledge Handbook of Multimodal Analysis*, 60–75. London: Routledge.
- 2020. Transposing Meaning: Translation in a Multimodal Semiotic Landscape. In: Boria, Monica; Carreres, Ángeles; Noriega-Sánchez, María; Tomalin, Marcus (eds.) *Translation and Multimodality. Beyond Words*. London & New York: Routledge, 24–48.
- Kring-Schreifels, Jake 2023. Sounds that cannot be unheard: Sound designer Johnnie Burn on The Zone of Interest. *Filmmaker Magazine*. Available online: <https://filmmakermagazine.com/123965-interview-sound-designer-johnnie-burn-the-zone-of-interest/>.
- Kull, Kalevi 2015a. Semiosis stems from logical incompatibility in organic nature: why biophysics does not see meaning, while biosemiotics does. *Progress in Biophysics and Molecular Biology* vol. 119, issue 3, 616–621.
- 2015b. Evolution, choice, and scaffolding: Semiosis is changing its own building. *Biosemiotics* 8(2): 223–234.
- 2016. Habits – semioses – habits. *Sign Systems Studies* 44(4), 623–629.
- Lakoff, George; Johnson, Mark 1980. *Metaphors We Live By*. Chicago and London: University of Chicago Press.
- Laos, Nicolas 2023. *Freemasonic Enlightenment in the Context of the Modern and Perfecting Rite of Symbolic Masonry*. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Lemke, Jay L. 1998. Multiplying meaning: visual and verbal semiotics in scientific texts. In: Martin, J. R.; Veel, Robert (eds.) *Reading Science: Critical and Functional Perspectives on Discourses of Science*. London and New York: Routledge, 87–113.
- Linask, Lauri 2024. *Autocommunication in the Semiotic Development of the Child*. [Dissertationes Semioticae Universitatis Tartuensis]. Tartu: Tartu University Press.
- Lipinski, Lisa 2019. *René Magritte and the Art of Thinking*. New York: Routledge.
- Lipkin, Steven N.; Paget, Derek; Roscoe, Jane 2006. Docudrama and Mock-Documentary: Defining Terms, Proposing Canons. In: Rhodes, Gary D.; Springer, John Parris (eds.) *Docufictions. Essays on the Intersection of Documentary and Fictional Filmmaking*. Jefferson: McFarland and Company, 11–26.
- Lobrutto, Vincent 2008. *Martin Scorsese. A Biography*. Westport: Praeger.
- Lotman, Juri 1976. The modelling significance of the concepts “end” and “beginning” in artistic texts. (Rosslyn, Wendy, trans.) In: O’Toole, Lawrence Michael; Shukman, Ann (eds.) *General Semiotics (Russian Poetics in Translation 3)*. Oxon Publishing, 7–11.
- 1977[1971]. *The Structure of the Artistic Text*. Ann Arbor: The University of Michigan Press.
- 1990a. Artistic space in Gogol’s prose. *Russian literature triquarterly* vol. 23, 199–241.
- 1990b. *Universe of the Mind*. London: LB. Tauris & Co Ltd.
- 2004. *Filmisemiootika*. Tallinn: Varrak.

- 2005[1984]. On the Semiosphere. *Sign Systems Studies* 33.1, 205–229.
- 2009 [1992]. *Culture and Explosion*. Berlin & New York: Mouton De Gruyter.
- 2011 [1967]. The place of art among other modelling systems. *Sign Systems Studies* 39(2/4), 249–270.
- Lotman, Juri; Ivanov, Vyacheslav; Pjatigorskij, Aleksandr; Toporov, Vladimir; Uspenskij, Boris 2013 [1973]. Theses on the semiotic study of cultures (as applied to the Slavic texts). In: Salupere, Silvi; Torop, Peeter; Kull, Kalevi (eds.) *Tartu Semiotics Library* 13. Tartu: University of Tartu Press, 53–77.
- Louhema, Karoliina; Zlatev, Jordan; Graziano, Maria; Van de Weijer, Joost 2019. Translating from monosemiotic to polysemiotic narratives: A study of Finnish speech and gestures. *Sign Systems Studies* 47(3/4), 480–525.
- Malpica, Sandra; Serrano, Ana; Gutierrez, Diego; Masia, Belen 2020. Auditory stimuli degrade visual performance in virtual reality. *Nature Scientific Reports* 10, 12363. Available online: <https://www.nature.com/articles/s41598-020-69135-3>.
- Maran, Timo 2008. *Mimikri semiootika*. Tartu: Tartu Ülikooli kirjastus.
- Marks, Laura 2000. *The Skin of the Film. Intercultural Cinema, Embodiment, and the Senses*. Durham and London: The Duke University Press.
- Martinet, André 1984. Double Articulation as A Criterion of Linguisticity. *Language Sciences*, Vol. 6, no. 1, 31–38.
- Massaro, Dominic W. 1999. Speechreading: Illusion or window into pattern recognition. *Trends in Cognitive Sciences*, Vol. 3, no. 8, August: 310–317.
- McGilligan, Patrick 2003. *Alfred Hitchcock: A Life in Darkness and Light*. New York: HarperCollins.
- Merleau-Ponty, Maurice 1962. *Phenomenology of Perception*. London: Routledge and Kegan Paul.
- Mesz, Bruno; Trevisan, Marcos A.; Sigman, Mariano 2011. The Taste of Music. *Perception*, Vol. 40, 209–219.
- Metz, Christian 1991[1974]. *Film Language. A Semiotics of Cinema*. Chicago: Chicago University Press.
- McGurk, Harry; MacDonald, John 1976. Hearing lips and seeing voices. *Nature* 264 (5588), 746–748.
- Mitchell, Thomas W.J. 1995. *Picture Theory: Essays on Verbal and Visual Representation*. Chicago and London: University of Chicago Press.
- Mittal, Aarush Mohit; Singh, Swikriti Saran; Gupta, Nitin 2018. Sensory coding: neurons that wire together fire longer. *Current Biology* 28, R599–R618.
- Müller, Jens; Wiedemann, Julius (eds.) 2018. *The History of Graphic Design. Vol. 2: 1960–Today*. Los Angeles: Taschen America.
- Murphy, Gillian; Groeger, John A.; Greene, Ciara M. 2016. Twenty years of load theory – Where are we now, and where should we go next? *Psychonomic Bulletin & Review* 23: 1316–1340.
- Nelson, Thomas A. 1982. *Kubrick, Inside a Film Artist's Maze*. Bloomington: Indiana University Press.
- Nichols, Bill 2001. *Introduction to Documentary*. Bloomington: Indiana University Press.
- Norgaard, Nina 2009. The semiotics of typography in literary texts. A multimodal approach. *Orbis Litterarum* 64/2: 141–160.
- O'Halloran, Kay; Lim Fei, Victor 2014. Systemic functional multimodal discourse analysis. In: Norris, Sigrid; Maier, Carmen Daniela (eds.) *Interactions, Images and Texts: A Reader in Multimodality*. Berlin: Mouton de Gruyter, 137–154.

- O'Halloran, Kay; Tan, Sabine; Wignell, Peter 2016. Intersemiotic Translation as Re-semiotization. *Signata: annales des sémiotiques* vol. 7: 199–229. <https://journals.openedition.org/signata/1223#text>.
- Oja, Martin 2014. Darkness on screen: subjectivity-inducing mechanisms in contemporary Estonian art film. *Baltic Screen Media Review*, 2: 76–95.
- Othman, Huda 2021. René Magritte as an Inspiration for Modern Advertising Design. *International Design Journal*, Vol. 11, Issue 4, 113–125.
- Paget, Derek 2011. *No Other Way to Tell it. Docudrama on Film and Television. Second Edition*. Manchester and New York: Manchester University Press.
- Palmer, Bill 2013. *Understanding Air France 447*. William Palmer.
- Parpia, Pasha 2011. Reappraisal of the Somatosensory Homunculus and Its Discontinuities. *Neural Computation* 23: 3001–3015.
- Partan, Sarah R. 2004. Multisensory animal communication. In: Calvert, Gemma A; Spence, Charles; Stein, Barry E. (eds.) *The Handbook of Multisensory Processes*. Cambridge: The MIT Press.
- Peirce, Charles 1998. *The Essential Peirce. Selected Philosophical Writings. Volume 2 (1893–1913)*. Bloomington: Indiana University Press.
- 1931–1966. In: Hartshorne, Charles; Weiss, Paul; Burks, Arthur W. (eds.) *The Collected Papers of Charles S. Peirce*, 8 vols. Cambridge: Belknap Press of Harvard University Press [Quoted as CP followed by volume and paragraph number].
- Pennartz, Cyriel M. A. 2009. Identification and integration of sensory modalities: Neural basis and relation to consciousness. *Consciousness and Cognition* 18: 718–739.
- Perlmutter, Daniel D. 2002. On incongruities and logical inconsistencies in humor: The delicate balance. *Humor* 15–2, 155–168.
- Perrone-Bertolotti, Maria; Kujala, Jan; Vidal, Juan R.; Hamame, Carlos M.; Ossandon, Tomas; Bertrand, Olivier; Minotti, Lorella; Kahane, Philippe; Jerbi, Karim; Lachaux, Jean-Philippe 2012. How Silent Is Silent Reading? Intracerebral Evidence for Top-Down Activation of Temporal Voice Areas during Reading. *The Journal of Neuroscience*, December 5, 32(49): 17554–17562.
- Pessoa, Luiz 2022. *The Entangled Brain. How Perception, Cognition, and Emotion Are Woven Together*. Cambridge and London: The MIT Press.
- Petkov, Christopher I.; Belin, Pascal 2013. Silent Reading: Does the Brain ‘Hear’ Both Speech and Voices? *Current Biology* Vol. 23 No. 4, R155–R156.
- Petrilli, Susan 2014. *Sign Studies and Semioethics. Communication, Translation and Values*. Berlin: Walter de Gruyter.
- Phillips, Gene D. (ed.) 2001. *Stanley Kubrick: Interviews (Conversations with Filmmaker Series)*. Jackson: University Press of Mississippi.
- Piaget, Jean 2001[1947]. *The Psychology of Intelligence*. London and New York: Routledge Classics.
- Porcu, Emanuele; Keitel, Christian; Müller, Mathias M. 2014. Visual, auditory and tactile stimuli compete for early sensory processing capacities within but not between senses. *Neuroimage* 97: 224–235.
- Portis-Winner, Irene; Winner, Thomas G. 1976. The Semiotics of Cultural Texts. *Semiotica* 18:2, 101–156.
- Purchase, Helen C. 1999. A semiotic definition of multimedia communication. *Semiotica* 123 (3/4), 247–259.
- Purves, Dale; Augustine, George J.; Fitzpatrick, David; Hall, William C.; LaMantia, Anthony-Samuel; McNamara, James O.; White, Leonard E. 2008. *Neuroscience. Fourth Edition*. Sunderland: Sinauer Associates Inc.

- Puumeister, Ott 2021. The autonomy of the aesthetic function. In: Sütiste, Elin; Gramigna, Remo; Griffin, Jonathan; Salupere, Silvi (eds.) *(Re)considering Roman Jakobson*. Tartu Semiotics Library 23. Tartu: University of Tartu Press, 149–163.
- Randviir, Anti 2010. *Ruumisemiootika: tähendusliku maailma kaardistamine*. Tartu: Tartu Ülikooli kirjastus.
- Ramachandran, Vilayanur S.; Hubbard, Edward M. 2003. Hearing colors, tasting shapes. *Scientific American*, May, 53–59.
- Raskin, Victor 1985. *Semantic Mechanisms of Humor*. Dordrecht: Reidel Publishing Company.
- Reinosa-Carvalho, Felipe; Gunn, Laura H.; Horst, Enrique; Spence, Charles 2020. Blending Emotions and Cross-Modality in Sonic Seasoning: Towards Greater Applicability in the Design of Multisensory Food Experiences. *Foods*, Vol. 9(12), 1876, 1–20.
- Richards, Rashna Wadia 2008. Unsynced: The Contrapuntal Sounds of Luis Buñuel's "L'Age D'Or". *Film Criticism*. Vol. 33, No. 2: 23–43.
- Rickberg, Merit 2023. *Towards Complexity Thinking with Juri Lotman: Modelling Cultural Dynamics in Educational Systems* [Dissertationes Semioticae Universitatis Tartuensis]. Tartu: Tartu University Press.
- Robinson, Christopher W.; Moore, Robert L. Jr.; Crook, Thomas A. 2018. Bimodal presentation speeds up auditory processing and slows down visual processing. *Frontiers in Psychology* 9, 1–10.
- Robinson, Douglas 2008. *Estrangement and the Somatics of Literature. Tolstoy, Shklovsky, Brecht*. Baltimore: The Johns Hopkins University Press.
- Sager Eidt, Laura M. 2008. *Writing and Filming the Painting. Ekphrasis in Literature and Film*. Amsterdam and New York: Rodopi.
- Salmela, V.R.; Moiala, M.; Alho, K. 2014. Working memory resources are shared across sensory modalities. *Atten Percept Psychophys* 76, 1962–1974.
- Salmon, Paul M.; Walker, Guy H.; Stanton, Neville A 2015. Pilot error versus socio-technical systems failure: a distributed situation awareness analysis of Air France 447. *Theoretical Issues in Ergonomics Science*, published online, DOI: 10.1080/1463922X.2015.1106618.
- Sawhney, Nitin 2012. Nitin Sawhney on the Lodger. *BFIEvents*. Youtube. Available online: <https://www.youtube.com/watch?v=NZBNGsu4zLk>. Accessed 08. sept. 2024.
- Schultz, Wolfram 2016. Dopamine Reward Prediction Error Coding. *Dialogues in Clinical Neuroscience* 18 (1): 23–32.
- Sebeok, Thomas 1991. In what sense is language a “primary modeling system”? In: Anderson, Myrdene; Merrell, Floyd (eds.) *On semiotic modeling*. New York & Berlin: Mouton de Gruyter, 327–229.
- Sebeok, Thomas; Danesi, Marcel A. 2000. *The Forms of Meaning. Modeling Systems Theory and Semiotic Analysis*. Berlin and New York: Mouton de Gruyter.
- Shams, Ladan; Kim, Robyn 2010. Crossmodal influences on visual perception. *Physics of Life Reviews* 7: 269–284.
- Shapiro, Matthew L.; Eichenbaum, Howard 1999. Hippocampus as a memory map: synaptic plasticity and memory encoding by hippocampal neurons. *Hippocampus* 9: 365–384.
- Shklovsky, Viktor 2016a[1917] Art as a Device. In: Berlina, Alexandra (ed.) *Shklovsky, Viktor. A Reader* New York: Bloomsbury Academic, 73–96.
- 2016b[1966/1983] Tales about prose. In: Berlina, Alexandra (ed.) New York: Bloomsbury Academic, 258–274.

- Shneiderman, Ben 1992. *Designing the User Interface*, 2nd ed. Wokingham: Addison-Wesley.
- Silverman, Marcus M. 2012. René Magritte and the Denial of Meaning. *Modern Psychoanalysis*, Vol. 37, Iss. 2, 69–98.
- Smith, Greg 2004. Moving explosions: Metaphors of emotion in Sergei Eisenstein's Writings. *Quarterly Review of Film and Video*, 21: 303–315.
- Sobchack, Vivian 1992. *The Address of the Eye. A Phenomenology of Film Experience*. Princeton: Princeton University Press.
- Spence, Charles; Senkowski, Daniel; Röder, Brigitte 2009. Crossmodal processing [Editorial]. *Experimental Brain Research*, 198(2–3), Berlin and Heidelberg: Springer-Verlag, 107–111.
- Spence, Charles; Etzi, Roberta; Zampini, Massimiliano; Gallace, Alberto 2016. When Sandpaper Is 'Kiki' and Satin Is 'Bouba': an Exploration of the Associations Between Words, Emotional States, and the Tactile Attributes of Everyday Materials. *Multisensory Research* 29, 133–155.
- Sporns, Olaf 2011. The non-random brain: efficiency, economy, and complex dynamics. *Frontiers in Computational Neuroscience*, February, Vol. 5, 1–13.
- Staiger, Janet 2003. Authorship Approaches. In: Gerstner, David A., Staiger, Janet (eds.). *Authorship and Film*. New York and London: Routledge.
- Stam, Robert; Burgoyne, Robert; Flitterman-Lewis, Sandy 1992. *New vocabularies in film semiotics. Structuralism, post-structuralism and beyond*. London: Routledge.
- Stein, Barry E.; Meredith, Alex M. 1993. *The Merging of the Senses*. Cambridge: The MIT Press.
- Stöckl, Hartmut 2004. In between modes. Language and image in printed media. In: Ventola, Eija; Cassily, Charles; Kaltenbacher, Martin (eds.) *Perspectives on Multimodality*. Amsterdam and Philadelphia: John Benjamins Publishing Company, 9–30.
- 2014. Semiotic paradigms and multimodality. In: Jewitt, Carey (ed.), *The Routledge Handbook of Multimodal Analysis* (2nd ed). London: Routledge, 274–286.
- Stroop, John Ridley 1935. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643–662.
- Struss, Jason 2023. Creating the sound of death: How The Zone of Interest brings the Holocaust to chilling life. *Digital Trends*. Updated December 29. Available online: <https://www.digitaltrends.com/movies/the-zone-of-interest-johnnie-burn-interview/>.
- Sütiste, Elin; Torop, Peeter 2007. Processual Boundaries of Translation: Semiotics and Translation Studies. *Semiotica* 163–1/4, 187–207.
- Sütiste, Elin 2021. The functional roots of Jakobson's plural concept of translation. In: Sütiste, Elin; Gramigna, Remo; Griffin, Jonathan; Salupere, Silvi (eds.) *(Re)considering Roman Jakobson*. Tartu Semiotics Library 23. Tartu: University of Tartu Press, 164–185.
- Tagg, Philip 2013. *Music's Meanings. A Modern Musicology for Non-musos*. New York and Huddersfield: The Mass Media Music Scholars' Press.
- Talsma, Durk; Doty, Tracy J.; Strowd, Roy; Woldorff, Marty G. 2006. Attentional capacity for processing concurrent stimuli is larger across sensory modalities than within a modality. *Psychophysiology* 43, 541–549.
- Thibault, Paul J. 2004. *Brain, Mind and the Signifying Body. An Ecosocial Semiotic Theory*. London and New York: Continuum.
- Thompson, Kristin 1981. *Eisenstein's "Ivan the Terrible": A Neoformalist Analysis*. Princeton, Princeton University Press.

- Tieber, Claus; Windisch, Anna K. (eds.) 2014. *The Sounds of Silent Films. New Perspectives on History, Theory and Practice*. New York: Palgrave Macmillan.
- Tieber, Claus; Windisch, Anna K. 2014. The Sound of Music in Vienna's Cinemas, 1910–1930. In: Tieber, Claus; Windisch, Anna K. (eds.) *The Sounds of Silent Films. New Perspectives on History, Theory and Practice*. New York: Palgrave Macmillan.
- Tihanov, Galin 2005. The Politics of Estrangement: The Case of the Early Shklovsky. *Poetics Today*, 26: 4, 665–696.
- Tikka, Pia; Kaipainen, Mauri 2023. The Unfolding now. Narrative sense-making from a neurocinematic perspective. In: Besser, Stephan; Lysen, Flora (eds.) *Worlding the Brain: neurocentrism, Cognition and the Challenge of the Arts and Humanities Vol. 3*. Leiden and Boston: Brill.
- Torop, Peeter 2003. Intersemiosis and Intersemiotic Translation. In: Petrilli, Susan (ed.) *Translation, Translation. Approaches to Translation Studies 21*. Amsterdam: Editions Rodopi BV, 271–282.
- 2023. Intersemiotics: A Conceptual Field. In: Kroo, Katalin; Juracsek, Kata (eds.) *Intersemiotics and Cultural Transfer*. Budapest and Tartu: Eötvös Loránd University Research Group and Tartu University Transmedia Research Group: 25–44.
- Truffaut, François 1985. *The Definitive Study of Alfred Hitchcock*. New York: Simon and Schuster Paperbacks.
- Uexküll, Thure von 1987. The Sign Theory of Jakob von Uexküll. In: Krampen et al. (eds.) *Classics of Semiotics*. Berlin: Springer 147–180.
- Van den Oever, Annie 2010. Introduction: Ostran(n)enie as an “Attractive” Concept. In: Van den Oever, Annie (ed.) *Ostrannen(e). On “Strangeness” and the Moving Image. The History, Reception, and Relevance of a Concept*. Amsterdam: Amsterdam University Press.
- Van den Oever, Annie; Gunning, Tom 2020. Viktor Shklovsky's Ostrannen(e) and the ‘Hermeneutics of wonder’. *Early Popular Visual Culture* 18:1, 15–28. Published online: <https://doi.org/10.1080/17460654.2020.1751915>.
- Van Leeuwen, Theo 2006. *Towards a semiotics of typography*. Information Design Journal 14, no. 2: 139–55.
- Vroomen, Jean; De Gelder, Beatrice 2000. Sound enhances visual perception: cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology: Human Perception and Performance* Vol. 26, No. 5, 1583–1590.
- Vygotsky 2012 [1934] *Thought and Language*. Cambridge and London: The MIT Press.
- Wahn, Basil; König, Peter 2017. Is attentional resource allocation across sensory modalities task-dependent? *Advances in Cognitive Psychology* 13(1): 83–96.
- Warschauer, Mark 2017. The pitfalls and potential of multimodal composing. *Journal of Second Language Writing* 38: 86–87.
- Watts, Duncan J; Strogatz, Steven H. 1998. Collective dynamics of “Small-World” Networks. *Nature* 393, 440–442.
- Wildfeuer, Janina 2014. *Film Discourse Interpretation: Towards a New Paradigm for Multimodal Film Analysis (Routledge Studies in Multimodality)*. New York and London: Routledge.
- Zlatev, Jordan; Möttönen, Tapani 2022. Cognitive linguistics and semiotics. In: Pelkey, Jamin; Copley, Paul (eds.) *Bloomsbury Semiotics Volume 4: Semiotic Movements*. London: Bloomsbury Academic.

SUMMARY IN ESTONIAN

Multimodaalse konflikti semiootika: moodsed, modaalsused ning auditiivse-visuaalse vastuolu mängufilmis

Nii looduskeskkonnas kui kultuuritekstide keskel satume tihti olukorda, mille tähenduse kujunemine sõltub korraga mitmest koodist, keelest või märgisüsteemist. Kui semiootika oma algusaegadel keskendus pigem üksikmärgi ja seejärel märgisüsteemi probleemide lahendamisele, on kaasaegses semiootikas ja kultuuriteoorias üheks tähtsaimaks ülesandeks mitme märgisüsteemi üheskoos toimimise aluste mõistmine ja kirjeldamine. Sellele probleemistikule lähenevad nii komplekssete süsteemide teooriad (vt nt Engelbrecht 2021, Rickberg 2023), multimeedia-uuringud (nt Elleström 2020) kui ka multimodaalsuse uuringud (Kress 2010, Bateman jt 2017). Kuna käesolev doktoritöö tõukub ennekõike filmisemiootika pinnalt, on selle kontekstiks soov ja vajadus paremini mõista tähendusloome mehhanisme filmis, mida võib nimetada “klassikaliseks” multimodaalseks tekstiks.

Film kasutab kahte tajumodaalsust (auditiivset ja visuaalset) ja ka hulgaliselt erinevaid väljendusvahendeid ning võtteid, mis pärinevad teatri, fotograafia, disaini, muusika, kirjanduse jt loominguliste praktikate “keeltest”. Erinevate komponentide koosseksiteerimine ühtses tervikus toob ühelt poolt kaasa olukorra, millele osutasid Juri Lotman ja ta kolleegid “Kultuurisemiootika teesides”: tekst on ühtaegu nii tervikmärk kui ka märkide kogum või järjestus (2013[1973]: 58). Teiselt poolt on erinevate tähendust kandvate üksuste koostoime puhul tegemist ka sünteesiga. Taolise sünteesi iseloomustamisel on mitmed teoreetikud kasutanud matemaatilisi metafoore nagu liitmine ja korrutamine, kuid sellised võrdlused on eksitavad. Multimodaalne süntees ei allu sellisele loogikale, vaid nõuab mõistmiseks iseseisvat raamistikku. Selles raamistikus mängib olulist rolli Roman Jakobsoni dominandi mõiste (Jakobson 1981[1935]), viidates ühele teksti komponendile, mis suunab nii teiste komponentide kui tervikteksti tähenduse interpretatsiooni. Puudutades semiootika ajalugu, on oluline märkida, et just Roman Jakobson rõhutas semiootika kontekstis esimesena vajadust uurida erinevate tajumodaalsuste rolli ja koosmõju, tõstes sealjuures esile tekstitüüpi, mis kasutab korraga näiteks auditiivset ja visuaalset modaalsust ning mida Jakobson ise nimetas “sünkreetiliseks teateks” (Jakobson 1971c [1964]: 339).

Multimodaalsust uurivates tänapäevastes teooriates valmistab sageli raskust multimodaalsete tekstide kahe põhitunnuse eristamine. Nagu eelpool öeldud, võib multimodaalne tekst väljenduda mitmes erinevas märgisüsteemis, teiseks võib selle vastuvõtt ja mõistmine sõltuda mitmest tajukanalist. Esimene tunnus tuleb hästi esile näiteks trükimeedias, kus esineb pilt kombineeritult verbaalse kirjaliku tekstiga, või ka kõigile tuntud René Magritte’i maalil piibust, mille alla on kirjutatud “See siin ei ole piip”. Antud tekstitüüp ei ole multimodaalne tajumodaalsuste lõikes, sest kasutab ainult visuaalset modaalsust; küll aga tekib

selles multimodaalsus erinevate märgisüsteemide kombineerumise tõttu. Teist tüüpi multimodaalsus eeldab mitme tajukanali kooseksisteerimist. Siinkohal on aga oluline rõhutada, et selle multimodaalsuse uurimine ei piirdu meediumit puudutavate küsimustega, vaid on suunatud iseäranis retseptiooni-, taju- ja tõlgendusprotsessidele.

Nende kaht tüüpi multimodaalsuse kirjeldamisel osutub üheks tähtsaimaks teoreetiliseks sammuks keeleteadlase Ellen Fricke eristus laia (mitu märgisüsteemi) ja kitsa (mitu tajumodaalsust) multimodaalsuse vahel (Fricke 2013: 740). Kuna looduskeskkonnas ning ka sotsio- ja kultuuritekstides toimub sageli mõlemat tüüpi multimodaalsuse põimumine (filmivaataja saab informatsiooni nii auditiiivse kui visuaalse modaalsuse kaudu, mis omakorda toetavad erinevaid “väljenduskeeli”, näiteks auditiiivses modaalsuses saab tähendust luua nii verbaalse dialoogi, muusika kui ka heliefektide märgisüsteemis), on taolist kompleksset teksti analüüsid esmatahtis mõistete täpne eristus. Siiski aetakse omavahel üpris sageli segamini mõisted, mis puudutavad väljendusviise ehk mooduseid (mida sotsiosemiotik Gunther Kress on kirjeldanud kultuurilis-sotsiaalselt konstrueeritud semiootiliste ressursidena) ning tajumodaalsusi.

Käesoleva doktoritöö esimene põhieesmärk on pakkuda nimetatud terminoloogilisele probleemile (ja sellest lähtuvatele metodoloogilistele väljakutsetele) võimalik lahendus, eristades selgelt mõisteid *moodus* (ingl. k. mode ehk semiotic mode) ja *modaalsus* (ingl. k. modality ehk sensory modality), luues samas multimodaalsete objektide käsitlemiseks kahetasandilise raamistiku. Selle raamistiku üks tasand koosneb – nagu neid kirjeldasid Gunther Kress ja ta kolleegid, hiljem aga oluliselt täiendasid John Bateman ja tema kolleegid – sotsiaalselt konstrueeritud, seetõttu ka paindlikest ja mõnevõrra ebastabiilsetest moodustest. Teise tasandi moodustavad aga võrdlemisi kindlalt fikseeritud tajumodaalsused, mida käsitledes semiootika ning ka teised kultuuri uurivad distsipliinid saavad toetuda kognitiivsele neuroteadusele ning evolutsioonibioloogiale. Taoline kaheosaline raamistik annab aluse interdistsiplinaarseks vaatenurgaks, mis võimaldab omavahel sünteesida kvalitatiivsete ja kvantitatiivsete lähenemiste tugevaid külgi; semiootikasiseselt aga teeb loomulikuks kultuuri-, bio-, ja sotsiosemiotika rakursside ühendamise.

Käesolev doktoritöö keskendub kahest multimodaalsuse liigist ennekõike “kitsale” (mitu tajumodaalsust) kui sotsiosemiotikas ja ka muudes kultuuriuuringute distsipliinides vähem käsitletule (selle võimalikuks põhjuseks on väljakutsed, mida pakub kognitiivsete neuroteaduse poolt kirjeldatud tajumodaalsuste toimeprintsipiide integratsioon humanitaarteadusliku kirjeldusega). Kirjeldades tajumodaalsuste omavahelisi suhteid ja lähtudes osaliselt ka dominandi mõistest, esitab doktoritöö tajumodaalsuse võimalike kombinatsioonide tüpologia, mis eristab toetavat, modifitseerivat, konfliktset, asendavat ja vastastikuselt aktiveerivat suhet tüüpi.

Erinevate tajumodaalsuste integratsioonil on oluline silmas pidada kaht fundamentaalset aspekti. Esiteks viitab aina rohkem teadustöid tajumodaalsuste sügavale ja tihedale omavahelisele seotusele ajus (vt nt Bertelson, De Gelder 2004); selle tagajärjel võib täheldada oluliselt suuremat ristmodaalsuse (cross-

modality) osakaalu kui varem seda tunnistati. Ristmodaalsuse puhul on tegemist olukorraga, milles ühe tajumodaalsuse kaudu vastu võetud informatsioon aktiveerib ajus teised tajumodaalsused ning võib käivitada kognitiivse protsessi, mille tulemiks ehk semiootika mõistes “tekkinud tähenduseks” on hoopis multimodaalne tervik. Selle tagajärjel võib ka traditsiooniliselt monomodaalseks peetud kultuuripraktikate interpretatsiooni näha multimodaalsena (kirjandusteoste kehalist ja multimodaalset vastuvõttu käsitlevad nt Hillesund jt 2022). Ristmodaalsust on käsitletud ka filmiteooria raames, näiteks Laura Marks kirjeldab haptilist visuaalsust, mis seisneb selles, et vaatajatena võime tajuda ekraanil nähtud puudutust otsekui reaalse puudutusena (Marks 2000: 22). Taolise reaktsiooniga on seotud vaatajas eelmotoorseid närvivõrgustikke aktiveerivate ja filmitegelasega toimuvate sündmuste kehalist simuleerimist võimaldavate peegelneuronite toime, mida seoses filmiteooriaga käsitlevad näiteks Vittorio Gallese ja Michele Guerra. Nad kinnitavad, et kaasaegse neuroteaduse toel on võimalik kujundada uus (filmi) tajumise mudel, milles on tegevus (motoorsus), taju (pertseptsioon), mõtlemine ning emotsioonid omavahel tihedalt integreeritud (Gallese, Guerra 2015: 151).

Tajumodaalsuse integratsioonile võib läheneda ka intersemiootilise tõlke mõiste kaudu; seda võimalust olen käsitlenud töö katuspeatükis. Ühes tajumodaalsuses esitatud informatsiooni esitamist teises tajumodaalsuses (näiteks ekraniseerimisel raamatu “tõlkimist” filmiks) saadab küsimus, millised elemendid või üksused sellist transpositsiooni võimaldavad; teisisõnu on küsimus variantide ja invariantide (ehk muutuva ja muutumatu) omavahelises suhtes. Nii intersemiootilise tõlke kui ka tajumodaalsuste integratsiooni kirjeldamisel võib oluliseks osutada *amodaalsete invariantide* mõiste, mis korreleeruvad teatud tüüpi (multimodaalsete) neuronitega ajus, mis sünteesivad erinevatest tajumodaalsustest saabuvaid sisendeid (Stein, Meredith 1993) kui ka taolises protsessis esile kerkiv “universaalne” ja võrdlemisi abstraktne märgisüsteem, mille tähendused on modaalsuse-üleised – näiteks rütm, liikumine, kontrastid jms. Siinkohal saab paralleelina viidata ka Timo Marani poolt osutatud liikidevahelise protokoodi võimalikkusele (Maran 2008: 22).

Doktoritöö teine põhieesmärk on lähemalt uurida multimodaalset konflikti, mis tekib siis, kui üks moodus annab edasi vasturääkivat informatsiooni võrreldes teise moodusega (Magritte'i maalil pildiline moodus kinnitab viitamist piibule, verbaalne moodus aga eitab seda) või üks modaalsus kannab vasturääkivat sõnumit, mõju või informatsiooni võrreldes teise modaalsusega. Modaalsuste konflikt ilmneb näiteks filmis, mille heliriba ehk auditiivne modaalsus kannab muusikat, mis interpreteerija hinnangul ei sobi kokku “pildiga” ehk visuaalse modaalsusega: selliselt toimib näiteks Stanley Kubricku “Kellavärgiga apelsinis” Beethoveni muusika, mis saadab vägivallastseene ning mille eesmärk on luua ironiat. Teiseks näiteks on filmides edasi antud tegelase mõttevoog või sisekõne, mida kujutatakse enamasti *sisemise diegeetilise helina* (internal diegetic sound ehk IDS). Filmis avaldub see visuaalses modaalsuses vaikiva/mittekõneleva tegelase ja temaga kombineeritud auditiivses modaalsuses väljenduva kõne kujul. Antud näide on ennekõike huvitav seetõttu, et tehniliselt

on tegemist tugeva multimodaalse konfliktiga, kuid filmikunsti väljendusvahenditega tuttav vaataja ilmselt selles konflikti ei näe, sest antud võte on pideval kordamisel kinnistunud ega mõju sedasi värskena või kummastavana; see on automatiseerunud ning saanud osaks koodist. Automatiseerumise ja deautomatiseerumise vaheldumise kui dünaamilise protsessi kirjeldamisel nii kultuuris kui looduses on olulised nii Charles Peirce'i harjumuse (habit) mõiste kui Juri Lotmani arusaam plahvatuslike ja kontinuaalsete protsesside vaheldumisest. Konfliktide uurimisel on süsteeme käsitlevates distsipliinides keskne roll, sest sageli võimaldab just konflikt kui näiline tõrge süsteemis mõista süsteemi ülesehitust ja süsteemi elementide omavahelisi funktsionaalseid suhteid. Käesolev töö teadvustab ka asjaolu, et semiootilise konflikti juured on sügaval tähendusloome põhialuste juures. Nagu juhib tähelepanu Kalevi Kull, on konflikt semioosi aluseks, sest seab elusolendi ette valikuid ja seeläbi kutsub üles semiootilisele tegevusele ja aktiivsusele (nt Kull 2015b: 225). Roman Jakobsoni, Juri Lotmani ja tema Tartu-Moskva koolkonda kuuluvate kolleegide seisukohti mõtestav Tyler Bennett viitab aga kokkuvõtvalt, et koosseksiteerivate, üksteisega vastusolus olevate tähenduste toimel oleme sunnitud looma uusi hüpoteese (Bennett 2021: 159). Seoses konfliktide aktiveeriva mõjuga osutub keskseks ka vene vormikoolkonnas, ennekõike Viktor Šklovski töödes käsitletud kummastamise (ostranenie) mõiste. Multimodaalset, sünteetilist tähendusloomet sobib iseloomustama ka Juri Lotmani idee kahest, osaliselt ühildamatust märgisüsteemist, mis on vajalik eeltingimus uue informatsiooni loomiseks.

Doktoritöö koosneb katuspeatükist ning kolmest teadusartiklist. **Esimeses artiklis “Mõisted *mode* ja *modality* multimodaalsuse semiootilises teoorias: komplementaarsusest ja eristamise vajalikkusest”** vaatlen multimodaalsuse uuringute ja kultuurisemiootika ristumiskohas esile kerkivat probleemi, nimelt mõistete *mode* (moodus) ning *modality* (modaalsus) ebajärjekindlat kasutust. Neid tarvitatakse sageli sünonüümselt, mis toob kaasa probleeme kategooriate eristamisel. Nende probleemide lahendamisel on abiks mõistete paigutamine sidusasse süsteemi, mida käesoleval juhul modelleerin kahekihilisena. Mooduste ja modaalsuste tasandid pole omavahel vastuolus, vaid komplementaarsed ehk teineteist täiendavad.

Esmalt käsitlen multimodaalse semiootika keskset haru, mis võrsub Gunther Kressi sotsioseemiootikast ja ulatub John Batemani põhjaliku lähenemiseni ning kasutab moodust (*mode*) baasmõistena. Nimetatud koolkonnas keskendutakse mooduse kui semiootilise ressursi materiaalsusele ning kultuurilistele praktikatele, mis mõjutavad materiaalsuse vormimist tähenduse loomisel ning ka diskursiivset semantikat, mis reguleerib mooduste abil loodud tähenduste mõistmist neid ümbritsevas kontekstis. Samal ajal peetakse tajumodaalsuste neurokognitiivseid omadusi tähendusloomes teisejärguliseks. Käesolevas artiklis väljendatakse seisukohta, et tajumodaalsuste roll tähendusloomes on keskne ning selle uurimisel on kindel koht multimodaalsuse semiootikas. Sellal kui moodus on terminina mitmeti tõlgendatav, tasakaalustab seda (taju)modaalsus, mis on psühholoogias konkreetselt määratletud. Et näitlikustada mooduste ja

modaalsuste käsitusvõimalust ühtses raamistikus, loon artikli lõpuosas katsetusliku tüpoloogia, mis kirjeldab modaalsuste võimalikke suhetüüpe ning võtab arvesse ka mooduseid. Eristan toetavat, modifitseerivat, konfliktset, asendavat ja vastastikuselt aktiveerivat suhetüüpi. Sellisena väljendab esimene artikkel püüdeid organiseerida terminivälja ja juhatada sisse doktoritöö järgnev osa.

Töö **teine artikkel “Modaalsuste konflikt mängufilmis: kontrapunktilisest montaažist karakteri sisekõneni”** moodustab terviku suhtes keskse osa. Kuna töö suund on üldiselt üksikule (teoorialt objekti suunas), siis jõuan siin põhilise uurimisobjekti, mängufilmini. Enne seda aga käsitlen sünteesi mõistet erinevate filmiteooria ning semiootika ristumisalal tegutsenud teoreetikute töödes. Vaatlen Sergei Eisensteini käsitlust kontrapunktilisest montaažist, peatun Roland Barthes’i kolmanda tähenduse mõistel, vaatlen põhjalikumalt dominandi ja sünkreetilise teate seost Roman Jakobsonil ning mitmekeelsete süsteemide käsitlust Juri Lotmanil (kus on vastamisi seatud osaliselt tõlkimatud märgisüsteemid, mis moodustavad uue informatsiooni loomise mehhanismi), lõpetan aga ülevaate kaasaegse kognitiivse filmiteooria panusega multimodaalsete konfliktide uurimisse. Evolutsioonibioloogiale toetudes toon välja, miks on multimodaalne konflikt kultuuritekstides erandjuhtum ning suur osa kultuuritekste sõltub vastastikku toetavast modaalsuste suhtest.

Järgnevalt esitan teises artiklis küsimuse, mis on teatud mõttes aluseks tervele multimodaalsete konfliktide uurimisprogrammile. Vaatlen erinevaid kriteeriumeid, mis võiks olla konflikti tuvastamise ja konflikti ulatuse mõõtmise aluseks. Heidan kõrvale võimalikena näinud sünkroonia ja liiasuse (eeldusel, et konflikt tekib nende eiramisel). Esmapilgul võib näida paradoksaalsena, et sünkroonia on hoopis konflikti märkamise aluseks. Siiski võimaldab just sünkroonia erinevaid modaalsusi tajuda samaaegsetena ning konfliktsetena. Nagu viitavad filmimuusika uurijad (nt Michel Chion ja Rick Altman), ei ole heli ja pilt filmis kunagi üksteise suhtes liiased. Leian, et erinevalt sünkroonial ja liiasusel rajanevatest raamistikest on konflikte võimalik modelleerida Juri Lotmani ja Tartu-Moskva koolkonna semiootikas tähtsal kohal olnud ruumi kui modelleeriva süsteemi mõiste abil. Viimaks analüüsin üht näidet audiovisuaalsest konfliktist Laurence Olivier’ filmis “Hamlet” (1948), kus Hamleti monoloogi keskel annab järsk üleminek vastuolulistele modaalsustele edasi murrangut tegelaskuju psüühikas ja narratiivi kulgemises.

Töö **kolmas artikkel “Eksitav trailer: filmitreiler kui paratekst ja filmi multimodaalne mudel”** on kõige rohkem suunatud uurimisobjektile, keskendudes tajumodaalsuste suhetüpoloogias konfliktse suhte ühele kindlale näitele. Vaatlen filmitreilerit, mis auditiivse modaalsuse kaudu viitab ühele, visuaalse modaalsuse kaudu aga teisele filmižanrile; sedasi toimides püüavad filmi turundajad kinno meelitada erinevate žanride austajaid. Käsitlen olukorda, milles trailer ehk paratekst hakkab tööle filmi ehk teksti mudelina, kujundades filmi retseptsiooni ja lõppkokkuvõttes põhjustades vaatajas pettumust, kui filmi žanr ei vasta traileris reklaamitule. Tuginedes Lisa Kernani ja Keith M. Johnstoni töödele, võib žanri käsitleda kui kõige tähtsamat traileri poolt edastatavat informatsiooni. Analüüsin Alejandro G. Inarritu “Lindmehe” ja Nicolas Winding

Refni “Ainult Jumal andestab” treilereid kui näiteid eksitavast multimodaalsest modelleerimisest. Teoreetilise alusmõistena rakendan taas Roman Jakobsoni dominandi mõistet, millel põhinev meetodika võimaldab määrata treileri (mudeli või tähistaja) ja filmi (objekti või tähistatava) dominandid ning nende kattuvust võrrelda.

PUBLICATIONS

CURRICULUM VITAE

Name: Martin Oja
Date of birth: June 22, 1978
E-mail: martino@ut.ee, martin.oja@ut.ee
Address: University of Tartu, Centre for Semiotic Applications, Jakobi 2-314, Tartu, 51005

Education:

2015–2024 University of Tartu, doctoral studies, Semiotics and Culture Studies
2005–2008 University of Tartu, master’s studies, Semiotics and Culture Studies, MA
2001–2005 University of Tartu, bachelor’s studies, Semiotics and Culturalology, BA
1996–2005 University of Tartu, bachelor’s studies, Literature and Folkloristics, BA
1994–1996 Tartu Forselius Gymnasium

Study visits at other universities:

09/2000–02/2001 University of Helsinki, exchange student

Career:

03/2024–... University of Tartu, Centre for Semiotic Applications, administrative manager

Teaching experience:

2015–2024 Tallinn University, Baltic Film and Media School, course “Mythology”
2019 University of Tartu, Department of Semiotics, course “Film Semiotics”
2016 University of Tartu, Department of Semiotics, supervising the seminars at the course “Introduction to Cultural Theories”
2008–2013 Tallinn University, Baltic Film and Media School, course “From Myth to Storytelling”

Organizational activities:

2024–... Estonian Semiotics Association, board member
2021–... Tartu Academic Tennis Club, member
2008–... Association of Estonian Film Journalists, member

Research topics:

semiotics of culture, multimodality, cognitive film theory, methods for the application of semiotics

Conference presentations:

- 2021 “Dualism and a severed head”. 16th autumn Conference of the Tammisaare Museum “Modern Myth”, Tallinn, Nov. 26.
- 2017 “How to internalize Myth?” Learning to understand, Tartu, May 6.–7.
- 2014 “Does cognitive film theory help to recognize a good film?”. Film conference „To see. To read. To understand”, Tartu, Elektriteater, 21.–22. nov. (co-organizer of the conference with Katre Pärn)

Acknowledgments:

- 2010 The journal’s “Theatre. Music. Cinema” award for the best writing in film criticism
- 2008 European League of Institutes of the Arts (ELIA) grant for participation in the project “Languages Through Lenses”, making a short film “Food Language” (2008) and taking part in workshops in Berlin, October 2008
- 2005 Diploma in the National Contest for University Students for the bachelor’s thesis “Jaan Kaplinski and the ecological view”
- 1998 Scholarship for the first price of Talinvest essay contest

Selected publications:

- Oja, Martin 2024. Conflicting modalities in feature film: from contrapuntal editing to internal diegetic sound. *Semiotica* 259: 95–126.
- Oja, Martin 2023. Semiotic mode and sensory modality in multimodal semiotics: Recognizing difference and building complementarity between the terms. *Sign Systems Studies* 51(3–4): 604–637.
- Oja, Martin 2022. Dualism ja raiutud pea: alusmüüdi lõhestamine ja lõhestamine alusmüüdi aluselt. *Acta Semiotica Estica* 19: 166–177.
- Oja, Martin 2019. On the concept of the deceptive trailer: Trailer as paratext and multimodal model of film. *Sign Systems Studies* 47(1–2): 177–204.
- Oja, Martin 2017. Kuidas internaliseerida müüti? *Acta Semiotica Estica* 14: 135–142.
- Oja, Martin 2014. Darkness on the screen: Subjectivity-inducing mechanisms in contemporary Estonian art film. *Baltic Screen Media Review* 2(1): 76–95.
- Oja, Martin 2014. Puuduolek jõudeelu südames. Ulrich Seidli kehatriiloo. *Teater. Muusika. Kino* 3
- Oja, Martin 2014. Reaalsuse ja representatsiooni vahel. Essee reaali- ja humanitaarteaduste põimumiskohtadest filmiteoorias. *Sirp* 16.04.2014

Oja, Martin 2009. Mällu tätoveeritud asjad (Sulev Keeduse filmiloomingust).
Teater. Muusika. Kino 3–4: 97–107.

Oja, Martin 2007. TalveAkadeemia. Jaan Kaplinski ning ökoloogiline vaade.
Akadeemia 11: 2402–2411.

Semiotic expert analyses:

2024 Centre for Semiotic Applications’s semiotic analysis for the products “Natural mineral water Haage” and water “Stellar”. Client: Haage Joogid OÜ

2024 Centre for Semiotic Applications’s expert analysis about the alleged employment of Nazi symbolics by Friends of Estonian Legion. Client: Estonian Police and Border Guard Board

ELULOOKIRJELDUS

Nimi: Martin Oja
Sünniaeg: 22. juuni 1978
E-mail: martino@ut.ee, martin.oja@ut.ee
Kontaktaadress: Tartu Ülikool, semiootika rakenduskeskus, Jakobi 2-314, Tartu, 51005

Haridus:

2015–2024 Tartu Ülikool, doktoriõpe, Semiootika ja kultuuriteooria
2005–2008 Tartu Ülikool, magistriõpe, Semiootika ja kulturoloogia, MA (teadusmagister)
2001–2005 Tartu Ülikool, bakalaureuseõpe, Semiootika ja kulturoloogia, BA
1996–2005 Tartu Ülikool, bakalaureuseõpe, Kirjandus ja rahvaluule, BA
1994–1996 Tartu Forseliuse Gümnaasium

Õppimine teistes ülikoolides:

09/2000–02/2001 Helsingi Ülikool, vahetusüliõpilane

Teenistuskäik:

03/2024–... Tartu Ülikool, semiootika rakenduskeskuse tegevjuht

Töökogemus õppejõuna:

2019 Tartu Ülikool, semiootika osakond, loengukursus “Filmi-semiootika”
2016 Tartu Ülikool, semiootika osakond, seminaride juhendamine loengukursuse “Sissejuhatus kultuuriteooriatesse” raames
2015–2024 Tallinna Ülikool, Balti Filmi- ja Meediakool, loengukursus “Mütoloogia”
2008–2013 Tallinna Ülikool, Balti Filmi- ja Meediakool, loengukursused „Müüdist loo jutustamiseni” ja „From Myth to Storytelling”

Organisatsiooniline tegevus:

2024–... Eesti Semiootika Seltsi juhatuse liige
2021–... Tartu Akadeemilise Tenniseklubi liige
2008–... Eesti Filmiajakirjanike Ühingu liige

Uurimisvaldkonnad:

kultuurisemiootika, multimodaalsus, kognitiivne filmiteooria, rakendussemiootika meetodid

Konverentsiettekanded:

- 2021 “Dualism ja raiutud pea”. Tammsaare muuseumi 16. sügiskonverents “Moodne müüt”, Tallinn, 26. nov.
- 2017 “Kuidas internaliseerida müüti?” Mõistma õppides, Tartu, 6.–7. mai
- 2014 “Kas kognitiivne filmiteooria aitab ära tunda väärtfilmi?”. Filmikonverents „Vaadata. Lugeda. Mõista”, Tartu, Elektriteater, 21.–22. nov. (Koos Katre Pärnaga selle konverentsi korraldaja).

Tunnustused:

- 2010 “Teater. Muusika. Kino” laureaat filmikriitika kategoorias
- 2008 European League of Institutes of the Arts (ELIA) grant osalemiseks projektis “Languages Through Lenses”, vastavateemalise lühifilmi “Food Language” (2008) tegemiseks ja osalemiseks töötubades Berliinis 2008 oktoobris
- 2005 Diplom Eesti üliõpilaste teadustööde riiklikul konkursil sotsiaalteaduste valdkonnas rakenduskõrghariduse ja bakalaureuseõppe üliõpilaste astmes teadustöö “Jaan Kaplinski ja ökoloogiline vaade” eest
- 1998 Talinvesti stipendium Talinvesti Kooliraha esseekonkursi võidu puhul

Tähtsamad publikatsioonid:

- Oja, Martin 2024. Conflicting modalities in feature film: from contrapuntal editing to internal diegetic sound. *Semiotica* 259: 95–126.
- Oja, Martin 2023. Semiotic mode and sensory modality in multimodal semiotics: Recognizing difference and building complementarity between the terms. *Sign Systems Studies* 51(3–4): 604–637.
- Oja, Martin 2022. Dualism ja raiutud pea: alusmüüdi lõhestamine ja lõhestamine alusmüüdi aluselt. *Acta Semiotica Estica* 19: 166–177.
- Oja, Martin 2019. On the concept of the deceptive trailer: Trailer as paratext and multimodal model of film. *Sign Systems Studies* 47(1–2): 177–204.
- Oja, Martin 2017. Kuidas internaliseerida müüti? *Acta Semiotica Estica* 14: 135–142.
- Oja, Martin 2014. Darkness on the screen: Subjectivity-inducing mechanisms in contemporary Estonian art film. *Baltic Screen Media Review* 2(1): 76–95.
- Oja, Martin 2014. Puuduolek jõudeelu südames. Ulrich Seidli kehatriлогия. *Teater. Muusika. Kino* 3

- Oja, Martin 2014. Reaalsuse ja representatsiooni vahel. Essee reaalsuse ja humanitaarteaduste põimumiskohtadest filmiteoorias. *Sirp* 16.04.2014
- Oja, Martin 2009. Mällu tätoveeritud asjad (Sulev Keeduse filmiloomingust). *Teater. Muusika. Kino* 3–4: 97–107.
- Oja, Martin 2007. TalveAkadeemia. Jaan Kaplinski ning ökoloogiline vaade. *Akadeemia* 11: 2402–2411.

Teostatud semiootilised ekspertiisid:

- 2024 Semiootiline analüüs toodetele “Haage looduslik mineraalvesi” ja vesi “Stellar”. Klient: Haage Joogid OÜ
- 2024 Semiootika rakenduskeskuse eksperthinnang seoses Eesti Leegioni Sõprade Klubi (ELSK) sooviga taastada Lihulas mälestusmärk, mis väidetavalt kannab Natsi-Saksamaa sümbolikat. Klient: Politsei- ja Piirivalveamet

DISSERTATIONES SEMIOTICAE UNIVERSITATIS TARTUENSIS

1. **Михаил Ю. Лотман.** Структура и типология русского стиха. Тарту, 2000.
2. **Елена Григорьева.** Эмблема: структура и прагматика. Тарту, 2000.
3. **Valdur Mikita.** Kreatiivsuskäsitluste võrdlus semiootikas ja psühholoogias. Tartu, 2000.
4. **Ирина Аврамец.** Поэтика новеллы Достоевского. Тарту, 2001.
5. **Ян Левченко.** История и фикция в текстах В. Шкловского и Б. Эйхенбаума в 1920-е гг. Тарту, 2003.
6. **Anti Randviir.** Mapping the world: towards a sociosemiotic approach to culture. Tartu, 2004.
7. **Timo Maran.** Mimikri kui kommunikatsiooni-semiootiline fenomen. Tartu, 2005.
8. **Элеонора Рудаковская-Борисова.** Семиотика пищи в произведениях Андрея Платонова. Tartu, 2005.
9. **Andres Luure.** Duality and sextets: a new structure of categories. Tartu, 2006.
10. **Peeter Linnap.** Fotoloogia. Tartu, 2006.
11. **Daniele Monticelli.** Wholeness and its remainders: theoretical procedures of totalization and detotalization in semiotics, philosophy and politics. Tartu, 2008.
12. **Andreas Ventsel.** Towards semiotic theory of hegemony. Tartu, 2009.
13. **Elin Sütiste.** Tõlke mõiste dünaamikast tõlketeaduses ja eesti tõlkeloos. Tartu, 2009.
14. **Renata Sõukand.** Herbal landscape. Tartu, 2010.
15. **Kati Lindström.** Delineating Landscape Semiotics: Towards the Semiotic Study of Landscape Processes. Tartu, 2011.
16. **Morten Tønnessen.** Umwelt transition and Uexküllian phenomenology. An ecosemiotic analysis of Norwegian wolf management. Tartu, 2011. 231 p.
17. **Anu Sarv.** Õppejõu eneserefleksioon ja professionaalne identiteet. Tartu, 2013, 141 lk.
18. **Vadim Verenich.** Semiotic models of legal argumentation. Tartu, 2014, 249 lk.
19. **Maarja Ojamaa.** The transmedial aspect of cultural autocommunication. Tartu, 2015, 173 p.
20. **Tiit Remm.** Sociocultural Space: Spatial Modelling and the Sociocultural World. Tartu, 2015, 142 p.
21. **Riin Magnus.** The Semiotic Grounds of Animal Assistance: Sign Use of Guide Dogs and Their Visually Impaired Handlers. Tartu, 2015, 154 p.
22. **Davide Weible.** Exaptation: Towards a Semiotic Account of a Biological Phenomenon. Tartu, 2016, 144 p.

23. **Mari-Liis Madisson.** The Semiotic Construction of Identities in Hypermedia Environments: The Analysis of Online Communication of the Estonian Extreme Right. Tartu, 2016, 195 p.
24. **Claudio Julio Rodríguez Higuera.** The Place of Semantics in Biosemiotics: Conceptualization of a Minimal Model of Semiosic Capabilities. Tartu, 2016, 148 p.
25. **Kadri Tüür.** Semiotics of Nature Representations: on the Example of Nature Writing. Tartu, 2017, 257 p.
26. **Gleb Netchvolodov.** Pictorially simplified images as machine vision design tool: semiotics approach. Tartu, 2017, 143 p.
27. **Сильви Салупере.** О метаязыке Юрия Лотмана: проблемы, контекст, источники. Tartu, 2017, 210 с.
28. **Remo Gramigna.** Augustine and the study of signs and signification. Tartu, 2018, 237 p.
29. **Nelly Mäekivi.** The Zoological Garden as a Hybrid Environment – A (Zoo)semiotic Analysis. Tartu, 2018, 172 p.
30. **Silver Rattasepp.** The Human Mirror. A Critique of the Philosophical Discourse on Animals from the Position of Multispecies Semiotics. Tartu, 2018, 156 p.
31. **Ott Puumeister.** On Biopolitical Subjectivity: Michel Foucault's perspective on biopolitics and its semiotic aspects. Tartu, 2018, 170 p.
32. **Lyudmyla Zaporozhtseva.** Structural Units of Mass Culture Mythology: A Cultural Semiotic Approach. Tartu, 2019, 193 p.
33. **Tatjana Menise.** Fairy Tales in Transmedia Communication: Fanfiction. Tartu, 2020, 142 p.
34. **Alexandra Milyakina.** Digitalization of Literary Education in the Context of Cultural Autocommunication. Tartu, 2020, 137 p.
35. **Tyler James Bennett.** Detotalization and retroactivity: black pyramid semiotics. Tartu, 2021, 238 p.
36. **Jason Mario Dydynski.** Semiotic Modeling of Cuteness in Cartoon Characters/Mascots. Tartu, 2021, 176 p.
37. **Tiina Hoffmann.** La traduction cinématographique en Estonie soviétique : contextes, pratiques et acteurs. Tartu, 2021, 489 p.
38. **Mirko Cerrone.** Interspecies Relationships: A zoosemiotic Analysis of Human-Ape Communication. Tartu, 2021, 160 p.
39. **Jonathan Grant Griffin.** A Semiotic of Motives: Kenneth Burke and Deely-Tartu Semiotics. Tartu, 2022, 244 p.
40. **Kyle Jonathan Davidson.** Semiotic Modelling of Identity and Communication in Virtual Reality, Augmented Reality, and Mixed Reality. Tartu, 2022, 221 p.
41. **Aleksandr Fadeev.** Lev Vygotsky's approach in the Contemporary Semiotic Research of Learning, Meaning-Making and Inner Speech. Tartu, 2022, 146 p.
42. **Eugenio Israel Chávez Barreto.** The Semiotic Theory of Luis Jorge Prieto. Tartu, 2022, 236 p.

43. **Katarina Damčević.** Semiotics of Hate Speech and Contested Symbols: The “Za dom spremni” Ustaša Salute in Contemporary Croatia. Tartu, 2023, 166 p.
44. **Merit Rickberg.** Towards Complexity Thinking With Juri Lotman: Modelling Cultural Dynamics in Educational Systems. Tartu, 2023, 233 p.
45. **Tuuli Pern.** Affective-imaginative modelling in semiotic context: A Vichian perspective. Tartu, 2023, 147 p.
46. **Auli Viidalepp.** The Expected AI as a Sociocultural Construct and its Impact on the Discourse on Technology. Tartu, 2023, 149 p.
47. **Lauri Linask.** Autocommunication in the Semiotic Development of the Child. Tartu, 2024, 193 p.
48. **Oscar Salvador Miyamoto Gómez.** The Forms of Memory: Biosemiotic Modelling of Alloanimal Episodic Semiosis. Tartu, 2024, 175 p.
49. **Andrew Mark Creighton.** Wolves as Signs of McDonaldization in Northern Manitoba. Tartu, 2024, 164 p.
50. **Daria Arkhipova.** How Artificial Intelligence recommendation systems impact human decision-making. Tartu, 2024, 327 p.