

How Semiotic Modes Work Together in Multimodal Texts: Defining and Representing Intermodal Relations

(Journal Article)

Martin Siefkes

This article investigates an area of multimodality research that has received relatively little attention: relations between semiotic modes in multimodal texts. It sketches an approach towards intermodal relations with the aim of categorising them, axiomatising them, and representing them as part of a formal discourse analysis. As a prerequisite, a model for text and discourse representation is developed that is suitable for this task, building on Segmented Discourse Representation Theory (SDRT). The proposed extension of SDRT considers three strata: (1) expression, (2) discourse semantics, and (3) style.

On the basis of this extended discourse representation, a number of Intermodal Relation Types (IRTs) are defined, such as Disambiguation, Typification, and Intermodal Predication. An important innovation for intermodality research is the possibility to represent how expression properties (e.g. the adjacency of two images or differences in font sizes) interact with content properties (e.g. who is depicted or what is mentioned), and with stylistic properties (e.g. the use of a recognisable layout style).

1. Representing Multimodal Texts¹

1.1 The Rise of Multimodality

In contemporary cultures, texts combining various semiotic modes are the norm, rather than the exception (cf. Bucher 2010). Websites are comprised of written language, hyperlinks, and layout, as well as images, embedded videos, and music; schoolbooks and brochures employ language, images, graphics, layout schemata, and symbols (Unsworth 2011; Hiippala 2015); science exhibitions involve language and images, but also 3-D models, video, and interactive displays. Literature has also begun experimenting with the inclusion of images, maps, and typographic experiments (Gibbons 2014). Since the 1980s, multimedia research has begun investigating these phenomena (Posner 2003; Hess-Lüttich & Schmauks 2004; Hess-Lüttich & Wenz 2006; Jäger 2010a; Thon et al. 2014).

Since the late 1990s, the new area of multimodality research has developed a range of theoretic approaches and methods

for the investigation of cultural artefacts whose sign functions are based on a number of different semiotic modes (e.g. Kress & van Leeuwen 2001; O'Halloran 2004; Bateman 2008). New approaches and transcription systems that consider the semiotic relations between the participating modes have been developed from different theoretical perspectives (Jäger & Stanitzek 2002; Stöckl 2004b; Matthiessen 2007; Mondada 2007; Jäger 2010b; Elleström 2011; Holly & Jäger 2011; Schneider & Stöckl 2011; Bateman 2014a).

Linguists have come to realise that language is neither the sole, nor even the dominant sign system. Other sign systems such as gesture, images, graphics, typography have been in use for centuries, yet they were marginalised by philosophic reflection and scientific research due to the influence of linguocentrism, the tendency of Western cultures to privilege language and downplay other sign systems and sign types.

The fact that language occurs regularly in close integration with these other sign systems is more problematic for traditional linguistics than usually acknowledged. Whereas it is current practice in linguistics to delegate these other modes of daily communication to other disciplines, the results of

¹ Corrections and helpful suggestions for this article were made by Claus Schlaberg, Daniel Schöller, and two anonymous reviewers.

decades of research in semiotics and multimodality ultimately imply that it is no longer realistic to study language on its own. Today, complex and multimodally intertwined sign processes are the rule rather than the exception (Stöckl 2004a: v). It therefore seems naïve to hope for adequate theories of language when it is theoretically modelled and empirically investigated in isolation from all the other sign processes that humans use to multimodally interact with each other, and interpret their environment.

Semiotics has traditionally investigated texts on the basis of sign systems and their combined use (e.g. Saussure 1916; Luccio 1997; Nöth 2000). Since the various sign systems used in multimodal texts are based on different sign types, media, and perceptual channels, their analysis is much more complex than that of monomodal texts. Semiotics has developed the means for describing many different aspects of these texts (an overview is given in Posner et al. 1997–2004), including uncoded sign processes with various degrees of intentionality (Posner 1993).

In the last years, multimodality research has searched for new answers to this challenge. Currently, multimodality is investigating a range of theoretical approaches which

concentrate on different aspects, problems, and text types, such as social semiotics (Hodge & Kress 1988; Kress & van Leeuwen 1996), multimodal discourse analysis (cf. Kress & van Leeuwen 2001; O’Halloran 2004; Royce & Bowcher 2007; Bateman 2011; Jewitt 2009, 2014), mediated discourse analysis (Scollon 2001; Norris 2004; Norris & Maier 2014), multimodal film analysis (Bateman & Schmidt 2011; Piazza et al. 2011; Wildfeuer 2012, 2014), and multimodal grammar (Fricke 2012, 2013). In experimental psychology, the focus lies on the relation between various perceptual modes (Calvert et al. 2004). In conversation analysis, an approach has been developed that examines speech, gestures, facial expression, proxemics, and eye gaze behavior as part of embodied multimodal interactions between participants (Goodwin 1981; Deppermann 2010, 2014; Deppermann & Linke 2014).

Before embarking on an investigation of intermodality, one basic terminological distinction should be made. There are two different meanings of “mode” that are currently in use:

(1) Multimodal texts and artefacts combine the use of various *semiotic modes* such as language, images, gesture, typography,

graphics, icons, or sound. Used in this sense, mode corresponds closely to the more traditional semiotic notions of “code” or “sign system”. Common misunderstandings about these terms should be avoided: sign systems are not simply sets of expression-meaning-pairings. Such an understanding would be a grave misrepresentation of Saussure’s terminology, which stresses the fact that signifier and signified are substances only formed by the respective other dimensions. A sign system is rather a set of resources that often belong to a specific sign type and for which combination or application rules exist.

(2) Semiotic modes are transmitted via different *perceptual modes* (= *sensory modes*), namely visual, auditory, haptic, olfactory, and gustatory perception.

1.2 Meanings between Modes

For multimodality research, understanding how semiotic modes work together in multimodal texts is of central importance. Although a number of approaches for describing and annotating relations between modes have been developed (Schriver 1997; Royce

1998; Oviatt 1999; Marsh & White 2003; Martinec & Salway 2005; Baldry & Thibault 2006; Liu & O'Halloran 2009; O'Halloran & Smith 2011; Bateman 2011, 2014a), the mechanisms that underlie various types of relations between semiotic modes are still insufficiently explained. Current multimodal analyses often proceed either by concentrating on the separate semiotic modes or by assuming that meaning is produced by all modes together; the latter approach allows the analyst to focus on modes at will when explaining the overall meaning. Furthermore, interactions between expressions are often disregarded entirely.

However, apart from mode-specific contributions and holistically produced effects, intermodal relations play a significant role in multimodal semiosis. An intermodal relation is present when one mode has a definable influence on the expression, semantic, and/or stylistic properties of another mode in a specific text. It is therefore important to understand which types of relations can be assumed and integrated into a general model of text/discourse analysis that adequately represents all modes and textual levels present.

In multimodality research, semiotic effects limited to one semiotic mode are some-

times called “intrasemiosis”, whereas relations between modes are termed “intersemiosis” or “intermodality” (cf. Forceville & Urios-Aparisi 2009; Wildfeuer 2012, 2013). Kay O'Halloran points out the importance of examining both intrasemiosis as the specific semiotic potential of each mode (determined by the sign types and the conventions for their use), and intersemiosis, which investigates the influence of modes on other co-occurring modes in theory as well as in analysis. She emphasises the role of metaphorical expressions in these relations (O'Halloran 2005: 16). However, intersemiosis is not limited to metaphor; a whole range of effects can be created by the interplay of semiotic modes, from general relations such as *Similarity* or *Contrast*, to more complex cases where expression, meaning, or style of various modes interact. This article takes a closer look at how to define specific types of intermodal relations, and how to include them into discourse representations.

2. Modelling Intermodals Relations

2.1 Previous Research on Intermodality

This section begins with a closer look at some of the accounts that have been given for intermodal relations. Marsh & White (2003) developed a taxonomy of image-text relations on the basis of a corpus study of 45 websites. The approach is applicable in analysis, but limited in its range since it doesn't examine other modes apart from text and images, and proposes rather general relations (such as *Elicit emotion* or *Control*, *ibid.*: 600) without defining conditions for the inference of these relations.

Martinec & Salway (2005) presented a detailed approach to text-image relations that combines “logico-semantic relations” as defined in systemic functional linguistics (cf. Halliday/Matthiessen 2004) with Barthes's (1964) foundational classification of image-text relations. Martinec & Salway (2005: 371) hypothesised that their account could be applied to gesture-speech-relations, but they did not elaborate this possibility further. The proposed categories are structured into a system of different levels and are precisely described; however, they concentrate

on ideational relations in the sense of Halliday, excluding textual and interpersonal relations between modes. The same focus on content is found in most other models of intersemiosis, a limitation that is addressed in this article by including a stratum of expression which roughly corresponds to Halliday's "textual function" and a stratum of style which includes at least part of Halliday's "interpersonal function", since style is closely connected with expressing and communicating personality and preferences. In recent years, style in multimodal texts has gained increasing attention (Fricke 2006; Holly 2011; Siefkes & Arielli 2015).

Another approach (Liu & O'Halloran 2009) developed a model based on the notion of Intersemiotic Cohesive Devices (ICDs). This model is, once again, firmly rooted in systemic functional linguistics and distinguishes relations between modes regarding the ideational (comprising the logical and experiential) and the textual metafunction. The authors remark that:

This analytical framework does not confine intersemiosis to these three metafunctions at the discourse stratum only. On the contrary, semantic interaction between different semiotic choices is seen to take place across ranks and levels on the expres-

sion plane, the content plane (i.e. grammar and discourse strata) and the context plane (i.e. register and genre) (Liu & O'Halloran 2009: 371).

The distinction between different levels in the participating semiotic modes permits a more detailed analysis of their combination and interaction in multimodal texts. The discourse model proposed in this article incorporates three strata of discourse which, in a certain sense, correspond to those proposed by Liu & O'Halloran, but with the added advantage of a fully specified, formalised description and therefore enhanced descriptive precision. For example, the model shows how the organisation of material on the expression plane (where meaning is not yet considered) in one mode influences the semantic or stylistic interpretation of another mode.

An extensive systematisation of image-text relations based on a comparison of existing approaches and their respective merits has been proposed by Bateman (2014a) and includes a discussion of the potential of various empirical methods and of using systems of categories for classifying intermodal relations. In his earlier book "Multimodality and Genre" (2008), Bateman explained the

connection between intermodal relations and text genres (cf. also Hiippala 2014).

Investigations of intermodal relations between modes other than spoken and written language, images, and gesture are still relatively scarce. Some examples are O'Halloran (2004), who investigated the relations between language, symbols and images in mathematical discourse, Wildfeuer (2012, 2014) and Markussen (2014), who both concentrated on intersemiosis in film.

Fricke (2012, 2013) investigated the relationships between gestures and speech, arguing for an integration of gesture into German grammar. On the basis of examples from a corpus-based experimental study, she demonstrated that gestures can be structurally integrated into language, resulting in multimodal noun phrases (Fricke 2012: 228-43, 2013: 745-747). On the basis of these observations, she argues for a grammatical model of mode integration for speech and gesture. Although she does not assume that a grammar-based approach can be generalised for all modes, her results place the notion of multimodal grammar (cf. also Kress & van Leeuwen 2002) back onto the agenda, at least as far as specific mode combinations are concerned.

2.2 A Formalised Approach towards Intermodal Relations

This article proposes a model that formalises and represents intermodal relations, and applies it to the analysis of a range of examples (cf. Section 2.4). It uses Segmented Discourse Representation Theory (SDRT) (Asher & Lascarides 2003), which builds on Discourse Representation Theory (DRT) (Kamp & Reyle 1993). SDRT is a formalised theory that allows the representation of discourse segments and the inference of relations between them. However, in order to represent all relevant aspects of intermodal relations, the discourse representation model taken from SDRT, which is limited to semantic aspects of discourse, has to be modified.

Section 2.3 introduces a model that draws on two approaches towards multimodal text analysis: an approach that bases meaning on processes of abductive inference (Bateman 2014b), and an approach that understands multimodal texts as based on various sign systems (codes) that are integrated into a matrix code (Fricke 2012: 47, 2013: 737). The model integrates both approaches. It assumes that multimodal meaning is based on meanings in the various participating

modes, which are complemented and partly modified by Intermodal Relation Types (IRTs). IRTs are context-sensitive influences between different strata of semiotic modes. They are specific processes that take place between semiotic modes and influence the formal, semantic, and/or stylistic structure of the multimodal text in a definable way.

The approach proposed here concentrates on specific logical relations where the mode M_1 influences the form, function, and/or style of M_2 , another mode used in the same multimodal text. In line with previous research on intermodality (cf. Section 2.1), a list of IRTs will be defined, and the following specifications will be developed for each of them:

- (a) an axiom that defines when the relation can be defeasibly inferred (corresponding to the axioms for inferring discourse relations in SDRT);
- (b) restrictions for the modes that can be inserted in the argument positions (some IRTs can only be inferred between specific modes or constellations of modes);
- (c) relations to other IRTs, which can result in a full or partial system network (cf.

Kress & van Leeuwen 1996; Halliday & Matthiessen 2004).

The developed approach to intermodality combines previous research in communication theory (Siefkes 2008, 2011), stylistics (Siefkes 2011, 2012a, 2013b), discourse analysis (Siefkes 2013a, Siefkes & Schöps 2013b), and formal semantics (Siefkes in review). Furthermore, it supplements work on multimodal discourse analysis conducted at the University of Bremen, both regarding the theoretical foundations of multimodality research and the use of SDRT for multimodal discourse representation (Bateman 2008, 2011, 2014; Wildfeuer 2012, 2013, 2014).

In the approach outlined in this article, SDRT is expanded into a three-strata model of discourse: the proposed inclusion of expression properties as well as stylistic properties into SDRT demands an extension of the language of information content (the first logic of SDRT). The second important expansion concerns the introduction of IRTs into SDRT, which already possesses a glue language that contains a logic for inferences on discourse structure (the glue logic; cf. Asher & Lascarides 2003: 189-194). The glue logic currently allows the inference of discourse relations such as *Narration* or *Back-*

ground between discourse segments. For our purposes, it will be expanded to include inferences across semiotic modes.

Both modifications of SDRT are only sketched here and will demand careful work. The current state of development of the preliminary discourse representation model points towards a successful integration. However, if it turns out that the limits of SDRT as a formal discourse representation theory are overly stretched by the proposed expansions, other solutions will be considered. A possible alternative is the use of RST (Rhetorical Structure Theory; Mann & Thompson 1988).

[Television documentary; excerpt] *m, n, o, p, q, r*

[v.i^{exp}] SHOTS:SIMILAR LENGTH (*m*)

[v.i^{exp}] SHOTS:SIMILARLY STRUCTURED (*n*)

[v.i] various_scientists (*o*)

[v.i^{sty}] [*colours*] elegant (*p*)

[v.i^{sty}] [*pan*] slow (*q*)

[a.m^{sty}] [*music*] ceremonial (*r*)

$v.i \wedge a.m \vdash \exists x \exists y. \text{persons}(x) \wedge \text{similar_importance}(x) \wedge \text{prestigious_context}(y)$

Figure 1: An example of an Extended Segmented Discourse Representation Structure (E-SDRS). Elements belonging to the expression stratum are capitalised, elements belonging to the semantic stratum are in regular font, and elements belonging to the style stratum are italicised with paradigms given in square brackets.

2.3 Representing Properties of Expression, Semantics, and Style

The proposed approach includes three different strata (levels) of description that can be applied to text and discourse: (1) aspects of expression, (2) aspects of semantics, and (3) aspects of style. Whereas the semantics stratum comprises what is usually called discourse semantics, and corresponds to the current discourse conditions given in SDRT, the strata expression and style are descriptive levels introduced for the first time with this approach.

The expression stratum describes properties of the expression plane (Hjelmslev 1974) which cannot be interpreted context-independently (e.g. as coded or iconic signs). Such properties, e.g. the spatial organisation of page elements, the length of shots in a film, the colours of an image, the size of a font etc., contribute to inference processes in interaction with other modes. The expression stratum can also include syntactic properties for those modes which possess syntax (e.g. natural and formal languages). For some semiotic modes (e.g. images), syntax cannot be assumed in a strict sense, even though

properties of spatial and temporal organisation exist for these modes as well.

The style stratum is another new aspect of the proposed approach towards intermodality. Although its inclusion is unusual and maybe counterintuitive in multimodality research, it is justified on the basis of extensive recent research: in my book “Stil als Zeichenprozess” (Siefkes 2012a), the importance of style for discourse interpretation, and the close integration of stylistic with semantic information, have been demonstrated in great detail. In this book, a partly formalised model of style was developed, which describes style as a phenomenon of selection-based production and interpretation that contributes to overall textual information. This line of work (cf. also Siefkes 2009, 2011, 2013b) forms the basis for the inclusion of style as the third discourse stratum into formal discourse representation structures. Style contributes to overall meaning and is often interpreted in combination with expression and discourse semantics (Siefkes 2012a: 236-238). Stylistic features are regularities of choice from paradigms which are not determined by function or content (Siefkes 2012a: 36-52), allowing for inferences about intentions and future choices of the style user.

Figure 1 shows an example, an excerpt from a fictitious television documentary. This example and all examples in Section 2.4, with the exception of the three film examples where the source is given, are invented for the purpose of demonstrating theoretical points. SDRT is used here in an extended version that is not yet fully developed and can therefore only be sketched here; besides semantic elements (discourse referents), extended SDRT (E-SDRT) additionally represents elements denoting properties of expression printed in capital letters, and stylistic elements printed in italics (including the respective paradigm in square brackets). For each element, mode markers are given in square brackets in the format [perceptual_mode.semiotic_mode]; such as v.i for “visual images” or v.g for “visual gesture”, and with an added superscript for expression, e.g. v.i^{exp} for “expression of visual images”, or a.l^{sty} for “style of auditory (= spoken) language”. The new format is called E-SDRS (“Extended Segmented Discourse Representation Structures”).

The E-SDRS in Figure 1 includes the two form-related discourse elements (*m*) and (*n*), the representation of the similar shot length and the similar shot structure (e.g. every shot begins with a portrait photo, followed

by historical film material, etc.). It also includes the discourse referent (*o*) and the stylistic discourse elements (*p*), (*q*), and (*r*), which represent the choice of elegant colours, slow pans, and ceremonial music. All discourse elements are the input for an inference process which results in the assumption that the presented persons are important and belong to a prestigious context. Apart from the visually presented information, this inference requires the expression-related aspects which describe a kind of visual parallelism between the shots, and it necessitates the stylistic aspects which support the inference of high importance. In the inference process, the connector \vdash expresses a relation of defeasible consequence, and the expression $v.i \wedge a.m$ is a gloss for all discourse conditions given in the modes v.i and a.m in the E-SDRS. The example shows that different semiotic modes, as well as different strata of these modes, can be involved in an inference process, and have to be adequately represented.

2.4 IRTs and Their Representation in E-SDRT

The following examples demonstrate the results of preliminary work on representation and formalisation of IRTs. The introduction of IRTs, which can be inferred under certain conditions, constitutes the second major advance for the expressive power of the proposed changes to SDRT (the first consists of the description and formalisation of the three strata of discourse, as outlined in the last section).

In a preliminary model, a list of IRTs has been defined and applied to different areas of multimodal texts (e.g., to film, websites, and multimodal printed documents). For the development of the final model, the arsenal of intermodal relations proposed in previous models, as well as the testing of further examples, will be used to complement the current list. Not all technical details can be explained here; the examples are intended to give an impression of the range, and descriptive possibilities, of the preliminary model. They demonstrate how the various proposed changes and expansions of SDRT allow for the description of a variety of examples. It should be noted that the examples illustrate

a provisional step of theory development. The proposed changes that result in what has tentatively been called Extended Segmented Discourse Representation Theory (E-SDRT) will demand extensive work on the formal foundations of the theory.

The glue logic currently used for defining discourse relations such as *Narration*, *Explanation* or *Background* in SDRT can be extended to cover intermodal relations by including axioms for the inference of IRTs. Inferred discourse relations are represented in an E-SDRS, the extended discourse representation format that was described in Section 2.3.

In DRT, a discourse representation structure K consists of a pair $\langle U_K, C_K \rangle$, where U_K is a set of discourse referents and C_K is a set of conditions on these discourse referents. SDRT additionally introduces labels such as π_1, \dots, π_n for discourse segments, e.g. verbal utterances or shots of a film, and rhetorical relations such as *Narration* or *Elaboration* that can be inferred to hold between discourse segments (Irmer 2013: 31). To investigate intermodal relations, we need to distinguish between conditions in C_K according to the mode in which they are expressed.

(1) Mode-specific Discourse Conditions

If M_i is a mode, x is a discourse referent and $\omega(x)$ is a proposition where x occurs as one argument, then $C_K^{M_i} \stackrel{\text{def}}{=} \{\omega(x) \mid x \in U_K \wedge \omega(x) \in C_K \wedge \text{expressed_in_mode}(\omega(x), M_i)\}$

Intermodal relations in extended SDRT have a status comparable to that of discourse relations in SDRT: they are regarded as an aspect of rhetoric structure that contributes to text cohesion. The glue logic of SDRT can therefore be used to formulate axioms that define the conditions for the defeasible inference, denoted by the symbol $>$, of each intermodal relation type.

2.4.1 Intermodal Predication

In regard to language, it is usually assumed that we can distinguish between reference and predication. In film, the distinction is less obvious, but it can still be assumed that individuals (persons or objects) are introduced as discourse referents and characterised in various ways through acts of predication. In the first example, a shot from the film “Vanilla Sky” (2001) (cf. Figure 2), a painted snowboard is shown visually, while the protagon-

ist David utters “this one was done by me”. Without the images, it would be impossible to identify which object he refers to with the indexical expression “this”; therefore, the property ‘painted by David’ is intermodally predicated about the snowboard in the filmic narrative.

[*Vanilla Sky*, 2001, C. Crowe, 0:23:35–0:23:39] d, s, u

[v.i] David (d)
 [v.i] snowboard (s)
 [v.i] points_towards (d, s)
 [a.l] was_done_by_me (u, d)
 $v.i \wedge a.l \vdash \exists x \exists y. \text{snowboard}(x) \wedge \text{David}(y) \wedge$
 $\text{painted_by}(x, y)$
Intermodal_Predication (v.i, a.l)

Figure 2: Example of the Intermodal Predication IRT.

Let’s take a closer look at the E-SDRS represented in Figure 2. In SDRT, newly introduced individuals are added to the set U_K of discourse referents, and assertions about these discourse referents are added to the set C_K of conditions on the discourse referents. These conditions can be formulated as propositions describing the content of the respective discourse segment: ‘ d is [the previously introduced protagonist] David’, ‘ s is a snowboard’, ‘ d points towards s ’, ‘ d utters

that he did [painted] s'. In the box-style representation, the elements of U_K are written above the horizontal line, and the elements of C_K below the horizontal line. Furthermore, inference processes that make use of these discourse conditions, as well as discourse relations that are inferred in the glue logic, are represented in the same box below the elements of C_K .

How is the instance of *Intermodal Predication* represented? The expression $v.i \wedge a.l$ is a gloss for all discourse conditions given in the modes “visual images” and “auditory language”, and the logical connector \vdash expresses a relation of defeasible consequence. The inference process can be paraphrased as follows: taken together, the visual and auditory modes allow the inference that David and a snowboard occur in this discourse segment, and that David painted the snowboard.

In order to describe relations between semiotic modes, it is obviously necessary to indicate which modes occur, and which aspects of the discourse are expressed in each of these modes. In E-SDRT, the discourse conditions C_K are differentiated according to the mode in which they are expressed into $C_K^{M_1}, \dots, C_K^{M_{|M|}}$, where M is the set of all modes

present in the analysed text, and $|M|$ the number of elements of this set. A similar differentiation is not possible for the set U_K of discourse referents, since many of them occur in different modes: the snowboard in the example in Figure 2 is both visually shown and talked about. This can be gleaned from the occurrence of the variable w in discourse conditions expressed in the mode “visual images” (v.i) and in the mode “auditory language” (a.l).

An *Intermodal Predication* can be defeasibly inferred if the following conditions hold for a discourse segment: an intermodal relation of unknown type exists between some modes, represented by the underspecified predicate $?(M_i, \dots, M_k)$. Furthermore, some new information about a discourse referent $x \in U_K$ is given in the propositions $\varphi(x), \dots, \psi(x) \in U_K$, which are expressed in modes M_i, \dots, M_k , respectively. At least two of these modes must be different.

These conditions can now be used to formulate an axiom in the glue logic of E-SDRT:

(2) Intermodal Predication

$$\begin{aligned} (? (M_i, \dots, M_k) \wedge x \in U_K \wedge \varphi(x) \in C_K^{M_i} \wedge \dots \wedge \psi(x) \in C_K^{M_k} > \text{Intermodal_Predication} (M_i, \dots, M_k) \\ \text{for } \{M_i, \dots, M_k\} \subset M, |\{M_i, \dots, M_k\}| > 1 \end{aligned}$$

This axiom can be read as follows: if an underspecified intermodal relation between at least two different modes exists, x is a discourse referent, and $\varphi(x), \dots, \psi(x)$ are discourse conditions expressed in the modes M_i, \dots, M_k , the intermodal relation type *Intermodal Predication* can normally be inferred to hold between the modes M_i, \dots, M_k . Since this inference is defeasible, it can be overturned by conflicting information which may arise later in the discourse. The glue logic of E-SDRT is a non-monotonic logic that allows for changes over time, as the discourse unfolds and new information becomes available.

2.4.2 Intermodal Inference

The previous example showed how various modes can be employed to characterise a discourse referent. We can now take one further step and look at examples in which this multimodally presented information contributes to the inference of a new pro-

[Gattaca, 1997, A. Niccol, 0:03:56–0:04:23] π_1, π_2
[e_{π_1} = exterior shot, π_1 : 0:03:56–0:04:01] p
[v.i] exterior_of_modernist_building (p)
[e_{π_2} = interior shot, π_2 : 0:04:02–0:04:23] q, r, s
[v.i] interior_of_modernist_building (q)
[v.l] [text insert] not-too-distant_future (r)
[a.l] [voice on PA system] Welcome_to_Gattaca (s)
$v.i \wedge v.l \wedge a.l \vdash \exists x. \text{futuristic_building}(x) \wedge \text{location_in_narrative}(x)$
Background (e_{π_1}, e_{π_2})
Intermodal_Inference (v.i, v.l, a.l)

Figure 3: Example of the Intermodal Inference IRT.

property about a discourse referent. Figure 3 provides an example, analysing a sequence from the film “Gattaca” (1997). In this scene, the first shot (π_1) shows the exterior of a modernist building; a second shot shows the interior, while a voice on the PA system says “Welcome to Gattaca!”, and the subtitle “The not-too-distant future” is shown. Here, the images (v.i), written language (v.l) and speech (a.l) modes work together to provide the viewer with some information about a newly introduced discourse referent: there is a modernist building called “Gattaca” that exists in the future and has certain visual prop-

erties. The visual presentation both from the exterior and interior, as well as the verbally given name, will facilitate recognition of the building in later scenes, and therefore warrant the inference that it functions as a location in the narrative.

In order to allow the inference of *Intermodal Inference*, a further axiom has to be introduced in the E-SDRT glue logic. The set of discourse conditions C_K can be distinguished into the subsets $C_K^{M_1}, \dots, C_K^{M_{|M|}}$ of conditions expressed in the modes $M_1, \dots, M_{|M|}$. We assume that a relation of yet unknown type between some modes exists. This fact is represented by the underspecified predicate $?(M_i, \dots, M_k)$. Regarding one discourse referent $x \in U_K$, the propositions $\varphi_1(x), \dots, \varphi_m(x)$ are given in modes M_i, \dots, M_k , at least two of which must be different modes. An example of such a discourse condition is interior_of_modernist_building (q), which is represented in the mode v.i (visual images).

Together, the discourse conditions $\varphi_1(x), \dots, \varphi_m(x)$ enable the inference of one or more propositions $\chi_1(x), \dots, \chi_n(x)$.

(3) Intermodal Inference

$$(? (M_i, \dots, M_k) \wedge x \in U_K \wedge \varphi_1(x) \in C_K^{M_i} \wedge \dots \wedge \varphi_m(x) \in C_K^{M_k} \wedge (\varphi_1(x) \wedge \dots \wedge \varphi_m(x) > \chi_1(x) \wedge \dots \wedge \chi_n(x)) > \text{Intermodal_Inference}(M_i, \dots, M_k) \text{ for } \{M_i, \dots, M_k\} \subset M, |\{M_i, \dots, M_k\}| > 1, n \geq 1$$

This axiom can be read as follows: if an underspecified intermodal relation between at least two different modes exists, x is a discourse referent, and $\varphi_1(x), \dots, \varphi_m(x)$ are discourse conditions expressed in the modes M_i, \dots, M_k , which enable the inference of one or more propositions $\chi_1(x), \dots, \chi_n(x)$, the intermodal relation type *Intermodal Inference* can normally be inferred to hold between the modes M_i, \dots, M_k .

It should be noted that the axioms given in this article do not represent different discourse segments, assuming instead that the analysis of intermodality is limited to one SDRS. Longer excerpts can be analysed in the same way, by dividing them into the discourse segments π_1, \dots, π_n , and investigating intermodal relations in each of these segments. For a segment π_i , it would be determined which modes occur, and an SDRS K_{π_i} constructed, resulting in a sequence of SDRSs which can be characterised as

$K_{\pi_1}: \langle U_{\pi_1}, \{C_{\pi_1}^{M_i}, \dots, C_{\pi_1}^{M_k}\} \rangle, \dots, K_{\pi_n}: \langle U_{\pi_n}, \{C_{\pi_n}^{M_p}, \dots, C_{\pi_n}^{M_r}\} \rangle$. For present purposes, the simpler formulation will be used.

2.4.3 Intermodal Typification

Reference is a central phenomenon in semantics (cf. the discussion in Lyons 1977, vol. 1: 177-197), and various IRTs describe situations in which one mode influences the reference of a discourse referent represented in another mode. The example in Figure 4 represents a fictitious newspaper article with an illustration that shows a giraffe in a zoo. In isolation, the giraffe might be interpreted either as referring to a specific individual, a token or element of a category, or to the whole category, a type. In the accompanying text, the giraffe is characterised as “the tallest land animal”. If it is assumed that both the visually represented and the verbally denoted giraffe refer to the same entity, the respective discourse referents can be unified: $p = q$. This situation can then be described as follows: The proposition $\psi(x) \in C_K^{M_j}$ (with $M_j = v.i =$ visual images in this example) only enables the inference that x refers either to a token or a type, whereas the

proposition $\varphi(x) \in C_K^{M_i}$ (with $M_i = v.l =$ visual language in this example) enables the more specific inference that x , in fact, refers to a type (category) of animal. This has further consequences for interpretation: individual properties of the visually shown giraffe, such as its size or the location indicated by the background, can now safely be disregarded.

[newspaper article] p, q
[v.i] giraffe_in_zoo (p)
[v.l] giraffe (q)
[v.l] tallest_land_animal (q)
$p = q$
$v.i \wedge v.l \vdash \exists x. \text{giraffe}(x) \wedge \text{type}(x)$
<i>Intermodal_Typification</i> (v.l, v.i)

Figure 4: Example of the Intermodal Typification IRT.

This IRT is called *Intermodal Typification*, and the following axiom for its inference can be assumed (where \wedge and \vee take precedence over \rightarrow):

(4) Intermodal Typification

$$(? (M_i, M_j) \wedge \{x, y\} \in U_K \wedge \varphi(x) \in C_K^{M_j} \wedge (\varphi(x) \rightarrow \text{token}(x) \vee \text{type}(x)) \wedge \psi(y) \in C_K^{M_i} \wedge (\psi(y) \rightarrow \text{type}(y)) \wedge (x = y) > \text{Intermodal_Typification} (M_i, M_j) \quad \text{for } \{M_i, M_j\} \subset M, M_i \neq M_j$$

This axiom can be read as follows: if an underspecified intermodal relation between two different modes M_i and M_j exists, x and y are discourse referents, $\varphi(x)$ is a discourse condition expressed in the mode M_j that enables the inference that x is a token or a type, $\psi(y)$ is a discourse condition expressed in the mode M_i that enables the inference that y is a type, and x and y refer to the same entity, the intermodal relation type *Intermodal Typification* can normally be inferred to hold between the modes M_i and M_j .

This and some of the following axioms assume only two interacting modes $M_i, M_j \in M$, which is sufficient for the examples that are analysed here. They can be generalised to more modes in a straightforward way; cf. (8) for the fully general formulation of an IRT.

It should be noted that *Intermodal Typification* presupposes the unification of reference across modes for the entity in question, namely the assumption that the giraffe in the image has the same reference as the word “giraffe” in the text. The unification of reference is expressed by the use of two discourse referents x, y in the SDRS, and the requirement that they refer to the same entity specified in the axiom. Without this assumption, the giraffe in the image retains its

ambiguous status: it might refer to the class of all giraffes, or to a specific giraffe with all its individual properties.

2.4.4 Intermodal Tokenisation

Tokenisation is complementary to *Typification*. Figure 5 modifies the previous example accordingly. Again, a giraffe in a zoo is shown, which could either refer to the category or to a specific animal, but the commentary talks about a giraffe with the name “Kurt”. If we look at the image again after having read the text, we will assume that it shows Kurt, an individual giraffe living in a zoo.

[newspaper article] p, q
[v.i] giraffe_in_zoo (p) [v.l] giraffe (q) [v.l] named_Kurt (q) $p = q$
$v.i \wedge v.l \vdash \exists x. \text{giraffe}(x) \wedge \text{token}(x)$
<i>Intermodal_Tokenisation</i> (v.l, v.i)

Figure 5: Example of the Intermodal Typification IRT.

The axiom for this IRT can be derived from (4), with some modifications:

(5) Intermodal Tokenisation

$$\begin{aligned} & (? (M_i, M_j) \wedge \{x, y\} \in U_K \wedge \varphi(x) \in C_K^{M_j} \wedge (\varphi(x) \\ & \rightarrow \text{token}(x) \vee \text{type}(x)) \wedge \psi(y) \in C_K^{M_i} \wedge (\psi(y) \\ & \rightarrow \text{token}(y)) \wedge (x = y)) > \\ & \text{Intermodal_Tokenisation} (M_i, M_j) \\ & \text{for } \{M_i, M_j\} \subset M, M_i \neq M_j \end{aligned}$$

This axiom can be read as follows: if an underspecified intermodal relation between two different modes M_i and M_j exists, x and y are discourse referents, $\varphi(x)$ is a discourse condition expressed in the mode M_j that enables the inference that x is a token or a type, $\psi(y)$ is a discourse condition expressed in the mode M_i that enables the inference that y is a token, and x and y refer to the same entity, the intermodal relation type *Intermodal Tokenisation* can normally be inferred to hold between the modes M_i and M_j .

2.4.5 Intermodal Focus

In a further IRT, elements in one mode are brought into focus by another mode. This is illustrated with an example where two modes occur that have been less frequently studied: “lighting”, all aspects of film, performative arts, and architecture that have to

do with the intentional use of light, and “visual performance”, all aspects of stage productions that can be visually perceived on a stage (as opposed to music and language, which are often additional parts of stage productions). In the example in Figure 6, lighting is used to pick out one person in a stage production, directing the viewer’s gaze to focus on her.

[stage production] b, r, s, p, q
[v.lighting] brighter_lit_than (b, r, s) [v.perf] person (p), person (q)
$v.\text{lighting} \wedge v.\text{perf} \vdash \exists x \exists y. \text{person}(x) \wedge \text{person}(y) \wedge$ $\text{more_important_than}(x, y)$
<i>Intermodal_Focus</i> (v.lighting, v.perf)

Figure 6: Example of the Intermodal Focus IRT.

(6) Intermodal Focus

$$\begin{aligned} & (? (M_i, M_j) \wedge \{x, y\} \in U_K \wedge \varphi(x) \in C_K^{M_i} \wedge \psi(y) \in \\ & C_K^{M_j} \wedge \text{focuses_attention_on}(\varphi(x), \psi(y)) > \\ & \text{Intermodal_Focus} (M_i, M_j) \\ & \text{for } \{M_i, M_j\} \subset M, M_i \neq M_j \end{aligned}$$

This axiom can be read as follows: if an underspecified intermodal relation between two different modes M_i and M_j exists, x and y

are discourse referents, $\varphi(x)$ is a discourse condition expressed in the mode M_i , $\psi(y)$ is a discourse condition expressed in the mode M_j and $\varphi(x)$ focuses attention on $\psi(y)$, the intermodal relation type *Intermodal Focus* can normally be inferred to hold between the modes M_i and M_j .

Other modes such as language, music, gesture, or typography, can be used to focus attention on persons, objects, or aspects of situations presented in other modes. It should be noted that *Intermodal Focus* takes place on various levels of intentionality (cf. Posner 1993 for this important distinction in regard to sign processes). Properties of visual perception can be directly influenced by making the focused-on person more visible; inferences about the producers' intentions to direct the gaze could occur; or it might even be inferred that the producers intended that viewers recognise their respective attention, and focus their attention accordingly.

2.4.6 Intermodal Disambiguation

A further IRT consists in the disambiguation of content given in one mode by elements of another mode. In the fictitious example in Figure 7, the layout places two pictures of

persons next to each other in a print medium or website; one of these is famous, whereas the other is less well-known. This configuration could be interpreted as implying that both persons are in fact of similar importance, but other reasons for the parallel visual configuration are possible, for example that both are represented in a similar social role. The interpretation will be directed towards this latter explanation provided the verbal text mentions that both persons are candidates in an election, which implies that they are depicted in their role as candidates, is compatible with the assumption of differing importance.

```
[newspaper article] a, b, c
[v.layoutexp] NEXT_TO (a, b)
[v.l] candidates for election (c)
v.layout  $\vdash \exists x \exists y. \text{similar\_importance}(x, y) \vee \text{similar\_role}(x, y)$ 
v.layout  $\wedge$  v.l  $\vdash \exists x \exists y. \text{similar\_role}(x, y)$ 
Intermodal_Disambiguation (v.l, v.layout)
```

Figure 7: Example of the Intermodal Disambiguation IRT.

Usually, disambiguation influences the interpretation of an expression that is, at least in a specific context, not connected with one

fixed meaning. This example necessitates the representation of aspects of the expression stratum, which was introduced in the extension towards E-SDRT in section 2.3. The act of decoding or interpreting an expression is presented as an inference process $\varphi^{\text{exp}}(x) \rightarrow \chi(x)$. If an expression is polysemous, it allows for various interpretations $\chi_1(x), \dots, \chi_n(x)$. In the present case, the mode layout could be interpreted in two ways: the two images could depict persons of similar importance or persons in similar roles.

If information gleaned from another mode enables the inference of one specific interpretation χ_h out of the set $\{\chi_1(x), \dots, \chi_n(x)\}$, it will normally be assumed that this is the most plausible interpretation, and competing interpretations will be disregarded.

In the example given above, the mode v.l disambiguates possible interpretations of the mode v.layout. The axiom below is formulated accordingly; this is not the most general case, since more than one disambiguating and more than one disambiguated mode could be involved.

(7) **Intermodal Disambiguation**

$$\begin{aligned}
 & (? (M_i, M_j) \wedge \{x, y\} \in U_K \wedge \varphi_1^{\text{exp}}(x) \in C_K^{\text{exp}(M_j)} \wedge \\
 & (\varphi_1^{\text{exp}}(x) \rightarrow \chi_1(x) \vee \dots \vee \chi_n(x)) \wedge \psi(y) \in C_K^{M_i} \wedge \\
 & (\psi(y) \rightarrow \chi_h(x)) > \text{Intermodal_Disambiguation} \\
 & (M_i, M_j) \\
 & \text{for } \{M_i, M_j\} \subset M, M_i \neq M_j, \chi_h \in \{\chi_1, \dots, \chi_n\}, n > 1
 \end{aligned}$$

This axiom can be read as follows: if an underspecified intermodal relation between two different modes M_i and M_j exists, x and y are discourse referents, $\varphi_1^{\text{exp}}(x)$ is a discourse condition belonging to the expression stratum of mode M_j that enables the inference of two or more propositions $\chi_1(x)$, ..., $\chi_n(x)$, and $\psi(y)$ is a discourse condition expressed in the mode M_i that enables the inference of one proposition $\chi_h(x)$ belonging to the set $\chi_1(x)$, ..., $\chi_n(x)$, the intermodal relation type *Intermodal Disambiguation* can normally be inferred to hold between the modes M_i and M_j .

2.4.7 *Intermodal Disambiguation*

This example demonstrates the inclusion of style, the third discourse stratum, into the analysis. Figure 8 analyses a film scene from “Gattaca” in which two protagonists have a

swimming competition in the open sea at night. Anton, one of the competitors, yells “Where’s the shore? We’re too far out!” Shortly afterwards, Anton, who is exhausted, is shown as he sinks below the surface and nearly drowns. The visual style of the scene employs shaky hand camera movements, presents the scene from angles both above and below the water, and changes between close-ups and distance views. The moon’s reflection on the waves provides erratic lighting.

[Gattaca, 1997, A. Niccol, 0:89:13–0:91:50]	
c, u, a, p, q, r, s	
[v.i]	swimming_competition_in_the_ocean (c)
[a.i]	too_far_out (u, a)
[v.i ^{sty}]	[camera movement] shaky, hand camera (p)
[v.i ^{sty}]	[perspective] quick changes (above, below, level) (q)
[v.i ^{sty}]	[field size] quick changes (distance, close-up) (r)
[v.lighting ^{sty}]	[distribution] erratic (reflections on water) (s)
v.i \wedge a.i \wedge v.lighting \vdash \exists x. dangerous_situation (x)	
<i>Intermodal_Emphasis^{sty}</i> ({v.i ^{sty} , v.lighting ^{sty} }, a.i)	

Figure 8: Example of the Intermodal Emphasis IRT.

In this example, the character’s speech already characterises the situation as danger-

ous. The visual style and the lighting style support and strengthen this interpretation. The *Intermodal Emphasis* IRT (cf. Marsh & White 2003: 653) can be inferred to hold between both the visual style $v.i^{\text{sty}}$ and the style of lighting $v.\text{lighting}^{\text{sty}}$ as emphasising modes, and auditory language a.i as the emphasised mode.

In this example, there are two emphasising modes and one emphasised mode. This case can be generalised to include an arbitrary number of modes:

(8) **Intermodal Emphasis^{sty}**

$$\begin{aligned}
 & (? (\{M_i, \dots, M_k\}, \{M_p, \dots, M_r\}) \wedge \{x_1, \dots, x_m, y_1, \dots, \\
 & y_n\} \in U_K \wedge \varphi_1^{\text{sty}}(x_1) \in C_K^{\text{sty}(M_i)} \wedge \dots \wedge \varphi_m^{\text{sty}}(x_m) \\
 & \in C_K^{\text{sty}(M_k)} \wedge \psi_1(y_1) \in C_K^{M_p} \wedge \dots \wedge \psi_n(y_n) \in C_K^{M_r} \\
 & \wedge \text{emphasise} (\{\varphi_1^{\text{sty}}(x_1), \dots, \varphi_m^{\text{sty}}(x_m)\}, \\
 & \{\psi_1(y_1), \dots, \psi_n(y_n)\}) > \text{Intermodal_Emphasis}^{\text{sty}} \\
 & (\{M_i, \dots, M_k\}, \{M_p, \dots, M_r\}) \\
 & \text{for } \{M_i, \dots, M_k, M_p, \dots, M_r\} \subset M, \{M_i, \dots, M_k\} \cap \\
 & \{M_p, \dots, M_r\} = \emptyset, m \geq 1, n \geq 1
 \end{aligned}$$

This axiom can be read as follows: if an underspecified intermodal relation between the modes M_i, \dots, M_k and different modes M_p, \dots, M_r exists, $x_1, \dots, x_m, y_1, \dots, y_n$ are discourse referents, $\varphi_1^{\text{sty}}(x_1), \dots, \varphi_m^{\text{sty}}(x_m)$ are discourse

conditions belonging to the style stratum of the modes M_i, \dots, M_k , which emphasise the discourse conditions $\psi_1(y_1), \dots, \psi_n(y_n)$ expressed in the modes M_p, \dots, M_r , the intermodal relation type *Intermodal Emphasis by style* can normally be inferred to hold between the modes M_i, \dots, M_k and the modes M_p, \dots, M_r .

2.5 Further Development Steps

The last sections outlined a model that allows the systematic description of intermodal relation types, as well as the representation of their occurrences (intermodal relation tokens) in specific discourse examples. In future research, the model that this article has sketched will be developed in more detail, in regard to its theoretical foundations and practical applications.

The first step is the extension of SDRT to include expression and stylistic discourse elements. This extension is an important step with major consequences for the reach and applicability of the proposed theory of intermodal relations. Currently, formal theories of discourse representation (SDRT as well as DRT) limit themselves to the representation of semantic and narrative aspects

of discourse. For a definition of intermodal relations, it is necessary to consider purely expression-related aspects in the sense of Hjelmslev (1936, 1974), as well as stylistic aspects of texts.

The second step is the compilation of a list of IRTs, and the formulation of formal axioms that define the conditions for the inference of each IRT. The list will focus on interactions where an influence of one mode on one or various other modes can be determined. This consequently results in a specifiable interaction between the formal, semantic, and/or stylistic properties of the respective modes. The theory will therefore exclude relations that are very general (for example similarity or difference) or cannot be precisely defined. All IRTs have to be sufficiently clear to enable strong intercoder agreement, which facilitates their application in a multimodal corpus analysis. The theoretical framework is applicable to all multimodal text types.

The third step is the application of the developed list of IRTs to film. Film can be defined as a specific type of intermodal text with high complexity, including spoken and written language (which can be used both extra- and intradiegetically), moving images, kinesics (gesture, body posture, facial ex-

pression, proxemics), sound, and music. Semiotic approaches to film research have dealt with this high complexity in different ways: sometimes they have limited themselves to specific questions (e.g. image composition or lighting), in other cases they have provided rather general semiotic accounts (e.g. Metz 1974; Branigan 1984; Bordwell 2004). However, precise accounts of semiotic relations between modes have only recently been developed (e.g. Bateman & Schmidt 2011; Wildfeuer 2012; Schöps 2013; Fricke 2014b). The proposed approach will enable an important advance of film semiotics, since it allows for a much more precise understanding of the way that modes influence each other in film.

From a specifically linguistic perspective, the relation of language (characters' speech, narrators' commentary, sub- and intertitles, and intradiegetic written language such as posters or letters), with other semiotic modes of film (moving images, gesture, facial expression, sound, and music) is important. For time-intensive research such as a large corpus analysis or an experimental study, the limitation to one specific type of multimodal text can help to ensure a thorough investigation. The selection of film as a challenging area of multimodality research,

in which intermodal relations are highly complex and comparatively little understood, will ensure that the theory is thoroughly tested and connects the present research with recent discussions of multimodal approaches, which have brought together film and gesture researchers (e.g. Kappelhoff & Müller 2011; Fricke 2014b).

3. Conclusion

This article introduced an approach towards intermodality that defines various types of intermodal relations (IRTs), integrates them in a formal discourse representation, and takes various textual strata into account. Currently, only a limited number of IRTs have been identified and formalised, seven of which were introduced in Section 2.4. The model is still in development. It aims at a formalised description of inference processes and context-dependent textual meaning production.

The proposed model for text and discourse representation considers three strata (or levels): (1) the stratum of expression or form; (2) the stratum of semantics (including discourse relations); (3) the stratum of style. Stratum (2) corresponds to the current se-

mantics of SDRT. The developed theory proposes the introduction of two additional types of elements, which necessitates a careful and thorough expansion of SDRT that is not yet completed. For this reason, it is only currently possible to give an outline of what examples in Extended Discourse Representation Theory (E-SDRT) may look like.

The approach defines a number of Intermodal Relation Types (IRTs), such as *Disambiguation*, *Typification*, and *Intermodal Prediction*, gives axioms for their inference, and shows how they can be represented. The theory will facilitate the recognition of intermodal relations, their categorisation on the basis of an axiomatised list of IRTs, their integration into multimodal text analysis, and their representation as part of formalised discourse representation structures.

As a further innovation, the new representation format of Extended Segmented Discourse Representation Structures (E-SDRSs) makes it possible to represent how expression properties of modes (e.g. the adjacency of two pictures, similar or different font sizes of headlines) interact with content properties (e.g. who is depicted on the images or what is mentioned in the headlines), and stylistic properties (e.g. the use of traditional or contemporary layout and/or verbal

expressions) of other modes. Examples such as the scene from *Gattaca* analysed in Section 2.4.7, where visual stylistic properties emphasise the semantic content of a verbal utterance, demonstrate the importance of differentiating between strata in the description of intermodal relations.

References

- Asher, N. & Lascares, A. (2003). *Logics of Conversation*. Cambridge: CUP.
- Baldry, A. & Thibault, P. (2006). *Multimodal Transcription and Text Analysis*. London: Equinox.
- Bateman, J. (2007). Towards a Grande Paradigm of Film: Christian Metz reloaded. *Semiotica*, 167.1-4, pp. 13-64.
- Bateman, J. (2008). *Multimodality and Genre: A Foundation for the Systematic Analysis of Multimodal Documents*. London: Palgrave Macmillan.
- Bateman, J. (2011). The Decomposability of Semiotic Modes. In K. O'Halloran & B. Smith (Eds.), *Multimodal Studies: Exploring Issues and Domains* (pp. 17-38). London/New York: Routledge.
- Bateman, J. (2014a). *Text and Image: A Critical Introduction to the Visual-Verbal Divide*. London/New York: Routledge.
- Bateman, J. (2014b). Dynamische Diskurssemantik als ein allgemeines Modell für Semiose: Überlegungen am Beispiel des Films. In M. Siefkes & D. Schöps (2013a).

- Bateman, J. & Schmidt, K.-H. (2011). *Multimodal Film Analysis: How Films Mean*. London/New York: Routledge.
- Barthes, R. (1964). Rhétorique de l'image. *Communications*, 4, pp. 40-51.
- Bordwell, D. (2004). *Film Art: An Introduction*. 7th edition. Boston: McGraw-Hill.
- Branigan, E. R. (1984). *Point of View in the Cinema*. Berlin: Mouton.
- Bucher, H.-J. (2010). Multimodalität – eine Universalie des Medienwandels. In H.J. Bucher et al. (Eds.), *Neue Medien – neue Formate* (pp. 41-79). Frankfurt/New York: Campus.
- Calvert, G. et al. (2004). *The Handbook of Multisensory Processes*. Cambridge MA: MIT.
- Deppermann, A. (Ed.) (2010). *Conversation Analytic Studies of Multimodal Interaction*. Special issue, *Journal of Pragmatics*, 46.1.
- Deppermann, A. (2014). Multimodal Participation in Simultaneous Joint Projects: Interpersonal and Intrapersonal Coordination in Paramedic Emergency Drill. In P. Haddington et al. (Eds.), *Multiactivity in Social Interaction: Beyond Multitasking* (pp. 247-282). Amsterdam: Benjamins.
- Deppermann, A. & Linke, A. (Eds.) (2010). *Sprache intermedial: Stimme und Schrift – Bild und Ton*. Berlin/New York: de Gruyter.
- de Saussure, F. (1916), *Cours de linguistique générale*. Lausanne: Payot.
- Elleström, L. (Ed.) (2011). *Media Borders, Multimodality and Intermediality*. London: Palgrave Macmillan.
- Forceville, C. & Urios-Aparisi, E. (2009). *Multimodal Metaphor*. Berlin/New York: de Gruyter.
- Fricke, E. (2006). Intermedialität, Stil und Mental Spaces: Das Visuelle als Dimension musikalischen Komponierens in Georg Nussbauers Installationsoper 'orpeusarchipel'. *Kodikas/Code*, 29.1-3, pp. 137-155.
- Fricke, E. (2012). *Grammatik multimodal: Wie Wörter und Gesten zusammenwirken*. Berlin/New York: de Gruyter.
- Fricke, E. (2013). Towards a Unified Grammar of Gesture and Speech: A Multimodal Approach. In C. Müller et al. (Eds.), *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction* (vol. 1, pp. 733-754). Berlin: de Gruyter.
- Fricke, E. (2014a). Deixis, Gesture, and Embodiment from a Linguistic Point of View. In C. Müller et al. (Eds.), *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction* (vol 2, pp. 1802-1823). Berlin: de Gruyter.
- Fricke, E. (2014b). Die (ab)geschnittene Hand in der Talkshow: Zur Fortschreibung antiker rhetorischer Traditionen in Bildwahl und Schnitt. In: H. Girth & S. Michel (Eds.), *Multimodale Kommunikation in Polit-Talkshows* (pp. 143-166). Stuttgart: ibidem.
- Gibbons, A. (2014). Multimodality in Literature. In S. Norris & C. D. Maier (Eds.), *Interactions, Images and Texts: A Reader in Multimodality* (pp. 373-379). Berlin/New York: de Gruyter.
- Goodwin, C. (1981), *Conversational Organization*. New York: Academic.
- Halliday, M. A. K. & Matthiessen, C. M. (2004). *An Introduction to Functional Grammar*. London: Arnold.
- Hess-Lüttich, E. W. B. & Schmauks, D. (2004). Multimediale Kommunikation. In R. Posner et al. (Eds.), *Semiotik / Semiotics. Ein Handbuch zu den zeichentheoretischen Grundlagen von Natur und Kultur* (vol. 4, pp. 3487-3503). Berlin et al.: de Gruyter.
- Hess-Lüttich, E. W. B. & Wenz, K. (Eds.) (2006). *Stile des Intermedialen*. Tübingen: Narr.
- Hiippala, T. (2014). Multimodal Genre Analysis. In S. Norris & C. D. Maier (Eds.), *Interactions, Images and Texts: A Reader in Multimodality* (pp. 111-124). Berlin/New York: de Gruyter.
- Hiippala, T. (2015). *The Structure of Multimodal Documents: An Empirical Approach*. London/New York: Routledge.
- Hjelmslev, L. (1936). *An Outline of Glossematics*. Copenhagen: Levin og Munksgaard.
- Hjelmslev, L. (1974), *Prolegomena zu einer Sprachtheorie*. München: Hueber.
- Hodge, R. & Kress, G. (1988). *Social Semiotics*. Cambridge: Polity.
- Holly, W. (2011). Social Styles in Multimodal Texts: Discourse Analysis as Culture Study. In C. Baraldi et al. (Eds.), *Hybrids, Differences, Visions* (pp. 191-206). Aurora: Davies Group.
- Holly, W. & Jäger, L. (2011). Transkriptionstheoretische Medienanalyse: Vom Anders-lesbar-Machen durch intermediale Bezugnahmepraktiken. In J. G. Schneider & H. Stöckl (Eds.), *Medientheorien und Multimodalität: Ein TV-Werbespot – Sieben methodische Beschreibungsansätze* (pp. 151-168). Köln: Halem.
- Irmer, M. (2013). Inferring Implicatures and Discourse Relations from Frame Information. *Lingua*, 132, pp. 29-50.
- Jäger, L. (2010a). *Transcriptivity Matters: On the Logic of Intra- and Intermedial References in Aesthetic Discourse*. In L. Jäger et al. (Eds.), *Media, Culture, and Mediality* (pp. 49-76). Bielefeld: transcript.

- Jäger, L. (2010b). *Intermedialität – Intramedialität – Transkriptivität. Überlegungen zu einigen Prinzipien der kulturellen Semiosis*. In A. Deppermann & A. Linke (Eds.), *Sprache intermedial: Stimme und Schrift – Bild und Ton* (pp. 301-324). Berlin/New York: de Gruyter.
- Jäger, L. & Stanitzek, G. (Eds.) (2002). *Transkribieren. Medien/Lektüre*. München: Fink.
- Jewitt, C. (2009). *The Routledge Handbook of Multimodal Analysis*. London/New York: Routledge.
- Jewitt, C. (2014). Multimodal Approaches. In S. Norris & C. D. Maier (Eds.), *Interactions, Images and Texts: A Reader in Multimodality* (pp. 127-136). Berlin/New York: de Gruyter.
- Kamp, H. & Reyle, U. (1993). *From Discourse to Logics: Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Dordrecht: Kluwer Academic.
- Kappelhoff, H. & Müller, C. (2011). Embodied Meaning Construction: Multimodal Metaphor and Expressive Movement in Speech, Gesture, and Feature Film. *Metaphor and the Social World*, 1.2, 121-153.
- Kress, G. & van Leeuwen, T. (1996). *Reading Images: The Grammar of Visual Design*. London/New York: Routledge.
- Kress, G. & van Leeuwen, T. (2001). *Multimodal Discourse*. London: Arnold.
- Kress, G. & van Leeuwen, T. (2002). Colour as a Semiotic Mode: For a Grammar of Colour. *Visual Communication*, 1.3, pp. 343-368.
- Liu, Y. & O'Halloran, K. (2009). Intersemiotic Texture: Analyzing Cohesive Devices between Language and Images. *Social Semiotics*, 19.4, 367-388.
- Luccio, R. (1997). Body Behavior as Multichannel Semiosis. In R. Posner et al. (Eds.), *Semiotik / Semiotics. Ein Handbuch zu den zeichentheoretischen Grundlagen von Natur und Kultur* (vol. 1, pp. 345-356). Berlin et al.: de Gruyter.
- Lyons, J. (1977). *Semantics*. Cambridge: CUP.
- Mann, W. C. & Thompson, S. A. (1988). Rhetorical Structure Theory: Toward a Functional Theory of Text Organization. *Text*, 8, pp. 243-282.
- Markussen, B. (2014). *Lydbilder: Om samspillet mellom musikk, tekst og visualitet*. Kristiansand: Portal Forlag.
- Marsh, E. E. & White, M. D. (2003). A Taxonomy of Relationships between Images and Text. *Journal of Documentation*, 59.6, pp. 647-672.
- Martinec, R. & Salway, A. (2005). A System for Image-Text Relations in New (and Old) Media. *Visual Communication*, 4.3, pp. 339-374.
- Matthiessen, C. (2007). The Multimodal Page: A Systemic Functional Exploration. In T. D. Royce & W. L. Bowcher (Eds.), *New Directions in the Analysis of Multimodal Discourse* (pp. 1-62). London: Erlbaum.
- Metz, C. (1974). *Film Language: A Semiotics of the Cinema*. Chicago: University of Chicago Press.
- Mondada, L. (2007). Multimodal Resources for Turn-Taking: Pointing and the Emergence of Possible Next Speakers. *Discourse Studies*, 9.2, pp. 195-226.
- Norris, S. (2004). *Analyzing Multimodal Interaction: A Methodological Framework*. London/New York: Routledge.
- Nöth, W. (2000). *Handbuch der Semiotik*. 2nd ed. Stuttgart: Metzler.
- O'Halloran, K. (Ed.) (2004). *Multimodal Discourse Analysis: Systemic Functional Perspectives*. London: Continuum.
- O'Halloran, K. (2005). *Mathematical Discourse: Language, Symbolism, and Visual Images*. London: Continuum.
- Oviatt, S. L. (1999). Ten Myths of Multimodal Interaction. *Communications of the ACM* 42.11, pp. 74-81.
- Piazza, R. et al. (Eds.) (2011). *Telecinematic Discourse*. London/New York: Routledge.
- Posner, R. (1993). Believing, Causing, Intending: The Basis for a Hierarchy of Sign Concepts in the Reconstruction of Communication. In R. J. Jorna et al. (Eds.), *Signs, Search and Communication: Semiotic Aspects of Artificial Intelligence* (pp. 215-270). Berlin et al.: de Gruyter.
- Posner, R. (2003). Sign Economy in Language and Related Sign Systems. *Linguistic Agency*, University of Essen LAUD. *Series A: General and Theoretical*, 574, pp. 1-25.
- Royce, T. (1998). Synergy on the Page: Exploring Intersemiotic Complementarity in Page-Based Multimodal Text. *JASFL Occasional Papers*, 1.1, pp. 25-50.
- Royce, T. D. & Bowcher, W. L. (2007). *New Directions in the Analysis of Multimodal Discourse*. London: Erlbaum.
- Schöps, D. (2013). Korpusgestützte filmische Diskursanalyse am Beispiel des DEFA-Films. In M. Siefkes & D. Schöps 2013a: pp. 321-352.
- Schriver, K. A. (1997). *Dynamics in Document Design: Creating Texts for Readers*. New York: Wiley.
- Scollon, R. (2001). *Mediated Discourse: The Nexus of Practice*. London/New York: Routledge.
- Siefkes, M. (2008). Unkooperativität und andere kommunikative Randphänomene in *Die Zeit und das Zimmer* von Botho Strauß. *Kodikas/Code*, 31.1-2, pp. 91-118.

- Siefkes, M. (2009). Zeichenmaterie und Zeichenträger bei stilistischen Zeichen. *Kodikas/Code* 32.1-2, pp. 63-83.
- Siefkes, M. (2011). Style: A New Semiotic View on an Old Problem. *Kodikas/Code* 34.1-2, pp. 15-25.
- Siefkes, M. (2012a). *Stil als Zeichenprozess: Wie Variation bei Verhalten, Artefakten und Texten Information erzeugt*. Würzburg: Königshausen & Neumann.
- Siefkes, M. (2012b). The Semantics of Artefacts: How We Give Meaning to the Things We Produce and Use. *Themenheft zu Image 16: Semiotik*. Retrieved from <http://www.gib.uni-tuebingen.de/image/ausgaben?function=fnArticle&showArticle=218>.
- Siefkes, M. (2012c). Grenzfälle der Kommunikation. In E. W. B. Hess-Lüttich (Ed.), *Sign Culture – Zeichen Kultur* (pp. 502-525). Würzburg: Königshausen & Neumann.
- Siefkes, M. (2013a). Wie wir den Zusammenhang von Texten, Denken und Gesellschaft verstehen: Ein semiotisches 4-Ebenen-Modell der Diskursanalyse. In: M. Siefkes & D. Schöps 2013a: pp. 353-391.
- Siefkes, M. (2013b). Stil und Gesellschaft: Plädoyer für eine Allgemeine Stilistik. Research report, IUAV Venice. Retrieved from: http://siefkes.de/dokumente/Siefkes_Stil_und_Gesellschaft.pdf
- Siefkes, M. (in review). Frames in Discourse: Connecting Frame Semantics and Discourse Analysis in an SDRT-Based Model.
- Siefkes, M. & Schöps, D. (Eds.) (2013a), *Neue Methoden der Diskursanalyse*. Special issue, *Zeitschrift für Semiotik* 35.3-4.
- Siefkes, M. & Schöps, D. (2013b). Von der Diskurslinguistik zur Diskurssemiotik. *Zeitschrift für Semiotik* 35.3-4, pp. 235-247.
- Siefkes, M. & Arielli, E. (2015). An Experimental Approach to Multimodality: How Musical and Architectural Styles Interact in Aesthetic Perception. In J. Wildfeuer (Ed.), *Building Bridges for Multimodal Research: International Perspectives on Theories and Practices of Multimodal Analysis*. Bern/New York: Peter Lang.
- Stöckl, H. (2004a). *Die Sprache im Bild – Das Bild in der Sprache: Zur Verknüpfung von Sprache und Bild in massenmedialen Texten. Konzepte – Theorien Analysemethoden*. Berlin/New York: de Gruyter.
- Stöckl, H. (2004b). In between Modes: Language and Image in Printed Media. In E. Ventola et al. (Eds.), *Perspectives on Multimodality* (pp. 9-30). Amsterdam: Benjamins.
- Thon, J.-N. et al. (Eds.) (2014). *Media Convergence and Transmedial Worlds. Special Issue, IMAGE 20* (July 2014).
- Unsworth, L. (Ed.) (2011). *Multimodal Semiotics: Functional Analysis in Contexts of Education*. London: Continuum.
- Wildfeuer, J. (2012). Intersemiosis in Film: Towards a New Organisation of Semiotic Resources in Multimodal Filmic Text. *Multimodal Communication*, 1.3, pp. 276-304.
- Wildfeuer, J. (2013). Trompeten, Fanfaren und orangefarbene Tage: Zur Intersemiose in 'Die fabelhafte Welt der Amélie'. In L. C. Grabbe et al. (Eds.), *Multimodale Bilder: Zur synkretistischen Struktur des Filmischen* (pp. 81-101). Darmstadt: Böhner.
- Wildfeuer, J. (2014). *Film Discourse Interpretation: Towards a New Paradigm for Multimodal Film Analysis*. London/New York: Routledge.

Filmography

- Gattaca* (1997, dir. Andrew Niccol)
Vanilla Sky (2001, dir. Cameron Crowe)