

# Semiotic mode and sensory modality in multimodal semiotics: Recognizing difference and building complementarity between the terms

Martin Oja<sup>1</sup>

**Abstract.** This article addresses an issue in multimodal cultural studies – the inconsistent use of the notions ‘mode’ and ‘modality’. As these notions are frequently employed interchangeably, making a clear distinction between them and positioning them in a coherent system will be helpful. To outline such a system, I envisage a two-layer framework where modes and modalities support each other. The central branch of multimodal semiotics (developing from Gunther Kress’ sociosemiotics towards John Bateman’s comprehensive approach) recognizes ‘mode’ as a pivotal research concept. While ‘mode’ as a semiotic resource is dependent on its materiality, culturally shaped practices and discourse semantics, the neurocognitive characteristics of sensory modalities are often seen as secondary to meaning-making. This article suggests that discussion of the semiotic potential of sensory modalities is complementary to the semiotic theory of multimodality. In order to illustrate this, I will construct an experimental typology of modality relations, which also takes modes into account. This typology distinguishes between supporting, modifying, conflicting, substituting and cross-activating relations.

**Keywords:** multimodality; semiotic mode; sensory modality; film

When in 2014 Kalevi Kull and Ekaterina Velmezova asked several leading semioticians questions about the main challenges in contemporary semiotics, quite a few scholars pointed out the usage of varying metalanguages and the problem of terminology lacking a systematic character. It was also observed that taking basic concepts as self-evident may cause lack of clarity in their definitions (Kull, Velmezova 2014: 532, 535). Academic research into multimodality encounters similar problems for two main reasons: first, akin to semiotics, it is

---

<sup>1</sup> Department of Semiotics, University of Tartu, Jakobi 2, 51005, Tartu, Estonia; e-mail: martino@ut.ee.

an interdisciplinary field; second, while it is likewise concerned with meaning-making, it closely intersects with semiotics, sharing its strengths and struggles.

John Bateman and Karl-Heinrich Schmidt (2011: 75) have noted about ‘mode’, one of the central concepts in multimodality, that its precise nature remains unclear and a variety of definitions circulate in the literature. A reason for the situation that Hartmut Stöckl (2014: 275) described as the “messy character of the field” at least partly lies in the co-existence of two basic viewpoints about the components that make up a multimodal unit. Lars Elleström (2020: 41) delineates these positions, saying that “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)”. More often than not, the efforts to align these two perspectives have led to various hybrid categories that contain elements of both types.<sup>2</sup> This, in turn, appears as an obstacle when clear discussion of mutual interaction of such types (textuality and materiality *vs* perception and sensorial processing) is required.

This article attempts to offer a contribution to organizing the methodological field of multimodality research, putting forward the argument that (semiotic) mode and (sensory) modality approaches should be seen as distinct, yet complementary. While pertaining to different aspects of meaning-making, these two sides are closely integrated in semiosis. Nevertheless, the mutual relationship between materiality and perception can be properly addressed only when modes and modalities are seen as characterizing different types of elements. After a brief discussion of the mode approach I will focus on the range of problems that can be addressed with the modality approach and, lastly, build a simple typology of modality relations that illustrate how modalities and modes can be effortlessly combined as different types of components in the same system.

## Complex sign systems and synthetic meaning-making

The semiotic theory of multimodality is concerned with complex objects that require a certain type of parallel processing. Multimodal texts employ various sign systems and, in order to be interpreted, call for simultaneous activity of multiple brain circuits that correspond to different sensory channels (see Anderson *et al.* 2004; Zacks 2015; Gallese, Guerra 2020). Although traditional theories of semiotics have predominantly been concerned with sign systems that are based on a single “grammar” or set of rules, meaning as resulting from synthesis has been

---

<sup>2</sup> It can be suggested that this is also a prominent reason why the terms ‘mode’ and ‘modality’ are frequently used interchangeably.

acknowledged throughout the development of modern semiotics, from Peircean compound signs (CP 2.441; Bateman 2018: 2) and Jakobson's syncretic messages (Jakobson 1971[1967]: 705) to Lotman's multiplicity of languages in an artistic text (Lotman 1990a: 211), Kress' sets of semiotic resources as modes (Kress 2010: 79) and O'Halloran's multisemiosis (e.g. O'Halloran 2004: 94). Hence, multimodality is about the synthesis of multiple components, and a successful approach to such a process will depend on fruitful demarcation of these components.

Comparing three major approaches to multimodality, namely conversation analysis, systemic functional linguistics, and social semiotics, Carey Jewitt and her colleagues see the issue of unit-naming as a crucial part of theoretical activity, which they call "conceptualization of the means for meaning making" (Jewitt, Bezemer, O'Halloran 2016: 11). Still, it should be noticed that all the approaches discussed tend to rely on materiality or, in other words, "the stuff in which communication is inscribed" (Bateman *et al.* 2017: 103). As Campbell and his co-authors notice, social semiotic approaches tend to neglect the role of the body in communication (Campbell, Olteanu, Kull 2019: 358). Nevertheless, meaning-making depends not only on the material, textual and social aspects of semiotic resources, but also on the neurocognitive characteristics of sign users. Although these have historically been poorly accessible as objects for observation,<sup>3</sup> processing information by specific neurological circuits has its own regularities and implications on meaning. Such embodied meaning-making and the significance of emotions and personal memories were anticipated, for example, in the phenomenology of Maurice Merleau-Ponty (1962) and Vivian Sobchack (e.g. 1992).

As Kathy A. Mills and Len Unsworth (2017) point out, modes as theoretical constructs are defined differently across various schools, their classification being somewhat contested. While modes tend to generate a variety of descriptions (Bateman, Schmidt 2011: 75), the notion of sensory modality offers an unambiguous classification: distinction of modalities is based on the type of physical stimulation – light for vision, sound for hearing, skin pressure for touch, molecules in air for smell, etc. (Bertelson, De Gelder 2004: 141). On the one hand, for evolutionary reasons it is helpful for organisms to identify the sources of different sensory inputs. On the other hand, inputs from different modalities are integrated into consistent perceptual experiences (Pennartz 2009: 720). Nevertheless, how the brain "binds" a multitude of stimuli into consistent

---

<sup>3</sup> It is interesting to notice with Bateman that Peirce himself was "quite antagonistic to individual, internalised, psychologised views of meaning and so does not, in general, talk of semiosis in this way at all" (Bateman 2018: 7).

perceptions or integrates the sensorial multitude into a coherent perceptual space, is still a widely disputed question (see e.g. Whitney 2008: 251). As the understanding of semiosis strongly depends on such issues, providing an explanation how sensory modalities mediate between distinctions and similarities is a pertinent goal.

The distinction between mode and modality approaches should also be considered against the background of (multi)media studies, as all three fields share partly overlapping spheres of interest as well as the task of negotiating the terms that concern the signifying substance and the practices of its structuration and interpretation. These paradigms can be aligned in several ways. For instance, Elleström embeds the discussion of both modes and modalities into his framework of media studies. While media modalities<sup>4</sup> can be divided into material, sensorial, spatiotemporal and semiotic ones (Elleström 2020: 20), modes can be seen as having both material and spatiotemporal characteristics (Elleström 2020: 41). Elleström interprets senses basically as channels – “media products must reach the mind through at least one sense” (Elleström 2020: 20).

A reversed perspective is also possible: characteristics of a medium can be approached in terms of properties of modes, or at least media discussed by the means of modes. This is done by Bateman, Wildfeuer and Hiippala (2017: 71), who assert that media must be seen as particular bundles of semiotic modes. Although analysing media platforms gives us knowledge about the participants in the communicative situation, it does not provide information about communicative strategies or why the situations are occurring at all (Bateman *et al.* 2017: 69). Thus, the notion of the medium can be reconstructed as “generalized materialities of modes or canvases” (Bateman, Wildfeuer, Hiippala 2017: 103). I suggest that, following Bateman and his colleagues, it is possible to discuss a mode-centred approach without directly addressing the problems of media research. Hopefully, it will also allow us to envisage a framework that is established both on modes as semiotic resources and sensory modalities as factors that have neurocognitive influence on meaning-making.

In the following, I will take a brief look at the development of the mode approach, as well as some points of criticism that have been aimed at it. The relative flexibility that characterizes the definition of ‘semiotic mode’ sometimes leads to the swapping of ‘mode’ and ‘modality’, i.e. their treatment as synonyms. This tends to cause categorizations that base themselves simultaneously on multiple grounds, taking into a single arrangement both textual components that

---

<sup>4</sup> In the context of this article, this is the first reference using ‘modality’ in an alternative and, for some, probably confusing sense, but it will not be the last one.

are distinguished by their materialities (modes) and components that are defined by their sensory properties (modalities). I will turn to several of such occurrences, insisting that such systematizations, especially addressing the fundamental levels of research object, work against methodological clarity. For an efficient discussion of how materialities interact with perception in the process of meaning-making, clear differentiation between (semiotic) modes and (sensory) modalities will benefit any approach or framework.

Next, I will discuss the relevance of sensory modalities to the process of meaning-making, suggesting that the focus on modalities should not be dismissed as less important, but seen as a necessary complement to the theory (foremost that of Bateman's) of semiotic mode. Acknowledging the semiotic potential of sensory modalities will contribute to the overall research of multimodality, making it possible to tackle the problems of multimodal integration and cross-modality, i.e. the impact of one sensory modality on the perception of another (Spence, Senkowski, Röder 2009: 108). As the conclusion of the first part of the article, I will contemplate a two-level framework in which the categories of semiotic modes and sensory modalities have equal importance and their relation is complementary, benefitting from the flexibility of modes and the concreteness of modalities. This kind of framework allows us to address specific issues whose interpretation strongly depends on the distinguishing sensory modalities. For instance, one such problem is the ways in which the meaning of a film is influenced when the translation of the dialogue is represented by subtitles (verbal mode in visual modality) or by voiceover (verbal mode in auditive modality).

In the final part of the article, I will build a preliminary typology of relations between different modalities, giving examples both from among simple cognitive actions that reside near the lower threshold of semiosis as well as intricate culture texts, especially film. In the typology, the supporting and modifying relations can be recognized as the most common ones for evolutionary reasons. Substituting, cross-activating and conflicting relations are the results of some specific, often purposefully misleading, artistic or synthesis-dependent semiotic practices. For instance, one modality may refer to a certain type of information and another undermine it, as was the case in the famous McGurk experiment where visual and auditive information contradicted each other and created a third meaning in speech perception (McGurk, MacDonald 1976). Last, but not least, an additional purpose of the typology is to illustrate the possibility of a clear demarcation between modes and modalities, as well as how this demarcation contributes to the analysis of complex semiotic objects.

## The semiotic mode: origins and critiques

A prominent branch of multimodal research, the sociosemiotics of Gunther Kress and his colleagues, was developed on the basis of Michael Halliday's linguistics. Halliday approached language as a context-dependent resource system for meaning-making (Lemke 2009: 284), noticing that a linguistic unit (e.g. sentence) embodies several semantic structures at the same time, each corresponding to a different function of language. Halliday differentiated three metafunctions: *the ideational function* expresses speaker's experience of the external world; *the interpersonal function* deals with the relations between the participants in communication; *the textual function* is about language's capacity to create text and relate it to the context (Halliday 1978: 45, 48). It is interesting to notice that Halliday, when speaking of resource, employs an environmentalist metaphor – language is seen as a resource for meaning (Halliday 1978: 17).

Discussions of the materiality in communication and media studies in the last decades have been called the 'material turn' and this can be seen as the context for the contemporary approaches to multimodality (Bateman 2021: 35, 36). Kress and van Leeuwen expanded the idea of a resource system beyond the borders of natural language to material objects, gestures, audiovisual entities, etc. Hence, the analysis of language's metafunctions led to the concept of the semiotic mode as reflecting the use of materiality for the achievement of these metafunctions (Bateman, Wildfeuer, Hiippala 2017: 49). From language, the frame extended to all kinds of meaning-making practices; 'mode' acquired the meaning of 'semiotic resource' (Boria, Tomalin 2020: 12, 13). Kress (2020: 28) explains that "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".

It is also significant how the notion of the semiotic mode is influenced by James Gibson's term 'affordance' (Gibson 2015[1979]: 119). Jeff Bezemer<sup>5</sup> points out that Kress adopted the term in the late 1990s, when it was suggested to him by Jon Ogborn. Bezemer also indicates that while Gibson's 'affordance' referred to physical objects, in the context of multimodal research the semiotic dimension was attached to materiality. In parallel, 'affordance' helped Kress to bring materiality into his theory, and advance the idea that communities developed semiotic resources over time by selecting how to use various kinds of 'stuff'

---

<sup>5</sup> Bezemer, Jeff 2022. The affordance of affordances in multimodality. Multimodality Talks (online event) is available at: [https://ucl.zoom.us/rec/play/H715XAXjldg95LHeG-lrBORv1\\_flTemR3j-7mZlBlcIQ4tboe4cPegPjfb7Rnx58sjygx9gwFDPSAlaM.yRPeP-glFkalHXxI](https://ucl.zoom.us/rec/play/H715XAXjldg95LHeG-lrBORv1_flTemR3j-7mZlBlcIQ4tboe4cPegPjfb7Rnx58sjygx9gwFDPSAlaM.yRPeP-glFkalHXxI). Last accessed 22.06.2023.

(Bezemer 2022).<sup>6</sup> All in all, the proximity of modes and affordances suggests constant interplay between material characteristics and creative social practices:

Foremost among the terms integral to this theory is that of *affordance*, a term which points to the potentials and limitations of specific modes for the purposes of making signs in representations. Affordance rests, on the one hand, on the materiality of the stuff, which work in social environments has fashioned into a cultural and semiotic resource on the other hand. (Kress 2010: 157)

Materiality also has the fundamental position in Bateman's model of the semiotic mode, functioning as its foundation. This model, consisting of three layers, provides one of the most comprehensive treatments of semiotic mode up to date. According to Bateman, every mode relies on its material component, on the grammar-like or lexicon-like principles that enable "drawing functionally motivated differentiations in form", and finally, on a more abstract stratum of discourse semantics that allows the mode work in its cultural context (Bateman 2016: 47). Nevertheless, it is still somewhat inconvenient to apply this model to a certain issue of the semiotic mode, namely the problem of its size or scope. As Bateman himself admits, discussing a semiotic mode of language and a semiotic mode of pictures, the delineations are pre-theoretical because of the work still required to be done in explaining to what extent single semiotic systems are operative (Bateman 2016: 48). This leaves open the issue if a specific semiotic mode should be further segmented according to the semiotic situation. As such, this problem calls for the notion of submodes.

Hartmut Stöckl, who offers one of the clearest systematizations of modes, acknowledges that modes and submodes are rather dynamic: they tend to mix, blend and overlap (Stöckl 2004: 9). For Stöckl, modes and submodes are strongly interdependent and can be seen functioning in a network. For example, there is a pictorial element in language and a linguistic element in images (Stöckl 2004: 10). Still, Stöckl's own division of modes is rather concrete and well-ordered. He distinguishes 'core modes' (*image* and *language*), 'medial variants' of core modes (*static* and *dynamic* for image), 'submodes' (*panning*, *tilting* and *cut* for 'dynamic image'), 'features' that characterize submodes (e.g. the *speed* of the submodal action) and lastly, 'peripheral modes'.<sup>7</sup>

<sup>6</sup> Bezemer (2022) also draws attention to a fundamental question: what are the affordances of modes? That is, what can a mode do and what it cannot, or, what can and cannot be done by a mode?

<sup>7</sup> Elsewhere, Stöckl also discusses depictive modes that characterize the relation between the signs used and the objects they refer to; depictive modes are especially relevant in digital art, raising semiotic issues of reference (Stöckl 2021: 190).

The importance of submodes<sup>8</sup> is revealed in the fact that they provide the building blocks of a mode's grammar (Stöckl 2004: 14). The notion of grammar has a central position in the functional approaches to mode; therefore, the study of various modes' grammars or "grammars" has often been one of the main tasks for researchers of multimodality:

Along with 'mode' comes the notion of a 'grammar', i.e. signs belonging to one mode are seen to be governed by a common set of rules that state how these signs can be combined to make meaning in particular situations. In practice, however, things turn out to be less straightforward than this. (Stöckl 2004: 11)

This "less than straightforward" is first and foremost caused by modes' semiotic asymmetry: some modes are oriented to iconic, some to indexical or symbolic representation. When modes shift and blend, as Stöckl acknowledges, the grammar of a mode can shift with the mode. Here, an important question is: what mechanisms cause modal asymmetry and the shifts in modes' grammar? In the first place, the answers would point to the material particularities and specific cultural practices that influence modes' characteristics, but as in semiosis the materiality and perceptual processing are entwined, also sensory modalities have an important role to play.

Before touching upon this subject, let us briefly highlight the link between the relativity of modes and the (self-)criticism that has been voiced towards the term. According to Kress (2010: 130), modes are shaped by the histories of their making in specific societies and therefore vary across cultures. Apparently, understanding of the mode also varies across research cultures, leading different theoreticians to divergent use of the term. In the words of Bateman and Schmidt (2011: 75):

For some, modes are "the 'stuff' which a culture uses as the means for the expression of [its] meanings" (Kress, 2000: 185) and form 'a regularized organized set of resources for meaning-making' (Jewitt and Kress, 2003: 1); for others, a mode is 'never a bounded or static unit, but always and only a heuristic unit' (Norris, 2004: 12). Some writers even suggest that it is not possible to define or list semiotic modes at all (e.g., Forceville, 2006: 382).

---

<sup>8</sup> Still, what to recognize as a mode and its feature, submode, or submode of a submode, depends on the sign user; the topic is open to dispute. Naturally, for Kress and his colleagues, these divisions are socially constructed: "The question of whether X is a mode or not is a question specific to a particular community. As laypersons we may regard visual image to be a mode, while a professional photographer will say that photography has rules and practices, elements and materiality quite different from that of painting and that the two are distinct modes." (Kress *et al.* 2000: 43) As the example illustrates, it is difficult to subordinate the "level of detail" to a mutual agreement.

Similarly, Jewitt, Bezemer and O'Halloran (2016: 12) point out:

There is, put simply, much variation in the meanings ascribed to mode and (se-  
miotic) resource. 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.

They also clarify the main reason behind this variety: according to them, the  
divergent opinions are rooted in the varying understandings of the relation of  
material and semiotic components in signs. While some theoreticians talk of  
modes in terms of materiality, others support the idea

[...] that a given mode may have different material manifestations. The common  
assumption that language is a semiotic mode that may appear in both written  
(i.e., visual) and spoken (i.e., aural) form is one consequence of this. Many have  
discussed this distinction before, as it relates broadly to Saussure's distinction  
between signifier and signified [...] to Hjelmslev's widely cited distinction between  
expression and content [...] and so on. (Bateman, Wildfeuer, Hiippala 2017: 113)

Thus, once again, we are hindered by the dialectic relationship of two fundamental  
components that can be called by different names – content and form, *signans* and  
*signatum*, information and its carrier. They blend into each other in the process  
of meaning-making, and, therefore, their connection remains too dynamic to be  
completely transparent. Still, conflicting opinions in the theory do not usually  
sprout from preferring one component over another, but from the adjustment  
of focus on the border between the material and the semiotic. This leads to such  
categorical doubles as Bateman and his colleagues referred to. For example,  
language can be called a mode, but comparing auditive language and visual  
language it remains somewhat inconvenient to acknowledge that the same mode  
can occur with radically dissimilar properties.

### **Mode or modality: interchangeable use**

Before addressing the role of sensory modalities in multimodal meaning-making,  
I would shortly illustrate a frequent consequence of the previously mentioned  
tendency towards the building of “hybrid categories” that incorporate both

material and sensorial aspects, leading to varying definition of modes and the interchangeable usage of ‘mode’ and ‘modality’. In a broad sense, this concerns multimodality research as the whole discipline, where navigating between the two approaches, or recognizing the benefits of clear categorical differences, has not been effortless even on highly professional levels. For example, media scholar Charles Forceville discusses both the verbal and the visual as modes: “I had to work within the limits of the two modes that I am reasonably knowledgeable about: visuals and language” (Forceville 2020: 3). Later in the text he provides a further explanation – he sees no need to differentiate between modes and modalities at all: “For present purposes, the main point is that ‘visuals’ contrast with other modes or modalities (which are the two interchangeable terms I will use for what in other approaches are called semiotic systems or semiotic resources).” (Forceville 2020: 64) The passage is revealing and symptomatic for a number of thinkers.

Cognitive linguist Neil Cohn (2016: 304) employs the term ‘modality’ in the sense of Kress’ ‘mode’: “Humans communicate through different modalities – whether through speech, bodily movements, or drawings” and occasionally mixes categories, e.g. adds visual as a sensory modality into the list of modes: “Multimodal interactions are characterized in terms of their component cognitive structures: whether a particular modality (verbal, bodily, visual) is present”. In a similar manner, narrative theorist Ruth Page, in an introduction to multimodal narrative analysis, comes up with the hybrid term ‘sensory mode’ and, perhaps as a result of a slippage, in the very next sentence, evokes ‘sensory modality’. It is difficult to understand if she means this as a synonym for the former: “[...] the role of sensory modes remains vital to a multimodal narrative analysis. Storyworlds routinely depict sensory modality, indexically evoked through verbal or visual resources, and, more generally, narrative communication is itself embodied.” (Page 2010: 7)

Even John Bateman’s work, despite its evocative engagement with the semiotic theory of multimodality, occasionally leads to an uncertainty between ‘mode’ and ‘modality’:

For this, we will proceed in terms of an ontological analysis of what constitutes a semiotic mode. Ontological is used here in the sense of determining what has to be available for there to be semiotic modalities of the kind that are of interest to us at all as well as what internal organizations such modalities must exhibit. (Bateman 2016: 39)

If a difference of meaning is considered here, it is left unexplained; therefore ‘semiotic mode’ and ‘semiotic modality’ appear as synonyms. Although Bateman’s definition of mode is clear-cut, his framework does not grant sensory channels

any defining role (Bateman 2016: 41) and, probably therefore, the marked or dominant notion overshadows the unmarked one.

Lastly, let us make a brief excursion into multimodal film analysis and observe the problem of definitions from yet another angle. Janina Wildfeuer, a leading specialist in the field, segments film into four basic unit types: (1) visual content; (2) audio track; (3) spoken language; and (4) music (see Wildfeuer 2014: 24). She does not call these four categories ‘modes’ or ‘modalities’ directly, but suggests the notion of ‘modality’: “[T]he interaction of the modalities in the discourse is of central interest, since it is the interrelationships between the different semiotic resources that are decisive for the overall structure and coherence of a film.” (Wildfeuer 2014: 24) When talking of semiotic resources, the use of the ‘mode’ derived from Kress’ tradition would be more precise.<sup>9</sup> Nevertheless, the problem does not arise so much from Wildfeuer’s conceptualization but from the research object itself. Among the four divisions, the first one (the whole visual content that can be perceived in a film shot) can be understood as an extensive super-mode that employs visual sensory modality or, in Stöckl’s typology, a visual sensory channel (Stöckl 2004: 13). The three others more or less correspond to Stöckl’s core modes and employ auditory modality or an auditory sensory channel.

Wildfeuer’s typology brings to mind two others: Christian Metz’s and Seymour Chatman’s classic divisions. Let us remind that Metz (1973: 90) saw five different

material ‘signifying substances’ in film: the moving photographic image, the recorded musical sound, recorded phonetic sounds, recorded noises, and graphical material – i.e., the graphic outline of written items (credits, [...] intertitles in the silent film, etc.).<sup>10</sup>

Both systems, at the deeper level, deal with culturally modified materiality. Chatman’s typology of the cinematic narrator’s communicating devices starts the division with two basic channels (auditory and visual) which more or less overlap with sensory modalities. Concerning the visual channel, Chatman differentiates between ‘nature of image’ and ‘treatment of image’. Treatment divides between cinematography and editing with their respective subdivisions (Chatman 1990: 135). While Chatman’s subcategories can, with certain reservations, be interpreted as modes, his division, when compared to Metz’s, is even more oriented towards

<sup>9</sup> Such usage of ‘modality’ for ‘mode’ is not a random slippage. Right from the introduction, Wildfeuer (2014: 1) employs the terms interchangeably: “Meaning in film arises out of the multiple interaction of various modalities such as images, sounds, music, gestures, camera effects, etc.”

<sup>10</sup> Here, we can also recall Umberto Eco’s typology of cinematic codes (see Eco 1976: 596).

functionality – actually so much so that it subjugates materiality or resources into a secondary position. Indirectly, it also helps to remind that materiality is affected by not only social forces, but also by evolutionary and neurocognitive factors.

Therefore, theories of meaning-making tend to encounter a dilemma: on the one hand, it is possible to omit cognitive components from the analyses and descriptions, which leaves the theories somewhat incomplete; on the other hand, integrating material and cognitive aspects tends to require interdisciplinary approaches with their specific inconveniences, e.g. the need of coordinating between multiple, yet not always compatible, metalanguages. Despite the fact that the mode approach and the modality approach seem to imply two different semiotics with their respective terminologies, I suggest that an integrated framework is conceivable and beneficial. Paraphrasing Peirce will help here: while modes are concerned with sign vehicles, modalities are oriented towards interpretants. In short, this means bringing together the semiotics of textual combinations and the meaningfulness of reception, cognitive processing and subjectivity.

### **Modalities and their role in the integrated framework**

It has been frequently noticed that a framework based on modalities alone would not provide enough explanatory power in the situation of practical textual analysis – and this position is convincing. Stöckl (2004: 11) has formulated the viewpoint and concisely indicated the complications:<sup>11</sup>

[...] there is justification to subdivide [between] visual, auditory [...] tactile, olfactory and gustative sign(-mode[s]). While this seems a valid and largely unchallengeable systematisation, it is too rough to be of much use to a theory of multimodality. Language or the linguistic code, for instance, which we justly feel to be a mode in its own right, would thus fall both into the visual and the auditory category as it can be medially realised as either speech or writing. Both are governed by the grammar of language, but as different medial or material varieties.

In a similar manner, Bateman and his colleagues claim that talking of, for example, the ‘visual’ is much too general to reveal differences in use. While written language and pictures are both visual, they are being used completely differently (Bateman, Wildfeuer, Hiippala 2017: 114). Evidently, this difference can be properly

---

<sup>11</sup> Stöckl does not use the term ‘sensory modality’, but this paragraph uses ‘sign-mode’ in that sense.

addressed by semiotic modes. For this purpose, Bateman (2016: 70) brings into his framework “[...] many, ‘smaller’ semiotic modes operating in a semiotic artefact or performance than typically revealed by broader, sensory channel views”. He writes:

Assigning priority to sensory channels when distinguishing modes is not then seen as methodologically helpful for the central multimodal task of explaining how combinations of modes function productively to produce meaning. Only when attention is turned to the micro-level of individual semiotic modes do we find the necessary level of detail for formalizing and empirically investigating how meanings are being combined. (Bateman 2016: 70)

It is easy to agree with Bateman about modes: we need a decent resolution to “map the territory” properly. Evidently, semiotic modes can address differences that sensory modalities cannot, and there is a *caveat*: the foci of modes and modalities do not overlap. That means, modes and modalities address *different* differences: the former are concerned with differences pertaining to materialities, the latter with those related to perceptual processing. At the same time, those two systems of differences make difference to each other, to paraphrase a famous passage by Gregory Bateson.<sup>12</sup>

Therefore, I do not think it is necessary to position mode view and modality view as somehow competing or mutually exclusive. Even if the practice of previous multimodal theory has outlined such tendencies, it is now time to consider an integrated framework where semiotic modes and sensory modalities are complementary. This framework would be bi-directional: the situations where meaning-making significantly depends on the co-existence of several modalities, e.g. the visual and the auditive in film, the analysis would begin with modalities as wide units, describing problems such as the domination of one modality over another and then moving to the system of modes as smaller units. The typology of modality relations in the last part of this article attempts to illustrate this viewpoint in a straightforward manner.

Such an approach is compatible with a view that distinguishes two types of multimodalities. Such demarcation has been elaborated by Ellen Fricke, who differentiates between multimodality in the narrow and the broad sense. According to Fricke, narrow multimodality employs more than one ‘sense modality’; broad multimodality relies upon multiple codes in a single sensory modality (Fricke 2013: 740). Whether the broad or the narrow multimodality is actualized will depend on the research object. While either one requires somewhat

---

<sup>12</sup> Bateson (2000[1972]: 272) has written, “A difference which makes a difference is an idea. It is a ‘bit’, a unit of information.”

specific methodology, it can be said that the broad is being embedded by the narrow, because broad multimodality is concerned with modes alone, while narrow multimodality has to do both with modes and modalities.

While the concepts of mode and modality are kept distinct, both can still be applied onto the same research object simultaneously. First of all, it should be asked what range of problems is suited for being addressed with a mode system and what with a modality system. It is important to notice here that modalities should not be understood as merely sensory channels,<sup>13</sup> but also as a category that is concerned with cognitive processes. Therefore, an extended model of a modality would entail a channel *and* a brain with their functions that relate to the processing of sensory information, including memory and emotional apparatus. Thus, modalities can be seen as organizing/interpreting principles that apply to meaning-making and stem from neurocognitive properties of the interpreters.<sup>14</sup>

Concerning the dual system of modes and modalities, the difference in scale is also beneficial. As modes are “smaller”, they can be developed for quite specific purposes – they allow more precise modelling of specific meaning-making devices, e.g. certain camera techniques or editing methods in film. Modes are also more adjustable in the hands of theoreticians. While a mode system is an adjustable toolkit, modalities function as a fixed reference system. Modalities can be seen both as carriers and features for modes. In other words, modalities provide an “environment” for modes with its conditions and properties.

First and foremost, differentiating between modes and modalities allows us to keep clarity in very simple things. As an example, let us look at one of the most celebrated multimodal works (in the broad sense!) of art history. René Magritte’s painting *The Treachery of Images* (1929) entails an intriguing mechanism – a conflict between semiotic modes. Under the image of a pipe, a text declares “*Ceci n’est pas une pipe*”. The artwork uses a single (visual) modality, which brings along its practices of production and reception and, most importantly, requires specific

<sup>13</sup> Here, caution about careless use of metalanguage is relevant: “[T]alking of ‘modes’ in terms of ‘sensory channels’ can be quite misleading. Sound is not just sound or the hearing of tones; it also gives information about space, hardness, distance and direction.” (Bateman, Wildfeuer, Hiippala 2017: 27)

<sup>14</sup> An extended model of sensory modalities allows us a possible connection with the interpretant, on the one hand, and discourse semantics, on the other hand. For instance, as Arthur Shimamura (2013: 21) points out, some avant-garde movies provide meta-level commentaries that are enjoyable only when the viewer has previous knowledge of film history, techniques and stylistics. This is just one example of the aspects of meaning-making that cannot be addressed with the mode approach alone, but require the consideration of how personal knowledge is activated and employed in the interpretation; such processes can be seen partly mediated by sensory modalities.

cognitive mechanisms for processing it. As there is only one sensory modality, there is no *modality conflict* in that painting, but a *conflict between modes*. (It should be noticed that talking about the internal textual conflicts in the terms of media would not be feasible, because such an artwork is, in a media-centred view, usually reduced to a single medium – the ‘painting’ or ‘visual image.’) Therefore, the painting features at least two core modes, which can be labelled ‘image’ and ‘text’<sup>15</sup> and are characterized by their affordances and materiality. Visual modality, due to its sensory properties, sets certain constraints to semiosis, affecting the affordances of the image and the text.

Due to their characteristics, modalities enable addressing general and robust problems that in turn influence device-specific or mode-specific discussions. A prominent sphere of such issues concerns the synthesis between meaningful components of multimodal texts or artefacts, and devices that facilitate intersemiotic cohesion or form semantic relations that has been called intersemiotic texture (Liu, O’Halloran 2009: 368).

Every complex semiotic system also poses the question of the dominant<sup>16</sup> (Jakobson 1981[1935]: 752). It does not matter whether the analysable components are modes or modalities, it is always relevant to ask which components dominate the system, i.e. provide the organizing principles for the other components and the whole. It is also important to acknowledge that any type of semantic coherence is dependent on sensory processing. When a research object is multimodal in the narrow sense (employing more than one sensory modality), it calls for an interdisciplinary dialogue with empirical disciplines, e.g. cognitive psychology and neurosciences. Also the discussion of modes enjoys more solid foundation when it is in accordance with the underlying cognitive mechanism, for example, the principles of multisensory integration.

---

<sup>15</sup> A situation where two modalities carry the same mode is also possible. In film, the textual (or verbal) mode can be represented by auditive modality (dialogue) and visual modality (subtitles) simultaneously. Such a situation entails amodal invariants that are discussed by Taberham (2013). Amodal invariants are common properties of objects that are represented in different modalities but perceived as carriers of the same meaning (Taberham 2013: 47). It is important to notice that Bateman, discussing similar situations, follows the opposite direction. As semiotic modes are primary for Bateman, he acknowledges that semiotic modes can be multisensorial (Bateman 2016: 56).

<sup>16</sup> Here, it can be noticed that Roman Jakobson acknowledged senses as the preconditions for any type of signification: “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 [...] abstracted from time.” (Jakobson 1971[1967]: 701)

## Modalities in the context of multisensory integration

For providing a coherent typology of modality relations, it is helpful to seek a foundation from the frameworks science has already built. I propose to look into the field of cross-modal or multisensory integration (MSI) as one of such foundations. It has been a research subject of experimental psychology for more than a hundred years (De Gelder, Bertelson 2003: 460). Although a more detailed classification of MSI situations is still being striven for, some outlines are well established. The neurophysiological mechanisms proposed behind MSI are *summation, suppression, inhibition* and *neural synchrony*. In other words, coupling of the stimuli of two (or possibly more) sensory modalities can be additive, cumulative, inhibitory, or lead to a new emergent percept (De Gelder, Bertelson 2003: 465). The latter, in a way, would provide an illustration to or even a neural correlate for Juri Lotman's mechanism for the generation of new information.<sup>17</sup>

MSI research has established a division between artificial and natural couplings of stimuli. The first type of stimulus pairs are outlined under laboratory conditions for the purpose of an experiment (for example, a low-frequency tone paired with the visual image of a square and a high-frequency tone with a rectangle). Conversely, naturalistic combinations consist of components that occur simultaneously in natural environments, such as an explosion that emits light, noise and heat (De Gelder, Bertelson 2003: 465, 460). In addition, cultural texts can be approached as the third category containing literature, film, multimedia, social texts, products, games, rituals, etc. Obviously, this category is more complex and diverse. It features the properties of both – the controlled synchronicity of modalities and the adjustability of modes, but nevertheless applies to organisms' disposition, to make better sense of the environment.

Events in natural environments tend to produce information of several modalities at the same time, as in the case of an explosion mentioned above. As the adaptive incentive of organisms is to locate dangers and resources, it can be claimed that using multiple modalities to create a coherent world model is the default survival tactics for animals. First of all, this means benefitting from redundancy and, if needed, compensating the lack of sufficient data in one sensory

---

<sup>17</sup> Lotman described a situation where two incompatible systems are juxtaposed. He has interpreted such systems as "signs relating to different registers" (Lotman 1990b: 51), "chorus of simultaneously speaking languages" where "one may occupy a dominant position" (Lotman 1990a: 211), or more precisely, in some situations, "discrete and continuous" languages (Lotman 1990b: 36, 37). According to Lotman (1977: 283), "juxtaposed units that are incompatible in one system force the reader to construct an additional structure in which the incompatibility is eliminated".

modality by the data from others. Therefore, the situation where modalities support each other can be considered the most frequent one. In cultural texts, this is consistently exploited by narrative cinema where a plausible synchronicity between audio and visual components helps to create an illusion of a seamless storyworld. On the other hand, in art-house or experimental cinema that illusion is occasionally broken on purpose, juxtaposing different modalities and modes.

However, in cultural texts, the combination of artificial and natural properties is not the only complicating factor. In multimodal synchronization and perception the parameters of time and space are crucial. The nervous system must detect which sensory signals were caused by the same object and should be combined, and which were caused by independent objects and should be kept apart (Shams, Kim 2010: 279). In natural environments this task is facilitated by the closeness of the multiple signals in time and space. It is more complicated in cultural texts, where different modes may follow each other in time. Previously experienced information can influence the meaning of later information, and *vice versa*. On the largest scale, that of transmedial discourse, various independent texts can be aligned or identified with specific modalities and, in a sense, combine into a multimodal whole (e.g. action figures, comic books and film). Where a text starts and ends in such a media environment is in part dependent upon the reader or viewer. All this highlights the importance of memory while approaching transmediality.

As cognitive theorist Luis Antunes notices, the understanding of the workings of human perception and mind has been developed from a modular concept (in the times of James Gibson and shortly afterwards, i.e. the 1970s–80s) towards a network-based one (Antunes 2016: 18). The question of the integrated whole has been touched upon above, but let us return to it for a moment. In an integrated network, a unimodal activity such as reading a book becomes multimodal or synaesthetic, e.g. when we imagine colours and smells, or develop an adrenaline rush when mentally simulating characters' actions while engaging in a narrative. Neuroscience has shown that while receiving information of a certain modality, the brain has the capability to co-activate regions that run other modalities. This can lead to various forms of modulation, where one modality weakens or strengthens another, and causes perception errors as the ventriloquist effect.<sup>18</sup> When memories are formed by modifying synapses, any knowledge or information can be seen as an emergent property in this integrated system, being

---

<sup>18</sup> While watching a ventriloquist performing with a puppet, the viewer tends to accredit the sound to the puppet's moving mouth, supposing that the source of the sound is where the visual cue is seen.

qualitatively more than the sum of its binary pattern of active/inactive neurons. Under some conditions, those patterns can be seen as constantly multimodal even in the sense (no pun intended) of sensory modalities. In that case, the semiotics of multimodality may look forward to a new type of integrated unit, similar to Antonio Damasio's 'image' (see Damasio 2012: 18).

The discussion of crossmodal modulation (see e.g. Driver, Spence 2000; Nuku, Bekkering 2010; Gobara, Yamada, Miura 2016) may also shed light on the dominance and hierarchy between modalities. For quite a long time, visual processing was treated as independent of other modalities, often as the most prominent one. While vision has the capability to alter other percepts, the influence of other sensory modalities on vision has been acknowledged later. Shams and Kim (2010: 271, 272) demonstrated that sound may significantly affect vision, too; it is due to the fact that auditory modality has much higher temporal resolution than visual one. This means that the features of dominance and subordination can be investigated empirically and the gained knowledge can be applied to the analysis of culture texts. The idea of the auditory modality as the primary processor of time and visual modality as the tool for space-related actions (see e.g. Jakobson 1971: 701; Kress 2010: 81) can be developed further by semioticians or other researchers and meaning generation.

### **A typology of relations between sensory modalities**

Although the current framework is meant to address cross-modal actions between sensory modalities, also semiotic modes can be placed into this scheme. The system of modes can be visualized as a second stratum above modalities. Each modality supports or carries at least one mode or feature, and possibly more, depending on the type of text or the practice of meaning-making. In some cases, a single mode can be expressed through different modalities (e.g. see 'supporting', Example 1) with no contradictions.

Acknowledging that in any type of MSI the integration can vary in strength and degree, the relation types are rather idealized in the following classification. In reality, a significance of a multimodal relation will rely upon various factors and parties, most importantly on the specific interpreter and the cultural context. It must also be emphasized that the types are not mutually exclusive.

For some relation types, I will give several examples. For the sake of clarity examples of modality and mode are provided separately: modes follow modalities and are given in *italics*. Under each category, a brief comment about the main characteristics of the relation type is given. Some comments are added to specific

examples, if needed. The examples are deliberately varied in their scope and characteristics, ranging from simple physical actions to complicated cultural texts. The latter are mostly drawn from film as a diverse multimodal medium that employs both types of multimodal relations (those between modes and between modalities). The main goal of this exploratory typology is to analyse how the modality relations enable and modify the construction of meaning.

## Supporting

The supporting relation is the most common one as it facilitates the development of coherent perceptions about the environment. It creates redundancy across sensory channels, compensating for noise in separate modalities. In this relation, different modalities carry the same or a similar meaning (although the notion of ‘meaning’ must be approached with caution here). This relation is the most straightforward and simple one.

### **Example 1: Picking berries in a forest or a garden**

modality 1: visual

*perceptual features: colour, shape*

modality 2: tactile

*perceptual features: surface texture, shape*

This example involves perceptual features instead of modes. While the example addresses multimodality in the sense of modalities, it can be argued that the situation is pre-semiotic, or more specifically, pre-socio-semiotic and does not contain socially shaped semiotic resources. Here, subjects interact with objects via perceptual features that functionally resemble modes because of their affordances.

### **Example 2: Mainstream narrative fiction film – creating a coherent storyworld**

modality 1: visual

*modes: colour, camera movement, lighting, written text, etc. Also objects and features that carry a straightforward and practical meaning in an environment. Alike to the actions that such objects afford to the inhabitants of an environment, as discussed by Gibson (e.g. 2015[1979]: 35), the potential functions of such objects and features are revealed by visual modality.*

modality 2: auditive

*modes: dialogue, music (with submodes of diegetic and extradiegetic music), foley art and other sound effects*

**Example 3: Computer game in 3D, a shooter such as *Doom* or *Call of Duty***

modality 1: visual

*modes: see the modes of Example 2 under visual modality*

modality 2: auditive

*modes: see the modes of Example 2 under auditive modality*

modality 3: (virtual) proprioceptive. This is a revealing example of a modality that can be absent but still be simulated by the means of other modalities. Here, the proprioceptive dynamics of the virtual character is conveyed by the means of visual and auditive modalities, as well as, to a limited extent, the tactile modality of the mouse. Still, the role of tactile modality is too minuscule here to consider it a separate one (as there is no reason to include proprioception as a modality of film experience during which the viewer's head or eyes slightly move). An alternative case would arise from the game controls with feedback mechanisms, such as reactive pedals and steering wheels for rally games. In such cases the tactile modality emerges independently. Simulating a modality by the means of other modalities allows us to consider it also under the type of 'substituting' (see below).

## Modifying

In a modifying relation, two or more modalities can support, adjust and modulate one another, but the functional purpose of the relation is different. Here a modality not only amplifies or inhibits another one, but also changes the meaning of the textual whole to some extent. Thus, it is more precise to speak about the meaning of the combination of modalities, as in a modifying relationship modalities cannot appear alone. Here, the desired meaning or function is not achieved by a single modality (as is possible in a supporting relationship). Nevertheless, adding one modality to another changes the meaning of its counterpart not in a conflicting, but a modifying manner. Out of this type of relation, emergent informational content or a synthetic meaning is prone to emerge.

**Example 1: A hunter shooting an arrow towards a target**

modality 1: visual

*perceptual features: distance, target properties*

modality 2: tactile + proprioceptive

*perceptual features: muscle activity, coordination, feedback of bow, arrow and string*

This example resembles ‘Supporting: Example 1’ due to the simplicity of the semiosis and the lack of modes. Here I would tentatively merge the tactile and the proprioceptive modalities, as they work almost inseparably and, in a sense, are unified in the interaction with visual modality.

**Example 2: Mainstream narrative fiction film – expressing the emotions of a character or a mood of a situation**

modality 1: visual

*mode: a shot conveying a neutral content, e.g. the main character walking down the street*

modality 2: auditive

*mode: sad (intense, playful etc.) music on the soundtrack*

Here, the soundtrack helps to create an emotionally marked shot in combination with the visual content. Following Stöckl’s typology (Stöckl 2004: 13), the modes carried by the auditive modality can be more precisely indicated in the form of submodes (melody, orchestration, rhythm) and their features (measure, tempo, phrase, pattern).

## Conflicting

While the previous relation types belong to everyday life, or the discourse of normality, the conflicting type draws attention to disjunction, disruption or error. In the structure of an artwork, conflicting relations can function as a device of defamiliarization (as defined by Russian Formalists), or, in a more general sense, fulfill some other artistic purpose such as deconstruction of a coherent meaning. Conflicting modalities and modes have the ability to direct attention from the message to the code (“This is not a sentence!”), revealing structural characteristics in the architecture of a sign system and in such a manner fit into the repository of self-reflexive or metafictional texts. Therefore, conflicting modalities are used with a more specific textual strategy in mind. As supportive and modifying relations are quite straightforward, a conflicting relation often associates with partially hidden and complex communication objectives. It can be in the service of propaganda, advertisement or other practice of goal-driven rhetorics. Still, to some extent, what is considered as conflicting is the matter of individual interpretation and cultural context.

**Example 1: Distant lightning observed and thunder heard later on**

modality 1: visual

modality 2: auditive

This simple example is presented to illustrate observer-induced subjectivity in the detection of a conflict. Thunder and lightning do not conflict with each other in any natural way, so they should naturally belong to the supporting type. Still, the position of the observer can set the visual and auditive modalities out of synchronization. However, this remains a quasi-example, because in reality it is usually a case of the conflict already solved by cultural experience and observation. Similarly to the examples of berry picking and bow shooting this case addresses the lower threshold of semiosis, there is no reason to talk about semiotic modes here.

**Example 2: Film subtitles out of synchronization**

modality 1: visual

*mode: written text of subtitles in the viewer's native language*

modality 2: auditive

*mode: incomprehensible or partially comprehensible dialogue in a foreign language*

This, in a sense, is a continuation of the previous example, because the cause of a conflict still lies in the trivial issue of asynchrony. Nevertheless, here the disruption will be in effect if a sufficient threshold of synchronization is lost. Again, the level of detail of the modes belonging under this example can be negotiated.

Alternatively, simply 'language' can be seen in the mode section as a core mode. In this case, language is carried both by visual and auditive modalities, but diverging in medial variants, peripheral modes, submodes and features, respectively.

**Example 3: Satyric mimicry in animal signalling as discussed by Howse and Allen**

Biologists Howse and Allen coined the notion of 'satyric mimicry', when drawing parallels between animal behaviour, puzzling artworks and a specific advertising technique in marketing. What can be employed in all those spheres is

a commonly used technique to increase the entropy of the message by breaking the convention for the purposes of shocking or confusing the observer. This may involve putting two signs in juxtaposition, or a familiar sign in a non-conventional setting. (Howse, Allen 1994: 111)

In the case of satyric mimicry, the prey animal simultaneously displays two messages, which are conflicting in their meaning. The main objective is to lengthen the perceptual process of the predator, allowing the prey animal to escape. Here, the most discussed examples refer to the conflict between modes under visual modality.

**Example 4: The deceptive trailer as an attempt to market a film to wider audiences than the film's original genre would suggest. An opening shot of the trailer for Alejandro Gonzalez Iñárritu's film *Birdman*.**

modality 1: visual

*modes: actor's costume, actor's position in space, lighting, camera angle*

modality 2: auditive

*modes: soundtrack music, voice-over*

In this case, an ironic social drama is presented, in an allusive manner, as a Marvel-style hero action film. This is not done by a straightforward lie, but by applying a strategy that resembles satyric mimicry: creating a group of confusing signals. While the film's genre is one of the main principal rhetorical appeals that is promoted with trailers (Kernan 2004: 14), the viewer is presented with various modes that express conflicting genre cues. Elsewhere, I have analysed the shot more closely as follows:

It is [...] unclear who is saying the opening words. The plural pronoun 'we' in the voiceover has an estranging effect: it makes us doubt that the monologue belongs to the protagonist. A moderate interaction of modes may emerge from the verbal mode and the code of costume: the character's old-fashioned underwear pants and the word 'balls' can give rise to a [...] comic effect. Still, the mysterious mood conveyed by the visual modality is prevailing (leading to the connotation of the protagonist's supernatural powers) and casting its meaning over the remaining part of the trailer. (Oja 2019: 176)

The single shot is discussed here in order to demonstrate the analysis of a minimal unit. For a better understanding of how the rhetorical function works in that case, I would suggest analysing the whole trailer.

In my opinion, the conflicting relation is especially interesting in the light of the works of Juri Lotman and the Tartu-Moscow School. According to Lotman, at least two different languages are needed to produce new information. Intersections of these languages are characterized by constant inadequacies of translation (Lotman 2013: 52), but it is this obstacle that impels the interpreter to fill in the blanks creatively. Lotman's understanding of language is remarkably

broad – in *Culture and Explosion* he defines language as a code with history or cultural memory built in it (Lotman 2009[1992]: 4). This makes it possible to apply Lotman’s language-centred model to a wide range of semiotic systems, including the system of sensory modalities and the system of semiotic modes. In multimodal conflicts, combinations of incongruent modalities (or modes) have the potential to instigate a synthesis in which a whole emerges that is qualitatively different from the sum of its parts.

### Substituting

The last two types of modality relations are even less common. The substituting relation may appear as a playful device in art or other types of communication. Here, one modality is present and the other one is absent. The former refers to the latter, which we can call ‘target modality’.

#### **Example 1: A scene from Ingmar Bergman’s film *Hour of the Wolf* (1968)**

modality 1: auditive

*mode: verbal text by a character*

modality 2: visual

*mode: a drawing inside a film (only referred to, not displayed)*

Here, the pictorial mode in visual modality is substituted by the verbal mode in auditive modality. A character shows a picture to another character, while describing it verbally. The picture is never seen on camera; only a verbal description is provided to the film’s viewer. Such a device is called ‘ekphrasis’, or “verbal representation of visual representation” (Mitchell 1995: 152), and it has also been pointed out that in Bergman’s poetics there is an apparent propensity towards the use of ekphrasis<sup>19</sup> (Koskinen 2021: 57).

It should be noted that while one modality replaces another, the outcome can be still multimodal on the level of multisensory integration inside the brain, if the brain’s visual processing areas are somewhat activated by verbal cues. This leads to the cross-activating relation, which is the last one in the typology. While both are similar in function, I propose that the criterion for differentiating between them is the degree by which the brain networks of the target modality are activated.

---

<sup>19</sup> Bergman’s ekphrastic representations can be traced back to his habit of taking notes that “are especially conspicuous in a number of ekphrastic descriptions of photographs and paintings found in Bergman’s notebooks, drafts, and manuscripts, although ultimately only traces of them may be visible in his finished works” (Koskinen 2021: 62).

If a reference to the target modality (which in this case is visual) is made just by symbolic means, without starting any significant processing activity in visual cortex, the relation is substitutive.

### Cross-activating

This type resembles the previous one, but relies upon an internal difference. If in a substitutive relationship ‘target modality’ was only referred to, here the target modality is activated in the interpreter’s consciousness; the object of the message is imagined, that means – simulated and virtually present. The cross-activating type is characteristic of synaesthetic experiences.

#### **Example 1: A book activates various modalities in reader’s brain**

modality 1: visual

*modes: written text*

modality 2: auditive

*modes: dialogue, voice, even music with its various submodes such as melody, timbre etc*

modality 3: olfactory + gustatory

*modes: smell, taste*

modality 4: tactile + proprioceptive

*modes: touch, the feeling of body in the space*

Here, the first modality activates others; 2, 3 and 4 are so-called “target modalities”.

Reading as a bodily mediated process would ideally require more exhaustive elaboration than the limits of this article would allow. In short, recent discussions have highlighted that “reading can be seen as a mental act of meaning-making that is partly grounded in a multitude of sensorimotor and social experiences, [being a] [...] multisensory activity” (Hillesund *et al.* 2022: 3). Research has also shown that while reading a verbal text “comprehenders use motor simulations to generate predictions about meaning. When participants comprehend the word for an action, the motor system represents its meaning” (Barsalou 2009: 1287). Evidently, the amount of simulation or cross-activation will also depend on the cultural context and the reader’s personal associations.

**Example 2: Haptic visuality**

Modality 1: visual

*mode: an interaction with a material substance, as observed visually*

Modality 2: tactile + proprioceptive

*mode: the bodily feeling*

In the film theory of Laura Marks, haptic or tactile visuality is the visuality that functions like the sense of touch (Marks 2000: 22). An excellent characterization of haptic visuality is given by Bateman and his colleagues: “[I]magine watching a film without sound where someone licks a very rough wall or drags their fingers against a blackboard – the fact that you probably have reactions to these that are not visual at all is what haptic visuality describes” (Bateman, Wildfeuer, Hiippala 2017: 36).

This ends the typology of modality relations. It was created with the hope to demonstrate some elementary possibilities for multimodal analysis. This analysis makes a clear difference between modalities and modes; due to that, it can deal with a powerful scope of research objects, without confusing the sensory perception of objects’ material properties with socially modified forms of that materiality. The specification and the limitation of the field of analysis is of utmost importance. As we saw in the two examples of *mainstream fiction film* (both second examples under supportive and modifying relations), the difference between a supporting and a modifying relation can remain negotiable when the scale of the research object is too general. The specification can be facilitated by considering the object’s dominants and semiotic functions.

To a degree, the typology revealed the dynamics of interaction between modalities and modes. However, this is still an area in need of further exploration. Negotiating between generality and specificity is the key. In some general objects where discriminating between modalities is easy, the question of modes causes problems. For example, should we include interaction into the set of modes in computer games? If multimodal analysis concerns complex objects such as films, games, performances, we must pay attention to the montage-like activities that are the most fundamental processes in the creation of such texts. Should we call editing a mode or not? Is a manipulation technique of a semiotic resource itself a resource? This problem illustrates the uncertainty that permanently follows modes. In case of such difficulties, it can be suggested that first of all we should pay close attention to what modalities these particular modes rely upon.

## **Concluding remarks on the perspectives of multimodal analysis**

The interchangeable use of the terms ‘mode’ and ‘modality’ is widespread even in the discussions of highly professional research projects of multimodality. Complex texts or semiotic objects whose meaning is synthetic, i.e. is dependent upon several sign systems or languages that are simultaneously perceived, inevitably complicate the metalanguage that is used to model them. For a division between research units or components, there are several options. Description of the units can be based on media, sensory channels and also on socially modified materialities or semiotic resources called ‘modes’ that as units were introduced by Gunther Kress and his colleagues. Difficulties will arise when theorists attempt to integrate these types carelessly or use distinct notions as synonyms.

As I indicated above, sensory modalities can be understood not only as channels that inherit their properties from a specific medium but also as a category that is concerned with perceptive and cognitive processes. Following Roman Jakobson, who pointed out the necessity to study the role of the senses in semiosis, I suggest that the research of multimodality via sensory modalities can address organizing and interpreting principles that stem from neurocognitive properties of interpreters. This helps to integrate a pre-cultural layer with the mainly sociosemiotic theory of multimodality, calling for a dialogue with psychology, neurosciences and biosemiotics. The aim of this article was to indicate that semiotic modes and sensory modalities are complementary: they can be integrated into a two-level system that benefits both from the fixedness of modalities as well as the flexibility of modes.

Modalities are concrete and few and they are limited by the structure of the brain and the sensory system, while the range of modes is theoretically unlimited. Furthermore, modes are always haunted by a certain relativity. What should be claimed as a mode may depend on the magnification scale or the level of detail. A single mode can be seen containing submodes or even smaller sub-units (e.g. for Stöckl, core modes contain medial variants, submodes and their features). Written language entails font use or letter spacing, and a film shot entails camera angles, colour, lighting etc. – should we call these modes, submodes or something else? I suggest that such relativity can be better dealt with when we pay attention to the layer of underlying modalities. In such situations, modalities function both as properties of modes as well as their carriers or supporters. It is also important to notice that a single mode can be expressed through different modalities. For example, written language (that uses visual modality) and spoken language (that uses auditive modality) are different semiotic systems due to the modalities.

The semiotic research of multimodality has to maintain awareness of its position in the context of other disciplines. On the one hand, providing general interdisciplinary explanations is necessary, on the other hand, tightening the framework for the practical culture-analytic tasks is also needed. While employing the two-level framework, it is crucial to discern which problems are better addressed with the system of modes, which ones by the system of modalities. There are some central topics and concepts that are suitable to be discussed, first and foremost, by the modality approach. One of such is multisensory integration, related to the much-discussed binding problem in neurosciences. The question of synthesis between components is relevant on any level of text or in any art form. Composite sign systems bring along the topic of dominance: multimodal analysis benefits from the description of how the dominant mode or modality organizes the structure of the text and the meaning-making in interpretation.

Discussing multisensory integration works as a good basis for building a typology of multimodal relations. In the last part of this article, I constructed a provisional version of such a typology. My first aim was to demonstrate an unproblematic integration between modes and modalities. Also, by including examples of very simple actions that rely on composite perception, the typology addressed the problem of a lower threshold of modes. In such actions (as berry picking and shooting a bow), at least two modalities are relied upon (visual, tactile), but it is impossible to speak of modes as socially modified semiotic resources. In such cases, we can speak of multimodality in the sense of modalities, but not of modes. Evidently, the opposite is also possible: some multimodal texts employ only one modality (e.g. visual) and multiple modes (e.g. image and text).

Perhaps the most thought-provoking categories in the typology are the cross-activating and the conflicting relation types. Cross-activation reflects the cognitive principles behind perception, suggesting that even monomodal texts can facilitate integrated, multimodal images. It suggests a view that any meaning-making is, to some extent, synaesthetic. The category of conflicting modalities describes a relation where the components of the meaningful whole contradict one another. Here we can notice a similarity to Juri Lotman's model of co-existing but partially intranslatable languages that form a mechanism for creating novel information. Lotman's general approach to 'language' allows us to create analogies with another semiotic systems, including modalities and modes. I suggest a further elaboration on this parallel would benefit the modelling of semiotic synthesis, where divergent components create the situation of code plurality, an open-ended system that calls for an active interpretation.

## References

- Anderson, John R.; Bothell, Daniel; Byrne, Michael D., Douglass, Scott; Lebiere, Christian; Qin, Yulin 2004. An integrated theory of the mind. *Psychological Review* 111(4): 1036–1060. <https://doi.org/10.1037/0033-295X.111.4.1036>
- Antunes, Luis Rocha 2016. *Multisensory Film Experience: A Cognitive Model of Experiential Film Aesthetics*. Bristol, Chicago: Intellect Books. <https://doi.org/10.2307/j.ctv36xw61s>
- Barsalou, Lawrence W. 2009. Simulation, situated conceptualization, and prediction. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364: 1281–1289. <https://doi.org/10.1098/rstb.2008.0319>
- Bateman, John 2016. Methodological and theoretical issues in multimodality. In: Klug, Nina-Maria; Stöckl, Hartmut (eds.), *Handbuch Sprache im multimodalen Kontext*. Berlin, Boston: De Gruyter, 36–74. <https://doi.org/10.1515/9783110296099-003>
- Bateman, John 2018. Peircean semiotics and multimodality: Towards a new synthesis. *Multimodal Communication* 7(1): 1–24. <https://doi.org/10.1515/mc-2017-0021>
- Bateman, John 2019. Transmediality and the end of disembodied semiotics. *International Journal of Semiotics and Visual Rhetoric* 3(2): 1–23. <https://doi.org/10.4018/IJSVR.2019070101>
- Bateman, John 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, Boston: De Gruyter, 35–64. <https://doi.org/10.1515/9783110725001-002>
- Bateman, John; Schmidt, Karl-Heinrich 2011. *Multimodal Film Analysis: How Films Mean*. New York, London: Routledge.
- Bateman, John; Wildfeuer, Janina; Hiippala, Tuomo 2017. *Multimodality: Foundation, Research and Analysis*. Berlin, Boston: Walter de Gruyter. <https://doi.org/10.1515/9783110479898>
- Bateson, Gregory 2000[1972]. *Steps to an Ecology of Mind*. Chicago, London: University of Chicago Press.
- 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. <https://doi.org/10.1093/acprof:oso/9780198524861.003.0007>
- Boria, Monica; Tomalin, Marcus 2020. Introduction. In: Boria, Monica; Carreres, Angeles; Noriega-Sánchez, María; Tomalin, Marcus (eds.), *Translation and Multimodality: Beyond Words*. London, New York: Routledge, 1–23. <https://doi.org/10.4324/9780429341557-1>
- 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. <https://doi.org/10.12697/SSS.2019.47.3-4.01>
- Chatman, Seymour 1990. *Coming to Terms: The Rhetoric of Narrative in Fiction and Film*. Ithaca, London: Cornell University Press.

- Cohn, Neil 2016. A multimodal parallel architecture: A cognitive framework for multimodal interactions. *Cognition* 146: 304–323. <https://doi.org/10.1016/j.cognition.2015.10.007>
- Damasio, Antonio 2012. *Self Comes to Mind*. London: Vintage Books.
- De Gelder, Beatrice; Bertelson, Paul 2003. Multisensory integration, perception and ecological validity. *Trends in Cognitive Sciences* 7(10): 460–467. <https://doi.org/10.1016/j.tics.2003.08.014>
- Driver, Jon; Spence, Charles 2000. Multisensory perception: Beyond modularity and convergence. *Current Biology* 10: R731–R735. [https://doi.org/10.1016/S0960-9822\(00\)00740-5](https://doi.org/10.1016/S0960-9822(00)00740-5)
- Eco, Umberto 1976. Articulations of the cinematic code. In: Nichols, Bill (ed.), *Movies and Methods*. Vol. 1. Berkeley, Los Angeles: University of California Press.
- Elleström, Lars 2010. The modalities of media: A model for understanding intermedial relations. In: Elleström, Lars (ed.), *Media Borders, Multimodality and Intermediality*. London: Palgrave MacMillan. <https://doi.org/10.1057/9780230275201>
- 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. <https://doi.org/10.1007/978-3-030-49679-1>
- Forceville, Charles 2020. *Visual and Multimodal Communication*. New York: Oxford University Press. <https://doi.org/10.1093/oso/9780190845230.001.0001>
- Freeman, Walter 1998. The neurobiology of multimodal sensory integration. *Integrative Physiological and Behavioral Science* 33(2): 124–129. <https://doi.org/10.1007/BF02688656>
- Fricke, Ellen 2013. Towards a unified grammar of gesture and speech: A multimodal approach. In: Müller, Cornelia; Cienki, Alan; Fricke, Ellen; Ladewig, Silva; McNeill, David; Tessendorf, Sedinha (eds.), *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction*. Berlin, New York: De Gruyter Mouton, 733–754. <https://doi.org/10.1515/9783110261318.733>
- Gallese, Vittorio; Guerra, Michele 2015. *The Empathic Screen: Cinema and Neuroscience*. Oxford: Oxford University Press.
- Gibson, James J. 2015[1986]. *The Ecological Approach to Visual Perception*. New York: Psychology Press.
- Gobara, Akihiko; Yamada, Yuki; Miura, Kayo 2016. Crossmodal modulation of spatial localization by mimetic words. *i-Perception* November–December, 1–9. <https://doi.org/10.1177/2041669516684244>
- Halliday, Michael A. K. 1978. *Language as Social Semiotic*. London: Edward Arnold.
- Hillesund, Terje; Schilhab, Teresa; Mangen, Anne 2022. Text materialities, affordances, and the embodied turn in the study of reading. *Frontiers in Psychology* 13(March): 1–9. <https://doi.org/10.3389/fpsyg.2022.827058>
- Howse, P. E.; Allen, J. A. 1994. Satyric mimicry: The evolution of apparent imperfection. *Proceedings of the Royal Society: Biological Sciences* 257: 111–114. <https://doi.org/10.1098/rspb.1994.0102>
- Jakobson, Roman 1971[1967] Language in relation to other communication systems. In: *Selected Writings II: Word and Language*. The Hague, Paris: Mouton, 697–708. <https://doi.org/10.1515/9783110873269.697>

- Jakobson, Roman 1981[1935]. *The dominant. Selected Writings III: Poetry of Grammar and the Grammar of Poetry*. The Hague: Mouton Publishers, 751–756. <https://doi.org/10.1515/9783110802122.751>
- Jewitt, Carey; Bezemer, Jeff; O'Halloran, Kay 2016. *Introducing Multimodality*. London, New York: Routledge. <https://doi.org/10.4324/9781315638027>
- Kernan, Lisa 2004. *Coming Attractions: Reading American Movie Trailers*. Austin: University of Texas Press.
- Koskinen, Maaret 2021. Bergman, writing, and photographs: The auteur as an ekphrastic ghost. In: Hedling, Erik (ed.), *Ingmar Bergman: An Enduring Legacy*. Lund, Manchester: Lund University Press, 56–69.
- Kress, Gunther 2006. *Reading Images: The Grammar of Visual Design*. New York, London: Routledge. <https://doi.org/10.4324/9780203619728>
- Kress, Gunther 2010. *Multimodality: A Social Semiotic Approach to Contemporary Communication*. London, New York: Routledge.
- Kress, Gunther 2014. What is mode? In: Jewitt, Carey (ed.), *The Routledge Handbook of Multimodal Analysis*. London: Routledge, 60–75.
- Kress, Gunther 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.
- Kress, Gunther; Jewitt, Carey; Ogborn, Jon; Tsatsarelis, Charalampos 2000. *Multimodal Teaching and Learning*. London: Continuum.
- Kull, Kalevi; Velmezova, Ekaterina 2014. What is the main challenge for contemporary semiotics? *Sign Systems Studies* 42(4): 530–548. <https://doi.org/10.12697/SSS.2014.42.4.06>
- Lemke, Jay 2009. Multimedia genres and transmedia traversals: Social semiotics and the political economy of the sign. *Semiotica* 173(1/4): 283–297. <https://doi.org/10.1515/SEMI.2009.012>
- Liu, Yu; O'Halloran, Kay 2009. Intersemiotic texture: Analyzing cohesive devices between language and images. *Social Semiotics* 19(4): 367–388. <https://doi.org/10.1080/10350330903361059>
- Lotman, Juri 1990a. Artistic space in Gogol's prose. *Russian Literature Triquarterly* 23: 199–241.
- Lotman, Yuri M. 1990b. *Universe of the Mind*. London: I. B. Tauris & Co Ltd.
- Lotman, Juri 2009[1992]. *Culture and Explosion*. Berlin, New York: Mouton De Gruyter. <https://doi.org/10.1515/9783110218473>
- Lotman, Juri M. 2013. The search for a path. *The Unpredictable Workings of Culture*. Tallinn: TLU Press, 39–53.
- Marks, Laura 2000. *The Skin of the Film: Intercultural Cinema, Embodiment and the Senses*. Durham: Duke University Press. <https://doi.org/10.1215/9780822381372>
- McGurk, Harry; MacDonald, John 1976. Hearing lips and seeing voices. *Nature* 264(5588), 746–748. <https://doi.org/10.1038/264746a0>
- Merleau-Ponty, Maurice 1962. *Phenomenology of Perception*. London: Routledge and Kegan Paul.

- Metz, Christian 1974. *Film Language: The Semiotics of the Cinema*. New York: Oxford University Press.
- Mills, Kathy A.; Unsworth, Len 2017. Multimodal literacy. *Oxford Research Encyclopedia of Education*. <https://doi.org/10.1093/acrefore/9780190264093.013.232>
- Mitchell, Thomas W. J. 1995. *Picture Theory: Essays on Verbal and Visual Representation*. Chicago, London: University of Chicago Press.
- Nuku, Pines; Bekkering, Harold 2010. When one sees what the other hears: Crossmodal attentional modulation for gazed and non-gazed upon auditory targets. *Consciousness and Cognition* 19: 135–143. <https://doi.org/10.1016/j.concog.2009.07.012>
- O'Halloran, Kay L. 2004. On the effectiveness of mathematics. In: Ventola, Eija; Cassily, Charles; Kaltenbacher, Martin (eds.), *Perspectives on Multimodality*. Amsterdam, Philadelphia: John Benjamins Publishing Company, 91–117.
- Oja, Martin 2019. On the concept of the deceptive trailer: Trailer as paratext and multimodal model of film. *Sign Systems Studies* 47(1/2): 159–186. <https://doi.org/10.12697/SSS.2019.47.1-2.07>
- Page, Ruth 2010. *New Perspectives on Narrative and Multimodality*. New York, London: Routledge. <https://doi.org/10.4324/9780203869437>
- Peirce, Charles Sanders 1965. *Collected Papers of Charles Sanders Peirce. Vols. 1–6*. (Hartshorne, Charles; Weiss, Paul, eds.), Cambridge: The Harvard University Press. [In-text references are to CP, followed by volume and paragraph numbers.]
- Pennartz, Cyriel M. A. 2009. Identification and integration of sensory modalities: Neural basis and relation to consciousness. *Consciousness and Cognition* 18: 718–739. <https://doi.org/10.1016/j.concog.2009.03.003>
- Shams, Ladan; Kim, Robyn 2010. Crossmodal influences on visual perception. *Physics of Life Reviews* 7: 269–284. <https://doi.org/10.1016/j.plrev.2010.04.006>
- Shimamura, Arthur P. 2013. Psychocinematics: issues and directions. In: Shimamura, Arthur P. (ed.), *Psychocinematics: Exploring Cognition at the Movies*. New York: Oxford University Press, 1–26. <https://doi.org/10.1093/acprof:oso/9780199862139.003.0001>
- Sobchack, Vivian 1992. *The Address of the Eye: A Phenomenology of Film Experience*. Princeton: Princeton University Press. <https://doi.org/10.1515/9780691213279>
- Spence, Charles; Senkowski, Daniel; Röder, Brigitte 2009. Crossmodal processing [Editorial]. *Experimental Brain Research* 198(2/3): 107–111. <https://doi.org/10.1007/s00221-009-1973-4>
- Stöckl, Hartmut 2004. In between modes: Language and image in printed media. In: Ventola, Eija; Cassily, Charles; Kaltenbacher, Martin (eds.), *Perspectives on Multimodality*. Amsterdam, Philadelphia: John Benjamins Publishing Company, 9–30.
- Stöckl, Hartmut 2014. Semiotic paradigms and multimodality. In: Jewitt, Carey (ed.), *The Routledge Handbook of Multimodal Analysis*. (2nd ed.) London: Routledge, 274–286.
- Stöckl, Hartmut 2021. Pixel surgery and the doctored image. In: Pflaeging, Jana; Wildfeuer, Janina; Bateman, John (eds.), *Empirical Multimodality Research: Methods, Evaluations, Implications*. Berlin, Boston: De Gruyter, 187–210. <https://doi.org/10.1515/9783110725001-008>
- Whitney, David 2008. Neuroscience: Toward unbinding the binding problem. *Current Biology* 19(6): 251–253. <https://doi.org/10.1016/j.cub.2009.01.047>

- Wildfeuer, Janina 2014. *Film Discourse Interpretation: Towards a New Paradigm for Multimodal Film Analysis*. (Routledge Studies in Multimodality.) New York, London: Routledge.
- Zacks, Jeffrey M. 2015. *Flicker: Your Brain on Movies*. New York: Oxford University Press. <https://doi.org/10.3167/proj.2015.090102>

### **Mode sémiotique et modalité sensorielle en sémiotique multimodale: faire la différence et construire une complémentarité entre les termes**

Cette article aborde une problème dans les études multimodales: l'utilisation incohérente des notions de 'mode' et de 'modalité'. Ces notions étant souvent abordées de manière interchangeable, il sera utile de les distinguer clairement et de les positionner dans un système cohérent. En esquissant une tel système, j'envisage un cadre à deux niveaux où les modes et les modalités se soutiennent mutuellement. La branche centrale de la sémiotique multimodale (issue de la socio-sémiotique de Gunther Kress en direction de l'approche globale de John Bateman) reconnaît le 'mode' comme le concept de recherche central. Alors que le 'mode' en tant que ressource sémiotique dépend de sa matérialité, des pratiques culturellement façonnées et de la sémantique du discours, les caractéristiques neurocognitives des modalités sensorielles sont souvent considérées comme secondaires. Cette article suggère que la discussion du potentiel sémiotique des modalités sensorielles est complémentaire à la théorie sémiotique de la multimodalité. Pour illustrer cela, je vais construire une typologie expérimentale des relations de modalités, qui prend également en compte les modes. Dans cette typologie, on distinguera les relations de soutien, de modification, de conflit, de substitution et d'activation croisée.

### **Mõisted 'mode' ja 'modality' multimodaalsuse semiootilises teoorias: komplementaarsusest ja eristamise vajalikkusest**

Käesolevas artiklis vaadeldakse multimodaalsusuuringute ja kultuurisemiootika ristumiskohas esile kerkivat probleemi – mõistete 'mode' (mida võib tõlkida nii 'laad', 'viis' kui ka 'moodus'; käesolevas resümees jään 'mooduse' kui samatüvelise sõna juurde) ning 'modality' ('modaalsus') ebajärjekindlat kasutust. Neid tarvitatakse sageli sünonüümselt, kuid see toob kaasa probleeme kategooriate eristamisel. Selle vastu aitab nimetatud mõistete paigutamine sidusasse süsteemi, mida võib modelleerida kahekihiliseks. Mooduste ja modaalsuste kihid pole vastuolus, vaid teineteist täiendavad. Multimodaalse semiootika keskne haru, mis võrsub Gunther Kressi sotsioseemiootikast ja ulatub John Batemani põhjaliku käsitluseni, kasutab mõistet 'mode' baasmõistena. Kui keskendutakse mooduse kui semiootilise ressursi materiaalsusele, kultuurilistele praktikatele, mis mõjutavad materiaalsuse vormimist tähenduse loomisel ning ka diskursiivset semantikat, mis reguleerib mooduste abil loodud tähenduste mõistmist nende kontekstis, peetakse tajumodaalsuste neurokognitiivseid omadusi tähendusloomes teisejärguliseks. Käesolevas artiklis väljendatakse seisukohta, et tajumodaalsuste roll tähendusloomes on oluline ning selle uurimisel on koht multimodaalsuse semiootikas. Sellal kui 'moodus' on terminina mitmeti tõlgendatav ja pluralistlik, tasakaalustab seda '(taju)modaalsus', mis on

psühholoogias konkreetselt määratletud. 'Modaalsust' võib käsitleda ka osana raamistikust, mis integreerib humanitaariat ja täppisteadusi, sealhulgas kultuuri- ja biosemiootikat. Lisaks võimaldab 'modaalsus' käsitleda tähendusloome alustasandi probleeme, näiteks olukordi, kus erinevate tajukanalite informatsioon on vastuoluline, ristuv või vastastikuselt modifitseeriv (neid nähtusi võib termini 'ristmeedia' eeskujul liigitada 'ristmodaalsuse' alla). Et näitlikustada mooduste ja modaalsuste käsitusvõimalust ühtses raamistikus, loon artikli lõpuosas katsetusliku tüpologia, mis kirjeldab modaalsuste võimalikke suhtetüüpe ning võtab arvesse ka mooduseid. Eristatakse toetavat, modifitseerivat, konfliktset, asendavat ja vastastikuselt aktiveerivat suhtetüüpi.