

Multimodal meaning as a blend? Critical discussion of integrating cognitive and social semiotic theories

SØREN VIGILD POULSEN

Centre for Multimodal Communication, Department of Language and Communication, University of Southern Denmark

ABSTRACT

Researchers, both inside and outside social semiotics, argue that the introduction of a cognitive perspective in this framework would improve an understanding of multimodal meaning (i.e., meaning created with multiple modalities). One theory that has been singled out as particularly suitable for this purpose is the cognitive theory of mental spaces and blending (hereafter MSBT) by Gilles Fauconnier and Mark Turner (2002a).

This paper aims to explore the potential of combining MSBT's concepts with a social semiotic framework, the social semiotic theory of multimodality by Gunther Kress and Theo van Leeuwen (2001). Special attention is given to how this may qualify the description of multimodal meaning. At the same time, the paper presents the argument that merging semiotic and cognitive perspectives gives rise to new theoretical and methodological issues that have yet to be properly addressed in the literature. Three issues are addressed and critically discussed: (1) MSBT's notion of modality and multimodality; (2) its apparent indifference to different modal input of meaning construction; (3) general inconsistencies in semiotic and cognitive descriptions of what may prompt cognitive processes of meaning construction. This way, the paper seeks to contribute nuances to the ongoing discussion of how cognitive and semiotic perspectives may enrich our understanding of meaning made multimodally. Furthermore, it reflects work done at the University of Southern Denmark's Centre for Multimodal Communication.

KEYWORDS

multimodality, blending, mental space, emergence, social semiotics

1. Introduction

This article aims to discuss the extent to which incorporating cognitive concepts in a social keywords framework may qualify our understanding of multimodal meaning, for example when meaning is created through many simultaneous modalities or semiotic systems like writing, images, colors, typography, etc.

vigild@sdu.dk

In the research literature, several authors argue in favor of introducing cognitive theories in semiotic studies of multimodality, especially in social semiotic research (e.g., Forceville 1999; Bateman 2014), and combining different complementary perspectives (e.g., Feng & O'Holloran 2013; Caponetto 2016). Some authors (e.g., Jewitt 2009; Forceville & Urios-Aparisi 2009; Gibbons 2011; Wong 2018) have claimed that combining social semiotic and cognitive frameworks may solve identified theoretical and analytical difficulties when studying multimodal data. In addition, such an approach could help explain key aspects of the process of multimodal meaning-making and provide a better understanding of multimodality as phenomenon, as well as contributing to appropriate methodologies (Pinar Sanz 2015).

One of the specific cognitive theories that have been discussed in this context is Fauconnier and Turner's (2002a) work on mental spaces and conceptual blending, a theory of how people make meaning by building and processing ('blending') conceptual packets of information ('mental spaces').

Alonso et al. (2013) suggest integrating social semiotics and Fauconnier's (1994, 1997) mental space theory as a framework for analyzing multimodal narratives. In my own work (Poulsen 2014), I have proposed that multiple modalities in a text may function as "prompts" or cognitive triggers for multimodal meaning construction. Wong demonstrates how Fauconnier and Turner's cognitive theory may complement the visual social semiotics of Kress and van Leeuwen (2006), describing it as "a perfect fit" (Wong 2018: 1). What these works have in common, is that they tend to focus on analytical points, while, to a lesser degree, discussing how the analyses can inform more advanced theoretical concepts of multimodality and multimodal meaning. Furthermore, they do not include reflections on deeper connections and differences between social semiotics and cognitive perspectives on these matters.

In this conceptual article, I want to explore how MSBT may specifically contribute to a more elaborated theoretical understanding of multimodal meaning within a social semiotic framework; yet this ambition gives rise to new unforeseen conceptual issues that have not been made explicit nor adequately addressed in the existing literature. Thus, on the one hand, this paper presents the claim that MSBT's unique contribution to multimodality studies may be to account for emergent meaning—i.e., the creation of a new (semantic) whole that is greater than its parts. On this view, Fauconnier and Turner's concept of *vital relations*, semantic structures between mental spaces appearing through a conceptualization process, might aptly describe such emergent meaning created through integration of multiple modalities. On the other hand, the paper points to three critical issues: (1) MSBT's concept of modality is underdeveloped when compared to a social semiotic notion of this key term, which has theoretical consequences for a conceptual framework for multimodality and meaning made multimodally. (2) Since MSBT is a cognitive theory, it appears unable to specify how meaning construction in one modality differs from meaning made with multiple modalities. As a general theory of conceptualization, I claim that MSBT does not offer clarification of what sets monomodal and multimodal meaning apart. It follows that this cognitive approach is inadequate as an articulated theory of multimodal meaning, beyond a simple understanding of the concept. (3) I point out the differences between MSBT's and social semiotic descriptions of the resources in a text that can function as prompts for multimodal meaning construction. Thus, in contrast to the claims made by existing research, a closer look at the two frameworks reveals that they are incompatible.

In exploring the cognitive concept of multimodal meaning, I revisit Fauconnier's early work on mental spaces (1994, 1997), Turner and Fauconnier's (2002a) theory of MSBT, as well as Turner's (e.g., 2015) later expansion of this work. The article's discussion and critique are inspired by Kress and van Leeuwen's work on social semiotics and multimodality (2001), as well as their individual works on these topics (van Leeuwen 2005; Kress 2010). I use examples from Danish insurance and medical websites to introduce the cognitive framework and illustrate its analytical application, as well as to support my critique.

The article is organized into four sections: In the first part, I introduce basic social semiotic multimodality theory and propose two metaphors (interaction and integration) of multimodal meaning that have historically underlined the development of theoretical concepts within this framework. This is followed in the second part by a critique of the social semiotic conceptualization of multimodal meaning, which primarily builds on an interaction metaphor, while paying less attention to the integration perspective. On this basis, I argue for the potential of MSBT. The third section introduces MSBT as a theory of multimodality and discusses how it can contribute to our understanding of multimodal meaning. The fourth and final section presents as critique of MSBT as a theory of "multimodal meaning as a blend" before concluding.

Overall, the paper offers an exploratory investigation of cognitive and social semiotic approaches to a one of the key terms in multimodality studies and connects it to ongoing research and discussion by members of the Centre for Multimodal Communication and the Department of Language and Communication.

2. Social semiotic theory of multimodal meaning

To present how MSBT can contribute to a conceptualization of multimodal meaning, I must begin by briefly introducing basic assumptions in a social semiotic theory of multimodality.

2.1 Social semiotic multimodality basics

Social semiotic multimodality is the study of the modalities¹ or semiotic systems like writing, speech, images, music, gesture, color etc. that people use to communicate and interact in social settings, including both face-to-face interaction and texts (Kress & van Leeuwen 2001). It examines both the processes and products (multimodal texts) of semiotic work in all social contexts.

According to Jewitt (2009), there are four basic assumptions in social semiotic multimodality theory: (1) Language is always embedded in non-linguistic semiotic modes in meaning-making, whether as face-to-face interaction or in any medium (a text). (2) In a multimodal ensemble (e.g. a text), each modality can perform different kinds of communicative functions depending on the affordances of that particular modality. (3) People create meaning through the selection and combination of modalities available to them. (4) Multimodal meaning-making is shaped by the interests of people as social actors and the social contexts of use.

¹ In the social semiotics of Kress and van Leeuwen, the term 'mode' is preferred over 'modality'; 'modality' is reserved for naming how true or real a representation, such as an image of a person or a thing, is perceived to be. However, in the next section (3.2.) and the remaining part of this article, I will consistently use the cognitive term 'modality' which is equivalent to the social semiotic term 'mode'.

In Kress' (2010) social semiotics, a modality is understood as an abstract semiotic system for making meaning. He points to metafunctionality as a criterion for calling something a modality, drawing on Michael Halliday's (1978) concept of metafunctions in language. Here, metafunctions are the overall functions of language. Kress claims that any modality may be described as a system for making meaning organized in three interrelated ways: As representation of some kind of human experience (called ideational meaning), as social interaction between people involved in the communicative act (called interpersonal meaning) and as flow of information that shapes and organizes human experience and interpersonal relations into a text (called textual meaning). Multiple modalities can contribute to each of these dimensions of meaning. A second criterion of a modality is that a social group must share knowledge of its meaning potential and use it to communicate and interact. It follows that the same modality, for instance color, may be understood differently depending on which social groups share knowledge of its meaning potential. To graphic designers, color is a fully developed, flexible semiotic system, while to lay people it tends to be a more rudimentary means of communication with fixed meanings. It must be pointed out that what counts as a *modality* in contrast to the concept of *medium* (the material resource for expressing meaning), is still debated in multimodal semiotics (e.g., Bateman 2008, 2014). Still, the definition by Kress is widely used by the research community.

A social semiotics analysis is concerned with the study of modalities (material traces/signifiers) and their meaning potentials (past, present, future), including design, use, histories, and discourses of meaning-making. The meaning potential of a modality (e.g. gaze or gesture) is dependent on its material qualities for expression and use in specific contexts, leading different modalities to provide different affordances and constraints for meaning construction.

2.2 Metaphors of multimodal meaning

As I pointed out in the introduction, several authors have suggested combining cognitive and social semiotic perspectives. While I acknowledge the usefulness of such an approach, existing work tends to focus on combining analytical tools. It is, to a lesser degree, concerned with how cognitive and semiotic perspectives may inform the conceptualization of multimodal meaning. However, I find this necessary in order to fully appreciate and evaluate a contribution of MSBT for the study of multimodality. Thus, in this section, I discuss how meaning as a multimodal phenomenon could be understood. To discuss the conceptualization of multimodal meaning, and the extent to which one can integrate MSBT into a social semiotic framework, I will depart from these two perspectives on multimodal meaning.

To do so, I need to give a brief overview of the historical development of social semiotic multimodality studies and the metaphors that have been used to describe multimodality as an abstract concept. Boeriis (2009) describes shifts in the social semiotic research interest from mono-modal, primarily linguistic, studies (Halliday 1978; Hodge & Kress 1988) to poly-modal studies (Kress & van Leeuwen 1996), in which researchers begin to study non-linguistic modalities and reflect on, for example, text-image relations (e.g., Martinec & Salway 2005; van Leeuwen 2005). In a final move, social semiotics fully embrace multi-modalities as an object of study (Kress & van Leeuwen 2001).

All of the above areas of interest can be said to fall under the umbrella of ‘multimodal studies’, even if their concerns are individual modes (cf. Halliday’s early notion of multimodality (1978)). In this literature, there is a general outlook to other modalities and an increasing attention to the concept of multimodality. Running through this development are different metaphors used to grasp multimodality in social semiotics literature. In this context, I want to focus on two overall metaphors that are particularly salient: “In a general perspective, multimodality is often conceptualized as either an interaction between semiotic modalities or an integration of semiotic modalities” (Poulsen 2015).

These perspectives result from my reading of both social semiotic multimodality theory (Poulsen 2014) and are an attempt at recording discernible trends in social semiotics research. The two perspectives, interaction and integration, also represent a historical development of multimodal research over the past fifteen years. Earlier, multimodality tended to be viewed from a perspective of interaction, whereas multimodality today is more frequently viewed from a perspective of integration. The two perspectives, however, are not mutually exclusive, and it is still possible to come across researchers working with one or both perspectives. I must stress, therefore, that these perspectives are the results of my simplified categorization of complex theories.

From an interaction perspective, multimodal meaning is a kind of coupling between individual mono-modalities; this connection of atomic parts gives rise to a separate multimodal meaning, which is a kind of extension of the monomodal meaning(s). The multimodal meaning itself is metaphorically described as a layer on top of the layers of monomodal meaning, or as a separate “bridge” between modalities (see e.g., van Leeuwen 2005; Forceville & Urios-Aparisi 2009). An illustrative model appears in Fei’s (2004) work on language-image relations, where the ‘multimodal’ meaning potential is depicted as a block (ironically called ‘space of integration’) separate from the meaning potentials of the involved monomodalities (see Figure 1).

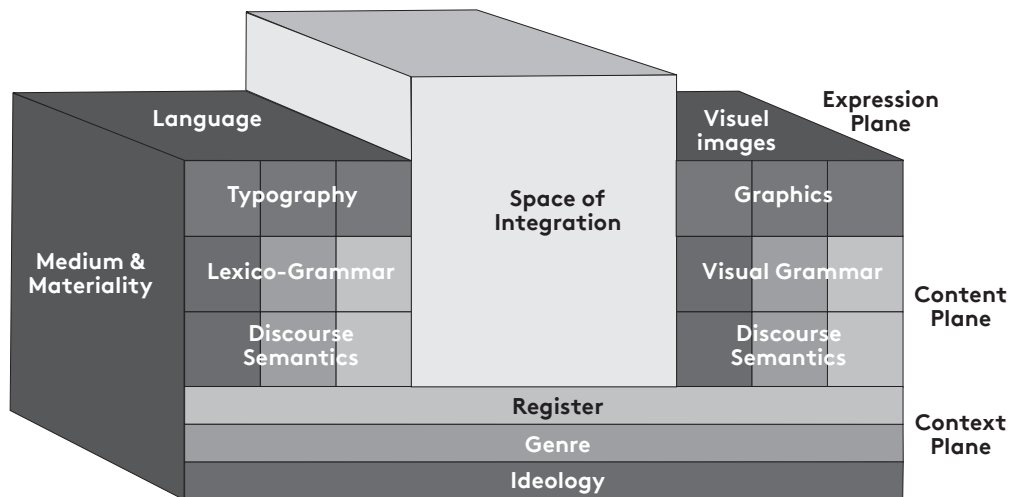


Figure 1. Fei's (2004) model of multimodal meaning.

From an integration perspective, however, multimodal meaning is rather thought of as a whole that is more than its parts (i.e., the resources of different semiotic systems). Multimodal meaning is not considered as somehow separate from individual monomodal meanings, as in the case of the interaction perspective. Metaphors of fusion, compression, chemical reaction, and emergence are used as descriptive analogies (see e.g., Baldry & Thibault 2006; Kvåle 2012; Poulsen 2014).

The notion of multimodal meaning as emergence would be just one among several possible varieties of the integration perspective. It is this frame of thinking I will pursue in this article. In order to do so, and before I discuss a concept of multimodal meaning that builds on the interaction metaphor, I must briefly present how social semiotics conceive meaning, and discuss why MSBT could be linked to it.

2.3 Problems with the interaction perspective on multimodal meaning (and an argument for the integration perspective)

While the interaction metaphor of multimodal meaning may be productive for analyses of multimodal texts, it introduces several problems. First, it creates an ontological problem of multimodal meaning potential because a multimodal phenomenon is conceptualized as a unity that can be analytically divided into atomic monomodal meaning potentials. However, from this does not necessarily follow that each of these monomodal parts has any individual significance. One might argue that it is only through their combination into a whole that they become meaningful. Thus, such a conceptualization relegates multimodal meaning to the backstage and focuses instead on the meaning potential of individual modalities and their individual contributions to the overall meaning. Second, the notion of multimodal meaning as combination seems to be a backward understanding of the multimodal phenomenon under study. I want to argue that the multimodal meaning should be the point of departure for analysis rather than its result.

Finally, many studies present robust analyses of single modalities but have less to say when it comes to questions of re-sampling individual modalities into a multimodal ensemble. These studies conceptualize multimodality as, for example, text-image relations (e.g., Martinec & Salway 2005; van Leeuwen 2005). These difficulties can be illustrated with a simplified analysis of the example shown in Figure 2.



Figure 2. Communication advisor Anne Anker. Retrieved from Tryg's website August 2013.

The example is taken from Danish insurance company Tryg's website. On the page, we are presented with communication advisor Anne Anker², her picture, and written information including her name, position, phone number, and e-mail address. Anne Anker, thus, is multimodally represented in image and writing. These both come with different modal affordances for expression as semantic parts-whole structure. Using social semiotic terminology, the image realizes the 'analytic' structure of Anne Anker as a whole, and her physical features as her parts (Kress & van Leeuwen 2006). In writing, her name, job position and contract information are displayed. These words also create an ideational parts-whole semantic structure (Martin 1992). The multimodal aspect would be described as an image-text relationship in which the writing anchors the image (e.g., van Leeuwen 2005), thereby identifying the person in the image as "Anne Anker, Communications Advisor".

The integration metaphor conceptualizes multimodal meaning as qualitatively different from modalities that contribute to its creation. It is a whole that is more or greater than its parts; in other words, it is a gestalt. However, theoretical work on multimodal meaning that builds on this integration metaphor is less developed than its interaction counterpart. This article seeks to advance the theoretical development further.

3. MSBT as a theory of (multimodal) conceptualization

This section presents the basic tenets of Gilles Fauconnier and Mark Turner's theory of mental spaces and conceptual integration (e.g., 1996, 1998, 2000, 2002a, 2002b, 2003, 2008a, 2008b) and discusses how it extends into multimodal studies.

3.1 MSBT basics

Mental space-blending theory is a cognitive, semantic theory based on the construction of meaning through the structuring, manipulation, and integration of conceptual spaces. This is Turner's concentrated description of the theory:

Blending is a process of conceptual mapping and integration that pervades human thought. A mental space is a small conceptual packet assembled for purposes of thought and action. A mental space network connects an array of mental spaces. A conceptual network is a mental space network that contains one or more "blended mental spaces". A blended mental space is an integrated space that receives input projections from other mental spaces in the network and develops emergent structures not available from the inputs. Blending operates under a set of constitutive principles and a set of governing principles. (Turner 2005)

2 After being in contact with Anne Anker, she has allowed me to use this example.

A classic example is the figure of speech ‘death as the Grim Reaper’, which according to Fauconnier and Turner requires the integration of multiple mental spaces to conceptualize:

- (1) a space with an individual human dying; (2) a space with an abstract pattern of causal tautology in which an event of a certain kind is caused by an abstract causal element: e.g., Death causes dying, Sleep causes sleeping, Smell causes smell, Sloth causes laziness, and so on; (3) a space containing a prototypical human killer; and (4) a space with reapers in the scenario of harvest (2002b: 476).

As Turner writes, the construction of blended meaning is achieved by projecting and condensing the selected elements and relations in and between input spaces within the blended space, whereby meaning is created that goes beyond what is signified by the individual elements of the various input spaces.³ To Fauconnier and Turner, structural relations across input spaces are responsible for establishing new relations within the blended space; these are what they call “vital relations” (2002a: 93-102). They distinguish between outer- and inner-space relations, with the former compressed into the latter within the blend. Only through blending is it possible to establish inner-space relations, and they only exist in the blended space, constituting emergent meanings⁴. In the Grim Reaper example, there are several vital relations of Cause-Effect: Death causes a chain of effects, e.g. from death of a human being to a corpse; from a corpse to its burial; from the decay of a living body into a skeleton. Some of these vital relations, not all, are structurally connected into another Whole-Parts vital relation that in the blend condensed into one and the same entity—a new, emergent entity labelled Uniqueness where e.g. the body and the skeleton become parts of the Grim Reaper (the example also includes other vital relations, cf. Fauconnier & Turner 2002a: 302-3).

In order to describe different types of emergence in blends, Fauconnier and Turner establish a typology of vital relations⁵ describing both outer- and inner-space relations, the latter denoting emergent structures (see Table 1).

3 In the following, I will only focus on the emergence of new semantic structures, which I will connect to the discussion of multimodality. In taking this approach, other parts of the complex MSBT are left out, as the discussion reflects my reading of MSBT as a theory of multimodality and not MSBT in all its aspects.

4 In this context I define emergence created in the blended space as a new manifestation of a meaning that cannot be deduced based on the parts contributing to the blended meaning (for a general discussion of the concept of emergence, see Stephan 1992).

5 In this paper, I decline from discussing the typology in itself because this would derail the point that I am trying to make about multimodality. That being said, a critical and evaluating discussion of ‘vital relations’ is needed. Some of the semantics that vital relations designate could arguable be more early explained within the cognitive paradigm as simple cases of e.g. semantic frames or metonymy. The concept has not been given much attention in the literature on MSBT (for publications on this subject, see <http://markturner.org/>)

VITAL RELATIONS BETWEEN MENTAL SPACES ("OUTER-SPACE VITAL RELATIONS")	VITAL RELATIONS COMPRESSED IN THE BLEND ("INNER-SPACE VITAL RELATIONS")
TIME	SCALED TIME SYNCOPATED TIME
SPACE	SCALED SPACE SYNCOPATED SPACE
REPRESENTATION	UNIQUENESS
ROLE-VALUE	UNIQUENESS
ANALOGY	IDENTITY CATEGORY
DISANALOGY	CHANGE UNIQUENESS
PART-WHOLE	UNIQUENESS
CAUSE-EFFECT	SCALED TIME UNIQUENESS PROPERTY
IDENTITY	UNIQUENESS CHANGE

Table 1. Overview of outer- and inner-space vital relations, from Fauconnier & Turner (2002a, 93-102).

3.2 A cognitive turn towards multimodality

When Fauconnier and Turner began to present MSBT in the 1990's and onwards, they illustrated their framework with both linguistic and nonlinguistic examples. Thus, one might argue, their theory has been geared towards multimodality from the outset. One such multimodal case taken under careful consideration in their book is a visual advertisement for better education in American schools. The advertisement shows three children dressed as doctors in an operating theatre and a caption stating: "Joey, Katie and Todd will be performing your bypass" (2002a: 65). With this example, Fauconnier and Turner demonstrate how a blend or conceptual integration network can be set up by multimodal—linguistic and visual—means, in order to create a blend in which the children are projected into a future scenario of acting as doctors operating on what may well be the reader of the advert. Thus, the attention to multimodal data is not new to MSBT. However, it is only in recent years that Turner and others have overtly declared an interest in multimodality and that researchers have begun to systematically study not only nonlinguistic but also multimodal blends. In this context, Turner and Francis Steen, a colleague cognitive scholar, set up the Red Hen Lab in 2013. Red Hen Lab is "a global enterprise whose goal is to create a massive systematic corpus of ecologically valid, multimodal data, along with new tools and practices to analyze this data" (Red Hen Lab 2013). The lab works towards three defined goals: a) to advance and refine a theory of multimodal communication; b) to develop

tools for collecting, filing and analyzing (large) multimodal data sets, and c) to create a pedagogy that functions as a learning environment. In Steen & Turner (2013), the authors describe the study of multimodal communication like this: “When people meet, they invariably communicate in multiple modalities: the eyes, gestures, and tones of voice merge with the perceived affordances of the surroundings into an integrated and partially shared experience” (2013: 1). Later in the same article, they write: “Red Hen data are highly multimodal, including speech, on-screen text, gesture (broadly defined), bodily stance, music, sound effects, graphics, and a range of other audiovisual expressions” (2013: 3). Turner and others mention that an example of such data could be TV news, but virtually all types of communication could be investigated (e.g., Sweetser 2017; Steen et al. 2018). While not giving a definition of modality, Steen and Turner embrace the concept, and the notion that people communicate using multiple modalities simultaneously. Red Hen Lab’s vision explicitly states that, as humans, we have developed an ability for multimodal communication:

Human beings are *evolved* [my emphasis, SVP] for elaborate multimodal communication. Cultures support this power. Communicating seems easy to human beings, just as seeing seems easy. But it is immensely complex, involving not only vision but also movement, sound, interpersonal interaction, dynamic coordination across agents, conceiving of the intentions of other agents, and so on. (Red Hen Lab 2013)

In the wake of Steen and Turner’s Red Hen Lab work, a broader turn in scholarly interest toward multimodality has occurred, including conferences on multimodal communication and an increasing number of articles and edited volumes.

3.3 MSBT’s contribution to an integration perspective on multimodal meaning

In the above section 3.1., the concept of vital relations, with special attention to inner-space relations, was introduced as a way to describe emergent meaning. As mentioned in the introduction, Wong (2018) (among others) have argued that blends with emergent vital relations can be created, not only linguistically, but also, for example, visually. I want to go beyond this claim to make an argument for MSBT as a theory of multimodal meaning that explicitly explains how this cognitive perspective complements social semiotic theory and addresses multimodal meaning in a novel way. Especially, this aspect of the theory – the concept of vital relations – is a significant contribution to a theoretical understanding of multimodal meaning from an integration perspective on multimodality (Poulsen 2014).

If we accept the premise that meaning created through integration of modalities may be compared to a blend, where a blend is created through integration of (selected parts of) the mental input spaces, multimodal meaning may then also be described as a conceptual integration process.

Among other things, this comparison opens up the use of blending theory to describe multimodal meaning in relation to the emergent relations that may be constructed in such a blend. Therefore, one could argue that multimodal meaning may be categorized using the typology of

emergent relations in blends proposed by Fauconnier and Turner; and if so, specifically multimodal meaning can be seen as an inner-space emergent relation in a blend. That fact makes it relevant for future multimodal research to venture into a description of different vital relations, see Table 1.

Using MSBT this way can describe two things: (1) Multimodality in terms of multiple modal inputs for blends, and (2) multimodal meaning as a blend containing emergent structures and relations that cannot be reduced to or deduced from the modal inputs contributing to the blend. As a result, MSBT makes it possible to designate various kinds of emergent meaning in terms of specific inner-space vital relations. This way, it is possible to go beyond the generic description of multimodal meaning as an extension of, or superstructure on top of, monomodal meaning and elaborate a framework of description for integrated multimodal meaning that uses MSBT to identify different types of emergence (new semantic structures), which have so far eluded description by other descriptive frameworks, including social semiotics.

3.3.1 MSBT AS A FRAMEWORK FOR ANALYZING (MULTIMODAL) TEXTS

If we return to the analytical example (Figure 2), we can use MSBT as a framework and the concept of vital relations to elaborate on the analysis of meaning in a (multimodal) text (Figure 3). This is followed by a more complex example.

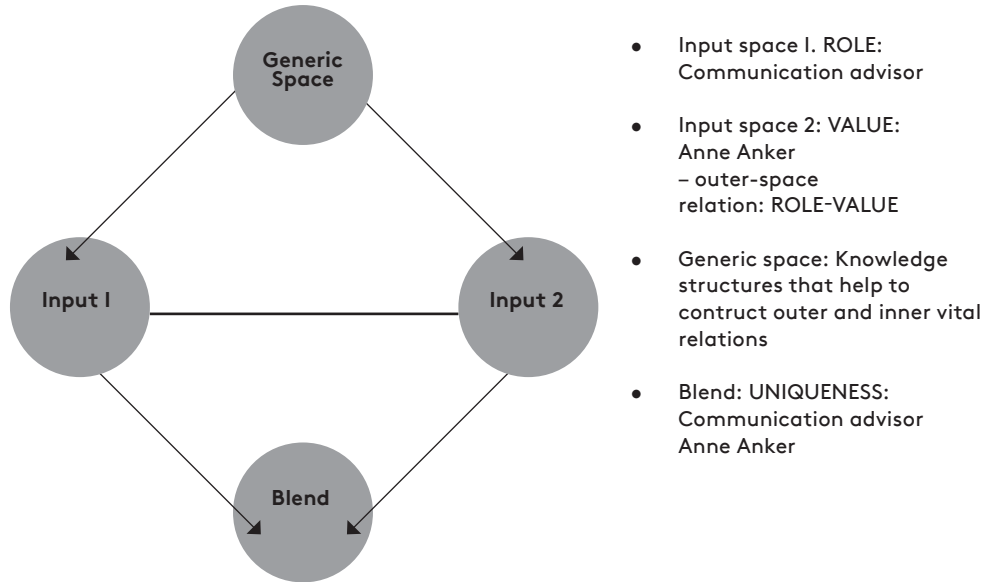


Figure 3. Blending model of Anne Anker's example with vital outer-space and inner-space relations.

In order to understand the information on the website about Anne Anker's job position, MSBT will describe this process of decoding by claiming that two mental spaces are constructed, based on the information about the individual Anne Anker, and on her job as a communication adviser, respectively. In order to construct these mental spaces and to create links between their elements and structures – as well as integrate selected qualities from the individual spaces into a new, joint space (the blended space) – a so-called generic space is set up: a space for background knowledge (encyclopedic knowledge). In the mental processing of these mental spaces, a relation between the input spaces is established, in this case a Role-Value relation. This is a relation between a given “role” or function that someone or something performs and the “value”, the person or entity assuming this role or function. MSBT will subsequently describe how these input spaces are integrated into a new, blended space with the Role-Value relation compressed into the vital relation of Uniqueness. This integration enables the web user (quite unproblematically) to perceive, think about, and refer to “Anne Anker” and “communication adviser” not as two separate pieces of information but as a unique, unified entity, “communication adviser Anne Anker”. This way, MSBT enables us to analyze more precisely the part-whole structure that the social semiotic analysis described. This is just one out of several vital relations that could be identified in the example.

A second example is taken from the pharmaceutical company Lundbeck's website and involves an interactive timeline telling the story of the company in text and pictures (Figure 4). This timeline is placed physically under a number of thematically relevant historical pictures: Lundbeck's founder, his wife, the first chief engineer of the company, a molecule from the first insulin product, etc. As one moves the mouse pointer along the timeline, which is divided into decades from the founding in 1915 to 2010, small images with captions appear.



Figure 4. Interactive timeline of historical events in the organisation, from the Lundbeck website. Retrieved from Lundbeck's website June 2013.

Using MSBT as an analytical framework, the individual images appearing for each decade can be analyzed together as a blend, specifically a compressed timespan, allowing the user to perceive each image as representative of the entire period of time (Figure 5). The individual historical events serve as input spaces, among which the vital relation of Time may be established. All events taking place within the company are integrated into a new blend in which time is compressed and a new structure is established—'Syncopated time' in Fauconnier and Turner's terminology (2002a: 314-320). The blend allows us to perceive the selected images as both related to a collective time in history and representative of all events taking place within each specific decade, for instance from 1920-1930.

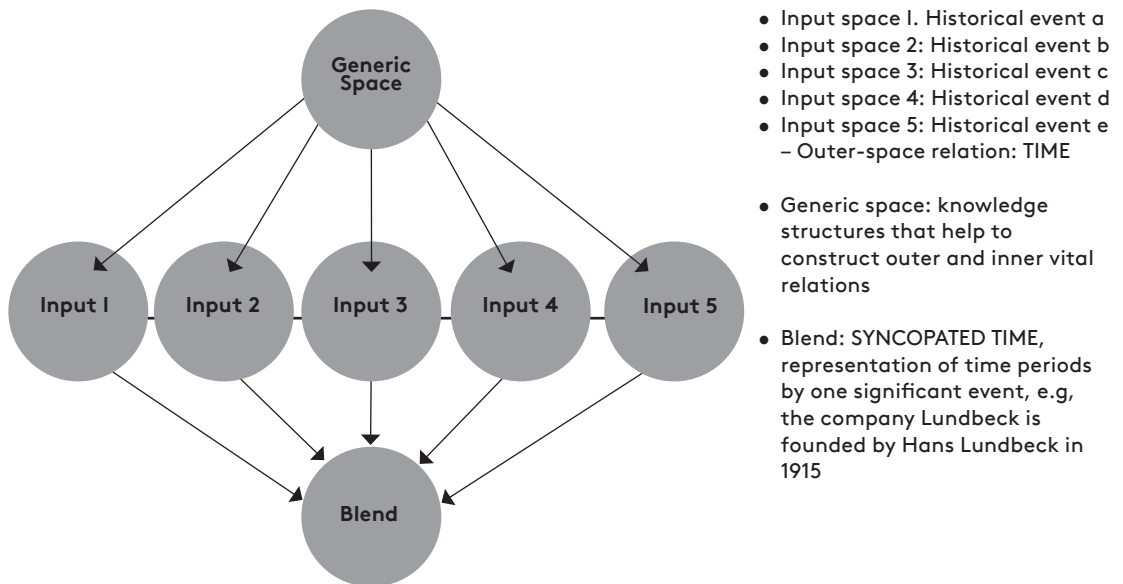


Figure 5. Blending model of Lundbeck's example with vital outer-space and inner-space relations.

4. A (social semiotic) critique of mental spaces-blending as a theory of multimodal meaning

In the previous sections, I have shown how MSBT may contribute to understanding multimodal meaning; in this section, I argue that MSBT's notion of multimodal meaning falls short on three issues.

4.1 First critical point: reflections on modality and multimodality as concepts

First, including MSBT into multimodality studies is still premature, because the theory lacks adequate reflection on the concept of modality. This shortcoming carries over into a vague definition of multimodality. This inevitably means that any notion of differences and similarities between modalities is also inadequately fleshed out, while evidently being needed. If we return to section 3.2., we find no definition of modality, nor of multimodality per se, only examples from which it can be inferred how these might be understood. Furthermore, MSBT seems to conflate two notions of the term: sensorial modality and abstract meaning-making system. In the first case, ‘modality’ has to do with vision (‘the eyes’), touch (‘gestures’), and sound (‘tones of voice’), as well as bodily movement, orientation, and experience (‘perceived affordances of the surroundings’) that links the description to perception psychology (Gibson 1979). In the same context, ‘modalities’ are closer related to what would also be a social semiotic understanding of the term, that is, as semiotic systems (e.g., speech, on-screen text, music, graphics, and other audiovisual expressions). Furthermore, there is no reflection on what makes different modalities different, or on what counts as a modality. It seems that the notion of multimodal communication is accepted a priori but not articulated or questioned.

It is, evidently, a problem that MSBT does not provide a working definition of modality, because, as Kress (2010) writes, it is essential to be able to discuss the gains and losses of different modalities amid or due to processes like transduction or meaning-making across modalities. In addition, reflections on the relationship between modalities are absent, cf. section 2.2. on the interaction and integration metaphors. Thus, in order to integrate this cognitive perspective of MSBT with the social semiotics, I consider it to be a prerequisite that MSBT would explicitly learn from the experience and discussions within the social semiotic paradigm. If MSBT adapted e.g. Kress’s two criteria (metafunctionality and socially shared knowledge of a modality’s meaning potential), the metafictional definition would bring more nuance. E.g. only ideational meaning construction seems to be accounted for in the concept of ‘vital relations’. Emergent interpersonal and textual meaning potentials created in a blend, for example figure-ground phenomena, would be interesting to explore in more detail. Also, MSBT could be informed by the discussion of modality in cognitive semiotics (see Stampoulidis, Bolognesi & Zlatev 2019).

4.2 Second critical point: Indifference to the modality of communication

Second, due to its focus on meaning as a cognitive phenomenon, MSBT does not account for how multimodal communication and meaning construction are distinct from conceptualizations of the same processes in single modalities—and how different modalities represent the “same” meaning potential.

A counterargument to the above critique could be that MSBT is still in the making and that its theoretical development is underway, though not yet complete. But as I see it, the problem runs deeper. In fact, the reason why the notion of modality is vague in MSBT is that MSBT is a cognitive theory primarily concerned with conceptualization as mental processes and to a less-

er degree with the question of which particular communicative modality might prompt conceptualization. Furthermore, it is also indifferent to whether communication is monomodal or multimodal. MSBT's indifference to the modalities of communication can be seen in Fauconnier and Turner's description of vital relations (2002a; 93-102), as well as in Turner's later work, for example on multimodal grammar constructions (e.g. Steen & Turner 2013). If we return to the two above-mentioned analytical examples, a by-product of the analysis was to demonstrate that emergent vital relations may be constructed multimodally. As also mentioned earlier, Fauconnier and Turner were already aware of this, and demonstrated it in their analysis of a visual ad (Fauconnier & Turner 2002a: 65). However, it is revealing that, when introducing the concept of vital relations in their book *The way we think* (2002a), Fauconnier and Turner's examples of the creation of emergent structures in the blended space are both linguistic and non-linguistic (2002a: 93-102). It seems that from this theoretical perspective, the text that sets up input spaces and, thus, establishes the conceptual blend, is indifferent to whether the input spaces are prompted by monomodal or multimodal texts. This supports my claim that blending theory is not attuned to a nuanced conception of multimodality. At its present level of evolution, MSBT is not able to specify how multimodal prompting differs from monomodal blends that create emergent structures. Its notion of conceptualization is clearly indifferent to the kind of modalities that may prompt a conceptual blend, and thus to what is meant by multimodality. In addition, if it is not possible to account for differences between monomodal and multimodal meaning construction, this also undermines the exploration of multimodal data.

For purposes of clarification, MSBT might (again) benefit from reflections found in social semiotic literature on the "gains and losses" of different modalities and differences in terms of the affordances that different modalities provide for making meaning (e.g., Kress 2010). An essential insight in this regard is that meaning is both dependent on resources and situationally anchored, so even if we would attempt to talk about similar meanings constructed in different modalities, they would in fact be different due to usage in different situational contexts.

4.3 Third critical point: Comparability of cognitive and semiotic frameworks for multimodal text analysis?

Thirdly and finally, existing attempts to integrate MSBT with social semiotics prove to be problematic when examined more closely. As an example, I want to highlight how the two frameworks disagree on what elements in a text may function as prompts for conceptualization.

At first glance, cognitive and semiotic frameworks work with a concept of meaning-making prompts. To Kress, "communication is the response to a prompt [in the social environment]; that communication happens only when there is 'interpretation'" (2010: 35). In the words of Wong "the interpreter's interest directs his or her attention to a prompt in the communicative event; the interpreter then engages with features of the prompt and forms his or her attention" (2018: 5). Semiosis, therefore, involves an active process of constantly interpreting meaning. Thus, Kress's notion of "prompt" could link social semiotics to a cognitive perspective on how a (multimodal) text triggers conceptual processing in the interpretation process. But when we analyze the concept of prompts, i.e. initiators of blending processes, I suggest that there are dif-

ferences in the grammatical devices that cognitive and social frameworks identify as prompts for (multimodal) meaning construction, and that therefore it is questionable if the combination of these frameworks is a “perfect fit” (Wong 2018: 1).

To make this argument, we first need to draw a parallel between a social semiotic visual and a linguistic grammar⁶ to be able to imagine how a social semiotic account of visual prompts might correspond to a social semiotic account of linguistic prompts. The argument will then be: Wong describes the visual structures in social semiotic terms and goes on to claim that these structures would be prompts in MSBT’s terms. These visual structures would either be ‘narrative’ or ‘conceptual’ (Kress & van Leeuwen 2006). In images, narrative structures are realized by vectors, or lines in the image indicating actions by participants; by contrast, conceptual structures have no vectors, but are statically displayed organizations of entities realized by a group of similar, aligned elements or as a part-whole structure. Next, if we turn to language, we can express similar ideational content, in terms of social semiotic grammar ‘material processes’ expressing actions (doing and happening), and ‘relational processes’ expressing structures and relations between elements (being and having). Both are realized by verbal groups (Halliday 2004: 170-2, 210). Participants involved in these processes are realized by nominal groups.

Thus, if we imagine material processes in language to be analogous with narrative structure (cf. Kress & van Leeuwen 2006: 59), and relational processes with visual conceptual structures, then following Wong’s line of thought, material and relational processes may be said to function as linguistic “prompts” for conceptualization.

However, if we compare the social semiotic description of grammatical elements that might serve as linguistic prompts to the grammatical devices described by Fauconnier as prompts, we can see that he is actually pointing to different elements. In his work on mental spaces (e.g., 1994, 1997, 2001, 2007, 2009), Fauconnier describes prompts in language as grammatical devices for construction and connection of mental spaces. As an aside, he also points to prompts in similar studies of non-verbal discourses, such as in images, music, human-computer interaction, theater, gesture, and drawing (for a review, see Fauconnier 2009), thereby demonstrating that these cognitive phenomena are not exclusive to language alone. Among others, Fauconnier (2007) lists the following linguistic categories:

Space builders are grammatical expressions that open a new mental space in a conceptual network or shift focus to an existing space. They include prepositional phrases, adverbials, subject-verb complexes, or conjunction+clauses; [...]

Names and descriptions enable one to either set up new elements or point to existing elements in mental spaces; [...]

6 Social semiotics also talk about grammars of non-linguistic modalities, e.g. a visual grammar (Kress & van Leeuwen 2006). The equivalence between grammars of different modalities in a social semiotic approach is important for the sake of my argument about what may be considered to function as modal-specific prompts in a multimodal text. Therefore, I highlight the expression ‘linguistic’ grammar.

Tense and mood determine what kind of space is in focus, its connection to other spaces, its accessibility, and the location of its counterparts in other spaces [...]

Presuppositional constructions signal that a structure is assigned to or placed within a space in the presuppositional mode. That is, they allow a structure in one mental space to be propagated into other spaces for counterparts of the relevant element. (2007: 40-1)

The point is that the grammatical devices that Fauconnier claims function as prompts are different from those to which a social semiotic description would assign such functions. While cognitive approaches do not provide a systematic, detailed description of how multimodal texts afford meaning construction, and while social semiotics does offer an analytical framework for this purpose, a closer look at blending theory must still lead one to consider how cognitive theory can qualify social semiotics in multimodal meaning construction.

To sum up this discussion, I would agree with e.g. Feng and O'Halloran (2013) who claim that MSBT does not provide a framework for systematic close text-detailed description of how blending processes are initiated in non-linguistic text, e.g. images. Still, I am skeptical about both their attempt as well as Wong's (2018) to join cognitive and social semiotic frameworks for analysis. Using the modality of language as an example, we may locate differences in the social semiotic and cognitive descriptions of linguistic prompts, and this points to disagreement about what might initiate conceptualization.

5. Conclusion

This paper has critically discussed Fauconnier and Turner's (2002a) work on mental spaces and blending as a theory of multimodality from the perspective of social semiotics. I have argued that MSBT could be a way to advance our understanding of multimodal meaning as an integration of modalities that create emergent semantic structures. But on these key terms, MSBT is still underdeveloped, entailing theoretical, methodological and analytical problems for analysis of multimodal data: First, the concepts of modality and multimodality, second, indifference to the specific modalities of communication used in a text, and finally, inconsistency in cognitive grammatical descriptions of multimodal prompts which makes MSBT incompatible with a social semiotic framework.

Putting the critique of this paper into a broader perspective, there are several issues that also need to be discussed if the frameworks of MSBT and social semiotics are to be combined. These might include the following:

- i) It is unclear how MSBT regards concepts of metafunctional meaning and stratification that constitute key terms in social semiotics. The examples discussed in this paper all concern ideational meaning construction. But it is paramount to

consider to what extent MSBT might complement the construction of interpersonal and textual meaning within a conceptual integration framework. Furthermore, social semiotics describe meaning in different modalities in relation to different levels or strata (Boeriis 2017). While Turner (2015) seems to operate with a notion of stratification when he talks about ‘forms’ and ‘meaning’, further research is needed to map similarities and differences to be able to argue that cognitive and semiotic frameworks are comparable.

- ii) The notion of ‘social’ as the engine of semiosis is an underlying assumption in social semiotics but MSBT is a theory of individual cognition. On this crucial issue, models that view blending not as an individual cognitive process, but as joint social activity has been proposed by e.g., Gibbs and Santa Cruz (2012), but how this understanding of the social relates to that of social semiotics still remains unanswered.
- iii) If we turn to a cognitive point of view, while this paper has focused on integration of modalities in multimodal meaning constructions, the concept of disintegration also needs to be considered in a multimodal perspective (Bache 2005; Hougaard 2005)
- iv) The so-called ‘third’ generation of cognitive theory also challenges basic assumptions of conceptualization in MSBT: “This could be done by pointing to studies of extended cognition (Clark 2008) or distributed cognition (Hutchins 1995, 2005) between individuals and their surroundings. In continuation of this, it would be nice to see a more in-depth discussion of what constitutes multimodality when it is linked to more recent definitions of cognition” (Poulsen 2017).

While MSBT holds great promise as a theory of multimodality, the issues pointed out in this paper need to be discussed in greater detail to advance multimodal studies further.

Acknowledgements

The author would like to thank the editors and the anonymous reviewer for the feedback and encouragement that helped improve the article.

References

- Alonso, Isabel, Silvia Molina & Maria Dolores Porto Requejo. 2013. Multimodal digital storytelling: integrating information, emotion and social cognition. *Review of Cognitive Linguistics* 11 (2). 369–387.
- Bache, Carl. 2005. Constraining conceptual integration theory: Levels of blending and disintegration. *Journal of Pragmatics* 37(10). 1615–1635.

- Baldry, Anthony & Thibault, Paul. J. 2006. Multimodal transcription and text analysis: A multimedia toolkit and coursebook. London: Equinox.
- Bateman, John. A. 2008. Multimodality and genre: a foundation for the systematic analysis of multimodal documents. Houndmills, Basingstoke, Hampshire: Palgrave Macmillan.
- Bateman, John A. 2014. *Text and image: a critical introduction to the visual/verbal divide*. London & New York: Routledge.
- Boeriis, Morten. 2009. Multimodal socialsemiotik og levende billeder [Multimodal social semiotics and moving images] (Unpublished PhD Thesis). University of Southern Denmark, Odense, Denmark.
- Boeriis, Morten. 2017. Modeling multimodal stratification. In J. Wildfeuer & O. Siezov (Eds.), *New studies in multimodality: Conceptual and methodological elaboration*. London: Bloomsbury. 65-90.
- Caponetto, Paolo G. 2016. Red Hen Lab: a further step towards multimodality. *Journal of Multimodal Communication Studies* 3 (1-2).
- Clark, Andy. 2008. *Supersizing the mind: Embodiment, action, and cognitive extension*. Oxford & New York: Oxford University Press.
- Fauconnier, Gilles. 1994. *Mental spaces: aspects of meaning construction in natural language*. New York: Cambridge University Press.
- Fauconnier, Gilles. 1997. *Mappings in thought and language*. Cambridge: Cambridge University Press.
- Fauconnier, Gilles. 2001. Conceptual blending and analogy. In D. Gentner, K. J. Holyoak & B. N. Kokinov (Eds.), *The analogical mind: Perspectives from cognitive science*. Cambridge, Mass.: MIT Press. 255-286.
- Fauconnier, Gilles. 2007. Mental spaces. In D. Geeraerts & H. Cuyckens (Eds.), *The Oxford Handbook of Cognitive Linguistics*. Oxford: Oxford University Press. 351-376.
- Fauconnier, Gilles. 2009. Generalized Integration Networks. In V. Evans & S. Pourcel (Eds.), *New directions in cognitive linguistics*. Philadelphia: John Benjamins Publishing Company. 147-160.
- Fauconnier, Gilles & Mark Turner. 1996. Blending as a Central Process of Grammar. In A. Goldberg (Ed.), *Conceptual Structure, Discourse and Language*. Stanford: CSLI Publications. 113-130.
- Fauconnier, Gilles & Mark Turner. 1998. Conceptual Integration Networks. *Cognitive Science*, 22(2). 133-187.
- Fauconnier, Gilles & Mark Turner. 2000. Compression and global insight. *Cognitive linguistics*, 11(3/4). 283-304.
- Fauconnier, Gilles & Mark Turner. 2002a. *The way we think: conceptual blending and the mind's hidden complexities*. New York: Basic Books.
- Fauconnier, Gilles & Mark Turner. 2002b. Metaphor, metonymy, and binding. In R. Dirven and R. Póring (Eds.), *Metaphor and Metonymy in Comparison and Contrast*. Berlin & New York: Mouton de Gruyter. 469-488.
- Fauconnier, Gilles & Mark Turner. 2003. Conceptual blending, form and meaning. *Recherches en communication* 19(19). 57-86.

- Fauconnier, Gilles & Mark Turner. 2008a. The origin of language as a product of the evolution of doublescope blending. *Behavioral and Brain Sciences*, 31(5). 520-521.
- Fauconnier, Gilles & Mark Turner. 2008b. Rethinking metaphor. In R. W. Gibbs (Ed.), *The Cambridge handbook of metaphor and thought*. New York: Cambridge University Press. 53-66.
- Fei, Victor L. 2004. Developing an integrative multi-semiotic model. In K. O'Halloran (Ed.), *Multimodal discourse analysis: Systemic functional perspectives*. London: Continuum. 220-246.
- Feng, Dezheng & Kay O'Halloran. 2013. The visual representation of metaphor: a social semiotic approach. *Review of Cognitive Linguistics* 11(2). 320-335. (Reprinted in María Jesús Pinar Sanz (ed.), *Multimodality and cognitive linguistics*. Amsterdam & Philadelphia: John Benjamins. 99-114.)
- Forceville, Charles. 1999. Educating the eye? Kress and Van Leeuwen's *Reading Images: The Grammar of Visual Design* (1996). *Language and Literature* 8(2). 163-178.
- Forceville, Charles & Eduardo Urios-Aparisi (Eds.). 2009. *Multimodal Metaphor*. Berlin & New York: Mouton de Gruyter.
- Gibbons, Alison. 2011. *Multimodality, cognition, and experimental literature*. Berlin & New York: Mouton de Gruyter.
- Gibbs, Raymond W. & Malaika J. Santa Cruz. 2012. Temporal unfolding of conceptual metaphor experiences. *Metaphor and Symbol* 27. 299-311.
- Gibson, James J. 1979. *The ecological approach to visual perception*. Boston, Houghton Mifflin.
- Halliday, Michael A.K. 1978. *Language as Social Semiotic: The Social Interpretation of Language and Meaning*. London: Edward Arnold.
- Halliday, Michael A.K. 2004. *An Introduction to Functional Grammar* (3rd ed.). London: Arnold.
- Hodge, Robert & Gunther Kress. 1988. *Social Semiotics*. Cambridge: Polity Press.
- Hougaard, Anders. 2005. Conceptual disintegration and blending in interactional sequences: a discussion of new phenomena, processes vs. products, and methodology. *Journal of Pragmatics* 37(10). 1653-1685.
- Hutchins, Edwin. 1995. *Cognition in the wild*. Cambridge, MA: MIT Press.
- Hutchins, Edwin. 2005. Material anchors for conceptual blends. *Journal of Pragmatics* 37(10). 1555-1577.
- Jewitt, Carey. 2009. Different approaches to multimodality. In C. Jewitt (Ed.), *The Routledge Handbook of Multimodal Analysis*. London & New York: Routledge. 28-39.
- Kress, Gunther. 2010. *Multimodality: A Social Semiotic Approach to Contemporary Communication*. London & New York: Routledge.
- Kress, Gunther & Theo van Leeuwen. 1996. *Reading Images: The Grammar of Visual Design* (1st ed.). London & New York: Routledge.
- Kress, Gunther & Theo van Leeuwen. 2001. *Multimodal Discourse: The Modes and Media of Contemporary Communication*. London: Edward Arnold.
- Kress, Gunther & Theo van Leeuwen. 2006. *Reading Images: The Grammar of Visual Design* (2nd ed.). London & New York: Routledge.
- Kvåle, Gundhild. 2012. Multimodalt samspill i bildeskriftkomplekser - En sosiosemiotisk undersøkelse av relasjoner mellom skrift og bilde. [Multimodal interaction in image-writing complexes. A social semiotic study of image and writing relationships] (Unpublished Ph.D. thesis). University of Agder, Norway.

- Martin, James R. 1992. *English text: system and structure*. Amsterdam: John Benjamins.
- Martinec, Radan & Andrew Salway. 2005. A system for image-text relations in new (and old) media. *Visual Communication*, 4(3), 337-371.
- Poulsen, Søren Vigild. 2014. *Mod en analysemetode til webstedet som multimodal tekst - en undersøgelse af socialsemiotikkens og den kognitive semantiks bidrag til udviklingen og udformningen af en metode til webstedsanalyse [Towards an analytical method for the website as a multimodal text - a study of social semiotic and cognitive semantic contributions to the development and design of a method for website analysis]*. Unpublished PhD Thesis. Department for Language and Communication. University of Southern Denmark.
- Poulsen, Søren Vigild. 2015. Multimodal meaning. In Nina Nørgaard (Ed.), *Key Terms in Multimodality: Definitions, Issues, Discussions*. www.sdu.dk/multimodalkeyterms. Retrieved 05.11.2019.
- Poulsen, Søren Vigild. 2017. Book review *Multimodality and Cognitive Linguistics*. María Jesús Pinar Sanz (Ed.). *Metaphor & Symbol* 32(2). 135-137.
- Pinar Sanz, María Jesús (Ed.). 2015. *Multimodality and Cognitive Linguistics*. Amsterdam, & Philadelphia: John Benjamins Publishing Company.
- Red Hen Lab. 2013. Mission statement. <http://www.redhenlab.org/>. Accessed on 26 May, 2019.
- Stampoulidis, Georgios, Marianna Bolognesi & Jordan Zlatev. 2019. A cognitive semiotic exploration of metaphors in greek street art. *Cognitive Semiotics* 12(1). 1-20.
- Steen, Francis & Mark B. Turner. 2013. Multimodal construction grammar. In M. Borkent, B. Dancygier & J. Hinnell (Eds.), *Language and the Creative Mind*. Stanford: CSLI Publications. 255-274.
- Steen, Francis F., Anders Hougaard, Jungseock Joo, Inés Olza, Cristóbal Pagán Cánovas, Anna Pleshakova, Soumya Ray, Peter Uhrig, Javier Valenzuela, Jacek Woźny & Mark Turner. 2018. Toward an infrastructure for data-driven multimodal communication research. *Linguistics Vanguard* 4(1). 1-9.
- Stephan, Achim. 1992. Emergence - A Systematic View on its Historical Facets. In A. Beckermann, H. Floor & J. Kim (Eds.), *Emergence Or Reduction?: Essays on the Prospects of Nonreductive Physicalism*. Berlin & New York: de Gruyter. 25-48.
- Sweetser, Eve. 2017. Metaphor and metonymy in advertising: Building viewpoint in multimodal multi-space blends. *Journal of Pragmatics* 122. 65-76.
- Turner, Mark. 2005. Blending and conceptual integration. <http://markturner.org/blending.html>. Accessed on 26 May, 2019.
- Turner, Mark. 2015. Blending in language and communication. In E. Dabrowska & D. Divjak (Eds.), *Handbook of cognitive linguistics*. Berlin: De Gruyter Mouton. 211-32
- Van Leeuwen, Theo. 2005. *Introducing Social Semiotics*. London and New York: Routledge.
- Wong, May L-Y. 2018. Analysing Visual Imagery: Connecting Semiotic and Cognitive Perspectives. *Cognitive Semantics* 4(1). 39-75.